

digital forum

WIOSNA 96

ROK 5 NUMER 18



MICROSOFT
WINDOWS NT

**System to podstawa
- dla każdego coś innego**

WYWIAD

- 4** OpenVMS - SYSTEM OPERACYJNY DLA NAJBARDZIEJ WYMAGAJĄCYCH
Howard Hayakawa, szef zespołu rozwijającego system OpenVMS dla komputerów Alpha mówi strategii Digitala w zakresie oprogramowania systemowego. Uważa, że mimo upływu lat VMS wciąż jest systemem młodym.

DIGITALinfo

- 8** Najpotężniejszy ośrodek Digitala w Europie
Technologia Alpha Generation
Digital i SAP w systemie Windows NT
- 9** Digital w bombowcuB-2
Miliardowy kontrakt Digitala
AlphaStudio - marzenie realizatora telewizyjnego
Microsoft - Digital polskie następstwa śwaitowego aliansu
- 10** Las Vegas na Wołoskiej
Komputery i artyści Infosystem '96

NOWE PRODUKTY

- 11** AlphaSerwer 4100 - NOWY OREŻ DIGITALA
6 maja ogłoszono o nowym fantastycznym rekordzie w teście TPC-C, który został osiągnięty dla cztero-procesorowego serwera w klasie komputerów przeznaczonych do obsługi oddziałowej - 6056 transakcji na minutę. Rekordowy rezultat osiągnięto posługując się 64-bitową bazą Oracle 7 Server oraz najnowszym serwerem Digitala AlphaServer 4100.

OPROGRAMOWANIE

- 14** SYSTEM OPERACYJNY OpenVMS v7.0
Wersja 7.0 jest jedną z najważniejszych udoskonaleń systemu VMS w ciągu ostatnich lat wnosząc kilka ważnych i nowoczesnych technologii. Są to przede

wszystkim - 64-bitowa przestrzeń adresowa, system plików nowej generacji oraz wątki procesów na poziomie jądra systemu.

- 24** AFFINITY - ŁĄCZY TO CO NAJLEPSZE
W sierpniu 1995 roku szefowie firm Digital i Microsoft ogłosili program o nazwie Alliance for Enterprise Computinf. Program ten wraz z programem OpenVMS i Windows NT Affinity ogłoszonym w maju 1995 wyznacza ramy współdziałania obu firm.

- 34** DIGITAL UNIX - STANDARD, KTÓREMU MOŻNA ZAUFAC
Digital UNIX wywodzi się bezpośrednio z wersji OSF/1, którego rozwój oparto na elementach różnych implementacji UNIXa. OSF/1 posiadał wybrane cechy takich technologii jak AIX, 4.3 BSD, System V i CMU Mach. Digital UNIX 4.0 jest pierwszym systemem, który jednocześnie uzyskał prawo do znaku UNIX 95 i CDE (Common Desktop Environment).

- 45** TECHNOLOGIA KLASTER - SPEŁNIENIE OCZEKIWAŃ
W roku 1983 Digital wypuścił na rynek unikalne rozwiązanie łączące w sobie walory odporności konwencjonalnych sieci komputerowych z wydajnością maszyn MPP. Produkt nazywał się VAXcluster. Od tamtej pory Digital jest uznawany za niekwestionowanego lidera w technologii klastrów.

DIGITAL PARTNER

- 57** REALIZACJA APLIKACJI DLA CENTRALNEGO REJESTRU SKAZANYCH
W 1993 roku Digital Equipment Polska podpisał z Ministerstwem Sprawiedliwości kontrakt skomputeryzowany system obsługi Centralnego Rejestru Skazanych (CRS). W celu realizacji oprogramowania systemu CRS, Digital podjął współpracę z firmą PentaComp. Prace nad aplikacją zakończyły się w marcu 1996. Projektowany system został przekazany Ministerstwu i rozpoczął się intensywny okres jego wdrażania.

Poniższe znaki są zastrzeżonymi znakami handlowymi firmy Digital Equipment Corporation:

AdvantageCluster, ALL-IN-1, Alpha AXP, AlphaGeneration, AlphaServer, AlphaStation, AXP, DEC, DECchip, DECnet DECsafe, DECUS, Digital, Digital UNIX, DSSI, ELECTRONIC LOCKER, FDDI, GIGAswitch, InfoServer, INTERNET, MAILbus, MailWorks, MicroVAX, NAS, OpenVMS, PATHWORKS, PDP, POLYCENTER, PowerStorm, RSTS/E, RSX/11, RT/11, TeamLinks, the AXP logo, the DIGITAL logo, TURBOchannel, ULTRIX, ULTRIX/SQL, UNIBUS, WPS, WPS PLUS, VAX, VAX ACMS, VAXBI, VAXcluster, VAXft, VAXstation, VAXsystem, VAX VTX, VAX 11/780, VAX 4000, VAX 6000, VAX 9000, VMS, VT.

Poniższe znaki są nazwami zastrzeżonymi przez Digital Equipment Polska:
DECforum, DECpartner, System Otwartych Możliwości, Wspomaganie Aplikacji Sieciowej.

AIM jest zastrzeżonym znakiem handlowym AIM Technology, X Window System i X Window System Version 11 są zastrzeżonymi znakami handlowymi Massachusetts Institute of Technology, MIPS jest zastrzeżonym znakiem handlowym MIPS Computer System, Sun, Solaris, NFS są zastrzeżonymi znakami handlowymi Sun Microsystems, Inc., Intel jest zastrzeżonym znakiem handlowym Intel Corporation, AT&T są zastrzeżonymi znakami handlowymi American Telephone and Telegraph Company, Motif, OSF i OSF/1 są zastrzeżonymi znakami handlowymi Open Software Foundation, POSIX jest znakiem handlowym Institute of Electrical and Electronics Engineers, XENIX, MS-DOS, MS, MS Windows, MS Word i Windows NT są zastrzeżonymi znakami handlowymi, a DOS znakiem handlowym Microsoft Corporation, AIX, IBM, IBM PC/AT, NetView są zastrzeżonymi znakami handlowymi International Business, Cray jest zastrzeżonym znakiem handlowym Cray Research, Inc., Ethernet jest znakiem handlowym Xerox Corporation, X/Open jest znakiem handlowym X/Open Company, Ltd, AppleTalk, LocalTalk, Macintosh i Apple są zastrzeżonymi znakami handlowymi Apple Computer, Inc., Ingres jest zastrzeżonym znakiem handlowym INGRES Inc., NetWare jest zastrzeżonym znakiem handlowym, a Novell i IPX są znakami handlowymi Novell, Inc., Inc. SPEC i SPECmark89 są zastrzeżonymi znakami Standard Performance Evaluation Corporation., HP i HP/UX są zastrzeżonymi znakami handlowymi Hewlett-Packard Corporation, Informix jest zastrzeżonym znakiem handlowym Informix Software, Inc., ORACLE jest zastrzeżonym znakiem handlowym Oracle Corporation, Sybase jest zastrzeżonym znakiem handlowym Sybase, UNIX jest zastrzeżonym znakiem handlowym licencjonowanym wyłącznie dla X/Open Company, Ltd.

Pozostałe nazwy produktów mają zastrzeżone znaki handlowe przez macierzyste firmy.

Wiosna '96
rok 5, numer 18
ISSN 0867-8782

Kwartalnik wydawany przez
Digital Equipment Polska

Rada programowa
Marek Racieski
Waldemar Calka
Włodzimierz Denis
Maciej A. Markowski
Jarosław Parliński
Magdalena Poklewska-Koziełł
Piotr Sobolewski
Artur Stefanowicz

Digital Equipment Polska Sp. z o.o.
ul. Wołoska 18 (d. Komarowa)
02-672 Warszawa
tel. 640-01-67
fax. 640-01-11
sat. 39.121801

Zamieszczone w piśmie informacje zostały opracowane na podstawie materiałów wewnętrznych i przedruków z pism Digitala. Digital jest przekonany, że informacje w tej publikacji są prawdziwe w chwili ich zamieszczenia, chociaż mogą się one zmienić bez ogłoszenia, stąd Digital nie odpowiada za problemy z tego faktu wynikające. W piśmie są też zamieszczane teksty przygotowane przez autorów niezależnych od Digitala. W takim przypadku treść publikacji nie zawsze musi być zgodna z opinią Digitala. Dla ostatecznego zweryfikowania podanych informacji prosimy o kontakt z naszym biurem w Warszawie.

**Redakcja Techniczna
i opracowanie graficzne**
"CLASSICS" sp. cyw.
ul. Niemcewicz 7/9
02-022 Warszawa
tel. 668-78-12

Przygotowanie techniczne
"B i W" oraz "PR-INFO"

Serwis fotograficzny
Leszek Putkowski oraz
materiały Digital Equipment Corp.

DIGITALforum
jest dostępny w prenumeracie rocznej

Egzemplarze archiwalne są dostępne w Redakcji w Digitalu do wyczerpania nakładu.

Reklamy i ogłoszenia przyjmowane są przez Redakcję, która zastrzega sobie prawo odzroczenia publikacji reklamy i ogłoszenia.

(C) Digital Equipment Polska
Wszelkie prawa zastrzeżone.

Wykaz zastrzeżonych znaków handlowych jest podany pod spisem treści. Przedruk dopuszczalny z podaniem źródła i poinformowaniem Redakcji.

Nakład 4000 egz.

Druk

Drukarnia Siostr Loretanek w Rembertowie

System to podstawa

Informatyka odciska coraz bardziej znaczące piętno na naszym codziennym życiu. Różne informatyczne określenia, nazwy i definicje wkradają się do naszego codziennego życia. Jednym z takich słów kluczowych jest „system”. Słowo to zdobyło prawo obywatelstwa pod koniec lat siedemdziesiątych. W latach następnych jego użycie doprowadzono do absurdu. Wszystko było systemowe, a kilka elementów, które miały ze sobą cokolwiek wspólnego nazywano systemem. Nawet preparaty do łatania dziur w zębach składały się na system.

Dzisiaj prawie wszyscy wiemy co to jest system. Samo słowo pochodzi z języka greckiego, w którym oznaczało pewne zestawienie, pewną całość złożoną z części funkcjonującą w określony sposób. Komputer jest doskonałym przykładem systemu. Jednakże ten system nie działałby w ogóle bez części, która sama jest także systemem. Ta niezbędna część to system operacyjny. Większość użytkowników dzisiejszych komputerów zwykle nawet nie zauważa tego niezwykle istotnego elementu systemu komputerowego. Digital jako jedyny już chyba producent systemów komputerowych przykładą ogromną wagę do tej części, która nadaje komputerowi charakter. Dlatego mimo wielu niedogodności stara się od kilku lat z powodzeniem rozwijać trzy systemy operacyjne - OpenVMS, Digital UNIX i Windows NT - przeznaczone do różnych zastosowań.

Tym systemom poświęcone jest bieżące wydanie DigitalForum. Chcemy przekonać czytelników, że wysiłek Digitala jest zasadny i potrójna oferta systemowa ma na celu zaspokojenie wszystkich wymagań naszych klientów. Co więcej, wszystkie trzy systemy współpracują ze sobą bez żadnych ograniczeń na poziomie aplikacji i komunikacji sieciowej. Digital jako jedyny daje klientom wybór, umożliwiając łączenie ze sobą tych elementów systemu informatycznego, które do tej pory były rozwijane samodzielnie.

O możliwościach Digitala świadczy nie tylko rozwijanie trzech systemów operacyjnych, ale również niezwykle szybkość wprowadzania na rynek nowych modeli serwerów i stacji roboczych. Praktycznie co kwartał pojawiają się nowe systemy. Ostatnim przebojem jest rodzina serwerów AlphaServer 4100, która stanowi pomost pomiędzy rodzinami serwerów 2100 i 8000 zbliżając się wymiarami do tych pierwszych równocześnie zachowując wiele cech systemowych, zwłaszcza opcję działania na wielkiej (ponad 2GB) pamięci, rodziny 8000. Jestem przekonany, że rodzina 4100 odniesie jeszcze większe sukcesy niż jej poprzednicy.

Życzymy przyjemnej lektury!

Redakcja

OpenVMS

- System operacyjny dla wymagających

Rozwiązywanie dużych problemów obliczeniowych dotarło do granicy możliwości systemów 16-bitowych

Konkurenci Digitala wyśmiewali się z potrzeby wprowadzania systemów 32-bitowych

Motto:

„VMS jest jak wino dobrej marki - im starszy, tym lepszy”.

W bieżącym numerze chcemy Państwu przedstawić bardzo interesujący wywiad z Howardem Hayakawą, człowiekiem, który pełni rolę szefa zespołu rozwijającego OpenVMS, jeden z trzech systemów operacyjnych oferowanych przez Digital. System ten zaprojektowany prawie 20 lat temu jest ciągle rozwijany i unowocześniany. Spełnia obecnie wszystkie wymagania nakładane na nowoczesny system operacyjny, taki jak np. UNIX. Co więcej, w dużych zastosowaniach, wymagających podwyższonej niezawodności i szybkości działania (systemy transakcyjne) jest uważany za system o niedoścignionej jakości. Warto przeczytać co mówi na ten temat specjalnie dla Digital Forum Howard Hayakawa, który z powodzeniem dba o siebie o to, aby OpenVMS był najlepszym systemem operacyjnym na świecie.

Cofnijmy się do początków. Pod koniec lat 70. tych Digital posiadał wspaniałe 16-bitowe systemy operacyjne rodziny RSX. Dlaczego właśnie wtedy firma zdecydowała się wprowadzić całkiem nowy system VMS?

Digital opracował wówczas nowy system operacyjny orientując się, że wymagania nakładane na procesy przetwarzania ulegają szybkim zmianom. Zauważono, że rosnąca moc przetwarzania prowadzi do rozwoju systemów komputerowych w takich kierunkach, gdzie wykorzystywanie i ulepszanie istniejących systemów 16-bitowych już nic nie daje. Rozwiązanie tak dużych problemów obliczeniowych, jak na przykład analizy sejsmiczne dotarło do granicy możliwości systemów 16-bitowych. Stwierdzono także, że metody symulacyjne stosowane w technice i analizach naukowych mogą generować prawie nieograniczonych ilości danych.

Konkurenci Digitala wyśmiewali się z po-

trzeby wprowadzania systemów 32-bitowych.

Wizja Digitala została precyzyjnie zrealizowana w postaci komputera VAX i systemu operacyjnego VMS, które stały się standardami w przemyśle komputerowym nie obejmując jedynie systemów typu „mainframe” i PC.

Digital nie tylko przewidział potrzebę rozszerzenia architektury systemu pod względem liczby bitów, ale także konieczność podniesienia jakości, wydajności i możliwości wielodostępu środowiska systemowego. Digital zaproponował również metodologię szybkiego rozwijania oprogramowania wprowadzając wizjonerski w owym czasie standard odwołań, umożliwiający prostą współpracę programów pisanych w różnych językach.

W zakresie sprzętowym Digital jako pierwszy zastosował specjalizowane magistrale do łączności w sieciach lokalnych i połączeń między komputerami i systemami. Magistrala oznaczona krótką nazwą NI, stała się znana później jako standard przemysłowy o nazwie Ethernet. Podobnie, magistrala CI stanowiła podstawy standardu w zakresie tworzenia klastrów.

W wyniku zespolenia wielu pomysłów sprzętowych i programistycznych oraz całkowicie nowej architektury magistral Digital wprowadził na rynek całą gamę produktów - od stacji roboczych do dużych komputerów - wymienionych pod względem systemowym i programowym.

System VMS rozwija się z powodzeniem już prawie 20 lat. Jakie kamienie milowe można wyróżnić w rozwoju tego systemu?

Oczywiście, początkowo największy nacisk kładziono na strukturę 32-bitową, która umożliwiła pobicie rekordów we wszelkich możliwych testach. Przemysł komputerowy musiał to szybko wziąć pod uwagę.

Kolejnym sukcesem była 4 wersja systemu uwzględniająca tworzenie systemów klastrowych.

Niewątpliwie system VMS jako jeden z pierwszych przełamał barierę otwartości, gdy już jako OpenVMS często i szybciej niż wiele systemów UNIX osiągał zgodność z kolejnymi wymaganiami specyfikacji XPG4.

Wersja 6 systemu dawała możliwość łączenia w tym samym klastrze komputerów VAX i Alpha. Wciąż nie słychać aby przemysł komputerowy poradził sobie z łączeniem w ramach tego samego klastra maszyn o różnych listach rozkazów.

Niedawno Digital zaoferował również klustry odporne na katastrofy (Disaster Tolerant Clusters). Ta cecha systemu, unikalna w przemyśle komputerowym pozwala użytkownikom uniknąć sytuacji katastrofalnej, kiedy zostaje zniszczony cały budynek, a nawet obszar, gdzie działa system. Będzie on funkcjonował bez przeszkód dalej w innym miejscu oddalonym nawet o 1000 kilometrów. Niezwykłe jest to, że system zapasowy może przejąć pełną kontrolę w ciągu kilku sekund lub minut bez potrzeby montowania taśm magnetycznych i startowania oprogramowania od początku.

VMS przez długie lata był widziany jako własny, specyficzny system Digitala. Co firma zrobiła i nadal robi, aby przekonać potencjalnych klientów, że jest inaczej?

Jak już podkreśliłem, kilka lat temu Digital zmodyfikował system VMS aby osiągnąć zgodność ze specyfikacją XPG4. Zgodność została osiągnięta, a fakt przyznania odpowiednich certyfikatów firma zaznaczyła zmianą nazwy systemu na OpenVMS. Cały czas uzupełniamy serwisy systemowe tak aby zapewnić maksymalną wymiennność pomiędzy systemami OpenVMS i UNIX. Obecnie OpenVMS doskonale odpowiada specyfikacji 1170. Przenoszenie oprogramowania pomiędzy systemem UNIX i OpenVMS nie sprawia żadnych trudności.

Jednakże, musimy przyjąć do wiadomości, że bez względu na to co ludzie mówią zgodność ze standardami może nie mieć w tym przypadku znaczenia. Po prostu używają oni pewnych uproszczeń i schematów myślowych deklarując chęć stosowania systemu UNIX.

Teraz, w odniesieniu do OpenVMS głosimy tak jak Microsoft, że na prawdę ważna jest

sprawność rozprzestrzeniania się systemu przede wszystkim zależna od użytkowników. Sednem Strategii Affinity odnoszącej się do OpenVMS/Windows NT jest zastosowanie warstwowej architektury klient/serwer pozwalającej użytkownikom na wykorzystanie wszystkiego co najlepsze w obu systemach wciąż zachowując pełną swobodę wyboru systemu docelowego w przyszłości. Użytkownicy mogą stosować tanie, szeroko rozpowszechnione aplikacje Microsoft i rosnące możliwości przetwarzania w systemie OpenVMS Digitala.

Architektura klient/serwer daje możliwość stworzenia środowiska, w którym użytkownicy mogą pracować w systemie OpenVMS tak długo jak jest to dla nich korzystne i przejść do systemów NT lub UNIX w momencie, w którym uznają to za stosowne. Oprogramowanie warstwy pośredniczącej (middleware) nie pozwala użytkownikom wpaść w tarapaty.

Kiedy rozwiązania własne są wartościowe i nie prowadzą w ślepą uliczkę, wtedy należy je stosować.

Niewątpliwie największą konkurencję w latach 90...tych dla VMS stanowią systemy UNIX, których otwartość podkreśla się na każdym kroku. Czy VMS, którego otwartość i zgodność z licznymi standardami jest nie do podważenia ma szanse w tej konkurencji?

W całej rozciągłości OpenVMS jest najlepszym systemem do realizacji niezawodnych, długotrwałych obliczeń i przetwarzania na wielką skalę. Gartner Group i inne instytucje analityczne rynek komputerowy porównywały OpenVMS z systemami UNIX i Tandem. OpenVMS uzyskał zdecydowanie najwyższą notę.

W sytuacji gdy wymagane jest przetwarzanie 24 godziny na dobę przez 365 dni w roku OpenVMS ma bardzo niewielu konkurentów.

W sytuacji gdy wymagane jest wyjątkowo intensywne przetwarzanie transakcyjne OpenVMS nie ma żadnych konkurentów.

Ponadto, współpraca z Windows NT firmy Microsoft jest wystarczającym warunkiem istnienia i rozwijania się OpenVMS. Oba systemy zostały zaprojektowane przez tych samych konstruktorów, tak że mają one ze sobą wiele wspólnego. W przedsiębiorstwach, które używają systemu OpenVMS jest dużo łatwiej wprowadzić Windows NT, gdy zajdzie taka potrzeba. Przeszkolenie programistów jest również

Wersja 6 systemu VMS dawała możliwość łączenia w tym samym klastrze komputerów VAX i Alpha

W sytuacji gdy wymagane jest wyjątkowo intensywne przetwarzanie transakcyjne OpenVMS nie ma żadnych konkurentów

*Jeśli
potrzebujesz
systemu
pracującego
24 godziny
na dobę przez
365 dni
w roku
to system
OpenVMS
jest zalecany*

*Jeśli
wymagasz
szybkiego
i złożonego
przetwarzania
transakcyjnego (TP)
OpenVMS
jest dobrym
wybohem*

znacznie łatwiejsze niż w przypadku systemu UNIX. Ponieważ, zaś OpenVMS jest zwykle systemem działającym w skali całego przedsiębiorstwa, a nie poszczególnych operacji biznesowych nie stanowi bezpośredniej konkurencji dla Windows NT. Stąd oba systemy są wykorzystywane w pełnej harmonii.

Czy Digital, który oferuje dwa, a nawet trzy typy systemów operacyjnych o podobnej funkcjonalności nie stwarza sytuacji tzw. wewnętrznej konkurencji, wpływającej negatywnie na wielkość ich sprzedaży?

Jeśli ktoś chce działać w dłuższej perspektywie w zakresie serwerów, a w roku 2000 planuje rozszerzenie biznesu, to strategia rozwijania trzech systemów operacyjnych staje się oczywista. Oszacowania wykazują, że 40% serwerów będzie działać z systemem UNIX, 40% z systemem Windows NT, a 20% z innymi systemami. OpenVMS mieści się w tych ostatnich. Sprzedając serwery z procesorami Intel i Alpha znajdujemy się w doskonałej sytuacji, aby przejąć dużą część rynku NT. Digital UNIX jest najlepszym systemem UNIX jaki znam dlatego wierzę, że przechwycimy również dużą część rynku systemów UNIX. Byłoby nonsensem nie skorzystać z nadarzającej się okazji przechwycenia dużego segmentu rynku gdy mamy na to szansę. I tu dochodzimy do sprawy systemu OpenVMS. W roku 1996 udział tych systemów w ogólnych obrotach będzie znaczny, a zyski będą stanowiły 50%. Teraz jest oczywiste, że nikt kto współpracuje z Digitem w tym zakresie nie będzie sugerował rezygnacji z takiego interesu.

Wes Melling zakłada stopniowe zmniejszanie się obrotów z tytułu sprzedaży systemu OpenVMS, ale już znacznie mniejszy spadek zysków ponieważ te systemy będą instalowane w znacznym procencie silnych, zaawansowanych serwerów.

To wszystko o czym mówiłem zobowiązuje każdego pracownika Digitala, a w szczególności handlowców do inteligentnego traktowania wszystkich trzech ważnych systemów operacyjnych. Nie można sobie pozwalać na uproszczone widzenie każdego z nich, a zwłaszcza wyróżnianie jednego kosztem pozostałych. To wymaga od handlowców niezwykle profesjonalnego podejścia i zrozumienia, że te systemy są różne i różnie mogą być stosowane w zależności od specyficznych warunków napotkanych u klientów.

Jeśli jednak funkcjonalność systemów - VMS,

Digital Unix i Windows NT - jest odmienna, to jakie podstawowe cechy odróżniają je od siebie? Gdzie można stosować je najlepiej?

O tym najlepiej powiedzieliby specjaliści takiej klasy jak Wes Melling i Don Harbert.

Ja mogę tylko dodać, że obecnie zazwyczaj użytkownicy stają wobec jednej lub dwóch możliwych technologii. Digital natomiast daje do wyboru najlepszą możliwość. Najlepszy system własny, najlepszy UNIX i najlepszy system NT. Najlepsze rozwiązanie w zakresie NT jest prawdą ponieważ Digital jest oceniany jako wiodący dostawca systemów NT (powołując się na porozumienie ACE z firmą Microsoft).

Jeśli użytkownik nie ma początkowo żadnych preferencji może na podstawie kilku prostych stwierdzeń wypracować sobie decyzję.

1. Jeśli potrzebuje systemu pracującego 24 godziny na dobę przez 365 dni w roku to system OpenVMS jest zalecany.
2. Jeśli wymaga szybkiego i złożonego przetwarzania transakcyjnego (TP) OpenVMS jest dobrym wyborem.
3. Jeśli potrzebuje dużych, wydajnych baz danych to jedynym rozwiązaniem jest Digital UNIX.
4. Jeśli jednak wydajności bazy danych ma towarzyszyć niezawodne i sprawne środowisko należy najpierw rozważyć OpenVMS bazą Oracle Rdb lub, być może, Oracle w wersji 7. Opcja wielkiej pamięci (VLM64) jest obsługiwana przez OpenVMS od początku 1996 r.
5. Jeśli spotykamy się z sytuacją współpracy wielu takich samych systemów to powinniśmy zdecydować się na Windows NT na platformie Intel lub Alpha. Na pewno oszczędzimy wiele milionów dolarów decydując się na tani system, który równocześnie jest dość wydajny i daje możliwości szybkiej modernizacji.
6. Jeśli użytkownik chce modernizować wykorzystywany system OpenVMS powinien zainteresować się strategią Affinity, która zapewnia ewolucyjne i bezpieczne przejście do współpracy z innymi systemami.

Obecnie Digital oferuje już 64-bitową wersję VMS, a także całkowicie zrekonstruowany system plików. Co jeszcze istotnego wprowadzono w nowym projekcie VMS?

W dodatku do 64-bitowej architektury i no-

wego systemu plików OpenVMS w swoim jądrze posiada cechę realizowania wątków upraszczając i zwiększając wydajność systemów SMP (o przetwarzaniu symetrycznym). Programiści tworzący aplikacje stosujący interfejs DE-Cthread osiągają wiele korzyści angażując niewiele sił w programowanie równoległe. Wewnętrzne zmiany systemu zapewniają wzrost wydajności wejścia/wyjścia, zwłaszcza w odniesieniu do baz danych. Specyficzny interfejs optymalizuje obsługę funkcji we/wy. Ta cecha umożliwiła osiągnięcie przez bazę Oracle Rdb niezwyklego rekordu wydajności TPC-A i TPC-C niewiarygodną szybkość archiwizowania (1 terabajt w ciągu 8 godzin). W systemie zintegrowano też pełnię kilka dodatkowych serwisów związanych z protokołem TCP/IP.

Jak rysuje się przyszłość systemu VMS na tle niezwykle dynamicznie rozwijającego się systemu Unix?

OpenVMS jest niezwykle zaawansowanym systemem operacyjnym, który równocześnie dopuszcza korzystanie z rozprzestrzeniających się aplikacji pisanych dla systemów Windows i Windows NT. OpenVMS będzie dalej rozwijany dopóki Windows NT nie będzie zapewniał tych samych możliwości. Mówi się, że stosunkowo łatwo można w systemie Windows NT osiągnąć 50% możliwości, które daje system OpenVMS w zakresie klastrów. Osiągnięcie 75% wszystkich możliwości, które obecnie daje OpenVMS jest możliwe, ale bardzo trudne. Jednak jak wszędzie pozostałe 10% jest praktycznie nie do osiągnięcia.

Digital UNIX jest na najlepszej drodze do osiągnięcia doskonałej niezawodności funkcjonowania w ciągły sposób nawet w okresach wieloletnich. Wykorzystanie tutaj technicznej doskonałości zespołu rozwijającego OpenVMS powoduje, że Digital UNIX wyprzedzi zdecydowanie wszystkie konkurencyjne systemy UNIX. Przewiduję, że następny raport Gartner Group uwidoczni zdecydowaną przewagę systemu Digital UNIX nad systemami innych producentów i uplasuje go już niedaleko OpenVMS.

Jest pan człowiekiem znajdującym się w samym centrum prac nad systemami operacyjnymi Digitala. W jaki sposób udaje się firmie zapewnić zgodną pracę nad wszystkimi trzema typami systemów operacyjnych?

Pracuję w grupie prowadzonej przez Wesa Mellinga, która w tej chwili rozwija koncepcję OpenVMS/Windows NT Affinity. Jestem przekonany, że wsparcie techniczne dla wszystkich krajów i

partnerów biznesowych w tym zakresie zakończy się sukcesem. Ponieważ Affinity integruje prawie całe istniejące oprogramowanie, musi też zapewniać interfejs z systemami tradycyjnymi.

Na poziomie technicznym pełnię rolę integratora. Chodzi mi przede wszystkim o to, aby klient dostał to co jest dla niego dobre. Oznacza to przede wszystkim dostarczanie odpowiednich informacji technicznych handlowcom i służbom wspierającym proces sprzedaży. Oznacza to także iż muszę zawsze być pod ręką i nie lekceważyć nikogo, ani żadnego produktu. Czasami jest to trudne.

Uczestniczyłem w procesie, który prowadził do przejmowania przez Digital UNIX części rynku systemu OpenVMS gdzie z kolei w zakresie nowych aplikacji był faworyzowany IBM. Wiedziałem, że rozwiązania Digitala są doskonałe, a najlepszą opcją jest Digital UNIX. Życie udowodniło prawdziwość moich przypuszczeń, które tak zaskoczyły kierownictwo korporacji, że musiało ono zweryfikować własne długoterminowe plany w zakresie technologii.

Musimy troszczyć się o naszych klientów. Osiągniemy sukces jeśli klienci dostaną rozwiązania konkurencyjne. Jeśli natomiast klient wyczuwa, że chcemy osiągnąć sukces jego kosztem wtedy zaczynamy ponosić straty.

Na zakończenie chcielibyśmy przybliżyć panu sylwetkę naszym czytelnikom. Proszę o garść informacji na swój temat oraz innych działań pozazawodowych.

Cieszy mnie narciarstwo alpejskie, jazda samochodem i uprawa Bonsai (miniaturowych drzewek japońskich).

Zostałem szefem zespołu rozwijającego VMS dla systemów Alpha, a potem szefem produkcji tego systemu dla komputerów Alpha. Zajmując się przez szereg lat tym systemem prowadziłem zespoły rozwijające koncepcję klastrów, mechanizmów we/wy oraz jakości i testowaniem systemu VMS dla maszyn Alpha.

Przez 1,5 roku uczestniczyłem także w pracach zespołu prowadzącego działalność marketingową systemu OpenVMS, skupiając się przede wszystkim na informowaniu o strategii Affinity i szkoleniach dotyczących tej tematyki. Gdy zaś jeszcze mam trochę czasu zajmuję się konsultacjami dla klientów, którzy pragną wytyczać strategiczne kierunki swojego działania.

Dziękujemy za wywiad.

*Digital UNIX
jest na
najlepszej
drodze do
osiągnięcia
doskonałej
niezawodności
funkcjonując
w ciągły
sposób nawet
w okresach
wieloletnich*

*Musimy
troszczyć się
o naszych
klientów*



Najpotężniejszy ośrodek Digitala w Europie

14 marca Digital Equipment Corporation otworzył nowy Ośrodek Technologiczny Digital/SAP, w którym zespoły inżynierów z obu firm zajmować się będą opracowywaniem najlepszych wdrożeń systemu R/3 dla zastosowań w różnych gałęziach gospodarki. Ośrodek Technologiczny znajduje się w biurach firmy Digital w Viernheim, niedaleko Walldorf - siedziby firmy SAP AG.

Nowy ośrodek wyposażono w siedem komputerów AlphaServer 8000 zawierających w sumie 70 procesorów Alpha. Wraz z 28 GBajdami pamięci operacyjnej, 1000 GBajdami pamięci dyskowej i bardzo szybkimi połączeniami FDDI tworzą zaawansowane środowisko umożliwiające przeprowadzanie różnych testów.

„Jesteśmy bardzo zadowoleni z tej inwestycji i zaangażowania firmy Digital. 64-bitowa technologia Alpha jest dla nas bardzo ważna. Od 2 lat Digital jest jedynym dostawcą w pełni 64-bitowego systemu operacyjnego dla naszego systemu R/3. Architektura 64-bitowa umożliwia zastosowanie olbrzymiej pamięci operacyjnej. A to się cieszy szczególnym zainteresowaniem dużych klientów pragnących zainstalować system, z którego będą korzystać tysiące użytkowników”, powiedział Christa Koppe, szef ośrodka kompetencyjnego firmy SAP.

„Ta inwestycja jest kolejnym silnym sygnałem o ścisłej współpracy Digital/SAP”

stwierdził Toni Steiner, szef działu współpracy Digital z SAP. „Klienci decydujący się na system R/3 chcą być pewni, że technologia będzie wspomagać ich operacje gospodarcze i umożliwiać będzie ciągły wzrost aplikacji odpowiadający wzrostowi ich organizacji. Jest to szczególnie ważne dla dużych firm, które muszą przebudować swoją strukturę. Chcą one stabilnej i perspektywicznej struktury sprzętowej dla systemu SAP. Dzięki nowemu ośrodkowi będą mogły one oszacować swoje wymagania pod względem sprzętu” dodał Steiner.

Technologia AlphaGeneration

Bazą nowego ośrodka są komputery AlphaServer 8000. Niedawno, Digital opublikował najlepszy na świecie wynik szybkości działania systemu SAP R/3 zostawiający daleko w tyle pozostałych dostawców sprzętu dla systemu R/3.

Moc i jakość platformy AlphaServer 8000 była ostatnio zaprezentowana w niezależnych testach TPC-C. Używając najszybszy na świecie mikroprocesor Alpha 21164 - 350 MHz osiągnięto na komputerze AlphaServer 8400 5/350 wynik 11,456 transakcji na minutę (tpmC) przy cenie 268 dolarów /tpmC. Rezultat ten osiągnięto dla systemu operacyjnego Digital UNIX z bazą danych Oracle.

„Jest to najszybszy komercyjny serwer, prawie 2 razy wydajniejszy niż najbardziej wydajny komputer konkurencji. Stosunek ceny do wydajności jest również najlepszy dla komputerów tej klasy” zauważyła Pauline Nist, wi-

ceprezydent ds. serwerów.

Podczas otwarcia ośrodka, reprezentujący bazodanowych partnerów firmy Digital, Fritz Dustehaus, szef Oracle odpowiedzialny za współpracę z SAP stwierdził: *„64-bitowa architektura w połączeniu z opcją VLM (bardzo duża pamięć) ukazuje znaczną przewagę nad tradycyjnymi środowiskami 32-bitowymi. Opcja VLM umożliwia przechowywanie w pamięci operacyjnej znacznie większych części baz danych. Wielogigabajtowe pamięci operacyjne pozwalają na znacznie szybsze sięganie do danych i rzadsze odwoływanie się do dysków, a w rezultacie dają znaczący wzrost szybkości działania aplikacji”*.

W lutym, po raz drugi z rzędu magazyn Datamation wyróżnił jakość i wydajność systemów AlphaServer nadając komputerowi AlphaServer 8400 miano „Wybrane przez czytelników w 1996 roku”. Poprzednio tytuł ten uzyskał AlphaServer 2100.

Także w lutym, Digital zebrał większość laurów w testach porównawczych prowadzonych przez niezależną instytucję - AIM Technology. Listę 11 nagrodzonych komputerów Digitala otwierał AlphaServer 8400 5/350, nazwany najwydajniejszym wśród systemów wieloużytkownikowych/wielodostępnych kosztujących od 150.000 dolarów.

Digital i SAP w systemie Windows NT

14 marca Digital Equipment Corporation i SAP AG ogłosiły, że system SAP R/3 będzie dostępny dla rodziny komputerów AlphaServer

pracujących nie tylko z systemem operacyjnym UNIX, ale także z Windows NT firmy Microsoft. Digital razem z SAP AG, Oracle i Microsoft będzie dostarczał jako pierwszy rozwiązanie R/3 na systemie operacyjnym Windows NT pracującym na mikroprocesorach typu RISC. Użytkownicy będą mogli wybierać pomiędzy bazami danych Oracle 7 Enterprise Server i Microsoft SQL Server. Planuje się możliwość zakupu tych rozwiązań na połowę 1996 roku.

Coraz większa liczba klientów jest zainteresowana wykorzystywaniem systemu operacyjnego Windows NT w swoich projektach wdrażania R/3. Do tej pory na komputerach z procesorami Intel wdrażano systemy tylko w małych i średnich przedsiębiorstwach. Teraz dzięki dostępności systemu R/3 dla komputerów AlphaServer z Windows NT, firma Digital rozszerzyła możliwości implementacji systemu SAP R/3 na Windows NT także dla dużych, złożonych organizacji gospodarczych.

Komputery AlphaServer dzięki zastosowaniu w nich najszybszego na świecie i pierwszego, który wykonuje ponad miliard operacji na sekundę (1000 MIPS) procesora Alpha 21164 umożliwiają uzyskanie wielkiej wydajności aplikacjom pracującym na Windows NT. Dodatkowo, aplikacje R/3 także uzyskują większą wydajność korzystając z bardzo dużych pamięci dostępnych jedynie w komputerach AlphaServer.

„Bazując na platformie Alpha dostarczanej przez Digital możemy rozszerzyć nasz

Najszybszy rezultat TPC na pojedynczej maszynie

Digital opublikował ostatnio rekordowy rezultat testu TPC-C na pojedynczej maszynie. Konfiguracja AlphaServer 8400 5/350 z systemem OpenVMS v7.1, bazą Oracle Rdb v7.0 oraz monitorem transakcyjnym DEC ACMS v4.1 osiągnęła rezultat ponad 14.227 tpm-C przy koszcie 269\$/tpm. Dotychczasowy rekord należał również do Digitala (13.646 tpm-C, 277\$/tpm, Digital UNIX, Informix, Tuxedo). Platforma OpenVMS-OracleRdb-ACMS po raz kolejny potwierdziła swą dominację w kategorii systemów dla przedsiębiorstw.

rynek dla średnich i dużych organizacji gospodarczych sięgając po większe implementacje R/3 na Windows NT. Takim rynkiem są na przykład tysiące użytkowników systemu operacyjnego VMS, którzy obawiają się systemu UNIX, a są zainteresowani Windows NT", powiedział Paul Wahl, Dyrektor Generalny SAP Ameryka.

Digital ma najszerszą ofertę certyfikowanych rozwiązań dla systemu SAP R/3: Windows NT dla komputerów z procesorami INTEL i Alpha oraz UNIX dla procesorów Alpha.

Toni Steiner, szef działu współpracy Digital z SAP, powiedział: „Nasi klienci obecnie mają niezwykle szeroki wybór platform dla systemu R/3. Jeżeli zdecydują się na Windows NT, mogą wybrać serwery Priors z procesorami Intela dla małych i średnich projektów lub serwery z procesorem Alpha dla średnich i dużych przedsięwzięć. Unikalna dla Windows NT jest łatwość zmiany platformy na komputery Alpha gdy serwery PC osiągną kres swoich możliwości.”

Dla rynku użytkowników systemu operacyjnego UNIX, Digital demonstrował swoją przewagę już wielokrotnie. Niedawno, firma opublikowała najlepszy na świecie wynik szybkości działania systemu SAP R/3 (1,700 użytkowników modułu SD dla baz danych Oracle i ADABAS-D firmy Software AG).

Digital w bombowcu B-2

Na początku bieżącego roku oddział korporacji Northrop Grumman zajmujący się budową bombowca strategicznego B-2 podpisał umowę z Digital Equipment Corporation wartą 4.7 miliona dolarów.

Nowy system zostanie zainstalowany w laboratorium pracującym nad niedostrzeżalnym bombowcem B-2

(Stealth Bomber) w zakresie symulacji lotów, badań w czasie rzeczywistym i integracji systemów lotniczych oraz sterowania lotem.

Komputery Alpha zapewnią niezbędną moc obliczeniową laboratorium, w którym opracowuje się aplikacje realizujące zadania pilota i analizuje się wydajność systemów lotniczych takich jak radar, nawigacja, sterowanie i komunikacja z pilotem. Laboratorium opracowuje i wdraża również systemy sterujące symulacją lotu.

W ramach kontraktu zostaną dostarczone stacje robocze AlphaStation 200 4/266 i AlphaStation 250 4/266 oraz serwery AlphaServer 8400 i AlphaServer 8200, a także zaawansowane terminale graficzne Multia, systemy przełączające GIGAswitch/ATM oraz urządzenia StorageWorks integrujące pamięci masowe.

Northrop Grumman jest znaną firmą zbrojeniową zlokalizowaną w Los Angeles. Firma opracowuje, projektuje i produkuje bombowce, myśliwce oraz inne konstrukcje lotnicze do zastosowań wojskowych i cywilnych, a także sterujące nimi systemy

elektroniczne i informatyczne.

Miliardowy kontrakt Digitala

W marcu Digital Equipment Corporation ogłosił, że 64-bitowe stacje robocze AlphaStation zostały wybrane przez Siły Lotnicze USA do realizacji kontraktu obsługiwanego przez Hughes Data Systems. Digital jako poddostawca dostarczy co najmniej 37.000 stacji roboczych Alpha. Kontrakt na dostawę części stacji roboczych dla lotnictwa wojskowego USA równolegle wygrała firma Sun Microsystems.

Wartość wszystkich wydatków poniesionych przez Siły Lotnicze może wynieść ponad \$956 milionów dolarów w ciągu 7 lat. Kontrakt dotyczy stacji roboczych, urządzeń peryferyjnych, oprogramowania i usług. Stronę US Air Force reprezentuje Ośrodek Systemów Elektronicznych Sił Lotniczych USA mieszczący się w bazie Hanscom w Bedford, Massachusetts.

Stacje robocze Digitala zostaną sprzedane z najnowszą, czwartą, wersją Digital UNIX. Wśród wielu nowych cech tego systemu można wyróżnić

podwyższoną wydajność i skalowalność, ulepszony interfejs użytkownika, lepszą wymienność aplikacji i zwiększone możliwości pracy w sieci.

MACIEJ A. MARKOWSKI
maciej.markowski
@rpw.mts.dec.com



z kraju

AlphaStudio - marzenie realizatora telewizyjnego...

28 marca br. realizatorów programów telewizyjnych zaprosiliśmy do krainy marzeń. Digital zaprezentował AlphaStudio Broadcast Server - cyfrowy system archiwizowania i emitowania obrazu. Dzięki wykorzystaniu sprzętu AlphaGeneration wyposażonego w superszybkie szyny transmisji danych i skalowalne zasoby dyskowe Alpha



Microsoft i Digital trzymają się razem - szefowie obu oddziałów podpisują umowę

Studio jest idealnym rozwiązaniem dla stacji telewizyjnych w Stanach Zjednoczonych i Europie. Drugim elementem rozwiązania jest AlphaStudio Production Server pozwalający na dowolne zabawy z animowaniem obrazów, kreowaniem efektów specjalnych i „bezszywowym” (słowo ze słownika montażysty) łączeniem obrazów. Tyle jeżeli chodzi o cyfrową obróbkę obrazu. Nieodłącznym elementem AlphaStudio jest archiwum skompensowanych i cyfrowo przetworzonych taśm filmowych. W modelu idealnym taśma bezpośrednio z produkcji jest wprowadzana do AlphaStudio, następnie przechodzi do archiwum i praktycznie nigdy już nie będzie użyta. Sprawa bezpieczeństwa takiego systemu nie pozostawia żadnej wątpliwości. Zabezpiecza go niezawodne rozwiązanie TMCluster. AlphaStudio jest systemem skalowalnym. Dla stacji telewizyjnej oznacza to, że może rozpocząć inwestycję od rozwiązania bazowego i z czasem je rozwijać.

Microsoft - Digital Polskie następstwa światowego aliansu

29 marca br. Dyrektor Generalny Microsoft Polska Waldemar Sielski i Dyrektor Generalny Digital Polska Marek Racieski podpisali dokument określający zasady współpracy na terenie Polski. Podobnie jak na całym świecie obie firmy współdziałać będą w dziedzinie usług informatycznych, integracji systemów, marketingu i sprzedaży. W praktyce oznacza to, że polski odbiorca najnowszych produktów obu firm będzie mógł korzystać z wiedzy i doświadczenia specjalistów z Microsoftu i Digitala wdrażających technologię Microsoft BackOffice na serwerach AlphaGeneration. Na terenie Polski obowiązują również wcześniejsze umowy:

- Microsoft Solution Provider Partner

- Microsoft Solution Provider Authorized Support Center
- Microsoft Solution Provider Authorized Technical Education Center
- Microsoft Large Account Reseller.

Digital Equipment Polska jest jedynym partnerem Microsoftu posiadającym wszystkie powyższe tytuły.

Na konferencji prasowej pokazywano również najnowszy model komputera przenośnego HiNote Ultra II - najcieńszego wśród komputerów typu notebook - grubość jego nie przekracza 25 mm, waga 1.8 kg, a w środku... może być procesor Pentium 133 MHz, 40 MB pamięci, dysk 1.35 GB. Dzięki wymiarom nie przekraczającym kartki formatu A4 doskonale się mieści w teczce każdego biznesmena.

Las Vegas na Wołoskiej

2 kwietnia br. w biurze Digitala na ul. Wołoskiej gościł Bill Gates. Nie osobiście i nie wirtualnie, a jedynie na telewizyjnym ekranie. Szef Microsoft korzystając z satelitarnej łączności Digitala zaanonsował z Las Vegas wprowadzenie do sprzedaży Microsoft Exchange Server. System zawiera pocztę elektroniczną oraz wbudowane oprogramowanie grupowe pozwalające na ustalenie harmonogramu prac dla zespołu, prowadzenie konferencji, wspólną pracę nad dokumentami i zabezpieczenia pozwalające wykorzystywać sieć internetową. Pierwszym klientem był sam Microsoft.

Komputery i artyści - Infosystem '96

W dniach 14-17 kwietnia br. Poznań w swoich halach targowych gościł komputery i komputerowych fanów. Digital Equipment Polska został zaproszony na wspólne stoisko SAP i jego partnerów.

Nie było to działanie przypadkowe. System SAP R/3 działający na 64 bitowej platformie AlphaGeneration osiągnął rekord testowy wynoszący 1700 standardowych użytkowników SD. O tym, że jest to jedyna 64 bitowa platforma sprzętowa R/3 wspominał już tylko z kronikarskiego obowiązku. Digital i SAP zawarły strategiczne przemyśle mające na celu przede wszystkim wspólne rozwiązywanie problemów klientów poprzez rozwój centrów kompetencyjnych. Digital jest również największym klientem firmy SAP. Decyzję o wyborze systemu SAP R/3 podjęto w grudniu 1993 roku. Kierowano się, poza obniżeniem kosztów działalności i poprawieniem wydajności również uniwersalnością systemu SAP R/3 zapewniającą odpowiednie wspomaganie przedsięwzięć międzynarodowych. Nie bez znaczenia była oczywiście platforma sprzętowa AlphaGeneration, której architektura doskonale spełnia wymogi skalowalności systemu. Pierwszym etapem wdrażania systemu było stworzenie idealnego modelu i wybór 1800 różnych aplikacji, które zostały przeznaczone do wymiany. Jako bazę danych wybrano rozwiązanie proponowane przez Oracle, a do kontroli funkcjonowania systemu sieciowego NetView. Konfiguracja składa się z czterech serwerów pracujących w klastrze, komunikujących się z klientami za pomocą superprzełączników GigaSwitch. Bezpieczeństwo systemu zapewnia zdublowanie konfiguracji. Swoimi doświadczeniami z procesu wdrażania dzielił się na konferencji prasowej towarzyszącej targom Clemens Clamer. Najpoważniejsze problemy stwarzała jednak zmiana sposobu myślenia użytkowników. System działa w Digitalu od 12 lutego br.

Do Poznania pojechaliśmy z dwoma serwerami AlphaGeneration. Na jednym AlphaServer 1000 działał system R/3 w wersji 3.0B obsłu-

gujący wszystkie komputery klienckie na stoisku SAP. Jest to całkowicie polonizowana, najnowsza wersja systemu. Dodatkowo prezentowaliśmy własne rozwiązania programowe stanowiące pomost pomiędzy platformą maszyn i urządzeń kontrolujących proces produkcji ciągłej a modulem PP-PI SAP R/3 obsługujący planowanie procesu produkcji i obrazujący na monitorach jej przebieg. Do tego celu służy nakładka PDAS. System BASEstar natomiast integruje miniprogramy sterujące pracą poszczególnych urządzeń. Wersja demo tego systemu była tak interesująca, że firma SAP odstąpiła nam własny czas prezentacyjny na jej sesjach programowych.

Targi to jednak nie tylko czas ekspozycji, rozmów biznesowych i seminariów. Każde targi to jeszcze duża ilość imprez towarzyszących. Organizatorzy tegorocznego Infosystemu przygotowali na tę okazję nie lada gratkę. Była nią przepiękna wystawa kompozycji fotograficznych Ryszarda Horowitza. Na specjalnym pokazie autor zdradzał sekrety swojej pracy. Nie ukrywał, że do tworzenia cyfrowych archiwów fotograficznych i wielokrotnego przetwarzania obazu używa komputerów o wielkich zasobach pamięci i skomplikowanych programów animacyjnych. Podkreślał jednak, że technika nie może zastąpić osobowości artysty, jego wyczucia piękna i dobrego smaku.

Na Targach spotkałam jeszcze jednego artystę. Był nim Ginter wystawowy nadworny kucharz SAPa. W jego małej, ale doskonale wyposażonej kuchni powstawały prawdziwe dzieła sztuki. Szaszłyki, gulasze, lizanie i sałatki mogły zadowolić najbardziej subtelne gusta.

Drogi SAPie, nasz Partnerze, dzięki Ci za to, że masz nie tylko doskonały system...

Magdalena
Poklewska-Koziełł
magda.poklewska@rpw.mts.dec.com

AlphaServer 4100

- nowy oręż Digitala

6 maja firma Oracle, strategiczny partner Digitala, ogłosiła kolejny fantastyczny rekord w teście TPC-C, który został osiągnięty dla cztero-procesorowego serwera w klasie komputerów przeznaczonych do obsługi oddziałowej - 6056 transakcji na minutę przy koszcie jednej transakcji 223 \$. Rekordowy rezultat osiągnięto posługując się 64-bitową bazą Oracle7 Server, którą Digital obecnie oferuje wraz z najnowszym serwerem AlphaServer 4100 i systemem operacyjnym Digital UNIX.

Warto przeczytać komentarze przedstawicieli obu firm na temat nowego rekordu ponieważ jego osiągnięcie bez najnowszego komputera byłoby po prostu niemożliwe. *"Osiągnięcie rekordu w teście TPC-C dowodzi, że Oracle już teraz proponuje niewiarygodnie szybko przetwarzanie transakcyjne (OLTP) na systemach średniej wielkości. Takie wyniki wraz z niepodważalnymi korzyściami jakie daje zastosowanie 64-bitowej bazy Oracle7 we wspomaganiu procesu decyzyjnego zapewniają rozwiązania, z którymi dzisiaj nikt w przemyśle komputerowym nie może się równać. Dla obu firm - Oracle i Digital - stwarza to niezwykłą okazję dostarczania ekonomicznie efektywnych 64-bitowych rozwiązań już na poziomie środowiska oddziałowego w dużych organizacjach gospodarczych"* stwierdziła Beatriz Infante wiceprezydent Oddziału Systemów Otwartych firmy Oracle.

Rekord został osiągnięty na nowej 4-procesorowej maszynie AlphaServer 4100 5/400 wyposażonej w pamięć o pojemności 4 GB. *"Oracle i Digital znowu w dramatyczny sposób podniosły poprzeczkę w dziedzinie skalowalności systemów. Wyniki testów wskazują na utrwalające się przodownictwo obu firm w dziedzinie systemów 64-bitowych, których stosowanie jest obecnie uzasadnione ekonomicznie już na poziomie oddziałów dużych organizacji gospodarczych i przedsiębiorstw"* powiedziała Pauline Nist, wiceprezydent Digitala ds. serwerów.

Jaki jest AlphaServer 4100?

Digital wprowadzając na rynek AlphaServer 4100 zainicjował kolejną po serwerach 2100 i 8000 rewolucję tym razem w zakresie serwerów średniej wielkości. Nowy serwer stanowi most pomiędzy rodzinami serwerów 2100 i 8000



zblizając się wymiarami do tych pierwszych równocześnie zachowując wiele cech systemowych, zwłaszcza opcję działania na wielkiej (ponad 2GB) pamięci, rodziny 8000. Systemy AlphaServer 4100 charakteryzują się następującymi możliwościami:

- pamięć o pojemności do 4GB. Jest to pierwsze w serwerach średniej wielkości wdrożenie przez Digital opcji wielkiej pamięci operacyjnej (VLM64). Opcja VLM64 pozwala stosować niezwykle szybkie i wydajne aplikacje bazo-danowe po przystępnej cenie;
- do czterech, obecnie najszybszych 64-bitowych procesorów Alpha o częstotliwości zegara 300 Mhz, wymieniających już w lipcu na jeszcze szybsze 400 Mhz;
- wydajna (z korekcją błędów), 1,1 GB/sek. szyna systemowa, dostosowująca automatycznie swoją szybkość do typu procesora, z którym współpracuje;
- dwie niezależne, 64-bitowe szyny PCI o przepustowości 250 MB/sek. każda. Szczytowa przepustowość szyny może dochodzić do 500 MB/sek. znakomicie zwiększając pasmo podsystemu we/wy i balansując całkowitą wydajność komputera;
- unikalne wdrożenie koncepcji Modułów Systemowych umożliwia elastyczne konfigurowanie systemów dla potrzeb użytkownika;
- zaawansowane funkcje zarządzania całością systemu i jego zasilaniem.

**6056
transakcji
na minutę
przy koszcie
jednej
transakcji
223 \$**

*W konstrukcji
wszystkich
modeli rodziny
AlphaServer
4100
zastosowano
unikalną
konceptję
Modułu
Systemowego*

	2100	4100	Efekty
Mikroprocesor	EV5	EV5/EV5.6	30 - 40% szybszy
Szybkość procesora	250/300MHz	300/400MHz	
Liczba procesorów	4	4	
Pamięć podręczna na procesor	8K/I + 8K/D 96K / I-D 4MB	8K/I + 8K/D 96K / I-D 0, 2MB, 4MB	2 razy więcej
Wydajność TPM	3200	>6000 transakcji	
Max. pamięć	2GB	4GB	2 razy więcej
Szyba systemowa	0,67 GB/sek	1,1 GB/sek	84% więcej
Szyba we/wy	0,12 GB/sek	0,44 GB/sek	4 razy więcej
Gniazda PCI	8 (4+4 32b)	8 (4+4 64b)	
Typ szyny we/wy	PCI, EISA (3)	PCI, EISA (3 Combo)	
Wewnętrzne kieszenie	14+2	21	

AlphaServer 4100 - zaawansowana konstrukcja

Rodzina serwerów 4100 składa się z trzech komputerów. Modele AlphaServer 4100 5/300 i 5/300E napędzane są zegarem o częstotliwości 300 Mhz. Jedyną różnicą jest brak zewnętrznej pamięci podręcznej w modelu 5/300E, która o pojemności 2MB występuje w modelu 5/300. Najbardziej zaawansowany jest model 5/400 działający z procesorem Alpha EV5.6 o częstotliwości 400 MHz i pamięcią podręczną o pojemności 4MB. Dla modelu 5/300 istnieje możliwość wymiany procesora na szybszy i utworzenia modelu 5/400. Wszystkie modele rodziny AlphaServer 4100 są oferowane w stojącej obudowie niskiej lub wysokiej z systemami Windows NT, Digital UNIX lub OpenVMS.

W konstrukcji wszystkich modeli rodziny AlphaServer 4100 zastosowano unikalną koncepcję Modułu Systemowego, która umożliwia łatwe konfigurowanie i rozwijanie całego serwera. Moduł Systemowy zawiera do czterech procesorów tego samego typu, do 4GB pamięci operacyjnej, 8 gniazd PCI oraz do trzech zasilaczy. Wszystko poza pamięciami masowymi i innymi urządzeniami we/wy znajduje się w Module Systemowym.

Z Modułów Systemowych składa się różne modele montowane w niskiej lub wysokiej obudowie. System w obudowie niskiej zawiera jeden Moduł Systemowy i nie więcej niż trzy urządzenia StorageWorks, natomiast w obudowie wysokiej dwa Moduły Systemowe i do sześciu urządzeń StorageWorks. W każdym Module Systemowym może działać inny system operacyjny. Model montowany w obudowie wysokiej może więc w jednym module realizować aplikacje działające w systemie Windows NT, w drugim zaś unixowe.

Miejsce rodziny AlphaServer 4100

Został on zaprojektowany jako kontynuacja serwera AlphaServer 2100, który osiągnął w

ciągu ostatnich dwóch lat niebywały sukces na świecie. Nowy serwer ma o 40% szybsze procesory, dwa do czterech razy większą pamięć i czterokrotnie większą przepustowość kanałów we/wy. Mając takie możliwości serwer 4100 zbliżył się do rodziny najbardziej wydajnych super-serwerów Digitala - AlphaServer 8000 zdecydowanie wyprzedzając konkurencyjne produkty mieszczące się w tej samej klasie. Zwłaszcza opcja VLM umożliwiającą korzystanie z pamięci o pojemności ponad 2GB, której nie posiada konkurencja wyróżnia serwer 4100.

Głównymi konkurentami nowego serwera Digitala są komputery serii 9000 K420 firmy HP, SS1000E firmy Sun, RS6000 J30 IBM oraz Proliant 4500 P5/166 firmy Compaq.

Jednakże chociaż serwery firm HP i Sun działają z 64-bitowymi procesorami to ich 64-bitowe, rzeczywiste systemy operacyjne w dalszym ciągu pozostają w sferze obietnic - być może będą dostępne pod koniec tego roku. Te systemowe perypetie stwarzają użytkownikom HP-UX (UNIX HP) i Solaris (UNIX Sun) wiele kłopotów opóźniając wdrożenia oprogramowania użytkowego. Ponadto użytkownicy systemów HP z niepokojem obserwują losy porozumienia pomiędzy firmami Intel i HP, którego wynikiem ma być powstanie wspólnego procesora o architekturze "Tahoe".

Digital posiada także dużą przewagę w zakresie architektury magistrali we/wy. Serwer 4100 jest wyposażony w otwartą, standardową magistralę PCI/64, podczas gdy zarówno Sun, jak i HP dysponują własnymi, "zamkniętymi" rozwiązaniami szyn. Jest to powodem kłopotów z implementacją systemu NT na tych maszynach. Podobno obie firmy usilnie pracują nad projektami magistral we/wy bazujących na koncepcji szyny PCI.

Serwer Proliant firmy Compaq jest wyposażony tylko w szynę EISA. Jednakże modele zawierające procesor P6 będą posiadały również szynę PCI.

System to podstawa (dla każdego coś miłego)

Rynkiem oprogramowania ciągle wstrząsają wybuchy nowych technologii i produktów. Niemal co roku pojawiają się fale produktów, które na pewien czas dominują rynek. Co najmniej od kilkunastu lat (od kiedy pojawiły się pecety) jesteśmy bombardowani wiadomościami o coraz lepszych i nowszych wersjach oprogramowania, bez którego nie możemy się obejść. Mieliśmy więc fale procesorów tekstu, arkuszy kalkulacyjnych, baz danych, narzędzi czwartej generacji, poczty elektronicznej i Internetu. W całym tym zgiełku jakby na uboczu znajdują się systemy operacyjne. Jeszcze kilka lat temu wydawało się, że chociaż tutaj sytuacja jest jasna. Pecety to Windows a maszyny wielodostępne, serwerowe to UNIX. Jednak pojawienie się Windows NT zburzyło ten klarowny obraz. Stoimy więc znów wobec nowych wstrząsów związanych z wojną jaką będzie toczyć Windows NT z całym światem UNIXa. Co wobec tego mamy począć, jak w tym wszystkim zachować zdrowy rozsądek i wybierając platformę systemową, wybrać to co dla nas najlepsze na dziś, przy zagwarantowaniu korzyści z naszej inwestycji na kilka lat do przodu? Cykl artykułów, które prezentujemy Państwu w niniejszym numerze Digital Forum nie daje na to odpowiedzi. Pomoże natomiast zorientować się w tym, co jest obecnie najważniejsze w technologii systemów operacyjnych i w jakim kierunku się one rozwijają. Digital jako jedyna firma oferuje na jednej platformie sprzętowej trzy systemy operacyjne do wyboru. Można powiedzieć 'dla każdego coś miłego'. Każdy z systemów posiada zestaw cech, które powodują iż może on być idealny dla określonych wymagań użytkowników. O tym, który wybrać decyduje nasz klient.

Pierwszy tekst cyklu dotyczy systemu operacyjnego OpenVMS. OpenVMS jest jednym z niewielu systemów operacyjnych tzw. proprietary, który ma się dobrze na rynku. Zawdzięcza to temu, iż oferując bogatą funkcjonalność (niespotykaną wśród innych systemów) jest systemem otwartym na współdziałanie z innymi systemami i zgodnym ze standardami. Zasadniczą zaletą OpenVMS jest jego wysoki stopień niezawodności i skalowalności udostępniany poprzez technologię klastrów.

Drugi artykuł opisuje program Affinity ogłoszony wspólnie przez Microsoft i Digital. Jest to przykład w jaki sposób dwie duże firmy o całkiem odmiennych korzeniach dążą do rozwiązań, które zapewniają klientom pełną integrację dużych systemów produkcyjnych działających w przedsiębiorstwie z nowymi technologiami klient/serwer.

W kolejnym artykule znajdziecie Państwo opis najnowszej wersji systemu Digital UNIX. Digital UNIX jest pierwszym w pełni 64 bitowym systemem UNIXowym. Zalety tego systemu to jego zgodność ze standardami (o czym świadczy sama nazwa), bogata funkcjonalność i łatwość obsługi dzięki nowemu środowisku graficznemu i zestawowi gotowych narzędzi.

Na koniec specjalność Digitala, czyli klastry. Do niedawna była to technologia, którą oferowaliśmy jedynie na platformie OpenVMS. Obecnie technologia ta dostępna jest dla Digital UNIXa, a już niedługo pojawi się pierwsza wersja dla Windows NT. Decydujące znaczenie dla rozwoju klastrów ma technologia Memory Channel, o której dowiecie się Państwo z lektury artykułu.

piotr.sobolewicz @rpw.mts.dec.com
artur.stefanowicz @rpw.mts.dec.com

Piotr Sobolewski
Artur Stefanowicz

System Operacyjny OpenVMS v7.0

*Mając
500.000
instalacji
i 10mln
użytkowników
OpenVMS
pozostaje
nadal jedną
z czołowych
platform
systemowych*

*OpenVMS jest
pozycjonowa-
ny jako wyso-
kowskądajna
i bezpieczna
platforma
serwerowa w
3-warstwowej
architekturze
klient/serwer*

System operacyjny OpenVMS jest dobrze znany naszym Czytelnikom (patrz „OpenVMS - wzorcowy system operacyjny”, Maciej Modrzejewski, DECforum rok 2, nr 7). Dlatego też niniejszy artykuł skupia się niemal wyłącznie na nowych cechach funkcjonalnych dostępnych w wersji 7.0 oraz dalszej strategii rozwoju produktu.

OpenVMS

System OpenVMS powstał w roku 1977 jako uniwersalny, wielozadaniowy i wielodostępny 32-bitowy system operacyjny dla procesorów VAX. Nowością było wówczas wprowadzenie pełnej implementacji pamięci wirtualnej (stąd nazwa systemu Virtual Memory System). Bogata funkcjonalność, przystawowa wręcz niezawodność i pewność działania sprawiły, że system OpenVMS był przez długi czas jedną z dominujących platform systemowych. Wiele rewolucyjnych idei, powstałych na OpenVMS znalazło trwałe miejsce na rynku komputerowym. Jedną z nich jest technologia typu cluster, w której od początku OpenVMS pełni wiodącą rolę.

Era systemów „otwartych” dokonała poważnej erozji w rynkowej pozycji OpenVMS. Mimo stałego wzbogacania funkcjonalności, implementacji wszystkich ważniejszych standardów i realnej przewagi technologicznej nad konkurencyjnymi rozwiązaniami, Digital nie potrafił prowadzić skutecznego marketingu na miarę tego produktu. Hasło „UNIX” górowało nad technologią.

Strategia rozwoju

Lekcja „Systemów Otwartych” zaowoco-

wała jednak wyjątkową koncentracją wysiłków nad wyposażeniem OpenVMS we wszelkie możliwe mechanizmy ułatwiające współpracę systemu z innymi platformami oraz implementacją światowych standardów. OpenVMS był jednym z pierwszych systemów (także UNIX-owych), który uzyskał certyfikat XPG (obecnie XPG4). Jest zgodny ze specyfikacjami SPEC 1170 i POSIX, NIST FIPS 151-2, OSF/Motif R1.2.3 (zawiera X Windows System X11 R5), OSF DCE 1.0.3, implementuje TCP/IP i OSI. Może to zabrzmieć niewiarygodnie, ale OpenVMS jest bardziej standardowym UNIX-em niż wiele renomowanych implementacji tego systemu.

Obecnie, z ponad 500.000 instalacji i 10mln użytkowników OpenVMS pozostaje nadal jedną z czołowych platform systemowych. Raporty porównawcze niezależnych firm konsultingowych (Gartner, IDC) jednomyślnie przyznają mu prymat technologiczny. To co skłania Klientów w stronę OpenVMS nie zmieniło się w ciągu ostatnich 25 lat: nadal decyduje niezawodność, wydajność, bezpieczeństwo - najmocniejsze strony OpenVMS. Strategia rozwoju systemu stara się wykorzystać właśnie te atuty. OpenVMS jest pozycjonowany jako wysokowydajna i bezpieczna platforma serwerowa w 3-warstwowej architekturze klient/serwer. Ma funkcjonować w środowisku heterogenicznym i obsługiwać kluczowe dla instytucji zastosowania. Program Affinity, opisany w artykule Piotra Sobolewskiego „Affinity - łączy to co najlepsze”, jest jednym z filarów tej strategii. Ścisła integracja z Windows NT, nacisk na jeszcze pełniejszą integrację ze środowiskiem Desktop przy jednoczesnym utrzymaniu prymatu najbezpieczniejszej platformy serwerowej - oto główne kierunki rozwoju OpenVMS na najbliższe lata.

Nowe cechy funkcjonalne wersji v7.0

Wersja v7.0 jest jedną z najważniejszych udoskonaleń systemu w ciągu ostatnich lat wnosząc kilka ważnych i nowoczesnych technologii.

Na uwagę zasługują szczególnie:

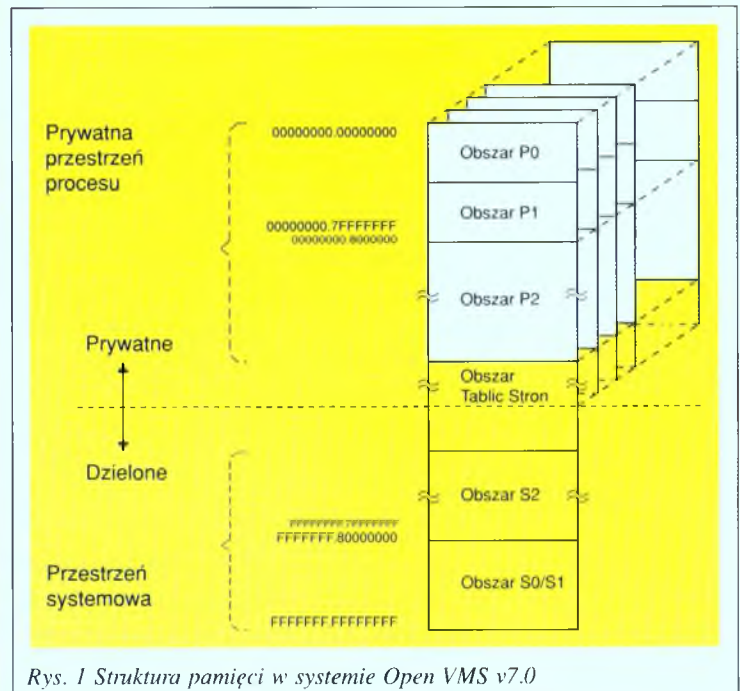
- 64-bitowa przestrzeń adresowa, umożliwiająca pełne wykorzystanie możliwości procesora Alpha.
- System plików nowej generacji, o wydajności ok. 10-krotnie większej od obecnie stosowanych.
- Wątki procesów na poziomie jądra systemu, skutecznie zwiększające wydajność aplikacji, w szczególności w zastosowaniach klient/serwer.

64-bitowa przestrzeń adresowa

Pomimo 64-bitowych możliwości adresowych procesora Alpha, OpenVMS do wersji v7.0 pozostawał systemem 32-bitowym. Oznaczało to, że każdy proces systemu operacyjnego mógł posiadać co najwyżej 4GB pamięci wirtualnej¹. Zastosowania typu Dużych Baz Danych (VLDB, pojemności przekraczającej 100GB) czy Magazynów Danych (Data Warehousing) charakteryzują się jednak wyjątkową zdolnością absorbowania każdego dostępnego Mabajta pamięci. Spektakularne osiągnięcia systemu Digital UNIX w testach TPC-C (30.390 tpmC, 305\$/tpm z bazą Oracle7 VLM), w pełni 64-bitowego systemu operacyjnego wspartego 8-14GB pamięcią fizyczną dowodzą, jak wielkie znaczenie ma wykorzystanie szerokiej przestrzeni adresowej i buforowania możliwie dużych porcji danych w pamięci.

Co oznacza 64-bitowa przestrzeń adresowa w wersji v7.0:

- Nowa struktura pamięci systemu, wprowadzająca dwa nowe obszary: P2 i S2 (patrz Rysunek 1).
- Drajwery urządzeń I/O wykorzystują 64-bitowe adresowanie np. drajwery dyskowe, taśmowe, LAN, mailbox.
- System plików Files-11/XQP umożliwia



Rys. 1 Struktura pamięci w systemie Open VMS v7.0

wykorzystanie 64-bitowych adresów buforów.

- RMS umożliwia wykorzystanie 64-bitowych adresów buforów.
- Rozszerzenie API polegające na dostosowaniu istniejących funkcji do mieszanej adresacji 32/64-bitowej oraz wprowadzenie nowych funkcji operujących wyłącznie na adresach 64-bitowych.
- Pełna zgodność z trybem 32-bitowym dla nieuprzywilejowanego kodu, minimalne modyfikacje dla kodu uprzywilejowanego.
- Fizyczna adresacja do 8TB RAM, powyżej - musimy poczekać na nową generację układów Alpha i pojemniejsze układy pamięci...
- Środowisko programistyczne, np. DEC C, DEC C RTL, debager przystosowane do adresacji 64-bitowej.

W skrócie oznacza to pełne przystosowanie OpenVMS do obsługi ekstremalnie wymagających aplikacji takich jak np. Data Warehousing. W ślad za możliwościami systemu operacyjnego powinny iść aplikacje czyniące z tych możliwości pożytek. Firma ORACLE ogłosiła już nową wersję systemu Oracle Rdb7 dla OpenVMS Alpha, wyposażoną w mechanizm VLM (Very Large Me-

*Myślą
przewodnią
było
stworzenie
systemu
plików
o radykalnie
odmiennym
od tradycyjne-
go profilu
wydajności*

*O ile
tradycyjne
systemy
plików są
zoptymalizo-
wane pod
kątem opera-
cji odczytu,
nowy system
plików powi-
nien optymalizować opera-
cje zapisu*

mory) i gotową do obsługi baz danych z buforowaniem ponad 14GB w pamięci.

Nowy system plików - Spirallog

System plików to coś więcej niż tylko fizyczna organizacja urządzeń dyskowych, a więc struktura bloków, sposób przechowywania wskaźników do plików czy reprezentacja wolnych bloków. System plików definiuje wszystkie aspekty zapisu, przechowywania i odczytu danych z nośników, udostępniając do tego celu mechanizmy kontrolne, struktury danych, funkcje systemowe i wiele innych elementów.

Praktycznie wszystkie współczesne systemy plików na platformach UNIX, MVS, VM, OS/2, MS-DOS a nawet Windows NT oparte są na modelach z lat 60/70-tych. Można je wspólnie określić mianem modeli blokowych, jako że reprezentują plik w postaci zestawu bloków rozrzuconych na dysku. Każda operacja odczytu pliku wymaga odszukania i skompletowania bloków tworzących plik. Każda modyfikacja oznacza odnalezienie właściwego bloku i wymianę na nowy. Dla uproszczenia, pomijam tutaj mechanizm odwzorowania logicznych bloków pliku do fizycznych bloków dysku.

Nowa idea systemu plików powstała na przełomie lat 80/90 na Uniwersytecie Berkeley w ramach projektu Sprite². Myślą przewodnią było stworzenie systemu plików o radykalnie odmiennym od tradycyjnego profilu wydajności. Założenia projektu określone zostały z myślą o perspektywie następnych 10 lat, jako efekt analizy tendencji rynkowych (patrz Tabela 1).

Jednym z najważniejszych wyników analizy była obserwacja o stałym przesuwaniu się środka ciężkości profilu I/O z operacji odczytu na operacje zapisu. Rozprzestrzenienie technologii buforowania (cache) zarówno pod kątem wielkości jak i ceny powoduje, że liczba faktycznych operacji I/O sięgających do urządzenia staje się coraz bardziej zredukowana. Dotyczy to szczególnie operacji odczytu. Przykładowo, przyjmuje się, że stosunek operacji odczytu do zapisu na platformie OpenVMS v5.x kształtował się na poziomie 80:20. W wersji v6.1, z wykorzystaniem buforowania, kontrolerów HSJ40 i dysków RZ28/29, stosunek ten zbliża się do 50:50. Tak więc większość operacji odczytu może być zrealizowanych na pewnym poziomie bufora. Jednak zapis do dysku, ze względu na integralność danych nie może być wyeliminowany w analogiczny sposób. Należy więc przypuszczać, że profil I/O przyszłych systemów dyskowych zostanie zdominowany przez zapis. O ile więc tradycyjne systemy plików są zoptymalizowane pod kątem operacji odczytu, nowy system plików powinien optymalizować operacje zapisu, zakładając, że odczyt i tak zostanie obsłużony przez szybki bufor.

Inne spostrzeżenie dotyczyło sposobów organizacji zasobów dyskowych. Na pierwszy plan wysuwa się tu tendencja łączenia tanich dysków SCSI w macierze RAID, posiadające zupełnie odmienną charakterystykę wydajnościową od pojedynczych urządzeń. Podczas gdy pojedyncze operacje I/O o niewielkiej objętości w obu przypadkach wykazują zbliżoną wydajność, to w przypadku wielu operacji I/O lub ich dużej objętości, macierze są zdecydowanie szybsze. Nowy system pli-

Tabela 1	lata 80.	lata 90. i dalej
Wydajność	Dyski 1G Ograniczone buforowanie	Dyski 100G Ogromne bufory
Skalowalność	od 1GB do 1TB adresowanie 32-bitowe	100TB adresowanie 64-bitowe
Niezawodność/ dostępność	archiwizacja off-line administracja off-line	praca non-stop 24x365 administracja on-line
Integralność danych	rekonstrukcja po awarii w trybie off-line	nieakceptowalna rekonstrukcja off-line

ków powinien uwzględniać tę ważną charakterystykę macierzy dyskowych.

Kronikowy system plików

W oparciu o powyższe spostrzeżenia, powstał projekt kronikowego systemu plików (log-structured file system). Posiada on formę kroniki, z zapisem wyłącznie w trybie dopisania (patrz Rysunek 2). Każda forma informacji: nowe dane, dane zmodyfikowane, metadane są dopisywane na końcu kroniki. Mechanizm ten praktycznie eliminuje narzut konwencjonalnych systemów plików związany z szukaniem bloku na dysku - w systemie kronikowym każdy zapis realizowany jest pod adresem logicznego końca kroniki, niemal zawsze „pod głowicą” i jako szybki, sekwencyjny zapis wielu bloków jednocześnie. Jest to bardzo istotna poprawa, okazuje się bowiem, że aż ok 85% czasu operacji I/O jest tracone w klasycznych systemach plików na pozycjonowanie głowicy. Ponadto, zapis sekwencyjny ma szybkość zbliżoną do maksymalnej wydajności urządzeń dyskowych.

No tak, ale co z odczytem? Przecież od czasu do czasu chciałoby się także coś przeczytać! A wówczas system będzie zmuszony odszukiwać odpowiednie bloki na dysku tracąc

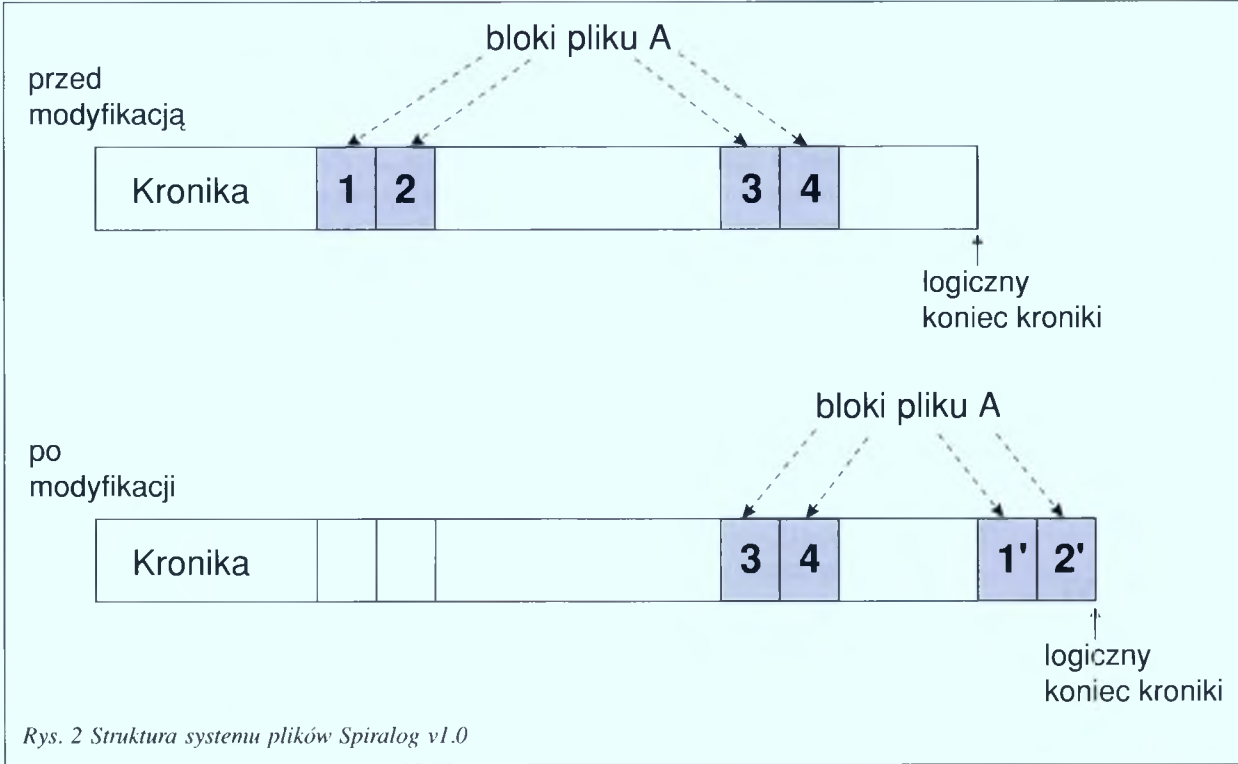
przy okazji mnóstwo czasu. Wielopoziomowe mechanizmy buforowania systemu operacyjnego, kontrolera macierzy czy ewentualnie samego dysku praktycznie eliminują większość operacji odczytu. Na przykład, system operacyjny OpenVMS, począwszy od wersji v6.0 udostępnia mechanizm Virtual I/O Cache (VIOC) istotnie redukujący liczbę fizycznych operacji I/O.

Jeśli operacji odczytu nie uda się obsłużyć na pewnym poziomie bufora, kronikowy system plików kolekcjonuje bloki pliku lokalizując je zapomocą wydajnego mechanizmu indeksów B-drzewa.

Zasada funkcjonowania kroniki

Rysunek 2 ilustruje proces modyfikacji pliku A w kronikowym systemie plików. Dwa zmodyfikowane bloki A1, A2 zapisywane są na koniec kroniki, natomiast oryginalne bloki stają się przestarzałe. Naturalnie, kronikowy system plików musi być wyposażony w mechanizm zbierania takich śmieci (garbage collection), przepisujący kronikę i odzyskujący nieużywane bloki. Wymaga to pewnych nakładów, ale struktura kroniki pozwala na realizację tego procesu jako szybkich sekwencyjnych operacji I/O. Przy obecnych szybkościach CPU, uwaga procesora

Digital rozpoczął prace nad nowym, kronikowym systemem plików przed czterema laty co zaowocowało produktem Spirallog v1.0



Rys. 2 Struktura systemu plików Spirallog v1.0

*W systemie
kronikowym
każdy zapis
realizowany
jest niemal
zawsze
„pod
głowicą”
jako
szybki,
sekwencyjny
zapis wielu
bloków
jednocześnie*

*Projekt Spira-
log zakłada,
że klientami
mogą być
różne systemy
operacyjne,
między innymi
UNIX,
MS-Windows
oraz
Windows NT*

poświęcona na odzysk śmieci jest minimalna. Najczęściej „odkurzacz” kronikowego systemu plików (jak np. Spirallog) działa jako proces w tle, aktywowany tylko wtedy, gdy liczba wolnych bloków na dysku spada poniżej ustalonego limitu.

Spirallog

Digital rozpoczął prace nad nowym, kronikowym systemem plików przed czterema laty co zaowocowało produktem Spirallog v1.0. Jest to obecnie druga implementacja kronikowego systemu plików na rynku (inne rozwiązanie to system FAServer firmy Network Appliance Corp.). Obok wymienianych celów wydajnościowych i maksymalego wykorzystania zalet RAID, zespół projektowy Digital'a miał także na uwadze:

- 64-bitowe adresowanie,
- pracę w konfiguracji VMSccluster,
- zgodność z wcześniejszym systemem plików Files-11/XQP,
- współpracę z oprogramowaniem Volume Shadowing, Disk Striping,
- skalowalność,
- wysoką dostępność (odporność na awarie).

Ogólną architekturę systemu Spirallog prezentuje Rysunek 3.

Model agent/serwer

Spirallog łączy najlepsze cechy architektury klient/serwer z zaletami środowiska VMSccluster oferując model o roboczej nazwie agent/serwer. Aplikacja klienta systemu Spirallog komunikuje się z jego agentem, działającym na tej samej maszynie, co klient (patrz Rysunek 4). Zadaniem agenta jest buforowanie danych i wstępne przetwarzanie żądań wszystkich klientów na danej maszynie oraz współpraca z serwerem Spirallog. Serwer działa na dowolnej maszynie konfiguracji VMSccluster i jest odpowiedzialny za fizyczne operacje I/O. Możliwe jest zdefiniowanie serwerów zapasowych na innych maszynach, które przejmą sterowanie na wypadek awarii maszyny dotychczasowego serwera Spirallog. Taki model przetwarzania umożliwia równomierne rozłożenie obciążenia na dostępne maszyny konfiguracji

VMSccluster (procesy agentów) oraz osiągnięcie wysokiej dostępności i odporności na awarie systemu Spirallog (serwery zapasowe)

Procesy agentów Spirallog wykorzystują mechanizm DLM (Distributed Lock Manager) do synchronizacji wzajemnych operacji.

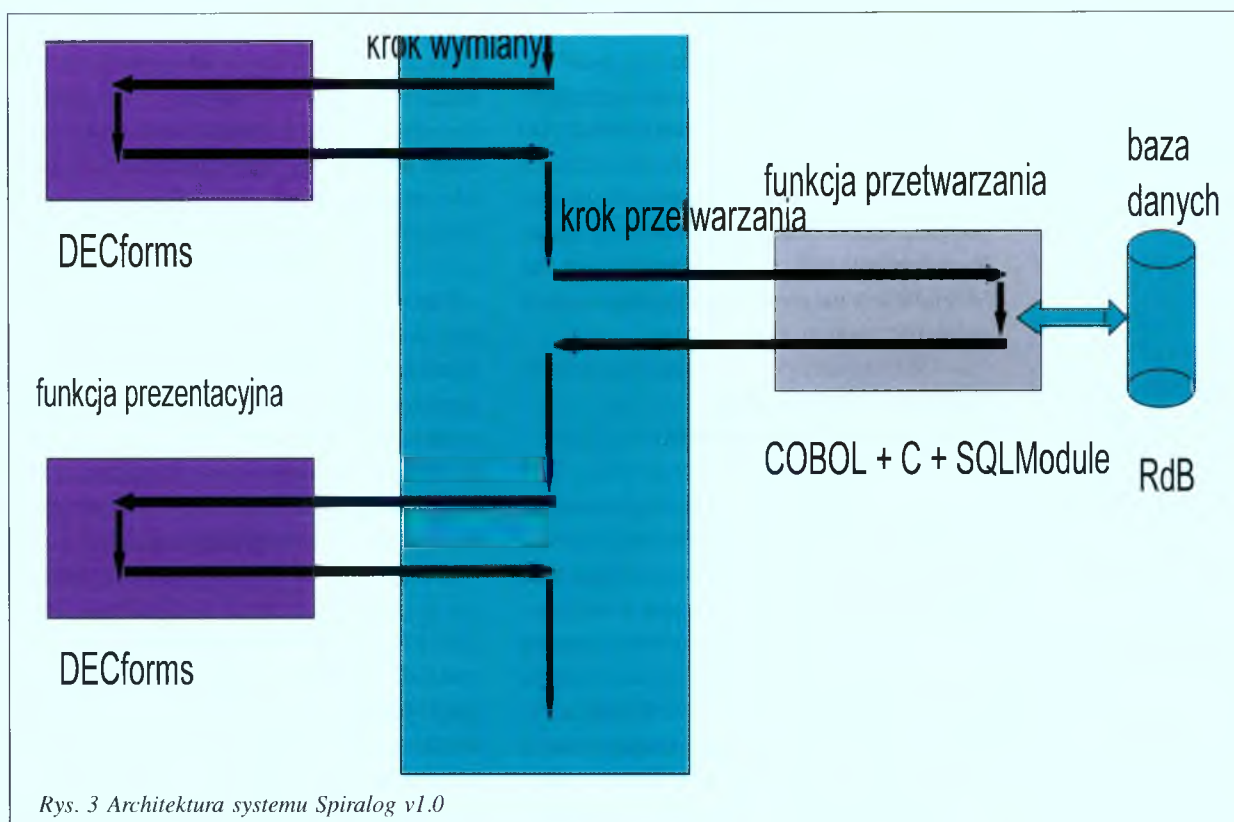
Wielu klientów

Projekt Spirallog zakłada, że klientami mogą być różne systemy operacyjne, między innymi UNIX, MS-Windows oraz Windows NT. Wysoki stopień niezawodności konfiguracji VMSccluster sprawia, że dla tych systemów Spirallog stanowi atrakcyjny mechanizm składowania i archiwizacji danych. Dzięki stosowanemu modelowi klient/agent, konkretna implementacja komunikacji między klientem a Spirallog nie stanowi problemu. Dzięki oprogramowaniu PATHWORKS dostęp do Spirallog będzie możliwy z platform MS-DOS, MS-Windows, Windows95, Windows NT i innych. Wykorzystując natomiast DFS (Distributed File System) lub NFS (Network File System) system OpenVMS/ Spirallog może stać się serwerem plików dla świata systemów UNIX-owych (patrz Rysunek 3).

Archiwizacja

Operacje archiwizacji i odtwarzania (backup/restore) w środowiskach produkcyjnych decydują o sprawności i dostępności systemów. Powinny być szybkie a co najważniejsze, działać w trybie on-line, czyli bez konieczności wyłączania systemu. W systemie Spirallog obie czynności mogą być realizowane w trybie on-line a ich szybkość jest jedną z głównych zalet systemu wielokrotnie wyprzedzając wyniki konwencjonalnych rozwiązań.

W wielu systemach plików operacja archiwizacji wymaga zaprzestania wykorzystywania urządzenia na czas backup'u. Dzieje się tak dlatego, że zapis zmian na dysk przez użytkowników w trakcie procesu archiwizacji utrudnia (a w wielu systemach wręcz uniemożliwia) stwierdzenie, czy dana modyfikacja miała miejsce przed rozpoczęciem archiwizacji, czy też w trakcie, a więc czy ma trafić do archiwum czy nie. Jest to kluczowe



zagadnienie dla integralności danych. Jasne jest, że zablokowanie dostępu do dysku w czasie archiwizacji istotnie obniża dostępność systemu.

Ponieważ Spiralog dopisuje modyfikacje danych zawsze na koniec kroniki, reszta danych w kronice nigdy nie ulega zmianie. Dlatego też Spiralog może realizować operację archiwizacji praktycznie w każdej chwili, po prostu kopiując zawartość kroniki aż do ustalonego punktu, stanowiącego migawkę stanu systemu plików. Szybkość archiwizacji Spiralog jest wysoka, rzędu 1TB/8h z niewielkim obciążeniem CPU. Odtwarzanie zasobów z archiwum może dotyczyć całości lub wybranych plików.

Archiwizacja inkrementalna

Możliwe jest również wykonywanie inkrementalnej archiwizacji i odtwarzania. Ponieważ wybrany punkt w kronice Spiralog reprezentuje określony moment czasu, możliwe jest zarchiwizowanie wszystkich zmian zaszłych w systemie od tamtej chwili po prostu poprzez skopiowanie zawartości kroniki od wybranego punktu. W ten sposób unikamy konieczności archiwizowania każ-

dorazowo całości urządzenia. Spiralog pozwala nam wykonać do 128 częściowych archiwizacji. Ciekawą cechą jest także możliwość scalania kilku archiwów wykonanych w różnym czasie w jedno archiwum. Ułatwia to zarządzanie archiwami (zawsze mniej do zgubienia) oraz przyspiesza ewentualne odtwarzanie zasobów.

Funkcje archiwizacji i odtwarzania Spiralog są ściśle zintegrowane z narzędziami zarządzania zasobami środowiska POLY-CENTER.

Szybki restart

Osoby znające środowisko UNIX'a wiedzą dobrze, co oznacza awaria dysku lub maszyny - rekonstrukcję systemu plików za pomocą niesławnego fsck, czyli kosztowną czasowo operację³.

Sposób organizacji kroniki Spiralog umożliwia radykalną redukcję czasu niezbędnego na weryfikację i rekonstrukcję systemu plików po awarii. Ponieważ każdy punkt w kronice reprezentuje pewien moment czasu, wystarczy cofnąć się do momentu, co do którego mamy pewność, że określa popraw-

Spiralog może realizować operację archiwizacji praktycznie w każdej chwili

*System
OpenVMS/
Spiralog
może stać się
serwerem
plików
dla świata
systemów
UNIX-owych*

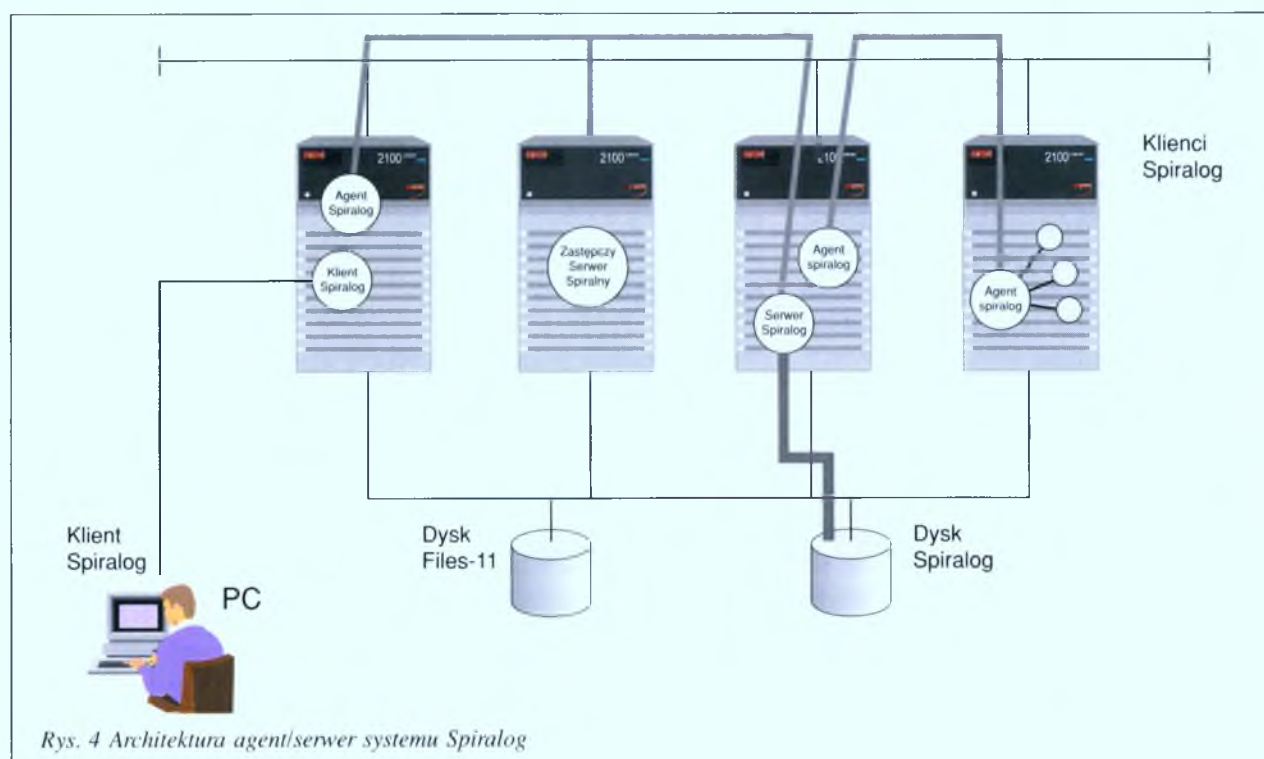
ny stan systemu plików. Aby to uzyskać, co pewien czas do kroniki trafia specjalny rekord zwany punktem kontrolnym (checkpoint). W trakcie restartu wystarczy więc cofnąć się do najbliższego punktu kontrolnego, co może się odbyć błyskawicznie. Dane eksperymentalne wskazują, że wolumen Spiralog potrzebuje 5s na restart po awarii serwera. Możliwe jest indywidualne skonfigurowanie redukujące restart nawet do 1s.

Technika zapisu w tle i mechanizm buforowania

System Files-11 stosuje strategię modyfikacji natychmiastowych: operacje zapisu trafiają najpierw na dysk, a dopiero potem odzwierciedlone zostają w buforach. Gwarantuje to bardzo wysoką niezawodność systemu plików i do minimum redukuje czas restartu po awarii. Systemy UNIX-owe działają zwykle odmiennie. Starają się optymalizować I/O poprzez buforowanie modyfikacji a fizyczny zapis do dysku odwołują do operacji „wymiatania” (*flush*) bufora. Prowadzi to w konsekwencji do znanych problemów konieczności rekonstrukcji systemu plików po awarii. Spiralog stosuje obie techniki. Agent buforuje modyfikacje zachodzące w ramach

danej maszyny. Utrzymuje porządek zmian i prowadzi politykę konsolidacji serii modyfikacji, o ile jest to możliwe. W ten sposób, dopiero ostateczny rezultat serii modyfikacji trafia do serwera, który zamiast wielu zapisów realizuje jeden. Technika ta nazywana jest zapisem w tle (*write-behind*).

Buforowanie modyfikacji, aczkolwiek cennie z perspektywy wydajności, może być niedopuszczalne dla pewnych zastosowań z uwagi na możliwość utraty danych w przypadku awarii. W przeciwieństwie do konwencjonalnych systemów plików, Spiralog nie wymaga jednak czasochłonnej rekonstrukcji - wystarczy cofnąć logiczny koniec kroniki do ostatniego punktu kontrolnego. Nie unikniemy jednak utraty tych modyfikacji, które pozostały w buforze maszyny. Dlatego Spiralog daje administratorowi możliwość decyzji, jaką strategię buforowania przyjmuje dla poszczególnych plików. Istnieje też możliwość programowej kontroli trybu buforowania z poziomu kodu programu. Dzięki temu serwer Netscap'a może z powodzeniem korzystać z buforowania w tle, natomiast baza danych żądać trybu zapisu natychmiastowego dla zapewnienia integralności danych.



Zgodność z Files-11/XQP

Zgodność z istniejącym systemem plików OpenVMS była jednym z kluczowych wymagań stawianych przed Spiralog. W efekcie, istniejące aplikacje pracujące z Files-11 wykorzystują Spiralog bez konieczności dokonywania jakichkolwiek zmian. Spiralog zachowuje pełną zgodność z Files-11 na poziomie podsystemu RMS oraz \$QIO(). Migracja jest wręcz banalna, wymaga zaledwie inicjacji urządzenia pod system Spiralog i skopiowanie danych istniejących z urządzenia Files-11. Ponadto, Files-11 i Spiralog mogą współzysztować w ramach jednej maszyny lub środowiska VMScluster.

Skalowalność

Spiralog umożliwia zarządzanie on-line zasobami o pojemności do 10TB. Dzięki 64-bitowej architekturze może obsługiwać bardzo duże pliki i wolumeny dyskowe, np. 100 GB zestawy RAID. Szacuje się że Spiralog może dokonywać zapisu 10GB nowych danych dziennie.

Dostępność produktu

Spiralog v1.0 jest dostępny dla systemu OpenVMS Alpha v7.0, licencje zawarte są w ramach licencji systemu operacyjnego, wymagany jest jedynie nośnik (płyta CD). Dla maszyn VAX możliwe jest wykorzystywanie Spiralog w mieszanych środowiskach VMScluster.

Wielowątkowość w trybie kernel

Wątki (threads) to niezależne konteksty wykonania kodu działające w ramach jednego procesu. Umożliwiają wydajne zwiększenie wydajności programów poprzez większą równoległość przetwarzania. Czy nie można byłoby osiągnąć identycznego efektu za pomocą kilku procesów? Niezupełnie. Wątki dzielą wspólną przestrzeń adresową, nie wymagają więc specjalnych mechanizmów komunikacji między-procesowej jak np. pamięć dzielona czy mailbox. Ponadto, wątki pochłaniają znacznie mniej zasobów systemowych niż procesy. Proces określa mapę pamięci wirtualnej, posiada informacje auto-

Szybkość archiwizacji Spiralog jest wysoka, rzędu 1TB/8h z niewielkim obciążeniem CPU



Rys. 5 Szata graficzna środowiska CDE na Open VMS v7.0

*Spiralog
umożliwia
zarządzanie
on-line
zasobami
o pojemności
do 10TB*

*Wersja v7.0
przynosi
kolejne
wzbogacenie
oferty opro-
gramowania
sieciowego z
zauważalnym
konsekwent-
nym
naciskiem
na usługi
TCP/IP*

ryzacyjne, statystykę. Wątki przechowują kontekst CPU, posiadają własne stopy i na podstawie indywidualnych priorytetów są szeregowane przez system w dostępie do CPU. Na maszynach SMP możliwe jest równoczesne wykonywanie kilku wątków danego procesu na różnych procesorach.

Wielowątkowość w ramach procesu można było dotychczas uzyskać na platformie OpenVMS dwojako: poprzez wykorzystanie AST (Asynchronous System Trap, czyli asynchronicznych procedur) lub za pomocą pakietu DECthreads (implementacja standardu POSIX 1003.1c, pthreads). Jednak bez jawnej ingerencji systemu operacyjnego takie formy emulacji wątków posiadały wiele ograniczeń. Dopiero OpenVMS v7.0 wprowadzając wątki w trybie kernel umożliwia pełną implementację tej ważnej technologii. Dla zachowania zgodności z istniejącym oprogramowaniem, jako API dla obsługi wątków pozostała biblioteka DECthreads operująca bezpośrednio na wątkach systemu.

Fast I/O

Jest to zestaw nowych usług systemowych, umożliwiających osiągnięcie wyższej wydajności operacji I/O. Stanowią one alternatywę dla standardowej funkcji I/O OpenVMS, SYS\$QIO(). Procedury Fast I/O posiadają 64-bitowy interfejs, początkowo oferując bezpośredni dostęp do urządzeń dyskowych i taśmowych. Idea Fast I/O polega na realizacji zakończenia operacji I/O na poziomie IPL (Interrupt Priority Level) 8, nie zaś na poziomie IPL 4 z dodatkowym AST w trybie kernel, uzyskując dzięki temu istotną redukcję czasu CPU. Dla maszyn jednoprosesorowych zysk jest rzędu 18% w porównaniu z klasycznym SYS\$QIO().

Usługi TCP/IP

Wersja v7.0 przynosi kolejne wzbogacenie oferty oprogramowania sieciowego z zauważalnym konsekwentnym naciskiem na usługi TCP/IP. Chyba najciekawszym objawem zmian zachodzących w tej dziedzinie jest odejście od wymagania protokołu DECnet dla oprogramowania VMScluster - obecnie możliwe jest pełne skonfigurowanie VMScluster na TCP/IP.

Usługi TCP/IP dostępne są na platformie OpenVMS v7.0 poprzez produkt *Digital TCP/IP Services for OpenVMS v4.0* implementujący standard BSD 4.3 i udostępniający między innymi:

- oprócz standardowych protokołów TCP/IP także RIP i BOOTP;
- serwery BIND i United;
- agenta SNMP;
- API w postaci gniazdek, RPC z mechanizmami XDR i Portmapper'a oraz SRIQIO i QIO;
- aplikacje Telnet, FTP, rlogin, rsh, rexec, LPD, SMTP;
- NFS v2.0.

DECnet/OSI v7.0

Wraz z OpenVMS v7.0 pojawia się także nowa wersja oprogramowania DECnet/OSI, umożliwiająca wykorzystanie 64-bitowej adresacji. Także w tym przypadku dostrzec można dominującą rolę TCP/IP - DECnet/OSI v7.0 umożliwia uruchamianie aplikacji wykorzystujących macierzysty interfejs DECnet na transporcie TCP/IP, czyli tunelowanie DECnet w TCP/IP. Podobnie w przypadku aplikacji OSI, które także mogą być uruchamiane na TCP/IP. DECnet/OSI v7.0 jest też zintegrowany z serwisami lokalizacyjnymi DNS/BIND. Ważną i zapewne długo oczekiwaną zmianą jest graficzny interfejs do języka NCL.

Bezpieczeństwo

OpenVMS Alpha/VAX v6.1 spełnia poziom C2 a SEVMS Alpha/VAX v6.1 (specjalna wersja OpenVMS, skoncentrowana na kwestiach bezpieczeństwa) poziom B1 standardu DoD 5200.28-STD TCSEC (Trusted Computer System Evaluation Criteria). Wersja v7.0 będzie poddana certyfikacji i oczekuje się identycznych rezultatów. W stosunku do wcześniejszych wersji, v7.0 nie wprowadza większych zmian w systemie bezpieczeństwa. Dodany został zestaw funkcji umożliwiających kontrolę prób włamań do systemu oraz dostępu proxy.

Nowy interfejs użytkownika - implementacja standardu CDE (Common Desktop Environment)

Obok pełnej implementacji standardu Motif v1.2.3, na OpenVMS v6.2 i v7.0 dostępny będzie niedługo pakiet CDE (Common Desktop Environment), popularny na wielu platformach UNIX-owych. Przykład szaty graficznej CDE ilustruje Rysunek 5. Przyglądając się jedynie interfejsom CDE na platformach UNIX i OpenVMS, niełatwo rozróżnić te systemy.

OpenVMS i Internet

Platforma OpenVMS, z jej wysoką niezawodnością, pracą non-stop i bezpieczeństwem wydaje się szczególnie dogodna do celów Internet'owych. Wychodząc naprzeciw takiemu zapotrzebowaniu, Digital przygotował zestaw produktów pod nazwą OpenVMS Internet Product Suite zawierający wszystko, co niezbędne do tworzenia i obsługi serwerów WWW. W skład zestawu wchodzi: Netscape Navigator (klient), Spyglass Enhanced Mosaic, lynx, mxrn, Purveyor Server, Netscape Communications Server, Netscape Commerce Server, serwer poczty IU-POP3, serwer gopher'a, Digital Firewall for OpenVMS, narzędzia do obsługi HTML i kilka innych produktów.

Hypersort

Jest to przenośna biblioteka funkcji realizujących wysoko wydajne operacje sortowania. Obecnie biblioteka ta jest dostępna na platformie Digital UNIX v3.2. Hypersort jest wykorzystywana między innymi przez DEC Cobol oraz ORACLE Rdb. Zawiera podzbiór funkcji narzędzi Sort/Merge z platformy OpenVMS, planowana jest pełna implementacja tej funkcjonalności w przyszłości.

Inne nowości

- Fast Path: restrukturyzacja drajwerów typu class i port zmierzająca do zwiększenia szybkości realizacji operacji I/O, głównie poprzez zredukowanie narzutów na alokację zasobów systemowych oraz zoptymalizowanie technik synchronizacji. W pierwszej fazie, jest dostępne dla dysków na magistrali CI z portami XMI CIXCD na platformie Alpha.

- Funkcja \$CREPRC dla konfiguracji cluster: Funkcja \$CREPRC służy w systemie OpenVMS do tworzenia procesów. Nowość polega na możliwości utworzenia procesu działającego na dowolnej maszynie konfiguracji VMScluster. Upraszcza to znacznie tworzenie aplikacji rozproszonych w tym środowisku.

Zakończenie

Wydaje się, że minął już etap utożsamiania otwartości z hasłem 'UNIX'. Najlepszym dowodem tego jest ekspansja Windows NT. Możemy dziś bez narażania na presję otoczenia przyjrzeć się uważniej systemom zaszklakowanym jako 'proprietary' czy 'legacy'. Na czym polega ich sukces? Jak to się stało, że skazane w wielu kręgach na zagładę unikają skutecznie losu dinozaurów? Odpowiedź wydaje się prosta - w kluczowych zastosowaniach są nadal niezastąpione.

Dewizą OpenVMS od zarania jego istnienia była i pozostaje niezawodność. Funkcjonalność VMScluster, bezpieczny system plików i nowości wersji v7.0 sprawiają, że OpenVMS może sprostać najbardziej wymagającym warunkom operacyjnym.

A że nie nazywa się UNIX...cóż, nikt nie jest doskonały ;-)

Artur Stefanowicz

Nowy interfejs użytkownika - implementacja standardu CDE (Common Desktop Environment)

Platforma OpenVMS, z jej wysoką niezawodnością, pracą non-stop i bezpieczeństwem wydaje się szczególnie dogodna do celów Internet'owych

¹ Pamięć fizyczna dla maszyn Alpha jest zależna od konkretnego modelu i waha się między 32MB i 14GB.

² „The Design and Implementation of a Log-Structured File System”, Rosenblum, Ousterhout. EE&CS Dept. University of California, Berkeley, 1991.

³ Nie dotyczy to systemu AdvFs na platformie Digital Unix i kilku innych implementacji, korzystającego z pokrewnej do kronikowego systemu plików idei logowania zmian.

Affinity¹ - łączy to co najlepsze

Programy Affinity i AEC oparte są na założeniu, iż dominującym trendem rozwoju systemów klient/serwer jest architektura 3 warstwowa

W sierpniu 1995 roku szefowie firm Digital i Microsoft ogłosili program o nazwie *Alliance for Enterprise Computing (AEC)*. Program ten wraz z programem *OpenVMS and Windows NT Affinity* ogłoszonym w maju 1995 wyznacza ramy współdziałania obu firm, których celem jest stworzenie infrastruktury informatycznej spełniającej potrzeby największych przedsiębiorstw. W ramach powyższej współpracy obie firmy dostarczą odpowiednich usług, produktów i technologii. Współdziałanie to rozciągnięte jest na kilka lat, a jego efekty będą pojawiały się etapami.

Cele strategii OpenVMS i Windows NT

Co jest celem strategii Affinity OpenVMS i Windows NT? Jakie korzyści przyniesie to naszym klientom? Celem strategii jest jak najlepsze połączenie krytycznych aplikacji produkcyjnych pracujących pod systemem OpenVMS z wielkim rynkiem aplikacji dla Windows NT i Windows. Strategia z jednej strony gwarantuje bezpieczeństwo inwestycji poczynionych w aplikacje dla OpenVMS, dając możliwość ich rozwoju w kierunku 3 warstwowej architektury klient/serwer. Z drugiej strony zapewnia użytkownikom Windows NT i Windows dostęp do nieograniczonych zasobów, tak w sensie przetwarzania jak i wielkości gromadzonych danych jakie może zagwarantować technologia klastrow OpenVMS.

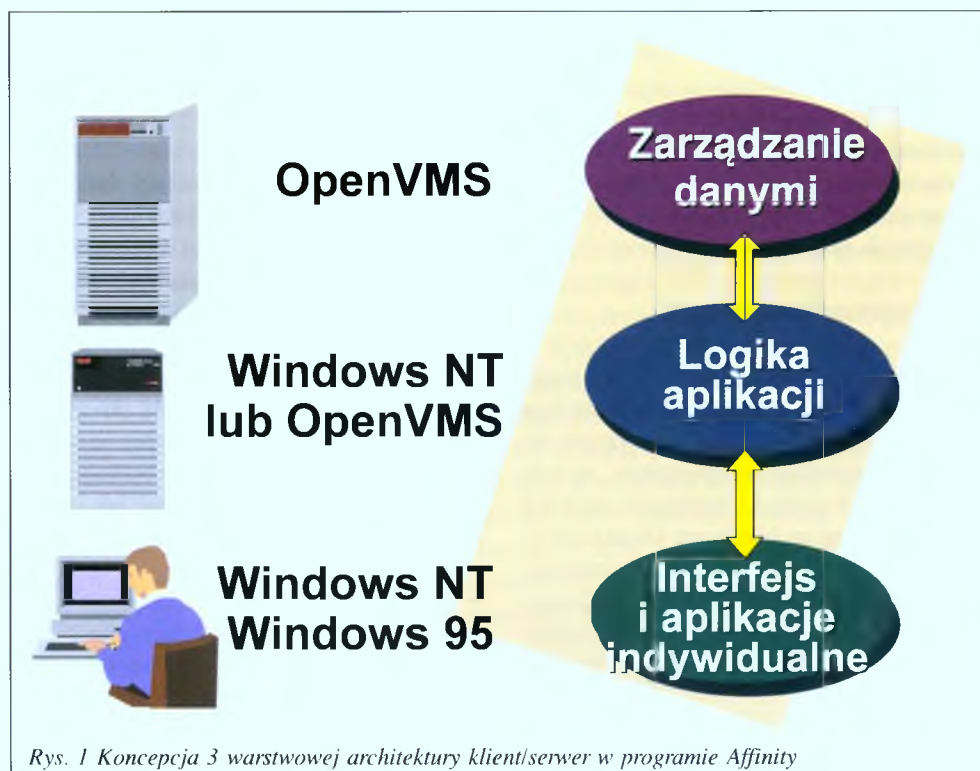
Programy Affinity i AEC oparte są na założeniu, iż dominującym trendem rozwoju systemów klient/serwer jest architektura 3 warstwowa. Architektura ta określa trzy warstwy oprogramowania tworzącego w sumie zintegrowane aplikacje obsługi przedsiębiorstw. Każda z warstw odpowiada za realizację różnych zadań i z tego też wynikają

wymagania stawiane przed platformą sprzętowo/programową, na której ma ona być implementowana. Idąc kolejno wyróżniamy następujące warstwy:

- Warstwa zarządzania danymi, na której pracują serwery baz danych zarządzające zasobami mającymi krytyczne znaczenie dla przedsiębiorstwa. Warstwa ta może dotyczyć także gromadzenia danych w postaci typowych plików, gwarantując odpowiednie mechanizmy ochrony dostępu oraz odpowiednie (tzn. bardzo dużo i bardzo szybko) możliwości archiwizacji. Na koniec mogą to być dowolne istniejące aplikacje produkcyjne, np. aplikacje transakcyjne. Częstokroć ta warstwa określana jest mianem serwera przedsiębiorstwa (*Enterprise Server*).
- Warstwa aplikacji obsługujących standardowe procedury przedsiębiorstwa określana mianem *Business Logic*. Mogą to być aplikacje, które koordynują dostęp do kilku serwisów poprzedniej warstwy, jak też serwery dla grup roboczych lub poczty elektronicznej. Ta warstwa określana jest zwykle jako serwer wydziałowy (*Department Server*).
- Warstwa aplikacji indywidualnych użytkowników pracujących w środowisku Windows z wykorzystaniem różnorodnych standardowych produktów typu MS Office, MS Exchange, LinkWorks czy też tworzonych z wykorzystaniem narzędzi typu VisualBasic i MS Access.

Realizacja programu

Całość programu Affinity rozciągnięta jest na kilka lat. Każdy kolejny rok przynosić będzie kolejną falę programu Affinity, czyli coraz większą integrację środowiska Open-



Rys. 1 Koncepcja 3 warstwowej architektury klient/serwer w programie Affinity

VMS i Windows NT. Kolejne fale przedstawiają się następująco.

Pierwsza fala - rok 1995

Zadaniem tej fali jest stworzenie podstaw do integracji aplikacji pracujących na obu platformach. Podstawą taką jest wspólna infrastruktura sieciowa i narzędzia programistyczne do tworzenia rozproszonych aplikacji. Tę falę mamy już za sobą. Wszystkie elementy tej fali zostały zrealizowane, a odpowiednie produkty pojawiły się na rynku. W ramach infrastruktury sieciowej mamy nowe serwisy TCP/IP w najnowszych wersjach systemu OpenVMS 7.0 oraz DECnet dla Windows NT. Infrastruktura narzędziowa to pakiety typu middleware. Pojawiły się nowe wersje pakietów ObjectBroker, DEC-messageQ, RTR, ACMSxp i DCE dla OpenVMS, Windows NT i Windows 95. Na koniec pierwsza wersja pakietu Management Station dająca pełną kontrolę nad kontami użytkowników OpenVMS z Windows.

Druga fala - rok 1996

Druga fala zakłada dalszy rozwój narzędzi integracyjnych ObjectBroker z technologią OLE Integration. W ramach infrastruktury poczty nastąpi integracja poczty ALL-IN-1, MailWorks i MS Exchange. Jest to także rok kiedy OpenVMS daje możliwości wykorzystania swoich 'nielimitowanych' możliwości

(bardzo duże bazy typu VLM i nowy, wysoce wydajny system plików). Dla środowiska programistycznego kluczowe jest wprowadzenie technologii API Win32s, MFC i OLE2 dla platformy OpenVMS.

Kolejna fala - rok 1997

Choć 2 lata do przodu w technologii softwarowej to bardzo dużo, już teraz można określić iż będzie to rok, gdy nastąpi dalsza integracja systemów poprzez pełną implementację klienckiej części oprogramowania systemu plików Spiralog na platformie Windows NT. Planowane jest także wspólne API dla obu systemów operacyjnych do obsługi takich cech technologii klastrów jak automatyczne rozłożenie obciążenia i odporność na awarię pojedynczej maszyny. W roku 1997 będzie też możliwe wspólne zarządzanie systemem operacyjnym OpenVMS i systemem Windows NT, włączając w to konta, zasoby drukarkowe i dyskowe. Pod systemem OpenVMS będzie też dostępne pełne API Win32 wraz z rozszerzeniami typu BackOffice.

W dalszej części artykułu przedstawię Państwu bardziej szczegółowo niektóre elementy programu Affinity. Wyróżniłem kilka obszarów w których pojawiły się lub pojawiają technologie mające zasadnicze znaczenie dla realizacji programu. Obszary te dotyczą kolejno:

Dla środowiska programistycznego kluczowe jest wprowadzenie technologii API Win32s, MFC i OLE2 dla platformy OpenVMS

- infrastruktura middleware do integracji aplikacji,
- wykorzystanie 'nielimitowanych' możliwości OpenVMSa,
- infrastruktura programistyczna,
- wspólne zarządzanie i administrowanie systemami w sieci,
- integracja systemów poczty.

Infrastruktura middleware

Infrastruktura middleware została zaimplementowana jako pierwsza, ponieważ daje ona firmom i użytkownikom tworzącym aplikacje odpowiednie wyprzedzenie, w celu rozpoczęcia tworzenia i wdrażania aplikacji w 3 warstwowej architekturze klient/serwer. Produkty tej kategorii zostały dość obszernie opisane w poprzednich numerach DECforum i Digital Forum. Dlatego też powierzchownie zaznaczę co wchodzi w skład tej bogatej oferty:

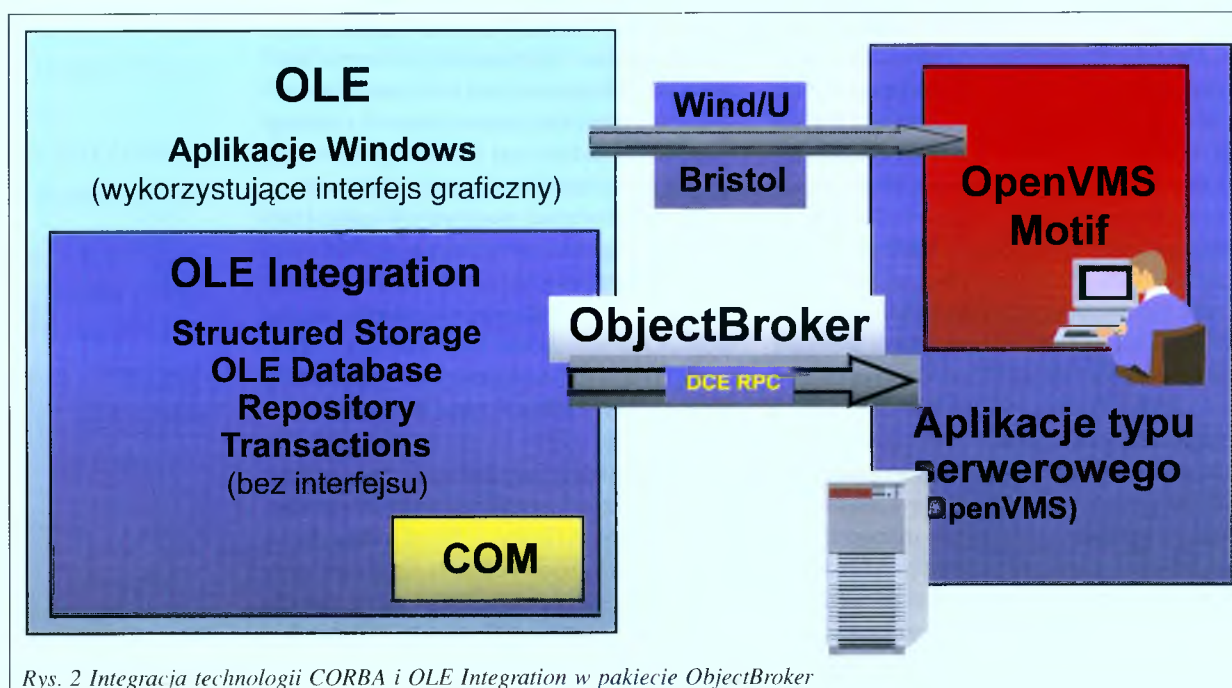
- ACMSxp - monitor transakcyjny o otwartej architekturze oparty o technologię DCE i implementujący standardy transakcyjne X/Open (język STDL). ACMSxp wywodzi się z monitora ACMS i stanowi rozwinięcie technologii używanej w kilku tysiącach dużych instalacji transakcyjnych na całym świecie.
- RTR (*Reliable Transaction Router*) - infrastruktura transakcyjna dająca możliwość tworzenia systemów odpornych na

wystąpieniu awarii (*software fault-tolerant systems*) przy wykorzystaniu zwykłego sprzętu komputerowego, systemów operacyjnych i baz danych.

- DCE (*Distributed Computing Environment*) - standardowa (OSF, X/Open) infrastruktura dla tworzenia rozproszonych systemów i aplikacji niezależnych od platformy sprzętowo-systemowej.

- DECmessageQ - jeden z najpopularniejszych pakietów klasy MOM (*Messaging Oriented Middleware*), który może być stosowany zarówno do tworzenia dużych rozproszonych aplikacji, jak i ich integracji.

- ObjectBroker - pakiet implementujący standard CORBA zdefiniowany przez organizację OMG (*Object Management Group*). ObjectBroker jest unikalny w klasie podobnych sobie produktów ponieważ umożliwia integrację standardu CORBA ze standardem OLE firmy Microsoft. W obecnej wersji jest to ograniczona integracja poprzez bramkę nazwaną OLE Portal. Jednak już w roku 1996 pojawi się wersja ObjectBrokera z implementacją technologii OLE Integration. Digital od ponad dwóch lat pracuje wraz z Microsoftem nad implementacją technologii OLE Integration dla platform innych niż Windows. Równolegle z wprowadzeniem technologii OLE Integration przez Micro-



Rys. 2 Integracja technologii CORBA i OLE Integration w pakiecie ObjectBroker

Krótką historia VMSa i Windows NT

System operacyjny VMS (*Virtual Memory System*) został wprowadzony na rynek przez Digital wraz z nowym procesorem rodziny VAX w 1978 roku. Architektura 32 bitowego procesora oparta została na doświadczeniach poprzedniej generacji komputerów serii PDP, oraz dostępnych wtenczas systemów operacyjnych takich jak RSX i UNIX. System operacyjny VMS został stworzony tak, aby zaspokoić potrzeby różnych grup użytkowników (tryb wsadowy, tryb interakcyjny i aplikacje czasu rzeczywistego) i już w roku 1984 posiadał wbudowane elementy technologii kłastrów (odpowiedni system plików, DLM, itp.). Na początku lat 80 kombinacja VMS/VAX była dominującą rodziną minikomputerów. Jednak wraz z pojawieniem się znacznie szybszych procesorów typu RISC (VAX jest procesorem typu CISC) i nowych wersji UNIXa, maszyny z VMSem zaczęły tracić rynek stacji graficznych. VMS z bogactwem swej funkcjonalności i architekturą nakierowaną na wielodostęp nie mógł konkurować w szybkości przetwarzania z nowymi stacjami typu RISC. Dlatego też, już w pierwszej połowie lat 80-tych, Digital rozpoczął pracę nad własnym procesorem RISC oraz nad nowym, 'lżejszym' systemem operacyjnym, który mógłby konkurować z UNIXem na rynku stacji. Tak rozpoczął się projekt PRISM. Szefem projektu PRISM został Dave Cutler, który wcześniej brał udział w tworzeniu VMSa. Niestety, po trzech latach pracy projekt PRISM został zawieszony. Dave Cutler wraz z kilkudziesięcio-osobową grupą najbliższych współpracowników przeszedł do firmy Microsoft, gdzie otrzymał zadanie stworzenia nowego systemu operacyjnego - Windows NT. Dla Microsoftu był to okres perturbacji związanych z tworzeniem systemu OS/2 wspólnie z IBM, a raczej faktem rozwodzenia się obu firm. Microsoft szukał więc alternatywy dla systemu OS/2.

OpenVMS i Windows NT mają więc wiele wspólnego. Ludzie, którzy współtworzyli VMSa, a potem zabrali się do tworzenia Windows NT odcisnęli piętno podobieństwa, które łatwo dostrzec w obu systemach. Rzeczy takie jak asynchroniczne I/O, mapowanie plików do pamięci, kwoty procesów, itp. sprawiają iż ludzie, którzy znają VMS 'czują', iż Windows NT ma coś w sobie coś znajomego. Oczywiście Windows NT jest całkiem innym systemem operacyjnym niż VMS. Wysoki stopień modularności, architektura jądra, centralne Registry stawiają ten system na czele najnowocześniejszych, dostępnych systemów operacyjnych. Z drugiej strony jego zalety stanowią też o jego ograniczeniach. Dlatego też w dalszym ciągu OpenVMS ma sporo do zaoferowania systemowi Windows NT, weźmy chociażby 64bity, nowy system plików i technologia kłastrów. I vice versa, Windows NT ma coś do zaoferowania VMSowi (program NT Connectivity).

soft, Digital zaoferuje OLE Integration dla OpenVMSa oraz kilkunastu różnych implementacji UNIXa.

Należy nadmienić iż każdy z powyższych pakietów jest dostępny nie tylko na platformie OpenVMS i Windows NT ale także na platformie Digital UNIX. W przypadku pakietów ObjectBroker i DECmessageQ są one dostępne na kilkunastu różnych platformach.

Nielimitowane możliwości OpenVMS

Głównym zadaniem programu AEC jest zapewnienie użytkownikom i aplikacjom pracującym na maszynach Windows NT i Win-

dows dostępu do 'nie-limitowanych' zasobów kłastrów OpenVMS. Oczywiście w sensie programistycznym dostęp taki zapewnią pakiety infrastruktury middleware, o których była mowa poprzednio. Co świadczy o tych 'nie-limitowanych' możliwościach Open VMSa. Trzy rzeczy, których nie ma system Windows NT (jak i większość innych systemów operacyjnych), a które mogą być niezbędne z punktu widzenia użytkownika:

- 64 bitowy system operacyjny z implementacją wątków na poziomie jądra systemu operacyjnego. Z jednej strony zapewnia to obsługę niewyobrażalnie dużej pamięci operacyjnej, obecnie do 14GB a już w tym roku do 28GB, z czego w

Głównym zadaniem programu AEC jest zapewnienie użytkownikom i aplikacjom pracującym na maszynach Windows NT i Windows dostępu do 'nie-limitowanych' zasobów kłastrów OpenVMS

Jednym z obszarów integracji obu systemów jest sfera zarządzania i administrowania systemami w sieci

praktyce korzystają bazy Oracle czy Sybase. Z drugiej to efektywność działania w architekturze wieloprosesowej typu SMP (*Symmetric Multi Processing*). Dzisiaj OpenVMS jest dobrze skalowalny do 12 procesorów, a dzięki nowej technologii wątków w roku 1996 pojawią się maszyny z większą liczbą procesorów. Nie każdy system operacyjny wykazuje dobrą skalowalność w rozwiązaniach SMP.

- Spiralog czyli nowy system plików, wykorzystujący nowatorskie rozwiązania fizycznej organizacji zapisów do dysku. Technologia zastosowana w nowym systemie plików pozwoli na wielokrotne skrócenie średniego czasu odczytu z dysku, przy jednoczesnym wielokrotnym poprawieniu szybkości archiwizacji danych dyskowych. Da to możliwość zarządzania systemem plików, którego pojemność idzie w Terrabajty. Nowy system plików jest w pełni zgodny z architekturą klastrów i posiada nowoczesną architekturę klient/agent pozwalającą na dostęp do dysków OpenVMSa z innych maszyn (np. Windows NT, UNIX) tak jakby to był ich własny system plików.

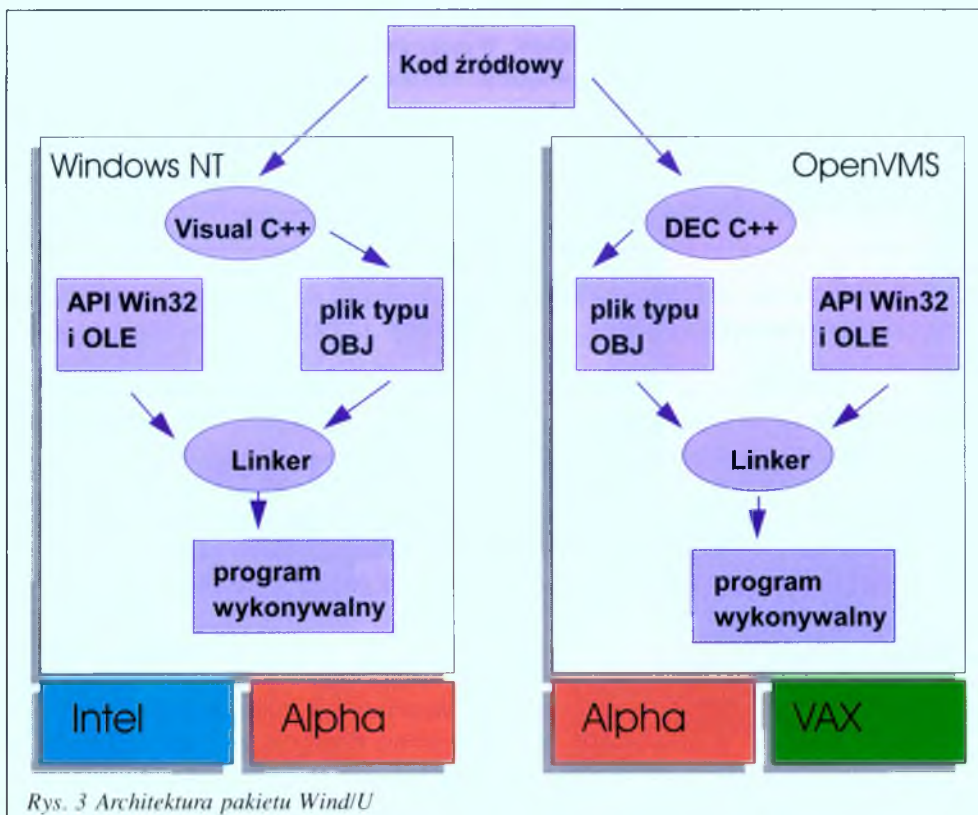
- Klastry gwarantują zwiększoną wydajność systemu i niezawodność, częstokroć bez

żadnych specjalnych zabiegów ze strony użytkowników aplikacji. Zastosowanie technologii MemoryChannel pozwoli na stworzenie jeszcze bardziej wydajnych konfiguracji klastrów, a nowe API dla klastrów, które częściowo będzie wspólne dla OpenVMSa i Windows NT pozwoli na tworzenie aplikacji, które łatwo będzie można przenieść pomiędzy obydwoma platformami.

Technologie powyższe zostały zaznaczone w sposób bardzo pobieżny, nieadekwatny dla ich nowatorstwa jedynie dlatego iż znacznie szersze opisy znajdziecie Państwo w artykule Artura Stefanowicza 'System OpenVMS v7.0'

Środowisko programistyczne

Jak wszyscy wiemy życie programisty nie jest lekkie. Skazany jest on na ciągłą naukę i ciągłe odkrywanie koła na nowo, coś na kształt ciągłego wpychania kamienia pod górę przez jednego faceta. Załóżmy, że chcemy tworzyć aplikację w rozproszonym heterogenicznym środowisku, powiedzmy 3 warstwowym czyli dla Windows 95, Windows NT i OpenVMS. Aplikację dla Windows 95 stworzymy, powiedzmy w Visual C++ (lub Visual Basicu), część aplikacji dla Windows NT znów piszemy w Visual C++, oczywiście medium pomiędzy maszynami stanowi technologia OLE



Integration. Została nam więc ostatnia warstwa na OpenVMS, logicznie jest ją także tworzyć w C++, który jest dostępny dla OpenVMS. Tyle tylko, że jesteśmy programistą Windows'owym i od zawsze korzystamy z dobrodziejstw i bogactwa API Win32 i MFC (*Microsoft Foundation Classes*). Oczywiście OpenVMS ma swoje własne API i bardzo bogate biblioteki systemowe (niestety nijak mające się do Win32), z drugiej strony OpenVMS jest zgodny z większością standardów X/Open (XPG4, itp) więc z łatwością można portować aplikacje z UNIXa, ale znów coś ma API standardów POSIX do Win32. Coż ma więc począć nasz biedny programista? Otóż, ma się nie martwić. Istnieje rozwiązanie, któremu na imię Wind/U. Wind/U jest produktem firmy Bristol i stanowi implementację API Win32, MFC i OLE na platformie OpenVMS. Bristol jest jedną z niewielu firm na świecie, która ma dostęp do źródeł firmy Microsoft i licencję na implementację powyższych technologii Microsoftu na innych platformach systemowych.

Każda aplikacja napisana pod Visual C++ i wykorzystująca klasy MFC, będzie mogła zostać przekompilowana na platformie OpenVMS z wykorzystaniem kompilatora DEC C++. Powstały w efekcie program będzie można bezpośrednio uruchomić pod OpenVMSem bez udziału jakiegokolwiek środowiska emulacyjnego (dotyczy to także pełnej obsługi ekranu, która zostanie wykonana w środowisku Motif). Aby w pełni urzeczywistnić tę ideę DEC C++ jest obecnie wyposażony w odpowiednie rozszerzenia standardu Microsoft Visual C++.

Powyższa technologia pozwala na realizację hasła 'twórz na jednej platformie, uruchom na wielu'. Oprogramowanie typu *Runtime* Wind/U firmy Bristol dostarczane jest wraz z systemem OpenVMS począwszy od wersji 7.0 (licencję za kilkadziesiąt dolarów trzeba dokupić). Oprogramowanie typu *Development* można zakupić bezpośrednio w firmie Bristol.

Wspólne zarządzanie i administrowanie systemami w sieci

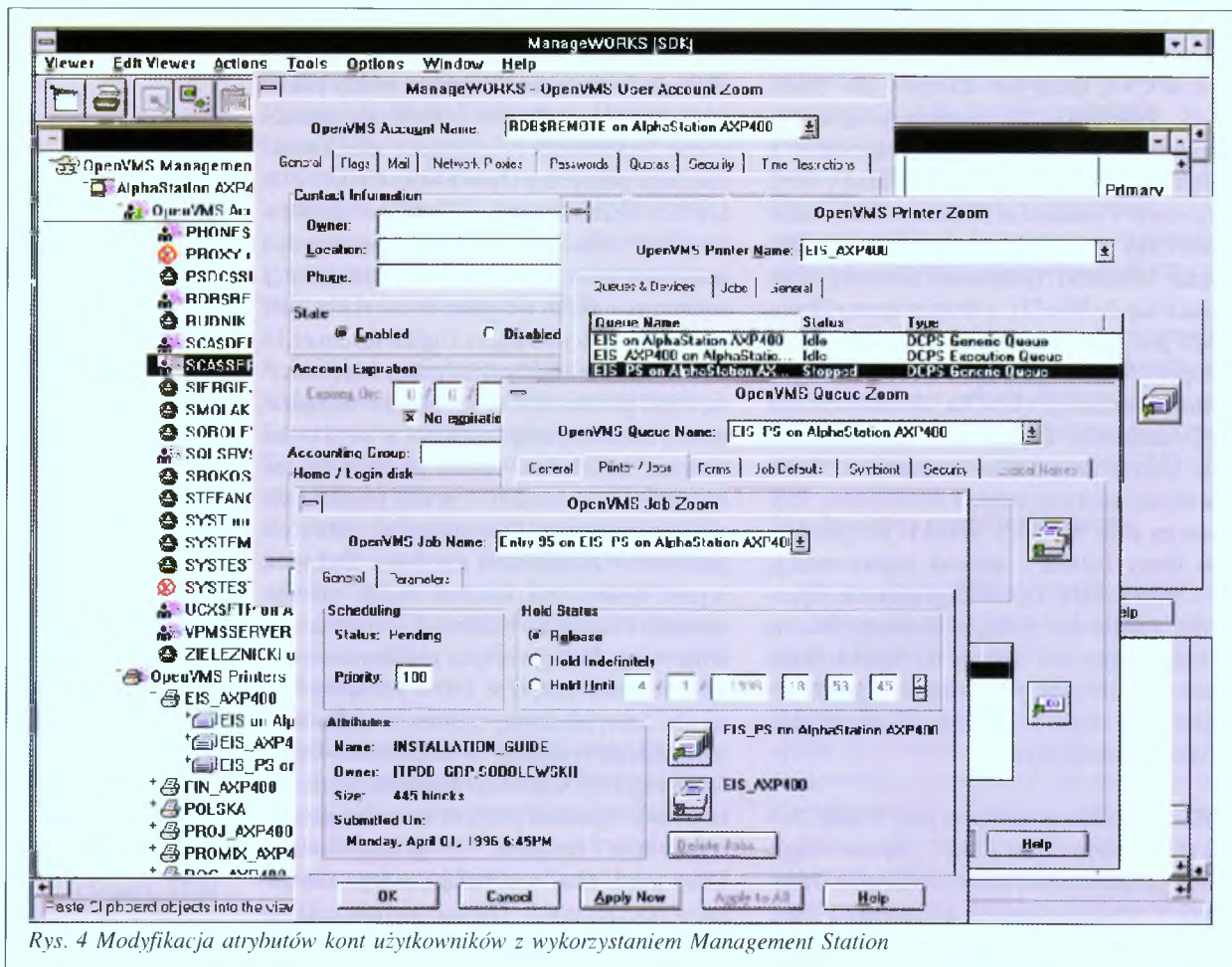
Jednym z obszarów integracji obu systemów jest sfera zarządzania i administrowania systemami w sieci. W sferze zarządzania elementami aktywnymi sieci dominującym standardem jest środowisko NetView, gdzie

Digital oferuje NetView dla Windows NT. W sferze administracji zasobami programowymi w środowisku Windows mamy pakiet SMS do którego Digital oferuje oprogramowanie Assetworks rozciągające możliwości SMSa na platformy OpenVMS i UNIX. POLYCENTER Manager on NetView dla Windows NT i Assetworks pojawiły się na rynku ponad rok temu, są sukcesywnie rozwijane, a zainteresowanych odsyłam do artykułu SMS Zygmunta Jerzyńskiego z Digital forum nr. 15 z 1995. Oba produkty choć pokrywają znaczną część problemów związanych z zarządzaniem i administracją zasobami w sieci to nie dotyczą zasadniczo takich problemów jak zarządzanie zasobami i pracą poszczególnych komputerów. Pozostaje więc cała sfera problemów zarządzania zasobami drukarkowymi, dyskowymi, kontroli zadań, monitorowania i badania wydajności, strojenia systemów, itp. Kategoria tych problemów jest o tyle niewdzięczna, iż w 100% jest specyficzna dla danej platformy systemowej. Budowa wewnętrzna systemów Windows NT, OpenVMS czy UNIX jest tak odmienna, że nie ma tu mowy o jakimś jednym modelu parametryzowania i strojenia. Oprogramowanie tej klasy musi 'znać' specyfikę każdego z systemów operacyjnych. Digital posiada całą rodzinę pakietów POLYCENTER specyficznych dla OpenVMSa i UNIXa. W ramach programu Affinity powstają nowe pakiety, które będą obejmować swym zainteresowaniem także Windows NT. Na uwagę zasługuje tu oprogramowanie Management Station i POLYCENTER Performance Expert oraz program pod roboczą nazwą *NT Connectivity*.

Management Station

Pakiet Management Station jest częścią oprogramowania OpenVMS od wersji 6.2. Jest to oprogramowanie, którego celem jest zapewnienie dostępu do mechanizmów administracji systemem OpenVMS z poziomu aplikacji pracującej w środowisku Windows. Oprogramowanie pracuje w architekturze klient/agent. Graficzny interfejs na maszynie Windows wykonuje odpowiednie akcje poprzez proces agenta działającego na docelowej maszynie OpenVMS. W wersji 1.1, dostępnej w systemie 7.0, użytkownik Management Station ma możliwość zarządzania kontami użytkowników systemu OpenVMS. Funkcjonalnie odpowiada to dostępowi do programu AUTHORIZE systemu OpenVMS. Management Station posiada wiele właści-

Performance Expert zakłada, iż z jednego stanowiska Windows będzie można zarządzać zasobami wielu maszyn Windows NT, OpenVMS i UNIX



Rys. 4 Modyfikacja atrybutów kont użytkowników z wykorzystaniem Management Station

wości ułatwiających zarządzanie kontami, pozwala na operowanie na podzbiorach kont z jednej lub z wielu maszyn, w sposób automatyczny tworzy katalogi dyskowe, wyświetla listę dostępnych zasobów dyskowych. Użyteczność pakietu jest bardzo duża i nie można go nie polubić. Management Station zaimplementowany jest w oparciu o technologię oprogramowania ManageWorks, znanego z pakietu PATHWORKS jako środowisko do zarządzania zasobami LAN Manager. Moduł zarządzania kontami jest pierwszym komponentem pracującym w ramach Management Station, następne moduły do zarządzania zasobami drukarkowymi i zarządzania zasobami dyskowymi pojawią się wkrótce.

Performance Expert

Rodzina produktów Performance Expert jest rozwinięciem idei produktów rodziny POLYCENTER Performance Solution. POLYCENTERps dostępny jest dla maszyn OpenVMS i DigitalUNIX. W skład rodziny wchodzi następujące pakiety:

- Data Collector - służy do kolekcjonowa-

nia informacji o obciążeniu systemu przez działające aplikacje, które są następnie agregowane i cyklicznie zapisywane do odpowiednich plików. Dodatkowo dostępny jest moduł monitora przedstawiającego na bieżąco informacje o obciążeniu systemu.

- Performance Advisor - na podstawie informacji zawartych w plikach stworzonych przez Data Collector produkuje odpowiednie zestawienia i proponuje akcje mające na celu poprawienie wydajności systemu.

- Capacity Planner - służy do modelowania przewidywanego obciążenia dla rzeczywistej lub zamodelowanej konfiguracji sprzętowej (pamięć, moc procesora, kontrolery i zasoby dyskowe). Jako dane wejściowe służą informacje zgromadzone przez Data Collector (rzeczywiste dane o obciążeniu i rzeczywista konfiguracja) oraz model aplikacji zdefiniowany przez użytkownika.

O ile dwa pierwsze pakiety są specyficzne dla OpenVMSa i Digital UNIXa to trzeci jest niezależny od platformy i potrafi pracować na danych wygenerowanych na obu systemach operacyjnych.

Performance Expert zakłada, iż z jednego stanowiska Windows będzie można zarządzać zasobami wielu maszyn Windows NT, OpenVMS i UNIX. Każda z maszyn w sieci musi mieć uruchomiony odpowiedni proces agenta typu Data Collector. Z jednej maszyny Windows istnieje następnie możliwość monitorowania na bieżąco zachowania się różnych systemów, przeprowadzania analiz obciążenia, alarmowania o 'niebezpiecznych' stanach systemu operacyjnego czy wreszcie przeprowadzanie długofalowych analiz trendów występujących w obciążeniu danej maszyny. W skład Performance Expert wchodzi następujące moduły:

- Collector - funkcjonalnie odpowiada pakietowi Data Collector z tym, że pracuje

jako serwis na Windows NT.

- Monitor - daje możliwość monitorowania wybranych parametrów systemu operacyjnego współpracując z kolektorami pracującymi na odpowiednich maszynach.

- Advisor - funkcjonalnie odpowiada pakietowi Performance Advisor, ale pozwala na analizę danych z różnych systemów operacyjnych.

- Alarms - na podstawie predefiniowanych i definiowalnych przez użytkownika reguł pozwala na wszczynanie alarmów w przypadku ich naruszenia. Może śledzić jednocześnie wiele maszyn a alarm może powodować odpalenie akcji takiej jak wysłanie wiadomości pocztą lub uruchomienie odpowiedniej aplikacji.

- Trends - daje możliwość analizowania pracy systemu na przestrzeni dłuższego okresu pracy (tygodnie, miesiące).



Rys. 5 Monitorowanie maszyny z systemem OpenVMS (góra) i Windows NT (dolna)

Obecnie w programie partycypuje kilkaset firm tworzących oprogramowanie narzędziowe i aplikacyjne między innymi takie jak: Oracle, Sybase, Forte, Magic, Progress i Software AG

Uzupełnieniem pakietu Performance Expert jest powstająca wersja Capacity Planner dla Windows.

Inicjatywa NT Connectivity

Urzeczywistnienie wizji zarządzania i administrowania systemem OpenVMS, czy też klastrami takich maszyn, z pojedynczej maszyny Windows NT wymaga głębszej integracji obu systemów. Temu celowi służy program NT Connectivity, którego celem jest implementacja takiej funkcjonalności pod systemem OpenVMS, iż będzie on widoczny dla aplikacji Windows NT jako jeszcze jedna maszyna Windows. Kluczem do tego jest implementacja następujących technologii:

- Pojedyncze logowanie i wspólny mechanizm bezpieczeństwa - gdzie użytkownik będzie mógł jednokrotnie logować się do sieci wielu maszyn. Obejmuje to też synchronizację kont na wielu maszynach, mechanizm potwierdzania tożsamości i autoryzacji.
- Integracja DCE RPC i mechanizmu Pipes - mechanizm komunikacji poprzez Pipes jest dominujący dla środowiska PATHWORKS i Windows i wymaga pełnej integracji z protokołem DCE RPC będącym standardem dla systemów heterogenicznych.
- Registry - spełnia kluczową rolę w systemie Windows NT gromadząc informacje o wszystkich parametrach systemu i aplikacji. Podobny mechanizm zostanie zaimplementowany dla OpenVMSa.
- Serwisy - są jedną z kluczowych technologii w Windows NT i podobny mechanizm uruchamiania i kontroli procesów zostanie zaimplementowany dla OpenVMSa.
- Zdarzenia - podsystem obsługi zdarzeń jest także niezwykle ważny przy wszystkich operacjach administracyjnych. Mechanizm kompatybilny dla technologii Windows NT zostanie więc zaimplementowany na OpenVMS.

Integracja systemów poczty

Systemy pracy grupowej idealnie wpasowują się w koncepcję 3 warstwowej architektury klient/serwer. Na poziomie serwe-

rów przedsiębiorstw możemy umieścić serwisy klasy X.400 i X.500, które będą koordynowały i pośredniczyły pomiędzy wieloma serwerami wydziałowymi czy biurowymi pracującymi niejednokrotnie z różnymi systemami poczty. Tak więc serwerom wydziałowym (biurowym) przypada rola serwerów poczty, konferencji i innych serwisów automatyzujących prace biura. Na końcu mamy użytkowników, którzy wykorzystując pakiety standardowe odbierają i wysyłają pocztę, faxy i dokumenty w różnej postaci, mają dostęp do informacji dystrybuowanej w ramach firmy.

Na poziomie serwerów przedsiębiorstwa Digital oferuje dwa produkty. Pierwszy to MAILbus 400, który gwarantuje niezawodne i wysoce wydajne dostarczanie komunikatów w ramach rozległych sieci komputerowych. MAILbus 400 MTA jest zgodny ze specyfikacją ITU 1992 i posiada bramki zapewniające łączność z takimi systemami jak SMTP/MIME, fax, telex i różnorakie systemy poczty dla maszyn PC. Drugi produkt to DEC X.500 Directory Services, który jest pierwszym pakietem X.500 zgodnym ze specyfikacjami GOSIP i OSTC. Oba produkty pozwalają na integrację pakietów takich jak ALL-IN-1 Office Server, MailWorks, MS Exchange, Lotus Notes, Lotus cc:Mail i wiele innych.

Co na to nasi Partnerzy?

Sukces każdego programu zależy od zaangażowania jego uczestników. W przypadku programu Affinity nie wystarczy zaangażowanie firmy Digital i Microsoft. Niezbędny jest także współudział firm partnerskich dostarczających oprogramowania na obie platformy i pasującego do koncepcji 3 warstwowej architektury klient/serwer. Obecnie w programie partycypuje kilkaset firm tworzących oprogramowanie narzędziowe i aplikacyjne, że wymienię tu choćby firmy takie jak: Oracle, Sybase, Forte, Magic, Progress i Software AG.

Podsumowanie

Jak nadmieniałem wcześniej program Affinity daje coś obu światom. Z jednej strony integruje świat istniejących systemów produkcyjnych z technologią klasy desktop. Z drugiej strony daje szansę na obsługę gigantycznych ilości danych, przy zachowaniu

FX!32, czyli aplikacje z Intela uruchom na Alphie

Magazyn Byte ogłosił podczas targów Comdex 95, że w kategorii nowych technologii zwyciężył produkt FX!32 firmy Digital. Czym jest ten tajemniczy FX!32? Oprogramowanie FX!32 pozwala na uruchamianie aplikacji zgodnych z Win32 i skompilowanych dla platformy x86 (zgodnej z Intel) na platformie Alpha Windows NT. Szybkość działania aplikacji jest porównywalna z działaniem aplikacji na oryginalnej maszynie x86 z najszybszymi dostępnymi procesorami. FX!32 stosuje dwie znane technologie wykorzystywane do uruchamiania aplikacji na różnych platformach: emulację i binarną translację. Każda z tych technik ma swoje wady i zalety. Emulacja jest w pełni przezroczysta dla użytkownika, ale wolna. Binarna translacja daje efektywny kod, ale wymaga specjalnych kroków translacji aplikacji. FX!32 łączy w sobie zalety obu technologii.

W skład FX!32 wchodzi trzy komponenty. Pierwszy to środowisko typu *run-time* dające możliwość uruchamiania aplikacji w sposób przezroczysty dla użytkownika. Drugi element to działający w cieniu proces optymalizatora czyli binarny translator. Trzeci, to specjalny serwer koordynujący pracę dwu poprzednich.

Działanie całości jest niewidoczne dla użytkownika. Uruchomienie aplikacji x86 Win32 na platformie Alpha Windows NT powoduje, iż system operacyjny automatycznie wywołuje FX!32. Jeśli jest to pierwsze uruchomienie tej aplikacji na danej maszynie to FX!32 rozpoczyna jej wykonanie w trybie emulacji, zapamiętując przetranslowany kod w wewnętrznej bazie optymalizatora. W ten sposób każde kolejne wykonanie aplikacji powoduje, iż aplikacja wykonuje się coraz szybciej. Oczywiście FX!32 w pełni wykorzystuje te biblioteki podsystemu, które dostępne są na platformie Alpha. I tak jeśli np. aplikacja wykorzystuje podsystem OLE2, który znajduje się na maszynie Alpha to podsystem ten będzie wprost wykorzystywany nawet przez aplikację wykonywaną w trybie emulacji. Kod generowany przez emulator jest następnie przetwarzany przez proces optymalizatora. O tym w jakim stopniu optymalizator będzie pracował nad kodem decyduje proces serwera. Proces serwera może być parametryzowany przez administratora. Istnieje możliwość określenia limitów dla zajętości dysku przez kod tworzony przez proces optymalizatora. Na uwagę zasługuje fakt, że emulator jest w pełni zoptymalizowany dla architektury procesora Alpha i wystarczająco mały, aby rezydować w pamięci typu cache.

FX!32 pojawi się na rynku w połowie 1996. Licencja na FX!32 będzie zawarta w licencji systemu Windows NT dla maszyn Alpha. Tak więc użytkownicy najszybszych dostępnych komputerów Windows NT, czyli komputerów z procesorem Alpha będą mogli uruchamiać dowolny program, który będzie posiadał etykietkę zgodności z Win32.

Dlatego też w dalszym ciągu OpenVMS ma sporo do zaoferowania systemowi Windows NT, chociażby 64bity, nowy system plików i technologię klastrów

łatwości w zarządzaniu tymi danymi i dostępu do nich z wykorzystaniem powszechnie używanych technologii. U podstaw wszystkiego leżą potrzeby użytkowników, które zazwyczaj są tak specyficzne iż nie jest łatwo zadowolić je stosując jedno rozwiązanie technologiczne, przysłowiowy *Silver Bullet*. Z systemami komputerowymi jest podobnie jak z ludźmi, należy wydobywać z nich to co najlepsze i próbować łączyć dto dla osiągnięcia jak najlepszych efektów. Nie należy natomiast dążyć za wszelką cenę do unifikacji i ujednolicenia. Tego typu eksperymenty tra-

gicznie kończyły się w przypadku ludzi i społeczeństw, i jak widać nie sprawdzają się też w przypadku komputerów:-)

Piotr Sobolewski

¹**affinity** [e'finiti] s 1. pokrewieństwo <koligacja> (**with** <to> **sb** z kimś) 2. przyciąganie 3. *chem* powinowactwo

Digital UNIX - standard, któremu można zaufać

Pierwszą platformą sprzętową na której w 1969 roku zaimplementowano system operacyjny UNIX była maszyna PDP-7 firmy Digital Equipment

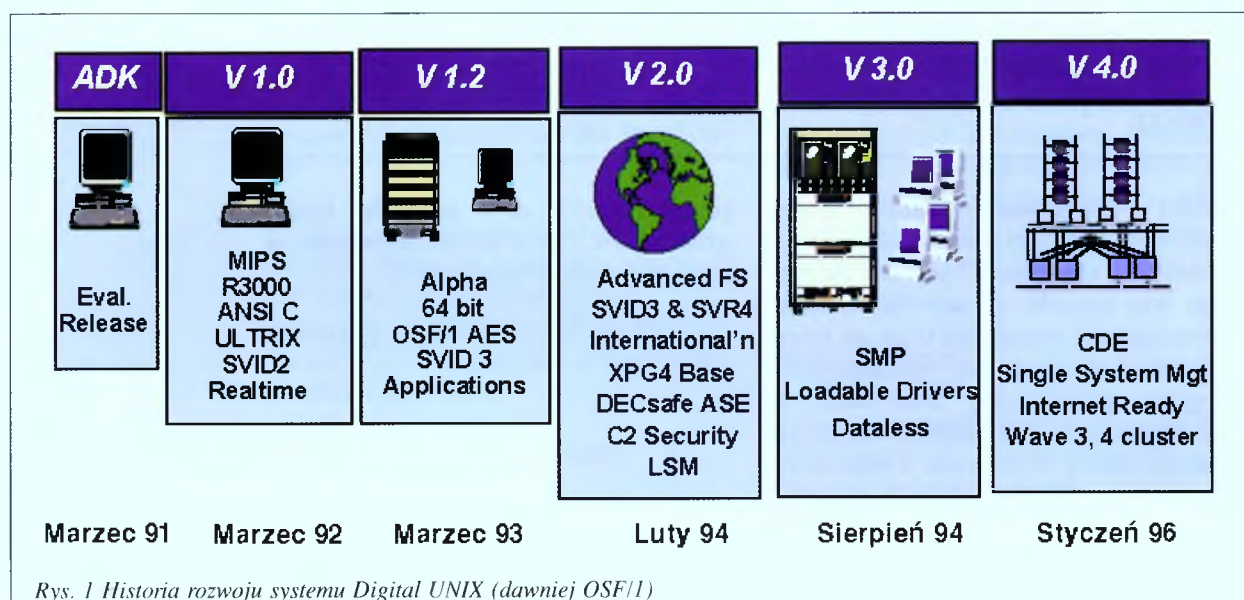
Historia

Pierwszą platformą sprzętową na której w 1969 roku zaimplementowano system operacyjny UNIX była maszyna PDP-7 firmy Digital Equipment. Następne wersje tworzone w języku C powstawały na początku lat 70-tych na maszynach PDP 11/45 i PDP 11/70. Lata 70-te to okres dojrzewania systemu UNIX i zdobywania szerokich rzesz zwolenników w ośrodkach akademickich. Natomiast lata 80-te to burzliwy rozwój i dojrzewanie systemu, który początkowo torował sobie drogę w zastosowaniach komercyjnych jako najlepsza platforma dla stacji roboczych, aby na koniec lat 80, a zwłaszcza na początku lat 90-tych ugruntować swą pozycję jako w pełni dojrzałego systemu komercyjnego na rynku serwerów. Digital początkowo oferował swoim klientom wersje UNIXa o nazwie ULTRIX na platformie maszyn VAX i maszyn z procesorem MIPS. W roku 1988 Digital wraz z IBM, HP i kilkoma innymi firmami

zainicjowały założenie organizacji OSF (*Open Software Foundation*), która obok UNIX Laboratories zajmowała się procesem standaryzacji i rozwoju UNIXa. Wynikiem prac organizacji OSF jest system operacyjny OSF/1, interfejs okienkowy OSF/Motif oraz infrastruktura DCE (*Distributed Computing Environment*). W roku 1991 roku, Digital jako pierwsza firma wypuścił na rynek system operacyjny oparty o technologię OSF czyli Digital OSF/1. Od roku 1991 system ten jest sukcesywnie rozwijany, a kamieniami milowymi jego rozwoju był rok 1992 - czyli implementacja na platformie Alpha, oraz rok 1995 - uzyskanie prawa do znaku UNIX (*branding*) od organizacji X/Open i co za tym idzie zmiana nazwy na Digital UNIX.

Kluczowe elementy funkcjonalności

Dla większości z nas naturalne jest skojarzenie UNIX = system otwarty = standard. Jeśli system ma być otwarty to znaczy, iż ma



być oparty na pewnych powszechnie uznanych standardach (standardach *de facto* i *de jure*). Standaryzacja systemu UNIX rozumiana jest jako zgodność interfejsu użytkowego, API programistycznego, poleceń systemu i większości podstawowych narzędzi z pewnymi określonymi normami definiowanymi przez organizacje takie jak OSF czy X/Open. Jeśli przyjąć takie założenie, to jaka jest różnica między jednym, a drugim UNIX-em? Dlaczego użytkownik ma wybierać pomiędzy jednym a drugim UNIXem, jeśli oba są na równi standardowe, a więc w zasadzie chyba identyczne. I tu jak to mówią 'jest pies pogrzebany'. Bo jeśli klient ma wybrać ten, a nie inny system to robi to dlatego, iż ten jeden ma coś czego ten drugi nie posiada. To coś, to jest ta wartość (funkcjonalność, cecha), którą każda firma stara się dodać, aby odróżnić swoją ofertę rynkową od oferty konkurencji. I w ten sposób dochodzimy do wniosku, że rozwojem systemu UNIX (tak jak rozwojem każdego innego produktu sprzedawanego na rynku) powodują dwa przeciwbieżne wektory: jeden to wektor standaryzacji i unifikacji a drugi to wektor odróżniania i wzbogacania o nowe cechy.

Dlatego więc potencjalny klient ma wybrać system Digital UNIX? Z kilku powodów, które można ująć w trzech hasłach:

- Digital UNIX jest systemem z unifikowanym i implementuje największą liczbę możliwych standardów.
- Digital UNIX jest systemem kompletnym i posiada cechy, które sprawiają iż równie dobrze sprawdza się w zastosowaniach typu stacje robocze jak i serwery bazodanowe.
- Digital UNIX jest systemem na wskroś nowoczesnym, wykorzystującym możliwości jakie daje zastosowanie najnowszych technologii, takich jak mikroprocesory 64 bitowe czy jądro oparte o technologię mikrokernels Macha.

Celem niniejszego artykułu nie jest opisywanie systemu operacyjnego UNIX. Zakładam, iż każdy z nas wie co to jest wielodostępny system operacyjny (trzydzieści kilka lat

rozwoju informatyki, dziesiątki jeśli nie setki implementacji i miliony instalacji). Dlatego też nie zamierzam opisywać rzeczy podstawowych i oczywistych. Opiszę natomiast to czym Digital UNIX odróżnia się od reszty UNIXów i w czym jest lepszy. Ważniejsze z nowych cech i właściwości podzielone zostały na kilka grup opisujących kolejno: standardy, łatwość użycia, bezpieczeństwo, technologię jądra systemu, obsługę pamięci masowej oraz niezawodność i dostępność. Treść artykułu dotyczy tylko tych składowych systemu, które wchodzi w skład standardowej dystrybucji Digital UNIX 4.0.

Standardy i otwartość

Digital UNIX wywodzi się bezpośrednio z wersji OSF/1, którego rozwój oparto na elementach różnych implementacji UNIXa. OSF/1 posiadał więc wybrane cechy takich technologii jak AIX, 4.3 BSD, System V i CMU Mach. Digital UNIX 4.0 jest pierwszym systemem, który jednocześnie uzyskał prawo do znaku (*branding*) UNIX 95 i CDE (*Common Desktop Environment*). Jest w pełni kompatybilny ze specyfikacją Single UNIX Specification (dawniej znaną jako Spec1170), nad którą czuwa organizacja X/Open Company, Ltd. i która określa definicje systemu operacyjnego UNIX. Wśród wielu standardów implementowanych przez Digital UNIX 4.0 na wyróżnienie zasługują:

- Implementacja CDE systemu Digital UNIX jest zgodna ze specyfikacją X/Open.
- Rozszerzenia dla aplikacji czasu rzeczywistego są zgodne ze standardem POSIX 1003.b, natomiast mechanizm wątków odpowiada normie POSIX 1003.c.
- Środowisko interfejsu graficznego X Windows jest zgodne z ostatnią wersją standardu X11, czyli z X11R6, zapewniając jednocześnie pełną kompatybilność dla aplikacji zgodnych z poprzednim standardem X11R5.
- System posiada rozszerzenia dające zgodność ze standardem UNICODE (ISO 10646).

*Digital
UNIX 4.0
jest pierwszym
systemem,
który
jednocześnie
uzyskał prawo
do znaku
(branding)
UNIX 95
i CDE
(Common
Desktop
Environment)*

Środowisko CDE zostało stworzone jako bezpośrednia odpowiedź na ekspansję świata pecetów i MS Windows

Ułatwienia w zarządzaniu systemem

Digital UNIX od samego początku otrzymywał wysokie oceny w kategorii niezawodności i zgodności ze standardami. Słabą stroną systemu był natomiast interfejs graficzny i niedostateczna liczba narzędzi graficznych ułatwiających administrowanie systemem. Wersja 4.0 niesie z sobą największą liczbę nowych cech właśnie w tej kategorii. Począwszy od instalacji systemu, poprzez jego konfigurowanie i użytkownię, aż do strojenia mamy doczynienia z przyjaznymi narzędziami pracującymi w trybie graficznym. Ułatwia to znacznie obsługę systemu pozwalając na dokonywanie większości operacji administracyjnych osobom, które wcale nie muszą być 'guru' UNIXa. Zapewniam, że to wcale nie jest żart. Dodatkowym ułatwieniem nauki systemu jest zawarty w nim program Netscape Navigator, który pozwala nie tylko na dostęp do świata World Wide Web, ale także na studiowanie CAŁEJ dokumentacji systemu Digital UNIX 4.0, dostępnej w

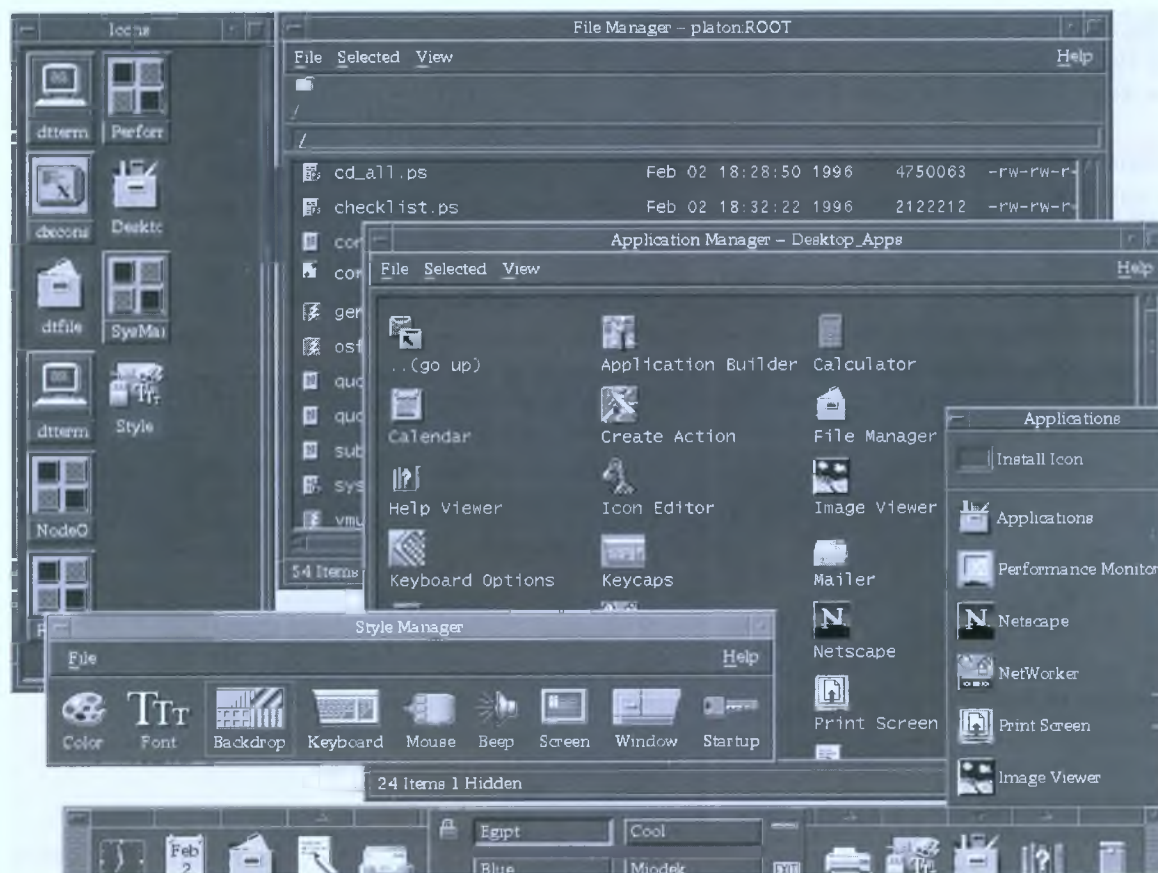
ramach dystrybucji systemu w formacie HTML na płycie CD.

Instalacja systemu

Proces instalacji systemu operacyjnego rozpoczyna się od interakcji z oknem, w którym można określić rodzaj instalacji i opcjonalnie przejść do specyficznego skonfigurowania partycji dyskowych. Interfejs graficzny pozwala na określanie liczby i rozmiarów partycji oraz rodzaju systemu plików dla każdej z nich. W podobny sposób można dokonywać selekcji oprogramowania, które ma zostać automatycznie zainstalowane wraz z systemem. W odpowiednie pola wpisujemy aktualną datę i czas oraz wybieramy kraj z listy - oczywiście Poland. Na koniec można określić hasło konta root i klikając na klawisz 'OK' rozpocząć instalację systemu. Po instalacji system zgłasza się oknem logowania. To wszystko.

Środowisko graficzne

Po zalogowaniu do systemu pierwszą rze-



Rys. 2 Środowisko CDE zapewnia użytkownikom różnych UNIXów spójny interfejs graficzny

Nagrody Hot Iron dla Digital UNIX

Digital UNIX i maszyny klasy AlphaServer i AlphaStation zdobywają co roku główne nagrody najbardziej prestiżowych 'konkursów' komputerowych. Jednym z bardziej znanych 'wyścigów' jest konkurs Hot Iron organizowany przez AIM Technology. Dwukrotnie w przeciągu roku (na wiosnę i na jesień) AIM Technology dokonuje testów różnych maszyn UNIXowych w trzech głównych kategoriach. Kategorie testów określają charakter ich zastosowań: współdzielony system wieloużytkowy, serwer plików i stacja graficzna. Dla każdej z kategorii oceniane są maszyny w kilku kategoriach cenowych. Wynik ogłaszany jest dla systemu o największej wydajności i dla systemu o najlepszym stosunku ceny do wydajności. Łącznie jest więc 20 kategorii szczegółowych. W lutym 1996 Digital zwyciężył w 11 z 20 kategorii, w 6 przypadkach były to maszyny z Alpha Generation z systemem Digital UNIX v3.2c, natomiast w 5 były to maszyny z procesorami Intel'a i systemem SCO UNIX 3.0. W konkursie jesiennym 1995 Digital zwyciężył w 6 kategoriach z 16 ogółem.

Kategoria typu - Multiuser Shared System MIX

System o największej wydajności

\$50,000-\$150,000 - Digital AlphaServer 2100 5/250

powyżej \$150,000 - Digital AlphaServer 8400 5/350

Najlepszy stosunek ceny do wydajności

mniej niż \$25,000 - Digital Prioris XL Server 5100DP

Kategoria typu - File Server MIX

System o największej wydajności

mniej niż \$25,000 - Digital Prioris LX 5120

\$25,000-\$50,000 - Digital Prioris ZX 5133MP/2

powyżej \$150,000 - Digital AlphaServer 2100 5/300 (3CPU)

Najlepszy stosunek ceny do wydajności

mniej niż \$25,000 - Digital Prioris LX 5120

\$25,000-\$50,000 - Digital Prioris ZX 5133MP/2

powyżej \$150,000 - Digital AlphaServer 2100 5/300 (3CPU)

Kategoria typu - General Workstation MIX

System o największej wydajności

mniej niż \$25,000 - Dec 3000 Model 600

powyżej \$25,000 - Digital AlphaStation 600 5/300



czą, którą zobaczymy jest Front Panel. Front Panel ma kształt listwy z kilkoma ikonami najczęściej używanych aplikacji i rozwijanymi menu. Sposób obsługi tego panela jest bardzo prosty i naturalny dla każdego kto pracował z jakimkolwiek środowiskiem okienkowym. W tym momencie, można przystąpić do nauki środowiska Common Desktop Environment typową metodą prób i błędów (czyli mniej lub bardziej przypadkowego klikania). Środowisko CDE zostało stworzone jako bezpośrednia odpowiedź na ekspansję świata pecetów i MS Windows. I trzeba przyznać, że jest to dobra odpowiedź. Od teraz użytkownicy różnych wersji UNIXa mogą mieć ten sam interfejs okienkowy, pozwalający na łatwy dostęp do systemu operacyjnego. Na wyposażenie CDE składa się kilka najszybszych standardowych aplikacji, takich jak:

kalendarz, kalkulator, interfejs poczty zgodny ze standardem MIME (Multipurpose Internet Mail Extensions), edytor tekstu i ikon, terminal, itp. Najważniejsze wydaje się jednak, że mamy menedżer plików (podobny do tego w Windows95) i menedżer aplikacji. Mamy też możliwość własnego definiowania akcji dla menu Front Panela i menedżera aplikacji. Front Panel pozwala nam na operowanie na wielu wirtualnych ekranach. Jeśli chcemy uruchomić wiele aplikacji jednocześnie i mieć do nich łatwy dostęp, wystarczy uruchomić je na różnych wirtualnych ekranach.

Konfiguracja systemu

Środowisko CDE organizuje nam codzienną pracę z UNIXem, tak jak biurko i jego podstawowe wyposażenie. Kiedy jednak

*Digital UNIX
i maszyny
klasy
AlphaServer i
AlphaStation
zdobywają
co roku
główne
nagrody
najbardziej
prestizowych
'konkursów'
komputero-
wych*

przyjdzie do konfiguracji sieci, dysków, drukarek, itp. potrzebne nam są dodatkowe narzędzia. W tym celu Digital UNIX 4.0 wyposażony jest w zestaw narzędzi objętych nazwą Single System Management.

Podczas pierwszego po instalacji logowania do systemu uruchamiany jest program o nazwie SysMan Configuration Checklist. Jest to rodzaj graficznego notatnika z listą kilku narzędzi do konfiguracji poszczególnych elementów systemu. Każde z narzędzi może zostać wywołane z tego miejsca, a fakt ten jest odpowiednio zaznaczony i odnotowany. Na szczególne wyróżnienie zasługują następujące narzędzia konfiguracyjne:

- Network Configuration Application pozwala na konfigurację adresów IP maszyny, listę HOST, adresy routerów IP, itp.

- BIND Configuration Application daje możliwość konfiguracji klienta i serwera serwisów BIND.

- NFS Configuration Application pozwala na konfigurację klienta i serwera serwisów NFS.

- License Manager pozwala na wprowadzanie, aktywowanie i likwidowanie licencji w systemie.

- Account Manager daje dostęp do operacji na kontach użytkowników, z pełną możliwością ich tworzenia, powielania, modyfikowania i usuwania.

- Mail Configuration Application pozwala na konfigurowanie różnych typów serwisów poczty elektronicznej.

- Disk Configuration Application jest tym samym narzędziem, które wykorzystywane jest w procesie instalacji.

Wszystkie narzędzia objęte w ramach Configuration Checklist są także dostępne z poziomu menedżera aplikacji.

Monitorowanie i strojenie systemu

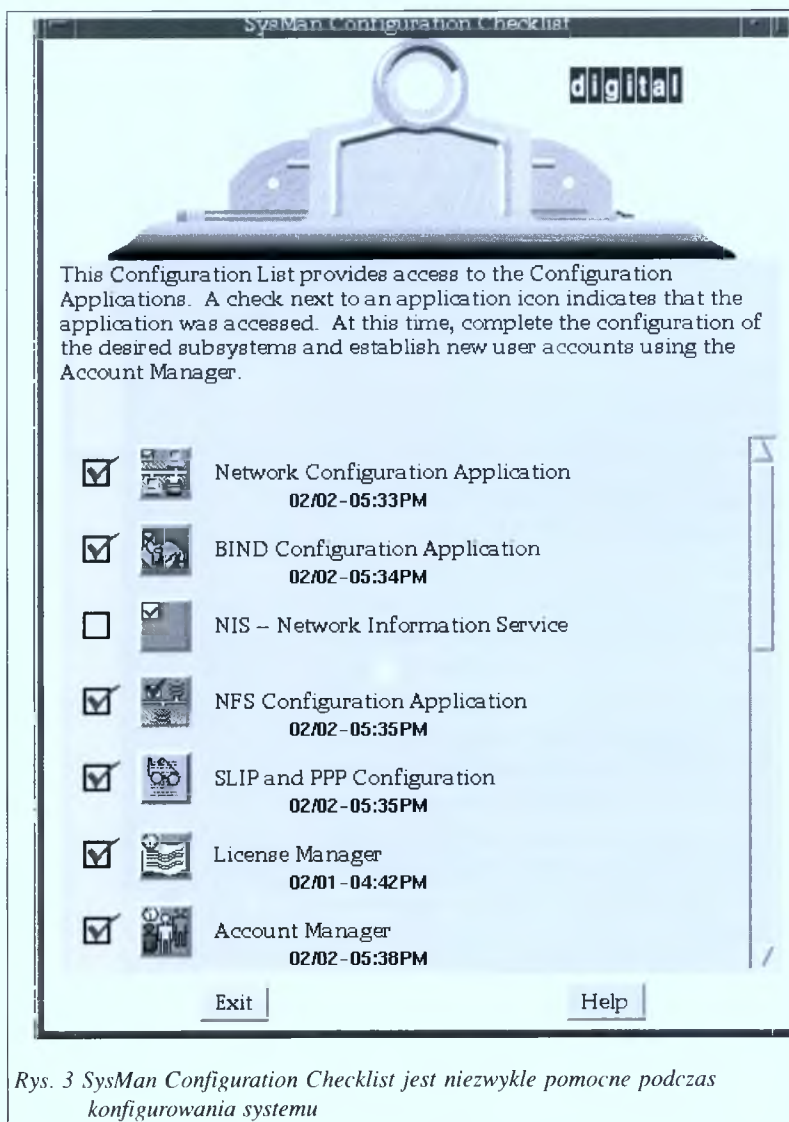
W codziennym kontakcie z systemem istotne staje się monitorowanie, a w szczególności określanie stopnia wykorzystania jego zasobów przez działające aplikacje i użytkowników. Digital UNIX posiada zestaw narzędzi graficznych ułatwiających ten proces.

Do grupy tych narzędzi można zaliczyć:

- System Information wyświetla podstawowe informacje o wykorzystaniu CPU, pamięci, pliku Swap i zajętości poszczególnych partycji dyskowych.

- Kernel Tuner pozwala na modyfikowanie atrybutów podsystemów jądra zgodnych z architekturą Configuration Management Facility systemu Digital UNIX.

- Process Tuner wyświetla listę procesów z określonymi charakterystykami, pozwala na modyfikowanie priorytetów procesów i wysyłanie różnych sygnałów (np. Terminate, Kill, itp.) do wybranych procesów.



Rys. 3 SysMan Configuration Checklist jest niezwykle pomocne podczas konfigurowania systemu

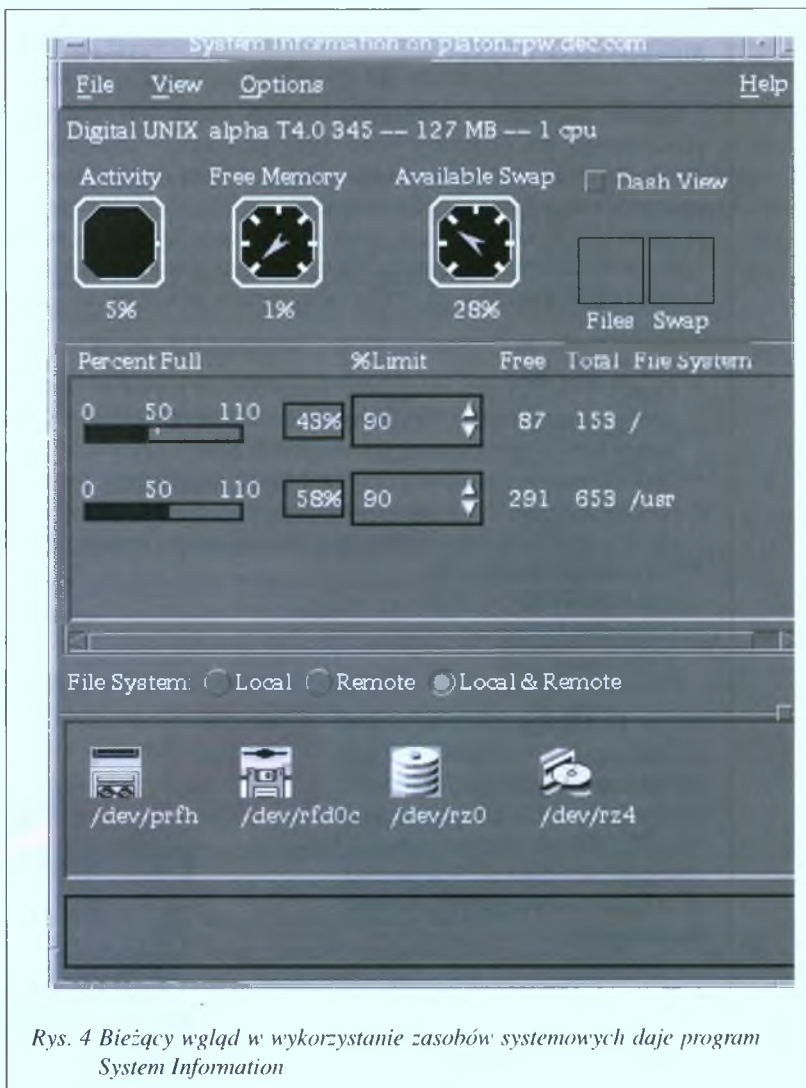
Powyższe narzędzia dają możliwość wglądu w system i bieżącego reagowania na pojawiające się problemy. W przypadku poważniejszych trudności wynikających z obciążenia maszyny przez aplikacje należy posłużyć się pakietem POLYCENTER Performance Manager. Performance Manager jest zawarty w dystrybucji systemu operacyjnego. Jest to bogaty pakiet narzędziowy pozwalający na bieżące monitorowanie systemu pod kątem dowolnie zdefiniowanych charakterystyk. Dodatkową zaletą jest możliwość tworzenia raportów z analiz dotyczących wykorzystania poszczególnych elementów systemu, np. CPU, pamięć, itp.

Mechanizmy zapewniania bezpieczeństwa

Digital UNIX 4.0 odpowiada normom bezpieczeństwa C2 określonym przez NCSC (*National Computer Security Center*). Digital oferuje także wersję MLS+ odpowiadającą poziomowi bezpieczeństwa B1/CMW. Z dodatkowych elementów związanych z zapewnianiem bezpieczeństwa w systemie operacyjnym na uwagę zasługuje mechanizm ACL (*Access Control List*), pozwalający na dokładne kontrolowanie dostępu do zasobów systemu operacyjnego. Mechanizm ACL znany od wielu lat i używany na wielu platformach systemowych nie doczekał się niestety jednoznacznego standardu w świecie systemów UNIXowych. Implementacja ACL w Digital UNIX może być wykorzystywana z poziomu systemu plików AdvFS, UFS i NFS. Drugim elementem zwiększającym bezpieczeństwo systemu operacyjnego jest możliwość definiowania kwot na zasoby jakie mogą być konsumowane (wykorzystywane) przez użytkownika. Kwoty te dotyczą takich zasobów systemu jak: czas CPU, zajętość pamięci operacyjnej, maksymalna liczba otwartych plików. Administrator wyposażony jest w odpowiedni interfejs graficzny pozwalający na określenie kwot dla użytkowników. Mechanizm ten daje możliwość uchronienia systemu operacyjnego przed pochłonięciem jego zasobów przez jednego (bardzo zachłannego) użytkownika.

Technologia i skalowalność

O szybkości i wydajności każdego systemu



Rys. 4 Bieżący wgląd w wykorzystanie zasobów systemowych daje program System Information

operacyjnego decyduje w dużym stopniu jego wewnętrzna architektura. W szczególności architektura jądra systemu odpowiada za skalowalność systemu i jego efektywność w środowisku wieloprocesorowym. Digital UNIX posiada dwie szczególne cechy powodujące, iż jest to system bardzo wydajny przy dużym obciążeniu charakterystycznym dla współczesnych serwerów.

Pierwsza cecha decydująca o szczególnie silnej pozycji Digital UNIXa to fakt, iż jest to w pełni 64 bitowy system operacyjny UNIX. Ta właściwość systemu operacyjnego pozwala na wykorzystanie maszyn AlphaServer z systemem Digital UNIX do zastosowań typu Data Warehousing, gdzie wykorzystywane są bazy danych takich firm jak Oracle, Informix i Sybase posiadające cechy VLM, a co za tym idzie możliwość przechowywania

Dobitnym przykładem możliwości jakie daje duża pamięć operacyjna jest Internet'owy serwer AltaVista

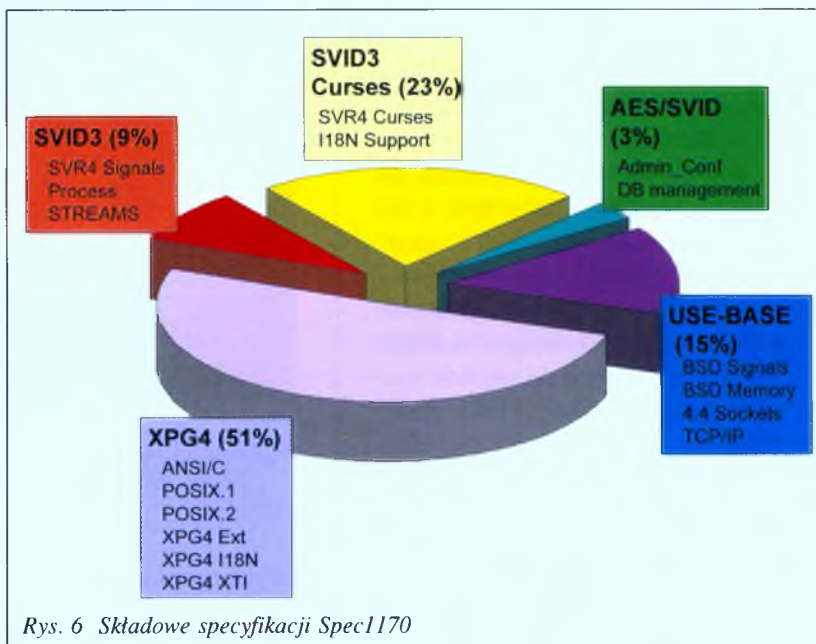
Standaryzacja UNIXa

Rozwój standardów technologii systemu UNIX wytycza organizacja X/Open powstała w 1984 roku jako przeciwwaga dla dominacji firmy IBM na rynku europejskim (założona przez Bull, ICL, Siemens, Olivetti, Nixdorf). Digital był pierwszą korporacją pozaeuropejską, która dołączyła do X/Open. Początkowo organizacja zajmowała się jedynie określaniem standardów zapewniających przenośność aplikacji pomiędzy różnymi wersjami systemu UNIX. Obecnie cele, które stawia sobie X/Open są znacznie szersze i dotyczą szeroko pojętej definicji systemu otwartego. W ramach X/Open działa kilka komitetów skupiających nie tylko producentów, ale także największych klientów na rynku oprogramowania. Obecnie w gestii X/Open znajduje się kilka kluczowych standardów:

- Znak UNIX. X/Open przejęła od firmy Novell prawa do znaku UNIX. Obecnie każda firma może uzyskać prawo do nazywania swojego systemu mianem UNIX o ile spełni wymagania specyfikacji określonej jako Single UNIX Specification (zwanej także Spec 1170).
- CAE (*Common Application Environment*) określa ramy i wizje niezbędnych interfejsów programistycznych systemu otwartego. W ramach CAE specyfikacja XPG (*X/Open Portability Guide*) określa standardy interfejsów i protokołów wraz z zestawem testów. Aktualne wersje to XPG3 i XPG4.
- Technologia standardów organizacji OSF, tj. Motif, CDE, i DCE staje się częścią specyfikacji CAE.

Single UNIX Specification znana też jako Spec1170, zawiera w sobie kompletną specyfikację 1170 interfejsów (w tym 926 programowych API, 70 plików nagłówkowych i 174 poleceń). Spec1170 wywodzi się z kilku wcześniejszych zbiorów technologii:

- X/Open Common Applications Environment, Issue 4 (XPG4),
- UNIX System V Interface Definition, Edition 3 (SVID3),
- OSF Application Environment Specification (AES),
- Use-base czyli BSD Signals, BSD memory, itp.



w pamięci operacyjnej baz rzędu 10 GB. Możliwości takie dają aplikacjom typu DSS (*Decision Support System*) szanse na działanie na bieżąco. Operacje, które jeszcze niedawno wymagały godzin lub dni, obecnie realizowane są w przeciągu minut. Dobrym przykładem możliwości jakie daje duża pamięć operacyjna jest Internetowy serwer AltaVista (<http://altavista.digital.com>), który obsługuje dziennie kilka milionów zapytań do indeksu zbudowanego z 22 milionów stron WWW. Index zajmuje 33 GB z czego 6GB jest buforowane w pamięci operacyjnej.

Modułarna budowa jądra systemu powoduje, że minimalna wymagana pamięć ope-

racyjna dla Digital UNIX to 32 MB, natomiast maksymalnie może to być 14GB, przy czym dotyczy to modelu AlphaServer 8400 dostępnego na rynku w pierwszej połowie 1995 roku. Na koniec 1996 roku można się spodziewać konfiguracji z podwojoną pamięcią operacyjną.

Drugim elementem decydującym o sile systemu jest skalowalność dla maszyn wieloprocessorowych. Celem jest tu osiągnięcie sytuacji w której dokładanie kolejnych procesorów do maszyny powoduje, iż moc obliczeniowa wzrasta jak najbardziej liniowo. Jądro systemu Digital UNIX jest oparte o technologię Mach z optymalizacją szeregowania wątków, co w rezultacie daje doskona-

łą skalowalność dla maszyn SMP. Dowodem na skalowalność systemu są wyniki takich testów jak SPECrate_int92, SPECrate_fp92 oraz nowsze testy SPECint_rate95, SPECfp_rate95. Pośrednio także wyniki testów TPC-C dowodzą skalowalności architektury SMP maszyn AlphaServer z systemem Digital UNIX.

Obsługa pamięci masowych

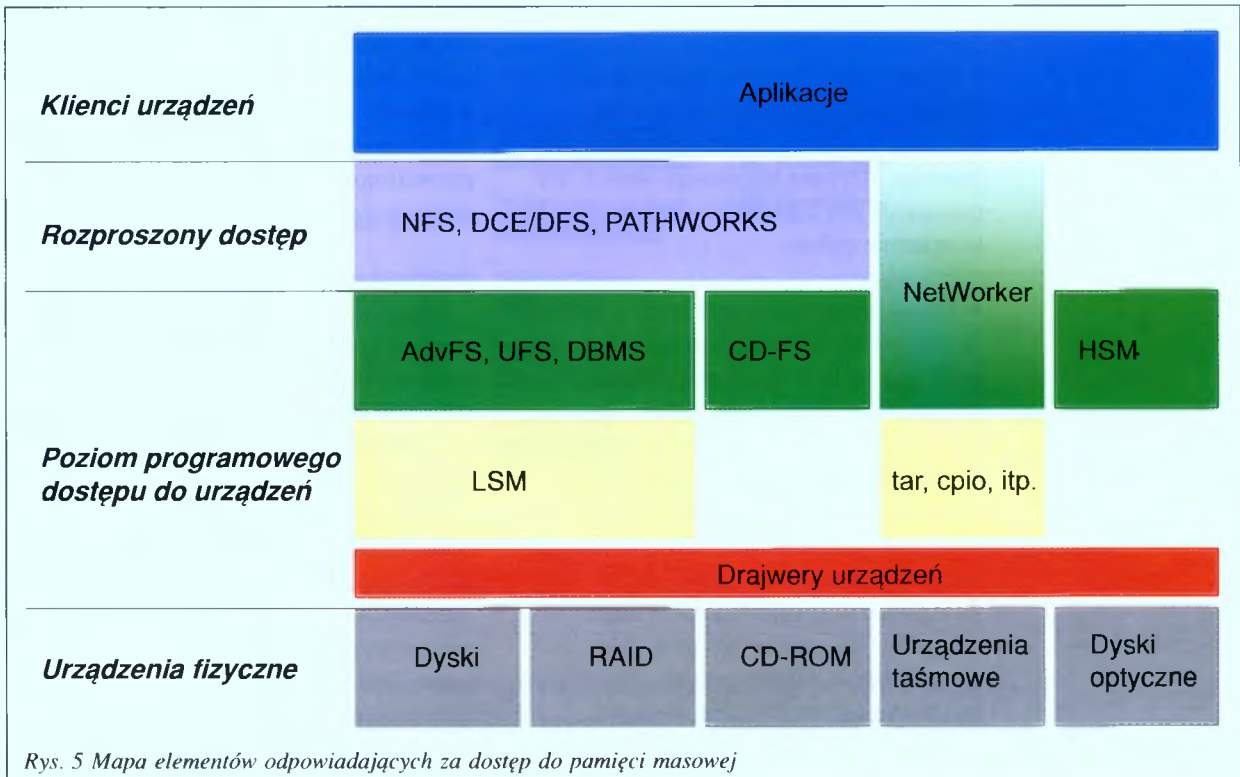
Gromadzenie dużych ilości informacji w trwały sposób jest jednym z podstawowych powodów istnienia systemów komputerowych. Z punktu widzenia użytkowników ważne jest, aby podsystem zarządzania pamięciami masowymi spełniał potrzeby użytkowników i administratorów w zakresie pojemności, integralności danych, szybkości działania i zgodności ze standardami. Dla spełnienia tych potrzeb niezbędne jest rozważne stosowanie różnych dostępnych technologii, poczynwszy od dysków półprzewodnikowych (szybkie i drogie), poprzez dyski magnetyczne i dyski optyczne, aż do urządzeń taśmowych (tanie i wolne). Rozwiązania zawarte w systemie Digital UNIX pozwalają na zaspokojenie potrzeb różnych klas użytkowników, poczynwszy od użytkownika pracującego na niezależnej stacji roboczej z kilkoma dyska-

mi, aż po serwery sieciowe z dziesiątkami dysków, szafami płyt optycznych i pamięci taśmowych. Pełna mapa elementów zapewniających w Digital UNIXie zarządzanie i dostęp do pamięci masowej przedstawiona jest na rysunku 5. Poziom fizycznych urządzeń nie wymaga specjalnego opisu, mogą to być typowe urządzenia oparte o standard SCSI i SCSI-2, takie jak dyski magnetyczne, konfiguracja RAID, CD-ROM, urządzenia taśmowe (np. DAT, DLT) i szafy dysków optycznych. Poziom drajwerów urządzeń wykorzystuje komponent nazwany Common Access Method pozwalający na łatwe tworzenie nowych drajwerów. Digital UNIX posiada możliwość dynamicznego konfigurowania i instalowania drajwerów urządzeń bez potrzeby relinkowania kernela systemu operacyjnego.

Logical Storage Manager

Typowym dla systemu UNIX jest instalowanie na określonej partycji dyskowej (dysk fizyczny lub jego część) systemu plików lub też wykorzystywanie jej bezpośrednio z aplikacji (system zarządzania bazą danych). Powoduje to, iż rozmiar pojedynczej partycji ograniczony jest do wielkości dysku. LSM pozwala na obejście tego ograniczenia i definiowanie większych partycji (logicznych dy-

Jądro systemu Digital UNIX jest oparte o technologię Mach z optymalizacją szeregowania wątków, co w rezultacie daje doskonałą skalowalność dla maszyn SMP



*UNIX
jest obecnie
dominującą
platformą
rozwiązań
serwerowych*

sków). LSM daje dodatkowo możliwość parcelacji danych na wielu dyskach (*striping*) i dublowania danych na wielu dyskach (*mirroring* lub *shadowing*). Bez względu na sposób skonfigurowania i wykorzystania powyższych właściwości dla oprogramowania wyższego poziomu w dalszym ciągu egzystuje partycja dyskowa. LSM jest bardzo wydajny i szybki, nie wnosi przy tym swoim działaniem opóźnień w dostępie do dysków. Dla ułatwienia konfigurowania logicznych dysków w ramach systemu dostarczane jest narzędzie z graficznym interfejsem do zarządzania LSM.

Advanced File System

System plików stanowi jeden z zasadniczych komponentów każdego systemu operacyjnego. Zwłaszcza dla systemów pełniących rolę serwerów (aplikacyjnych, sieciowych czy bazodanowych) posiadanie dobrego systemu plików staje się elementem podstawowym. Obok tradycyjnego dla UNIXów systemu plików o nazwie UFS, Digital UNIX oferuje bardziej zaawansowany system o nazwie AdvFS. AdvFS jest w pełni 64 bitowym systemem plików charakteryzującym się następującymi właściwościami.

- Wysoka dostępność, dzięki możliwości dodawania i usuwania dysków z systemu plików oraz archiwizowania danych bez potrzeby zatrzymywania pracy systemu.
- Elastyczność w powiązaniu plików w logicznej strukturze katalogów z ich fizyczną lokalizacją. Administrator może zmieniać fizyczną lokalizację danych bez naruszenia ich logicznego umiejscowienia w systemie plików.
- Elastyczność konfiguracji bez względu na stosowaną technologię urządzeń fizycznych, czy będą to dyski magnetyczne, optyczne czy też półprzewodnikowe.
- Szybkie uruchamianie, dzięki wykorzystaniu mechanizmu podobnego do tego jaki stosowany jest przez bazy danych. Informacja o strukturach dyskowych jest zapisywana w formie transakcyjnego dziennika. Ta właściwość powoduje, iż system plików AdvFS 'powstaje' w przeciągu sekund bez względu na wielkość zasobów dyskowych. AdvFS jest 'nieczuły' na wyłączanie maszyny bez zatrzymania

systemu i gwarantuje przy tym pełną integralność danych.

- Wysoka wydajność gwarantowana jest z jednej strony poprzez mechanizm elastycznej konfiguracji fizycznej lokalizacji danych, z drugiej natomiast AdvFS posiada mechanizm automatycznej defragmentacji.

UNIX File System

UFS to 64 bitowa implementacja tradycyjnego systemu plików wywodzącego się z BSD. Wszędzie tam, gdzie nie trzeba zarządzać dużymi zasobami dyskowymi UFS jest alternatywą dla znacznie potężniejszego systemu jakim jest AdvFS.

CD-ROM File System

Nośniki CD-ROM stały się głównym medium dystrybucji oprogramowania i dokumentacji. Praktycznie każda maszyna UNIXowa jest wyposażona w urządzenie CD-ROM. CD-ROM File System systemu operacyjnego Digital UNIX jest zgodny ze standardem ISO 9660 w obu jego wariantach "High Sierra" i "Rock Ridge".

Memory File System

Digital UNIX, jak wiele innych systemów w automatyczny sposób przechowuje w pamięci operacyjnej aktywne bufor I/O w celu zwiększenia szybkości obsługi urządzeń zewnętrznych. Częstość jednakże okazuje się, że świadome buforowanie całych plików w pamięci operacyjnej może znacznie wpłynąć na szybkość działania aplikacji. MFS pozwala na mapowanie wybranych plików w pamięci.

NetWorker

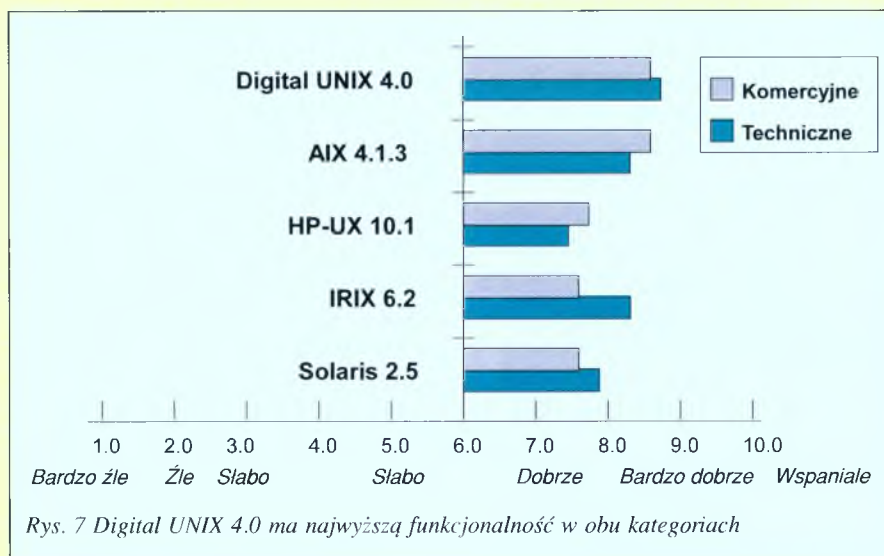
Preferowanym rozwiązaniem dla archiwizowania i odtwarzania danych w środowisku sieciowym Digital UNIX jest pakiet POLY-CENTER NetWorker. NetWorker powstał w kooperacji z firmą Legato Systems. Pakiet daje możliwość archiwizowania i odtwarzania danych na wielu platformach UNIXowych i innych (pecety, OpenVMS, itp.), wykorzystując do tego celu różne urządzenia taśmowe (4mm, 8mm, 0.25" lub DLT), dyski magnetyczne i optyczne. NetWorker wyposażony jest w interfejs graficzny, który ułatwia wykonywanie operacji administracyjnych związanych z archiwizacją. Do głównych zalet pakietu można zaliczyć:

Digital UNIX - opinia rynku

Digital UNIX zdobył na przestrzeni ostatnich lat wiele wyróżnień i bardzo pochlebnych opinii różnych organizacji zajmujących się badaniami rynku komputerowego. Wśród ostatnich ocen na uwagę zasługuje raport firmy D.H. Brown Associates, Inc. z marca 1996. Raport nosi tytuł *"Digital UNIX 4.0 Secures Meaningful Differentiation"* i stanowi ocenę funkcjonalności Digital UNIXa w odniesieniu do innych dostępnych na rynku systemów operacyjnych klasy UNIX. Systemy oceniano w dwóch kategoriach: komercyjnej i technicznej. Dla każdej z kategorii określono 9 następujących cech:

- przenośność aplikacji,
- zarządzanie zasobami pamięci masowej,
- bezpieczeństwo,
- budowa jądra systemu,
- architektura SMP,
- właściwości sieciowe,
- serwisy aplikacji,
- interfejs użytkownika,
- zarządzanie systemem.

W kategorii systemów komercyjnych brano pod uwagę te czynniki systemu, które są najważniejsze z punktu widzenia systemu pełniącego rolę serwera. Kategoria techniczna natomiast to typowe zastosowania dla stacji graficznych. Choć w obu kategoriach badano te same cechy funkcjonalności to jednak przykładano dla nich różne wagi, i tak np. z punktu widzenia zastosowań komercyjnych ważniejsze jest zarządzanie pamięciami masowymi niż interfejs użytkownika, natomiast z punktu widzenia zastosowań technicznych jest dokładnie odwrotnie. W obu kategoriach systemów Digital UNIX jest lepszy od średniej (wyliczonej z ocen wszystkich systemów) w ośmiu z dziewięciu powyższych cech.



Autorzy raportu wysoko ocenili zwłaszcza takie cechy systemu Digital UNIX jak bezpieczeństwo, skalowalność dla maszyn SMP, przenośność aplikacji (tzn. zgodność ze standardami) i możliwość zarządzania zasobami pamięci masowej

*Digital UNIX
oferuje
bardziej
zaawansowa-
ny system
o nazwie
AdvFS*

*AdvFS
jest w pełni
64-bitowym
systemem
plików*

Digital UNIX zachowując wszystkie zalety otwartości i standardowości, zapewnia swym użytkownikom kompletne środowisko do tworzenia niezawodnych, wysoce wydajnych systemów

- możliwość predefiniowania operacji, które wykonują się cyklicznie,
- właściwość dyskretnego określania danych, które mają zostać odtworzone,
- zdalna koordynacja operacji dla wielu maszyn w sieci,
- wykonywanie archiwizacji bez przerywania dostępu do dysków,
- możliwość inkrementacyjnego archiwizowania danych.

Network File System

NFS jest *de facto* standardem w świecie UNIXa pośród systemów umożliwiających rozproszony (w sieci komputerowej) dostęp do plików. Digital UNIX może pełnić rolę serwera i klienta w sensie NFSa i jest zgodny ze standardem protokołu w wersji 2 i 3. Wykorzystanie serwerów z systemem Digital UNIX jako serwerów NFS daje możliwość wykorzystania wszystkich dobrodziejstw tego systemu (64 bity, AdvFS) z poziomu maszyn UNIXowych nie posiadających tych cech.

DCE Distributed File System

NFS choć jest powszechnie stosowany to nie spełnia jednak wymagań jakie stawiają przed nim duże, rozległe instalacje sieciowe. NFS sprawdza się dobrze w sieciach typu LAN, natomiast dla konfiguracji WAN dużo lepszy jest system DCE DFS. Architektura systemu DFS jest przystosowana do działania w dużych instalacjach sieciowych i wykorzystuje w pełni inne składowe infrastruktury DCE, takie jak DCE Security i DCE CDS/GDS (*Cell Directory Services/Global Directory Services*). DFS zapewnia bramki dla systemów wykorzystujących NFS. Opcja serwera i klienta DFS zawarta jest standardowo w systemie Digital UNIX 4.0.

PATHWORKS

Integracja maszyn klasy PC ze światem UNIXa może przebiegać w dwojaki sposób. Po pierwsze można implementować na PC-tach rozwiązania z UNIXa, czyli zapewnić serwisy NFS wykorzystujące protokół TCP/IP. Rozwiązanie to jednak nie jest preferowane przez użytkowników PCtów. Bardziej po-

wszechnie i popularne jest więc rozwiązanie, które wykorzystuje naturalne dla PCtów protokoły Novell NetWare lub Microsoft LAN Manager. Oprogramowanie PATHWORKS daje możliwość pełnej integracji maszyn Digital UNIX z systemami klienckimi, takimi jak DOS, Windows 3.1, Windows 95, OS/2 i Macintosh.

Niezawodność i dostępność

Jedną z silniejszych stron systemu Digital UNIX jest stopień jego niezawodności i dostępności gwarantowanej w przypadku wystąpienia awarii. Digital jest twórcą współczesnej technologii klastrów, rozwijanej od kilkunastu lat na platformie OpenVMS. Logiczną konsekwencją jest więc implementacja technologii klastrów w systemie Digital UNIX. Od kilku lat realizowany jest program o nazwie AdvantageCluster, który zakłada rozwój systemu Digital UNIX w kierunku klastrów o funkcjonalności podobnej do tej jaką obecnie posiadają klastry OpenVMS. Szczegóły dotyczące tego programu oraz technologii znajdziecie Państwo w artykule Artura Stefanowicza "Klastry - specjalność Digitala".

Podsumowanie

Świat użytkowników komputerów, a w szczególności programistów i administratorów dzieli się na frakcje zwolenników tego lub innego systemu. UNIX od początku swego istnienia miał tyle samo wiernych (często wręcz fanatycznych) zwolenników co przeciwników (równie fanatycznych). To co jednych pociągało w tym systemie, czyli jego standardowość i niezależność od jednego producenta, drugim odstraszało. Wiele cech systemu, które powodowały iż można go łatwo modyfikować i integrować z innymi systemami powodowało, że użytkownicy komercyjnych systemów trzymali się z daleka od UNIXa. Te czasy odchodzą w niepamięć. UNIX jest obecnie dominującą platformą rozwiązań serwerowych. Digital UNIX jest pośród tych systemów liderem, zachowując wszystkie zalety otwartości i standardowości, zapewnia swym użytkownikom kompletne środowisko do tworzenia niezawodnych, wysoce wydajnych systemów. Digital UNIX to standard, któremu można zaufać.

Piotr Sobolewski

Klastry

- specjalność Digitala

W roku 1983 Digital wypuścił na rynek unikalne rozwiązanie łączące w sobie walory odporności konwencjonalnych sieci komputerowych z wydajnością maszyn MPP (Massively Parallel Processing), ale działające na bazie standardowego sprzętu. Produkt nazywał się VAXcluster i dał początek rozwojowi tej ważnej technologii informatycznej. Od tamtej pory, Digital jest powszechnie uznawany za niekwestionowanego lidera technologii klastrów, a system VMScluster z 35.000 instalacji jest niedoścignionym wzorcem rynkowym (patrz Ramka "Porównania technologii klastrów").

Ostatnio, technologia klastrów przeżywa szybki rozwój. Wiele firm oferuje produkty w większym lub mniejszym stopniu zbliżające się funkcjonalnie do VMScluster. Jakie czynniki powodują rosnące zainteresowanie tą technologią? Do najczęściej wymienianych należą niezawodność, wydajność i skalowalność systemów komputerowych.

Niezawodność

W potocznym rozumieniu system jest niezawodny wtedy, gdy *nie zawodzi* czyli działa zawsze wtedy, kiedy jest potrzebny. Systemy są jednak nieuchronnie narażone na awarie, z czym musimy się raczej pogodzić. Skoro więc muszą się psuć, to niech chociaż nie zdarza się to zbyt często i nie trwa zbyt długo! Tak więc kolejnym przybliżeniem znaczenia niezawodności może być procent czasu, w którym system działa poprawnie w stosunku do całości czasu pomiaru. Jest to dość powszechnie stosowana miara dostępności systemu (patrz Tabela 2). Zauważmy, że na *niedostępność* systemu, obok awarii wpływają wszelkie czynności administracyjne realizowane w trybie off-line. I jest to najzupełniej logiczne - w obu wypadkach system jest po prostu niedostępny.

Okazuje się, że niedostępność może być kosztowna. Badania prowadzone przez firmę Oracle wśród 400 dużych firm zachodnich wykazały, że średni koszt niedostępności systemu wynosi tam 1400\$/minutę. Powszechnie przyjmuje się, że pojedynczy system UNIX może dostarczyć 99.5% dostępności. Oznacza to ok. 2600 minut niedostępności w roku co daje okrągłą sumkę 3.6mln\$ strat. Dla porównania, przeciętne rozwiązanie typu klastrów pozwala zredukować niedostępność do ok. 50 minut/rok, co oznacza zaledwie 70 k\$ kosztów związanych z unieruchomieniem systemu.

Porównania technologii klastrów

Eksperti są zgodni co do dominacji technologii VMScluster na rynku. Raport Gartner Group (listopad 1993) stwierdza wręcz, że w swych opracowaniach do określenia standardów konfiguracji klastrów posługuje się funkcjonalnością systemu VMScluster.

W przeprowadzonej przez Aberdeen Group (raport z grudnia 1992) analizie dostępnych rozwiązań klastrowych 6 wiodących dostawców, VMScluster zwycięża w 23 na 24 kategorie:

Niezależne porównanie technologii klastrów

- | | |
|---------------------|------------------------------|
| - OpenVMS Clusters | - HP9000 HP-UX SwitchOver UX |
| - IBM MVS Sysplex | - HP3000 MPE/IX Shareplex/IX |
| - IBM RS/6000 HACMP | - Sun Solaris |

... VMS cluster numer 1 w 23 na 24 Kategorie

Konfiguracja i eksploatacja	Zarządzanie	Niezawodność i dostępność	Dzielenie zasobów
• Modyfikacje • Przezroczystość • Standardy • Nazwy globalne • Scalbilność	• Administracja • Konsola • Klastrowe kolejki • Wsadowe i drukarkowe • Klastrowy system plików • Konfiguracje wielo-klastrowe • Praca z różnymi systemami operacyjnymi	• Automatyczne obejście awarii • Brak pojedynczych punktów awarii • Aplikacje • Klastrowe OLTP • Odporność na katastrofy (disaster-tolerance)	• Klastrowy system plików • Klastrowe kolejki • Niezależne od maszyn urządzenia dyskowe • Optymalizacja rozłożenia obciążenia • Wspólna baza autoryzacyjna

Z kolei w raporcie WorkGroup Technologies, Inc. (kwiecień 1995) VMScluster zdobywa 52 punkty na 57 możliwych. Podsumowanie analizy ilustruje tabela 1.

Digital jest powszechnie uznawany za niekwestionowanego lidera technologii klastrów

Klaster to grupa maszyn pracujących wspólnie jako jednolity system

Wydajność

Może się to wydawać niemożliwe, ale są zastosowania, dla których niełatwo dobrać pojedynczą maszynę odpowiedniej mocy. Systemy masowego przetwarzania transakcyjnego z dziesiątkami tysięcy równoczesnych użytkowników i bazami rzędu 100GB, zaawansowane aplikacje symulacji inżynierskich - oto tylko niektóre z takich zastosowań. Wymagają olbrzymich mocy obliczeniowych, przewyższających możliwości nawet największych *mainframe'ów*.

Skalowalność

Co zrobić, kiedy maszyna osiąga kres wydajności a obciążenie stale rośnie? Najprostszym rozwiązaniem jest naturalnie wymiana na inną, większą. Niestety, może się to wiązać z równoczesną wymianą szefa IT na bardziej przewidującego. Zakup maszyny posiadającej moc wielokrotnie przewyższającą początkowe zapotrzebowanie także nie wydaje się łatwy do ekonomicznego uzasadnienia.

Idealnym rozwiązaniem byłaby możliwość stopniowej rozbudowy konfiguracji sprzętowej, tak aby rosnąca moc obliczeniowa równoważyła rosnące obciążenie. Innymi słowy chcielibyśmy, aby system był skalowalny - a więc dawał się rozbudować wtedy, gdy jest to konieczne.

Co to jest klaster

Klaster to grupa maszyn pracujących wspólnie jako jednolity system, maksymalizując w ten sposób wydajność i minimalizując niedostępność usług z powodu awarii poszczególnych komponentów. Technologia klastrów jest rezultatem racjonalnego kompromisu między różnymi koncepcjami zwiększania wydajności konfiguracji z jednej, a metodami zagwarantowania jak najlepszej niezawodności z drugiej strony.

Maszyny wchodzące w skład klastra dzielą swoje zasoby: dyski, drukarki, kolejki wsadowe (patrz Rysunek 1). Aplikacje, wykorzystując odpowiednie mechanizmy synchronizacji mogą wykonywać się równolegle na wielu maszynach. Dla użytkownika, klaster jest przezroczysty: wszystkie zasoby postrzega jako lokalne, dzięki centralnej bazie autoryzacji loguje się na dowolną maszynę. Dla administratora, klaster jest zarządzany

Tabela 1

Produkt	Łączna liczba zdobytych punktów
Digital Equipment Corp. OpenVMS Cluster (VMScluster)	52
IBM Parallel Sysplex	43
IBM RS/6000 HACMP	36
AT&T GIS LifeKeeper FRS & DLM	33
HP MC/Service Guard & Lock Manager	30
Sun Microsystems SPARCcluster PDB	24

Raport Gartner Group z lipca 1995 również podkreśla przewagę VMScluster. W szczególności, w kategorii systemów wysokiej dostępności (czyli działających w trybie 24x365) nasze rozwiązanie zdecydowanie wyprzedziło systemy takich firm jak HP, IBM, Tandem i Sun.

centralnie, bez zmyru odrębnych zbiorów konfiguracyjnych na każdej z maszyn.

Klaster jako forma sieci

Technologia klastra wygląda zatem na szczególnie, zoptymalizowany przypadek sieci. Maszyny komunikują się za pośrednictwem określonego medium, mają protokoły, adresy itp. Oto kilka istotnych różnic, dzięki którym optymalizacja jest możliwa:

- Konfiguracje klastrów są zazwyczaj homogeniczne, z jednym typem maszyn i tym samym systemem operacyjnym (wyjątkiem jest VMScluster, w skład którego mogą wchodzić zarówno systemy OpenVMS Alpha jak i OpenVMS VAX). Sieci łączą z reguły rozmaite typy maszyn.
- Liczba węzłów klastra nie jest zwykle duża (VMScluster może mieć co najwyżej 96 maszyn, inne implementacje najczęściej do 8 węzłów). Sieci obejmują dziesiątki tysięcy maszyn.
- Ponieważ klastry nie muszą do wzajemnej komunikacji implementować generycznych protokołów takich jak np. TCP/IP, możliwe jest radykalne uproszczenie formy komunikacji. Oczywiście, do celów zewnętrznej komunikacji klastry najczęściej dostarczają standardowe protokoły komunikacyjne.

Klaster a niezawodność

Podstawą niezawodności klastra jest redundancja sprzętu obsługiwana przez specjalizowane oprogramowanie. Awaria dowolnej maszyny w klastrze nie powoduje przerw w dostępności systemu, o ile istnieje jeszcze choć jedna sprawna. Oprogramowanie klastra dynamicznie dokonuje rekonfiguracji uwzględniając utratę jednostki przetwarzania i rozkładając obciążenie na pozostałe maszyny. Digital oferuje rozwiązania klaszrowe umożliwiające implementację różnych strategii *disaster-tolerance*, dzięki połączeniom opartym na FDDI. W takiej konfiguracji, poszczególne jednostki klastra mogą być od siebie oddalone o 40km. Obok redundancji maszyn, bardzo ważne jest także zapewnienie odporności na awarie układów dyskowych. Rozwiązania Digitala w tym zakresie opierają się na technologii RAID oraz ich softwarowym ekwiwalentom (oprogramowanie *Volume Shadowing* oraz *Disk Striping*). Pozycję klastrów na tle innych architektur prezentuje Rysunek 2.

Klaster a skalowalność

Technologia klastrów oferuje efektywne i ekonomiczne rozwiązanie problemu skalowalności. Dodanie nowej maszyny do konfiguracji to nie tylko większa moc obliczeniowa, ale i większa wydolność podsystemu I/O. Jest to podstawowa różnica w porównaniu ze skalowalnością SMP, gdzie dodanie nowego procesora nie dodaje nowych kanałów I/O. Wieloletnie badania konfiguracji VMScluster pokazują, że dodając nową maszynę do klastra można oczekiwać wzrostu wydajności rzędu 75-85%. Wraz z liczbą maszyn, zysk ten powoli maleje, głównie ze względu na nakłady ponoszone na wzajemną komunikację i synchronizację. Dotykamy tutaj kluczowego dla technologii klastrów zagadnienia, jakim jest efektywność utrzymania zbieranych maszyn w jednolitej i inteligentnej całości.

Technologie komunikacji konfiguracji klastrów

Jak to się dzieje, że maszyny tworzące klaster nie pogubią się, że wspólne zasoby są obsługiwane w sposób gwarantujący ich logiczną spójność, a w przypadku awarii nie tracą integralności? Otóż zajmuje się tym specjalne oprogramowanie tworzące fundament konfiguracji klastrów, odpowiedzialne między innymi za globalną synchronizację

Tabela 2 Stopnie dostępności systemów

Architektura	Maksymalna dostępność	Niedostępność podczas awarii	Niedostępność na rok (w minutach)
Praca ciągła (ideał)	100%	zero	zero
Fault-tolerant	99.9999%	kilka cykli maszynowych	0.5-5
Klaster	99.9-99.999%	sekundy do kilku minut	5-500
Systemy podwyższonej dostępności	99.9%	minuty	500-10,000
Pojedynczy system	99.5%	minuty do kilku godzin	2,600-10,000

maszyn i zasobów. Jednym z centralnych komponentów oprogramowania klastra jest DLM (Distributed Lock Manager), komponent regulujący dostęp do dzielonych zasobów takich jak pliki, bazy danych, kolejki BATCH czy drukarki. Tego rodzaju rozproszona synchronizacja wymaga pewnej formy fizycznego połączenia maszyn, umożliwiającej implementację protokołu DLM. Jasne jest, że przy rosnącym obciążeniu, DLM będzie wносił znaczny wkład do ogólnego protokołu konfiguracji klastra, będąc zmuszonym do dyktowania dostępu do dziesiątek tysięcy zasobów na kilku czy kilkadziesiąt maszynach równocześnie. Komunikacja musi posiadać więc dostateczną szybkość, aby wszystko działało sprawnie.

Digital oferuje wiele technologii łączenia maszyn w konfiguracji klastrów: Ethernet, FDDI, CI (Computer Interconnect), DSSI (Digital Storage System Interconnect) oraz najnowszy szlagier: Memory Channel on PCI (MC-PCI). Niedługo dostępne będzie połączenie ATM. Każda technologia ma swoje zalety i wady: Ethernet jest najtańszy ale i najwolniejszy, FDDI oferuje sporą szybkość i umożliwia separację maszyn na odległość do 40km (konfiguracje disaster tolerant), ATM daje dużą przepustowość i niezależność geograficzną za cenę skomplikowanej infrastruktury, MC oferuje doskonałą szybkość i przystępną cenę jednak na ograniczonym obszarze.

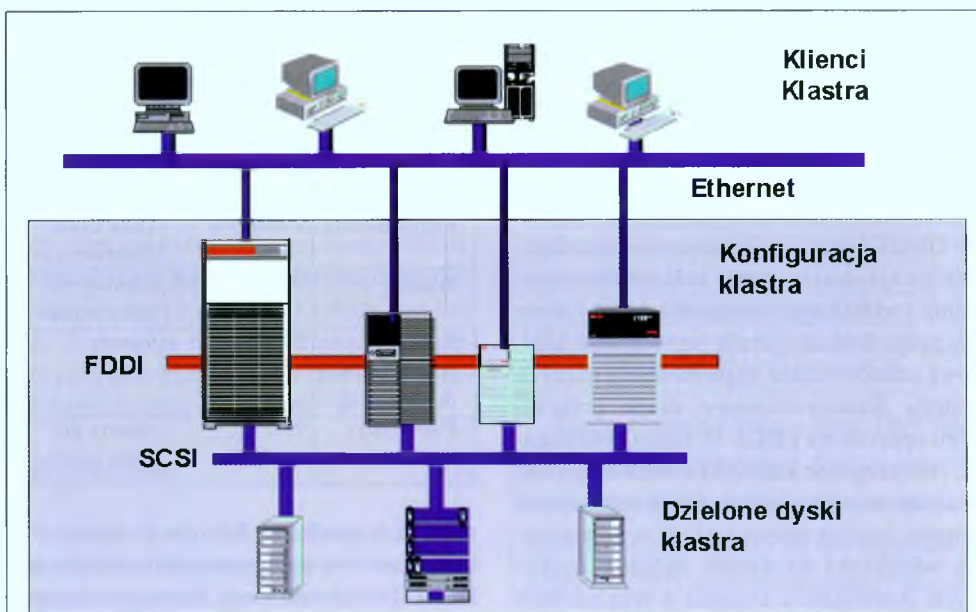
Technologie urządzeń dyskowych

Obok technik łączenia maszyn, ważną rolę w konfiguracji klastrów odgrywa sposób podłączenia urządzeń dyskowych. Aby spełnić wysokie wymagania niezawodnościowe, ja-

Podstawą niezawodności klastra jest redundancja sprzętu obsługiwana przez specjalizowane oprogramowanie

Wieloletnie badania konfiguracji VMSccluster pokazują, że dodając nową maszynę do klastra można oczekiwać wzrostu wydajności rzędu 75-85%

Jednym z centralnych komponentów oprogramowania klastra jest DLM (Distributed Lock Manager)



Rys. 1 Przykładowa konfiguracja Klastra

sne jest, że dyski nie powinny być lokalne w ramach pewnej maszyny, jej awaria uniemożliwiłaby bowiem dostęp do takiego urządzenia. Dyski w klastrze są zewnętrzne i posiadają niezależne zasilanie. Digital proponuje w tym zakresie rodzinę produktów StorageWorks. Wśród dysków, na pierwszy plan wysuwają się aktualnie urządzenia SCSI, tanie i powszechnie dostępne. Dlatego też Digital wśród kilku standardów podłączania urządzeń dyskowych do klastra obsługuje także SCSI.

Technologia Memory Channel

Stosowanie konwencjonalnych technologii sieciowych takich jak Ethernet, FDDI czy ATM wydawałoby się wystarczającą formą połączenia maszyn klastra. Jednak rosnąca szybkość procesorów nieustannie pogłębia dysproporcję między wydajnością CPU a wydajnością sieci. Dotyczy to zwłaszcza dwóch parametrów pomijanych często w podobnych analizach: opóźnienia komunikatu (latency) i narzutu CPU (overhead). Opóźnienie komunikatu to czas, jaki upłynął od momentu zainicjowania operacji komunikacyjnej do jej zakończenia. Narzut CPU to czas procesora skonsumowany przez operację komunikacyjną. Szczegółowe badania wykazały¹, że opóźnienie i narzut CPU są dla FDDI 1000-krotnie większe niż dla pamięci RAM (patrz Tabela 2).

Aby osiągnąć dobry poziom skalowalności klastra, niezbędne jest radykalne poprawie-

nie szybkości komunikacji między-maszynowej, we wszystkich jej aspektach.

Podstawowy problem komunikacji systemów klastrów poprzez konwencjonalne systemy sieciowe polega na ogromnym narzucie konsumpcji czasu CPU, wprowadzanym przez poszczególne warstwy oprogramowania sieciowego, głównie warstwę aplikacyjną, właściwy protokół sieciowy oraz drajwer. Dwukierunkowa transmisja 32-bajтового komunikatu po TCP/IP na FDDI trwa tyle co wykonanie 150.000 instrukcji przez procesor. Jeśli weźmiemy pod uwagę, że opóźnienie i narzut są zbliżone, to okaże się że właśnie wielowarstwowość oprogramowania sieciowego stoi na przeszkodzie uzyskania wysokich szybkości komunikacji. Dodatkowo, komunikacja protokołów klastra ma dla wydajności konwencjonalnych sieci destrukcyjną charakterystykę obciążenia: jest to bardzo duża liczba małych komunikatów.

Memory Channel

Technologia Memory Channel PCI (MC-PCI) to efekt współpracy Digital'a i Encore Computer Corp. Celem projektu sieci MC-PCI w Digital'u było stworzenie infrastruktury sieciowej radykalnie redukującej opóźnienie i narzut klasycznych rozwiązań przy równoczesnym wykorzystaniu niedrogich, standardowych komponentów. Sieć ta nie musi być uniwersalna, a wręcz przeciwnie - ściśle skoncentrowana na potrzebach komunikacji klastra. Dzięki temu, można było przyjąć awangardowe rozwiązania i techniki

optymalizacyjne. Uzyskane wyniki są imponujące - uzyskano 1000-krotną redukcję narzutu CPU i 100-krotną redukcję opóźnienia (w stosunku do standardowych sieci).

MC-PCI implementuje formę rozproszonej pamięci dzielonej na bazie technologii *reflective memory* i PCI. Rysunek 3 prezentuje sposób połączenia przestrzeni adresowych dwóch maszyn w klastrze. MC działa analogicznie do architektury SMP (Symmetric Multi Processing) gdzie zwykły zapis do pamięci wystarcza do komunikacji między poszczególnymi procesorami. Podstawowym mechanizmem MC jest wirtualny obwód umożliwiający połączenie w trybie zapisu między stroną w pamięci maszyny transmitującej a stroną w pamięci maszyny odbierającej. W ten sposób, aplikacje mogą wysyłać informacje do innych maszyn w klastrze posługując się zwykłą instrukcją STORE, bez jakiegokolwiek interwencji systemu operacyjnego. Bieżąca implementacja MC ma granulację strony (8Kb na platformie Alpha) i oferuje jednostronną komunikację: strona transmitująca-strona odbierająca. Ważną własnością MC jest możliwość mapowania wybranej strony transmitującej przez wiele stron odbierających. Dzięki temu modyfikacja zawartości pewnej strony w pamięci (np. struktur kontrolnych DLM) w ciągu 2.5-5us może być odzwierciedlona na wszystkich maszynach klastra.

Szybkość technologii MC-PCI sprawia, że

oparte na niej klastry Digitala posiadają wszelkie cechy wydajności **konfiguracji MPP** i wirtualnych maszyn równoległych, górując nad nimi odpornością na awarie oraz - ceną.

Jak działa MC-PCI?

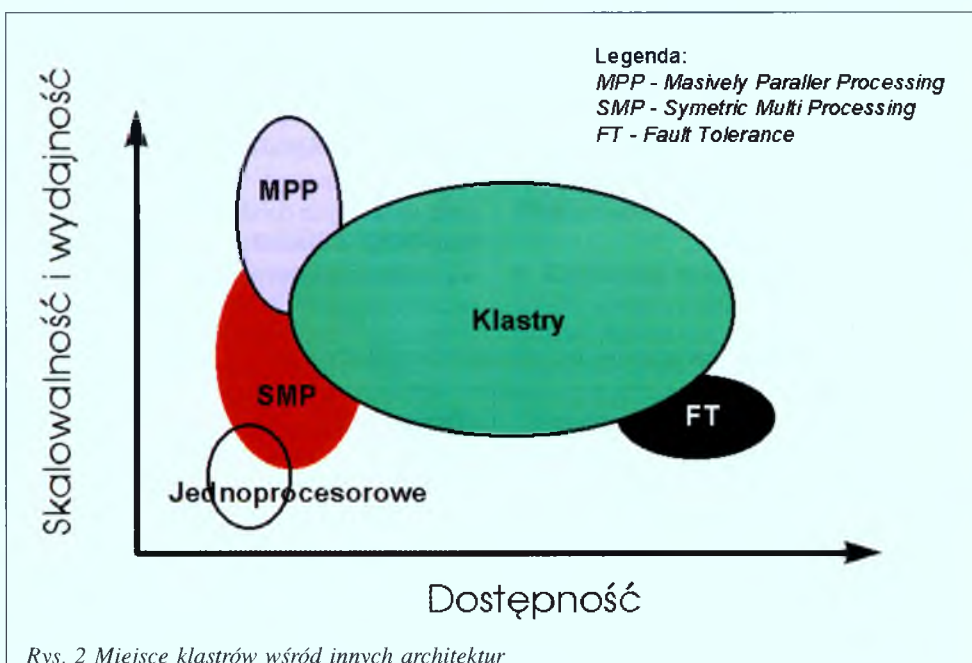
Prześledźmy sposób funkcjonowania MC (patrz Rysunek 3). Pewien proces na maszynie 1 dokonuje zapisu 32-bitowej wartości. Procesor realizuje operację LOAD pod odpowiednim adresem, który znajduje się w zakresie okna transmisji MC, czyli w przestrzeni adresowej I/O. Adres operacji należy do MC, w związku z czym operacja przechodzi z magistrali systemowej na magistralę PCI i trafia do kontrolera MC. Tam następuje translacja operacji do postaci pakietu MC, zawierającego reprezentację operacji, nagłówek adresata i CRC. Kontroler MC na maszynie 2 odbiera pakiet, interpretuje operację zawartą w pakiecie i wysyła ją na magistralę PCI. Adapter magistrali PCI przekazuje ją na magistralę systemową, gdzie ostatecznie trafia do docelowej strony pamięci operacyjnej, traktowana jak zwykła operacja zapisu.

Bezpośredni transfer między urządzeniami I/O

Ważną własnością MC jest to, że analogicznie jak procesor, standardowe urządzenia I/O mogą także bezpośrednio wykorzystywać MC w transferach urządzenie-pamięć. Pozwala to na realizację transferów danych z dysku na inną maszynę zupełnie tak, jakby docelowa pamięć była lokalna. Własność ta

*Mult.
Processor
Parallel*

*Digital
oferuje wiele
technologii
łączenia
maszyn
w konfiguracji
klastr*



*MC-PCI
implementuje
formę
rozproszonej
pamięci
dzielonej
na bazie
technologii
reflective
memory i PCI*

*Szybkość
technologii
MC-PCI
sprawia,
że oparte na
niej klastry
Digitala
posiadają
wszelkie
cechy
wydajności
konfiguracji
MPP i
wirtualnych
maszyn
równoległych*

jest podstawą efektywnej implementacji mechanizmu DRD (Distributed Raw Disk) w systemie Digital UNIX, dającego w środowisku klastra globalny dostęp do urządzeń typu *raw disk* (konieczne do bezpiecznej implementacji baz danych). Podczas bezpośredniego transferu poprzez MC, maszyna serwująca dysk inicjuje jedynie proces przesyłania danych, podobnie jak to ma miejsce w przypadku DMA, resztę realizuje samodzielnie MC-PCI. W efekcie dane trafiają bezpośrednio pod odpowiedni adres pamięci maszyny docelowej.

Konfiguracja

MC wykorzystuje standardową magistralę PCI. Połączenie dwóch maszyn Alpha w konfigurację klastrów za pomocą MC wymaga jedynie dwóch kart PCI-MC i odpowiedniego kabla (patrz Rysunek 4).

Aby połączyć więcej niż dwie maszyny, oprócz odpowiedniej liczby kart PCI-MC potrzebujemy dodatkowo urządzenia MChub (patrz Rysunek 4)

Porównanie SMP - MC

Porównując technologie SMP i MC należy na wstępie podkreślić, że mogą one, i są w rzeczywistości wykorzystywane równolegle w jednej konfiguracji. Rekordowy rezultat TPC, 30.390tpmC z ceną 305\$/tpm został osiągnięty właśnie na 4-maszynowej konfiguracji TruCluster, w której każda maszyna AlphaServer 8400 wyposażona była w 12 procesorów a całość spięta siecią MC.

Architektura SMP oferuje naturalnie większą szybkość komunikacji między procesorami niż MC. Opóźnienie jest tu ok. 10 razy mniejsze niż w przypadku MC. Jednak narzut CPU jest już identyczny. Tak więc pod względem szybkości propagacji komunikatów, SMP nieznacznie góruje nad MC.

Zupełnie inaczej wygląda porównanie z

perspektywy niezawodności. Maszyna SMP stanowi pojedynczy punkt awaryjny. Unieruchomienie oznacza przerwę w przetwarzaniu bez względu na liczbę zawartych w niej procesorów. Konfiguracja klastrów spięta MC jest znacznie odporniejsza na awarie: wystarczy praca choćby jednej maszyny, aby kontynuować przetwarzanie. Szybkość MC jest zaś wystarczająca by zapewnić skalowalność klastra nawet przy ekstremalnym obciążeniu. Optymalnym rozwiązaniem jest naturalnie klastr MC-PCI maszyn SMP.

Koszty

Zwykle pierwsze pytanie Klienta po poznaniu własności MC dotyczy ceny². A jest ona chyba jeszcze bardziej szokująca niż rezultaty techniczne. Koszt karty PCI-MC wynosi aktualnie (22-04-1996) 3.923\$. Tak więc budowa niezawodnej i wydajnej konfiguracji 2-maszynowego klastra Digital UNIX kosztuje nas niewiele ponad 7.800\$ oprócz kosztów standardowego sprzętu i oprogramowania.

Poniżej prezentowane są cechy funkcjonalne rozwiązań klastrów Digitala na trzech platformach systemów operacyjnych: Digital UNIX, Windows NT oraz OpenVMS.

Technologia klastrów dla Digital UNIX

Technologia TruCluster na platformie Digital UNIX miała swój oficjalny debiut w kwietniu 1996, a odbyło się to w bezprecedensowy sposób. System TruCluster 4xAlphaServer 8400 SMP z bazą Oracle7 osiągnął absolutny w historii TPC rekord wydajności: 30.390 tpmC przy cenie 305\$/tpm.

Klastr na Digital UNIX to zintegrowane środowisko maszyn i zewnętrznych urządzeń dyskowych dzielących wspólną magistralę SCSI. Komunikacja między-maszynowa realizowana jest poprzez połączenie MC-

Tabela 3. Porównanie parametrów technologii komunikacyjnych

Typ komunikacji	Przepustowość	Opóźnienie	Narzut CPU
tradycyjna sieć (FDDI)	10 MB/s	250,000 ns	200,000 ns
pamięć RAM	300 MB/s	250 ns	250 ns
Memory Channel	100 MB/s	2,500 ns	250 ns

PCI (patrz Rysunek 5).

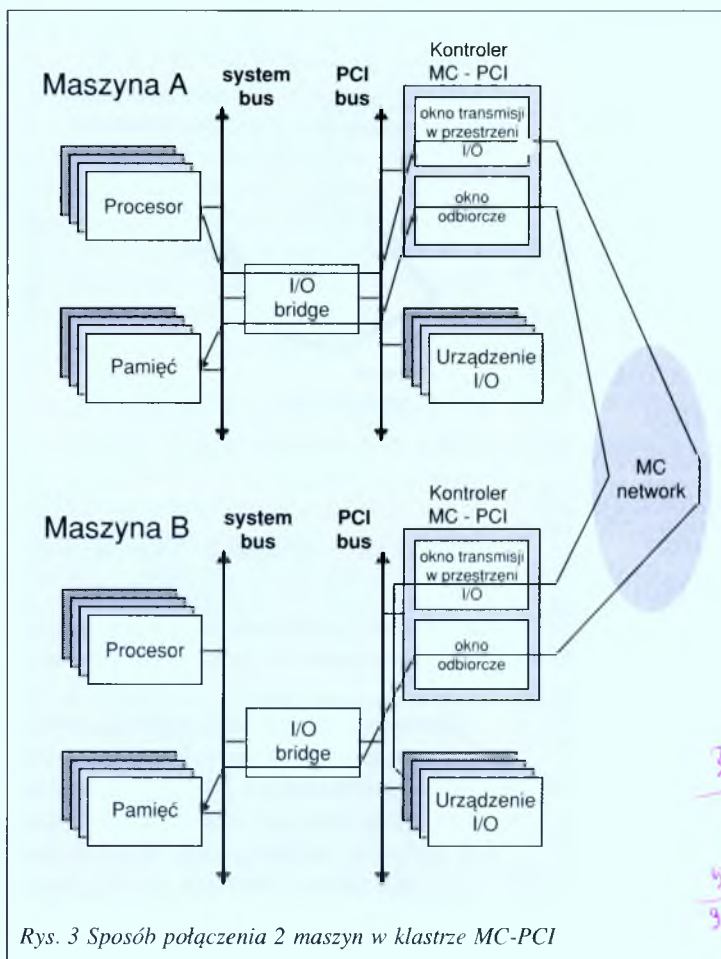
Obecnie TruCluster umożliwia połączenie do 8 maszyn Alpha. Technologia ta, jak na klastery przystało, oferuje bardzo wysoką skalowalność, niezawodność, centralne zarządzanie i przezroczystość dla użytkownika końcowego. Obok równoległego dostępu do urządzeń SCSI, TruCluster oferuje generyczny mechanizm obejścia awarii (failover) dla dysków i aplikacji.

Specyficzne serwisy oferowane przez TruCluster zawierają (patrz Rysunek 6):

- *Distributed Lock Manager (DLM)*, kluczowy element systemu, udostępniający globalny mechanizm koordynacji dostępu do dzielonych zasobów klastra. DLM dla Digital UNIX powstał przy współpracy zespołu VMScluster Engineering posiadającego 14 lat doświadczeń z analogicznym komponentem na platformie OpenVMS. Podobnie jak „starszy brat”, DLM na Digital UNIX posiada ważną zdolność asynchronicznego powiadamiania procesu posiadającego blokadę danego zasobu o tym, że blokuje dostęp innemu procesowi właśnie zgłaszającemu żądanie dostępu do tego zasobu. Funkcjonalność DLM była też konsultowana z jednym z najważniejszych aktualnie odbiorców tego serwisu - firmą ORACLE, której Parallel Server całkowicie opiera się na DLM przy regulacji synchronizacji w klastrze.

- *Connection Manager (CM)*, podsystem aktywnie kontrolujący bieżący stan konfiguracji, dodający i usuwający węzły klastra. CM utrzymuje także szybkie łącza komunikacyjne (MC-PCI) między wszystkimi maszynami klastra, wykorzystywane głównie przez DLM.

- *Distributed Raw Disk (DRD)*, mechanizm udostępniania „surowych” (raw) dysków w ramach klastra. Wiele systemów baz danych na platformie UNIX wymaga dostępności „surowych” dysków dzięki czemu mają one pełną kontrolę nad fizycznym I/O, bez zniekształceń buforowania wprowadzanych przez system plików. Obok DLM, jest to druga podstawowa funkcjonalność potrzebna do implementacji baz danych na klastrze. Dla zapewnienia maksymalnej wydajności, serwowanie DRD odbywa się w klastrze poprzez MC-PCI, czyli z przepustowością



Rys. 3 Sposób połączenia 2 maszyn w klastrze MC-PCI

ponad 100MB/s. DRD posiada ponadto mechanizm obejścia awarii maszyny.

- Usługi TCP/IP udostępniające standardowe środowisko programowania sieciowego na MC-PCI. Umożliwia to tworzenie wysoko-wydajnych programów komunikacyjnych w ramach klastra.

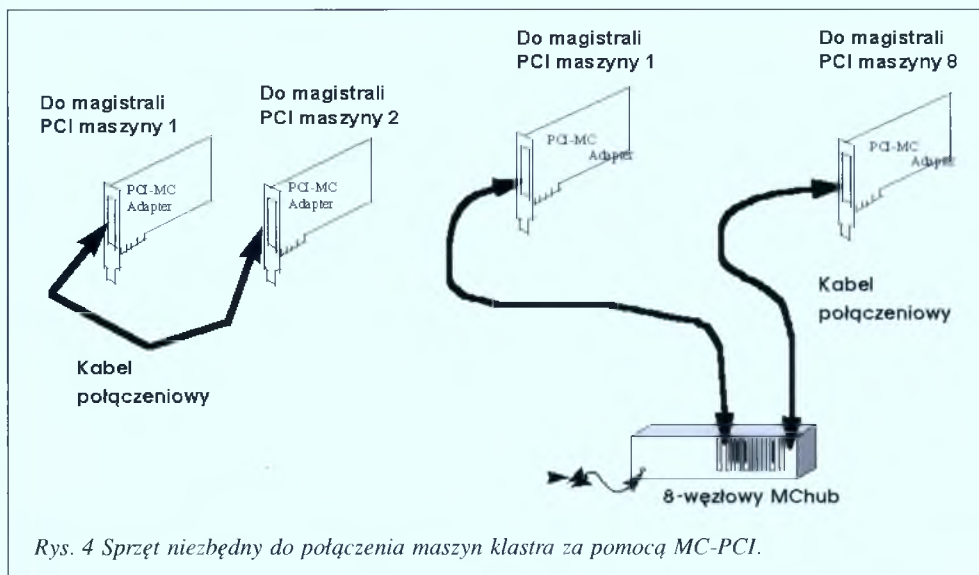
- Klastrowy system plików oparty na NFS, z zapewnieniem automatycznego obejścia awarii.

- Szybki restart AdvFS: w klastrze Digital UNIX niedostępność serwisów systemu plików AdvFS jest ograniczona do minimum dzięki mechanizmowi automatycznego obejścia awarii (failover, przejęcie obsługi przez inną maszynę). Dodając do tego właściwy dla AdvFS krótki czas rekonstrukcji można uzyskać czas restartu rzędu 15-30s.

- Graficzny interfejs administratora, ułatwiający zarządzanie i centralne monitorowanie stanu konfiguracji.

System
TruCluster
4xAlphaServer
8400 SMP
z bazą
Oracle7
osiągnął
absolutny w
historii TPC
rekord
wydajności:
30.390 tpmC
przy cenie
305\$/tpm

Administrator TruCluster może ręcznie skonfigurować usługi dyskowe i automatyczny restart aplikacji



- Zdolność przeprowadzania upgrade'ów bez konieczności wyłączania konfiguracji.

Ponieważ nie wszystkie aplikacje potrafią automatycznie w pełni wykorzystać możliwości klastra (praca z DLM), administrator TruCluster może ręcznie skonfigurować usługi dyskowe i automatyczny restart aplikacji w sposób gwarantujący ich wysoką dostępność dla systemów klienckich. Przykładowo, usługi eksportowanych dysków NFS, "surowe" dyski, aplikacje poczty czy pewne programy baz danych mogą posiadać własność obejścia awarii maszyny serwującej daną usługę. Serwis jest wtedy samoczynnie restartowany na innej maszynie klastra bez interwencji administratora. Redukuje to czas niedostępności usługi do minimum.

Oczywiście największy zysk z konfiguracji klastrów płynie z aplikacji i podsystemów świadomych funkcjonowania w tym środowisku. Oznacza to, że dany podsystem wykorzystuje DLM do wewnętrznej synchronizacji dostępu do dzielonych zasobów. Stosowanie DLM umożliwia w ten sposób pełne zrównoleglenie pracy na kilku maszynach, a w konsekwencji większą wydajność i niezawodność. Aktualnie, najważniejszym z takich „świadomych” podsystemów dla klastra na Digital UNIX jest Oracle Parallel Server (OPS) w pełni zintegrowany z TruCluster (tj. wykorzystujący DLM, patrz Rysunek 7). Stosowanie opcji OPS czyni klastery przezroczystym dla każdej aplikacji Oracle co wydaje się być niezmiennie wygodne dla Klientów.

Połączenie technologii MC-PCI, klastra oraz wydajności maszyn Alpha z 64-bitowym systemem Digital UNIX tworzy plat-

formę o bezkonkurencyjnej charakterystyce wydajności-niezawodność. UNIX, chronicznie cierpiący dotychczas na brak solidnych rozwiązań zwiększających jego niezawodność może w wydaniu Digital UNIX bez kompleksów pretendować do roli platformy dedykowanej do krytycznych zastosowań biznesowych.

Technologia klastrów na platformie Windows NT

Microsoft uznaje za strategiczne uczynienie z systemu Windows NT platformy serwera przedsiębiorstwa, nie zaś sprowadzanie go wyłącznie do roli doskonałego serwera LAN. Wizja ta nie jest łatwa z dwóch powodów. Po pierwsze, trudno przekonać VIP'ów do tego, żeby cały swój biznes powierzyli tak młodej platformie. Czynnikiem ten, wobec dynamicznego tempa migracji oprogramowania biznesowego na platformę NT (np. niedawne ogłoszenie dostępności systemu SAP R/3 na Windows NT) zaczyna pozwolić tracić na znaczeniu. Po drugie, Windows NT faktycznie potrzebuje radykalnego wzrostu niezawodności i skalowalności dla sprostania wymienionemu celowi. Dlatego też Microsoft zawarł z Digital'em umowę o współpracy na mocy której Digital dostarczy technologię klastrów dla Windows NT.

Projekt klastrów Digital'a dla NT nie jest prostym przeniesieniem technologii VMScluster. System klastra wkomponowany jest w warstwową architekturę NT oraz bierze pod uwagę unikalne cechy zarówno samego systemu operacyjnego jak i specyficznego klimatu dostępnych na nim aplikacji.

Stosowanie opcji OPS dla bazy ORACLE czyni klastery przezroczystym dla aplikacji

Poniżej omawiam podstawowe własności produktu o nazwie *Digital Cluster for NT v1.0*. Jest on aktualnie w fazie testów Beta, z planowanym wejściem na rynek w drugiej połowie 1996.

Konfiguracja

Klaster NT składa się z maszyn Intel 486/Pentium i/lub Alpha z systemem Windows NT Server oraz właściwym oprogramowaniem klastrowym (patrz Rysunek 8).

Wewnątrz klastra, istnieje wiele wymiarów redundancji (CPU, dyski, komunikacja), na której rozbudowano właściwości obejścia awarii (failover). Przykładowo, możliwy jest redundantny dostęp do urządzeń SCSI, ze zdefiniowanym serwerem głównym i zapasowym, przejmującym obsługę urządzenia na wypadek awarii tego pierwszego.

Połączenie maszyn klastra korzysta ze standardowych rozwiązań sieciowych Ethernet, FDDI, ATM lub Token Ring z transportami TCP/IP, IPX/SPX i NetBEUI. Digital planuje w przyszłości dodanie wysoko-wydajnego połączenia MC-PCI. Opcje łączenia zasobów dyskowych są również elastyczne i otwarte. Aktualnie, możliwe jest dwu-serwerowe połączenie typu SCSI. Digital zakłada dalszą ewolucję w kierunku Fiberchannel.

Asymetryczny model klastra

W swych założeniach, klaster NT odbiega od kanonu VMScluster. Tradycyjnie, model klastra był symetryczny w tym sensie, że obciążeniu systemu nie nadawano żadnej struktury - każda maszyna mogła wykonywać każdą aplikację i było obojętne, gdzie

jest serwowany dany fragment kodu. Zauważono jednak, że współczesne środowisko klient/serwer charakteryzuje się dobrze definiowalną dekompozycją funkcjonalności. Jedna maszyna serwuje pliki, inna obsługuje pocztę, jeszcze inna - oferuje serwisy aplikacyjne. Powyższy podział funkcjonalności na dyskretne fragmenty jest fundamentem modelu klient/serwer. Ta wyraźna asymetria rozłożenia obciążenia stała się podstawą strategii implementacji klastra NT.

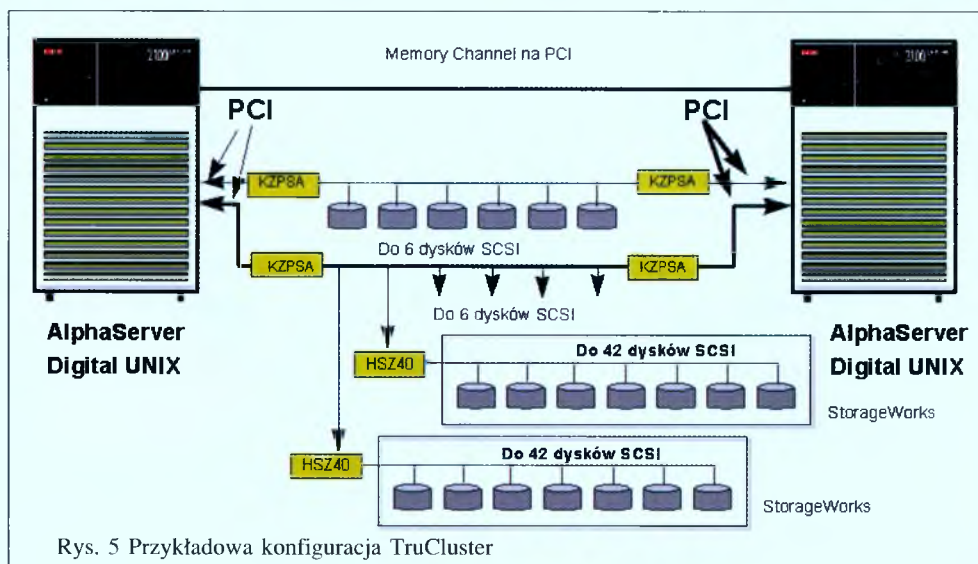
Model klastra NT jest oparty na *parcelacyjnym modelu danych* (partitioned data model). Obciążenie rozbite jest na funkcjonalnie izolowane fragmenty, z których obsługiwany jest niezależnie przez pewną maszynę w klastrze. W ten sposób, każda maszyna „specjalizuje się” w określonej klasie aplikacji.

Tradycyjny model klastra wymaga z racji swej symetryczności podstawowych usług synchronizacji typu DLM. Model klastra NT potrzebuje natomiast przede wszystkim mechanizmów parcelacji funkcjonalności i automatycznego obejścia awarii. Nie oznacza to jednak, że klaster NT nie może pracować w trybie synchronicznym - jest dodatkowo wyposażony we wszystkie potrzebne serwisy synchronizacyjne.

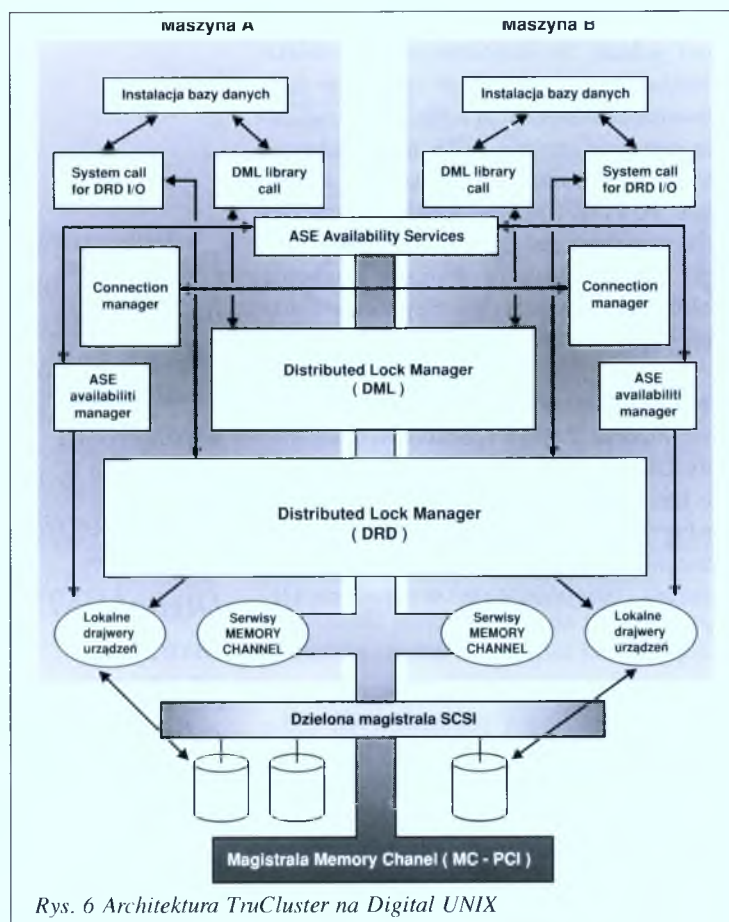
Niezawodność

Niezawodność w omawianym modelu uzyskuje się dzięki obejściu awarii (failover), czyli samoczynnym restarcie serwisu na zapasowym serwerze. Przykładowo, klient PC otwiera sesję Microsoft Word edytując plik obsługiwany przez klaster. Jeśli w trakcie pracy

Połączenie technologii MC-PCI, klastra oraz wydajności maszyn Alpha z 64-bitowym systemem Digital UNIX tworzy platformę o bezkonkurencyjnej charakterystyce wydajności - niezawodność



Rys. 5 Przykładowa konfiguracja TruCluster



serwer obsługujący edytowany plik ulega awarii, następuje automatyczne przejście na inny serwer w sposób niezauważalny dla klienta.

Skalowalność

Skalowalność to w klastrze NT kwestia granulacji parcelacji funkcjonalności - w miarę wzrostu obciążenia, poszczególne aplikacje będą migrować na dodawane stopniowo do klastra nowe maszyny. Ponadto, klaster stara się rozłożyć obciążenie wybierając dla klienta najodpowiedniejszy serwer.

Wewnętrzna architektura klastrów NT

Podstawowe elementy architektury klastrów NT są następujące (patrz Rysunek 9):

- Podstawowa infrastruktura klastra z mechanizmami komunikacyjnymi, rejestrem konfiguracji, kontrolą rozszczepienia (*cluster partitioning*, stan polegający na utracie spójności klastra i grożący zachwianiem integralności danych), synchronizacją dostępu (DLM).
- Komponenty mechanizmów obejścia awarii dla różnych grup obiektów

• Drajwery dodawane do oprogramowania podstawowego I/O i sieci dla zapewnienia funkcjonowania klastra.

• Usługi lokalizacyjne oraz rozłożenia obciążenia.

Klaster NT udostępnia własne API (rozszerzenie Win32) dając interfejs do poszczególnych usług oraz interfejs administratora (patrz Rysunek 10)

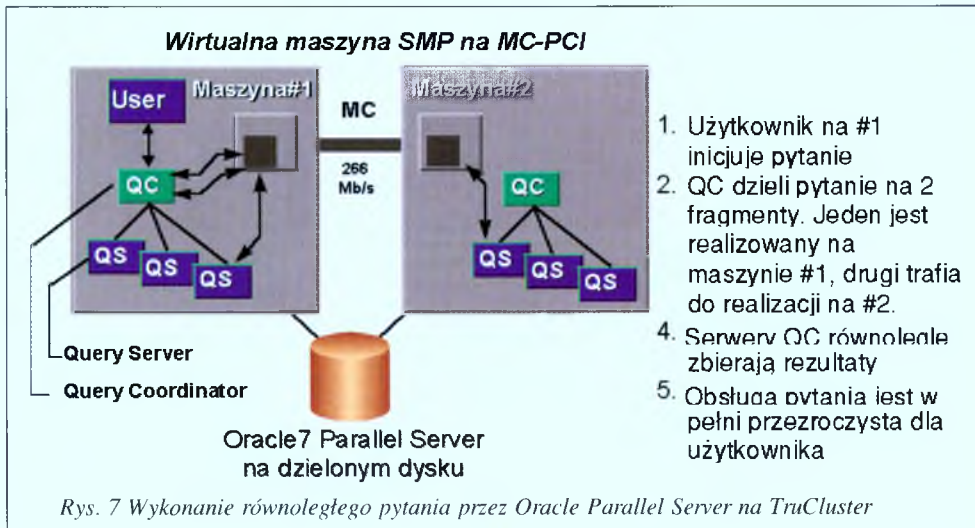
Osobiście uważam, że nasza implementacja klastrów na platformie NT będzie swoistym uwieńczeniem wieloletniej idei w Digital'u: stworzyć z ogólnodostępnych, statystycznie zawodnych komponentów systemu niezawodne i dobrze skalowalne. W wymiarze NT, oznacza to tworzenie odpornych na awarie i dowolnie wydajnych konfiguracji z najpowszechniej dostępnego sprzętu - komputerów PC.

Co nowego w VMScluster

VMScluster to bezsprzecznie najlepsza oferta klastrów na rynku. Decyduje o tym zarówno bogata funkcjonalność i stała integracja najnowszych technologii komunikacyjnych jak i dojrzałość technologiczna. Systemy klastrów są bowiem z natury złożonym i trudnym oprogramowaniem, gdzie stabilność osiąga się wieloletnimi doświadczeniami eksploatacyjnymi. Do podstawowych cech VMScluster należą:

- *Distributed Lock Manager* wkomponowany w system operacyjny. Blokowanie zasobów odbywa się z definicji za pośrednictwem DLM, dzięki czemu dowolna aplikacja korzystająca z jego usług może bez przeszkód pracować równolegle na kilku węzłach klastra.
- Klastrowy system plików. W przeciwieństwie do TruCluster i klastrów na NT, gdzie system plików (lub jego podzbiór) serwowany był przez jedną maszynę, system plików OpenVMS jest w pełni zintegrowany z VMScluster. Synchronizacja jest realizowana poprzez automatyczne i niewidoczne dla klienta wykorzystanie mechanizmu DLM.
- Bogate możliwości komunikacji klastrowej: Ethernet, FDDI, DSSI, CI. Wersja OpenVMS v7.1 przyniesie dodatkowo połączenia MC-PCI oraz ATM.

Model klastra NT jest oparty na parcelacyjnym modelu danych (partitioned data model)



Począwszy od wersji OpenVMS v6.2 możliwe jest konfigurowanie urządzeń SCSI jako zewnętrznie dzielenych dysków klastra. Pozwala to na tworzenie małych i ekonomicznych rozwiązań klastrowych posiadających wszystkie atrybuty niezawodności klastrow. Istnieje np. możliwość budowy konfiguracji VMScluster w oparciu o dwie stacje AlphaStation 200/250 i macierz dyskową BA350/364 SCSI. Skonstruowany w ten sposób system, przy koszcie ok 45tys\$ posiada doskonałą charakterystykę niezawodnościową.

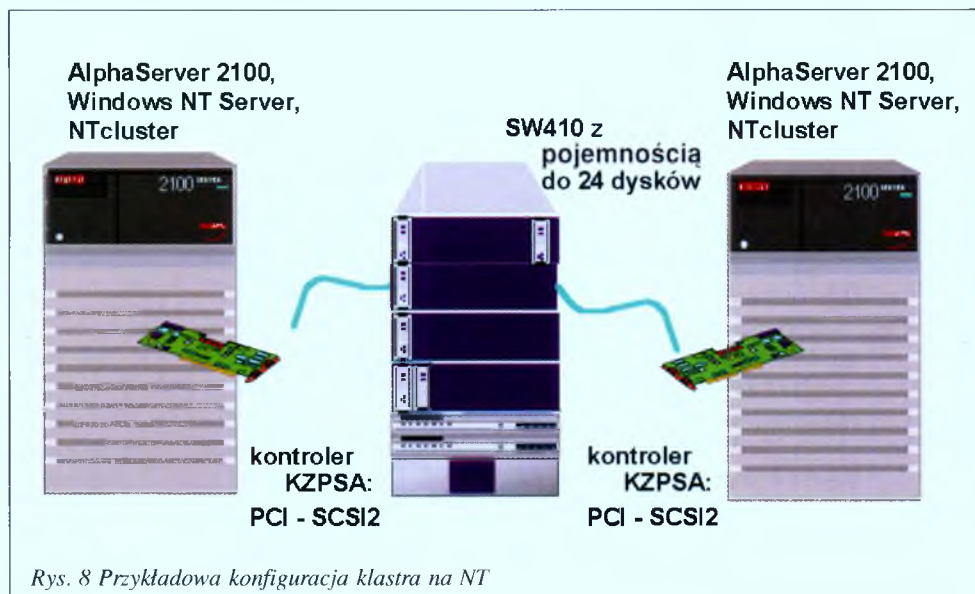
Innym kierunkiem rozwoju oprogramowania klastrowego jest przeniesienie funkcji administracyjnych do środowiska PC. Jest to część programu ManagementStation, którego celem jest zbudowanie infrastruktury administracyjnej dla całości systemu OpenVMS na platformie PC. Poszczególne funkcjonalne administracyjne klastra będą sukcesywnie

implementowane w ramach ManagementStation. Ułatwi to dodatkowo proces obsługi VMScluster.

Zakończenie

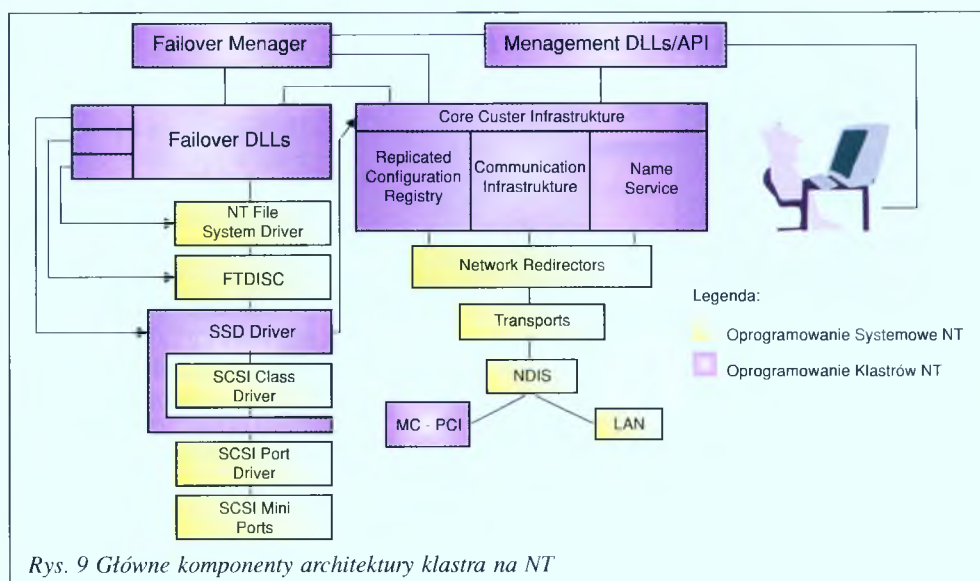
Technologia klastrow oferuje skuteczne i ekonomiczne rozwiązanie problemów niezawodności i skalowalności systemów. Wobec rosnących wymagań ciągłości pracy krytycznych aplikacji biznesowych, rola systemów klastrowych będzie najprawdopodobniej wzrastać. Równocześnie zwiększać się będzie liczba aplikacji "świadomie" korzystających z zalet tego środowiska. Z perspektywy polskiego rynku, szczególnie cenne wydaje się to, że klastr może wzrastać wraz z biznesem. Rozpoczynając od małej, niedrożej maszyny konfiguracja może ewoluować do wielo-maszynowego systemu o

Implementacja klastrow na platformie NT będzie swoistym uwieńczeniem wieloletniej idei w Digital' u: tworzyć z ogólnodostępnych, statystycznie zawodnych komponentów systemy niezawodne i dobrze skalowalne



Począwszy od wersji OpenVMS v6.2 możliwe jest konfigurowanie urządzeń SCSI jako zewnętrznie dzielonych dysków klastra

Digital Equipment jest jedyną firmą oferującą rozwiązania klastrowe na wszystkich proponowanych przez siebie systemach operacyjnych



ogromnej mocy obliczeniowej i doskonałej dostępności.

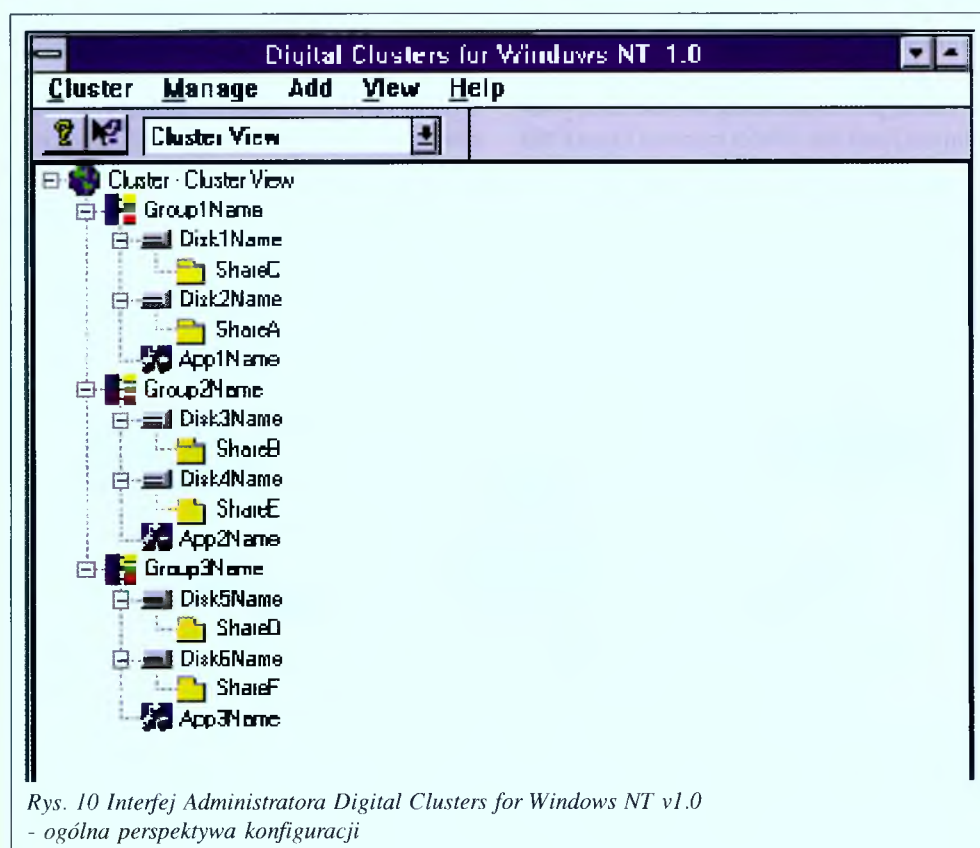
pod względem wydajności, ceny i niezawodności.

Artur Stefanowicz

Digital Equipment jest jedyną firmą oferującą rozwiązania klastrowe na wszystkich proponowanych przez siebie systemach operacyjnych: OpenVMS, Digital UNIX oraz Windows NT. Najnowsze osiągnięcia w dziedzinie komunikacji klastrowej, w szczególności technologia MC-PCI czynią z rozwiązań Digital'a platformę bezkonkurencyjną

¹ "Memory Channel for PCI", Richard Gillett, IEEE Micro, Feb 1996

² Digital zastrzega sobie prawo zmiany ceny podanej w tekście artykułu



Realizacja aplikacji dla Centralnego Rejestru Skazanych

W ostatnich dniach 1993 roku Digital Equipment Polska podpisał z Ministerstwem Sprawiedliwości kontrakt na realizację skomputeryzowanego systemu obsługi Centralnego Rejestru Skazanych (CRS). Digital zajął się zarządzaniem przedsięwzięciem, dostawą i instalacją sprzętu w centrali i oddziałach, budową sieci lokalnej i sieci rozległej z wykorzystaniem sieci zdalnej POLPAK do komunikacji z oddziałami. Menedżerem projektu ze strony Digitala został Piotr Sobolewski, zaś menedżerem kontraktu - Waldemar Całka. W celu realizacji oprogramowania aplikacji systemu CRS, Digital podjął współpracę z firmą PentaComp. Prace nad aplikacją zakończyły się w marcu 1996. Projektowany system został przekazany Ministerstwu i rozpoczął się intensywny okres jego wdrażania. Biorąc pod uwagę zakres i czas przedsięwzięcia można powiedzieć, że był to projekt informatyczny średniej wielkości. Czas ponad dwu lat spędzonych przy realizacji aplikacji dla projektu CRS pozwala na zaprezentowanie wielu, naszym zdaniem ciekawych doświadczeń dotyczących zarówno strony technicznej (ocena zastosowanych technologii), jak i samego przebiegu projektu.

Cele projektu

Celem opisywanego projektu było:

- radykalne usprawnienie pracy CRS, polegające na szybszym przetwarzaniu i wyszukiwaniu dokumentów w archiwum oraz automatyzacji wielu rutynowych czynności, takich jak udzielanie odpowiedzi klientom, rozliczanie pracy, kontrola poprawności dokumentów,
- zapewnienie większego bezpieczeństwa archiwum,
- zagwarantowanie poprawności danych w nim przechowywanych,
- udostępnienie części danych archiwum oddziałom w Polsce,
- umożliwienie zbierania danych statystycznych dotyczących osądzeń i skazań dorosłych i nieletnich, na podstawie informacji wprowadzanej z dokumentów oraz tworzenie zagregowanych opracowań statystycznych w postaci wydruków tabelarycznych,
- umożliwienie przejścia archiwalnych danych statystycznych i wprowadzenia ich do systemu.

Cel projektu miał być osiągnięty poprzez informatyzację archiwum dokumentów (obecna zawartość informacyjna archiwum to ponad trzy miliony dokumentów papierowych). W procesie informatyzacji przewidziano zbudowanie systemu pozwalającego na pełną obsługę danych w nowej, elektronicznej formie oraz przeniesienie dokumentów papierowych do bazy danych systemu. Ponieważ CRS działa w bardzo ściśle określonych ramach prawnych, przy tworzeniu systemu konieczne było zapewnienie takiej funkcjonalności, która będzie w pełni zgodna z dotychczasowymi zasadami działania Rejestru. Z zewnętrznego punktu widzenia system musiał w najdrobniejszych szczegółach odpowiadać określonym przez klienta założeniom.

Archiwum powinno zawierać kilka rodzajów dokumentów o ściśle określonej strukturze. Na każdym rodzaju dokumentu należało umożliwić wykonywanie pewnych dokładnie zdefiniowanych operacji. Konieczne było uzależnienie zakresu dostępu do dokumentów od uprawnień określonych dla klasy, do której dany użytkownik należy. Najistotniejszym jednak zagadnieniem w projekcie aplikacji było zapewnienie maksymalnego stopnia weryfikacji wprowadzanych informa-

W celu realizacji oprogramowania aplikacji systemu CRS, Digital podjął współpracę z firmą PentaComp

Obecna zawartość informacyjna archiwum to ponad trzy miliony dokumentów papierowych

Etap analizy przeprowadzonej zgodnie z metodologią DPM (Digital Project Methodology) umożliwił dokładne określenie zakresu projektu

cji, przy czym weryfikacja ta musiała być uzależniona od stanu prawnego obowiązującego w określonym czasie.

Od strony technicznej rozwiązanie polegało na utworzeniu centralnej bazy danych obsługiwanej przez komputer główny systemu. Wszystkie operacje przetwarzania dokumentów (wprowadzanie, edycja itp) miały się odbywać lokalnie. Założono możliwość jednoczesnej pracy kilkudziesięciu użytkowników zadających zapytania i wprowadzających oraz edytujących dokumenty. Możliwość zadawania zapytań dotyczących rekordów osobowych istniejących w bazie danych miała też być dostępna dla użytkowników zdalnych (20 stanowisk zlokalizowanych w całej Polsce). W kontrakcie narzucono stosunkowo ostre wymagania wydajnościowe: czas odpowiedzi na zapytanie z punktu zdalnego ma być krótszy niż 1 sekunda (bez uwzględnienia czasu transmisji). Poza tym założono, że system powinien umożliwić obsługę kilku tysięcy dokumentów dziennie, a dokładniej:

- wprowadzanie ok. 1500 dokumentów (lokalnie),
- usuwanie również ok. 1500 dokumentów (lokalnie),
- udzielanie ok. 1000 - 3000 odpowiedzi dziennie (lokalnie i zdalnie).

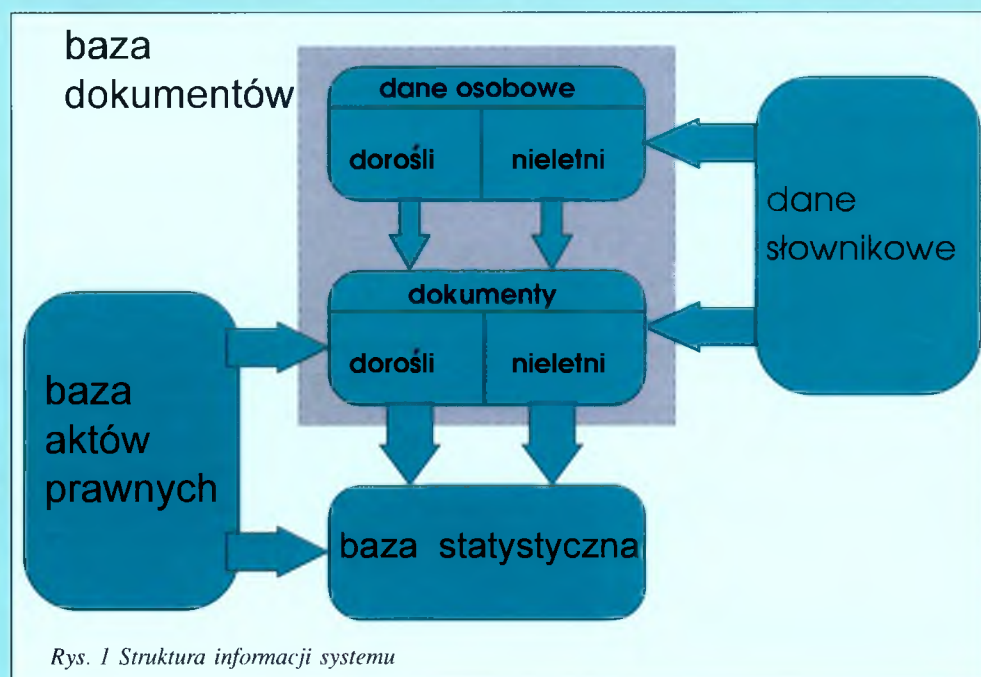
Ze względu na zakres merytoryczny oraz wymagania zamawiającego co do szybkiej dostępności podstawowej funkcjonalności, projekt został podzielony na dwie fazy. Pierwsza faza polegała na dostarczeniu klientowi podstawowego oprogramowania umożliwiającego działanie CRS. Druga faza miała się zakończyć wdrożeniem systemu w pełnym zakresie. W ramach każdej z faz wyróżniono kilka kamieni milowych, które dotyczyły oddania kolejnych dokumentów projektowych, czy też wersji aplikacji z dodaną nową funkcjonalnością.

Etap analizy - zakres merytoryczny projektu

Etap analizy przeprowadzonej zgodnie z metodologią DPM (*Digital Project Methodology*) umożliwił dokładne określenie zakresu projektu. System został zdekomponowany na poszczególne podsystemy, o szczegółowo zdefiniowanej funkcjonalności.

System, najogólniej mówiąc, miał obsługiwać rekordy osobowe oraz związane z nimi dokumenty. W trakcie analizy wyodrębniono 8 rodzajów dokumentów. Były one szczegółowo ustrukturalizowane, z wyraźnie wyodrębnionymi polami. Założenie dostępności dokumentów wyłącznie poprzez zadawanie zapytań dotyczących danych osobowych pozwalało na zastosowanie bazy relacyjnej jako składnicy informacji. W wypadku konieczności przeszukiwania pól opisowych

System, najogólniej mówiąc, miał obsługiwać rekordy osobowe oraz związane z nimi dokumenty



najprawdopodobniej należałoby zastosować serwer przeszukiwań pełnotekstowych. Dokumenty mają różny stopień złożoności, przy czym, co istotne, najbardziej skomplikowany dokument jest równocześnie podstawowym dokumentem w systemie. Pojęcie „skomplikowany” ma co najmniej trojakie znaczenie:

- dokument o dużej zawartości informacyjnej,
- dokument o rozbudowanej, hierarchicznej strukturze, co implikuje zastosowanie do jego opisu wielu tabel, a to z kolei powoduje znaczną komplikację operacji bazodanowych i obsługi interfejsu,
- dokument o skomplikowanych zależnościach merytorycznych pomiędzy poszczególnymi danymi - konieczność zastosowania silnych i złożonych mechanizmów walidacyjnych.

Zadaniem systemu jest umożliwienie pełnej obsługi dokumentów - wyszukiwania na podstawie danych osobowych, wprowadzania, edycji, usuwania. Do tych standardowych operacji dochodzą inne: generowanie i drukowanie w określonej, sformalizowanej formie i z określonymi zależnościami czasowymi rozmaitych dokumentów wynikowych (kilkanaście rodzajów). Ponadto dla obsługi statystyki resortowej, należało zapewnić generowanie raportów statystycznych (ponad sto rodzajów). Do tego dochodzą rozbudowane mechanizmy kontroli pracy i rozliczania użytkowników systemu. Osobno należy wspomnieć o module walidacji danych. Przy wprowadzaniu i edycji dokumentów należało bowiem uwzględnić silne zależności logiczne pomiędzy poszczególnymi elementami - podsystem walidacji ma wspomagać wykrywanie niezgodności w takich zależnościach.

W trakcie analizy stało się jasne, że ogromna większość operacji w systemie musi być wykonywana w trybie on-line. Oczywiście, tam gdzie to było możliwe, starano się uwzględnić możliwość przetwarzania wsadowego poza godzinami największego obciążenia systemu.

Etap projektowania - technologia

Na podstawie analizy określona została te-

chnologia niezbędna do realizacji systemu. Były to produkty firmy Digital, gdyż zdecydowana większość potrzebnych narzędzi znajdowała się wśród oprogramowania oferowanego przez tę firmę.

Zacznijmy od komputera głównego. Przyjęto, że będzie to VAX 4100A z systemem operacyjnym OpenVMS w wersji 5.5. Zarówno sprzęt jak i system operacyjny gwarantują, oprócz wydajności, wyjątkowo wysoki stopień niezawodności i dostępności.

Jeśli chodzi o serwer bazy danych, wybrano system DEC Rdb/VMS wersja 6.0 (Obecnie już Oracle Rdb) - baza danych znana z silnego mechanizmu optymalizacji zapytań. Miało to istotne znaczenie ze względu na zawarte w kontrakcie wymagania dotyczące czasu odpowiedzi na zapytania. System ten daje też pełne mechanizmy zapewniania integralności referencyjnej poprzez reguły spójności bazy (*constraints*) i mechanizmy procedur wyzwalanych (*triggers*), dając przy tym wysoko wydajne środowisko przetwarzania transakcyjnego. Do narzędzi dostępnych dla bazy Rdb należy język modułowy SQLModule pozwalający wygodnie definiować proceduralne wywołania poleceń języka SQL.

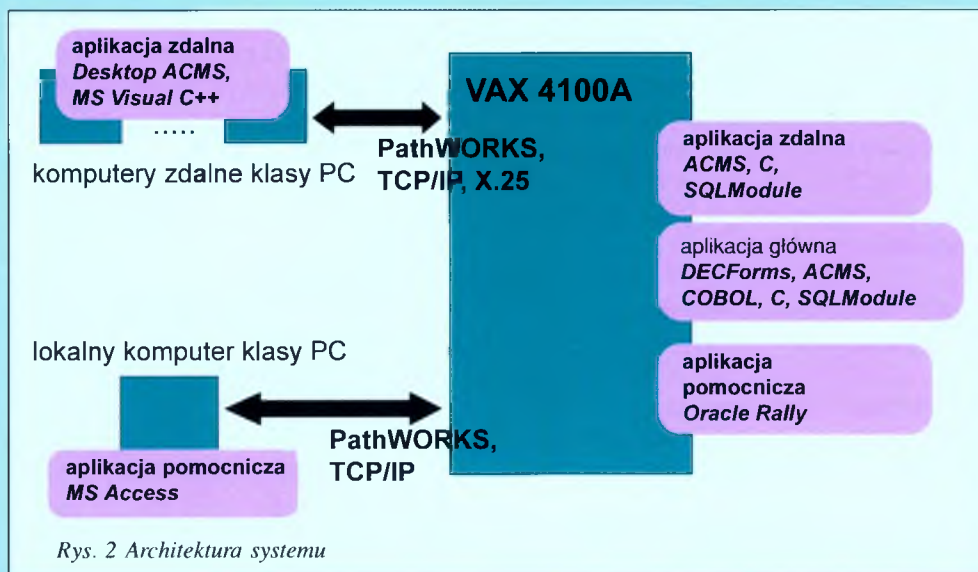
Jako języki tworzenia procedur obsługi bazy danych zastosowano C (kompilator VAX C wersja 3.2) oraz COBOL. Interfejs użytkownika miał zostać zrealizowany za pomocą pakietu DECforms wersja 1.4. Kluczowym elementem zastosowanej technologii miało być wprowadzenie warstwy pośredniej (*middleware*) pomiędzy procedury bezpośrednio realizujące obsługę operacji bazodanowych, a moduły realizujące obsługę interfejsu użytkownika. Ze względu na duże wymagania co do liczby operacji w trybie on-line przeprowadzanych na bazie danych, konieczne było zastosowanie monitora transakcyjnego. Monitor ACMS (wraz z mechanizmem Desktop ACMS do tworzenia aplikacji zdalnych) firmy Digital był tutaj naturalnym wyborem. Całość aplikacji w procesie tworzenia spinała wspólna składnica metainformacji, czyli CDD/Repository. Jako narzędzi pomocniczych użyto również generatora aplikacji DEC Rally oraz systemu bazy danych Microsoft Access v.2.0 wraz z mechanizmem ODBC do komunikacji z bazą podstawową Rdb/VMS.

Zadaniem systemu jest umożliwienie pełnej obsługi dokumentów

Jeśli chodzi o serwer bazy danych, wybrano system DEC Rdb/VMS wersja 6.0

W procesie analizy, projektowania i implementacji używane były narzędzia programistyczne typu CASE firmy Digitala

Prowadzenie prac projektowych oparto na metodzie DPM (Digital Project Methodology)



W procesie analizy, projektowania i implementacji używane były narzędzia programistyczne i typu CASE firmy Digital: DECDesign umożliwiające graficzne projektowanie struktury systemu wraz z generacją schematu bazy danych Rdb/VMS, pakiet narzędzi DECSet, zawierający m.in. edytor sterowany składnią LSE (*Language Sensitive Editor*), analizator kodu SCA (*Source Code Analyzer*), narzędzie kontroli wersji CMS (*Code Management System*), system generacji kodu wynikowego systemu MMS (*Module Management System*) oraz narzędzie do tworzenia i wykonywania pakietów testów DTM (*DEC-Test Manager*). Ponadto wykorzystywane były narzędzia wspomagające tworzenie struktury fizycznej bazy danych: GSE (*Graphical Schema Editor*) i Rdb Expert. Szczególną pomocą na etapie wdrożenia okazały się narzędzia do monitorowania i analizowania obciążenia systemu.

Podsumowaniem fazy analizy i projektu z punktu widzenia klienta było przygotowanie dokumentu z proponowanymi rozwiązaniami, wraz z proponowanym wyglądem i funkcjonalnością ekranów. Od strony zespołu projektowego rezultatem tej fazy było zdekomponowanie systemu na poszczególne moduły i określenie zadań (*task* w terminologii ACMS) realizujących szczegółową funkcjonalność systemu. Powstał też dynamiczny model systemu w formie diagramów przepływu danych oraz model danych, czyli diagramy ER.

Metodyka prowadzenia projektu

Prowadzenie prac projektowych oparto na

metodzie DPM (*Digital Project Methodology*). Konieczne było jednakże dostosowanie jej do wymogów konkretnego przedsięwzięcia, które charakteryzowało się m.in. bardzo istotnymi ograniczeniami czasowymi. Spowodowało to, że stosowanie pewnych elementów metodologii należało ograniczyć lub pominąć. Innymi słowy - podjęta została próba modyfikacji ścisłej metody DPM w kierunku opracowania szablonu metodologicznego bardziej przydatnego dla średnich projektów, o krótkich czasach realizacji.

Zespół realizujący projekt oprogramowania został zorganizowany następująco: na czele szef projektu, czyli osoba mająca obejmować całość merytorycznych zagadnień projektu, nadzorująca realizację poszczególnych zadań w ramach projektu i koordynująca pracę poszczególnych podzespołów; do wykonywania określonych zadań tworzone były niewielkie podzespoły podległe szefowi projektu.

Na etapie analizy i projektu zespół składał się jedynie z analityków i projektantów systemu. W fazie implementacji, zakres merytoryczny projektu oraz zastosowane trójwarstwowe podejście do konstrukcji aplikacji z wydzieloną warstwą pośrednią (*middleware*), pozwoliły na dobre wydzielenie poszczególnych modułów aplikacji i podział na realizujące je podzespoły. Tak więc w ramach zespołu implementującego aplikację powstały grupy odpowiedzialne za powstanie: zadań ACMS, interfejsu użytkownika, procedur serwerów obsługi bazy danych oraz wydruków i raportów. Osobne podzespoły realizowały aplikację zdalną i podsystem walidacji.

Osobno trzeba wspomnieć o grupach do "zadań specjalnych". Trzeba tu wymienić podzespoły odpowiedzialne za przygotowanie dokumentacji i przeprowadzenie szkoleń, osoby odpowiedzialne za strojenie bazy danych, systemu operacyjnego i ACMS oraz *last but not least* - zespół testujący.

Liczebność poszczególnych zespołów zmieniała się w trakcie realizacji projektu, w zależności od spodziewanego nakładu pracy na realizację zadań i wymagań czasowych.

Oprócz koniecznej bieżącej współpracy poszczególnych zespołów ważną rolę w organizacji prowadzenia projektu pełniły cotygodniowe spotkania z udziałem menedżera projektu z ramienia firmy Digital na których dokonywano podsumowania prowadzonych prac i ustalano priorytety oraz zakres działania na najbliższy czas. Inne procedury stosowane w zarządzaniu projektem oprogramowania, zgodne z metodologią DPM, to:

- przeglądy bieżące, na których rozważano merytoryczne zagadnienia związane z bieżącymi pracami;
- formalne przeglądy techniczne, stanowiące formę wewnętrznego odbioru wykonanych kroków, faz i etapów projektu.
- kontrola jakości, umożliwiająca bieżące kontrolowanie prowadzonych prac pod względem jakości i zgodności z założonymi standardami;
- kontrola wersji umożliwiająca utrzymywanie spójności kolejnych wersji całości systemu złożonego z wielu wersji poszczególnych komponentów;
- kontrola zmian pozwalająca rejestrować żądania dokonywania zmian w systemie, szacować ich wpływ na realizację przedsięwzięcia i kontrolować ich implementację.

Problemy fazy implementacji systemu

Z racji niewielkiej złożoności logicznej, stosunkowo najprostsze okazało się zrealizowanie aplikacji zdalnej, pomimo konieczności wykorzystania różnorodnego oprogramowania i heterogenicznego środowiska systemowego. Użyto tutaj po stronie serwera języ-

ka SQLModule, C i ACMS, natomiast po stronie klienta Desktop ACMS i Microsoft Visual C++. Komunikacja pomiędzy aplikacjami klienckimi i serwerem odbywała się poprzez sieć Polpak z użyciem protokołów X.25 i TCP/IP. Ze względu na poufność przesyłanych danych stworzono węzły dołączone do sieci tworzą zamkniętą grupę użytkowników (*Closed User Group*). Zastosowano też szereg innych mechanizmów ochrony bezpieczeństwa danych.

Mechanizm Desktop ACMS, dostarczany w postaci biblioteki procedur dla komputera klienckiego, wykorzystujących w swym działaniu warstwę transportu sieciowego, umożliwia wywoływanie zadań ACMS na serwerze poprzez mechanizm zdalnego wywołania procedury (*Remote Procedure Call*).

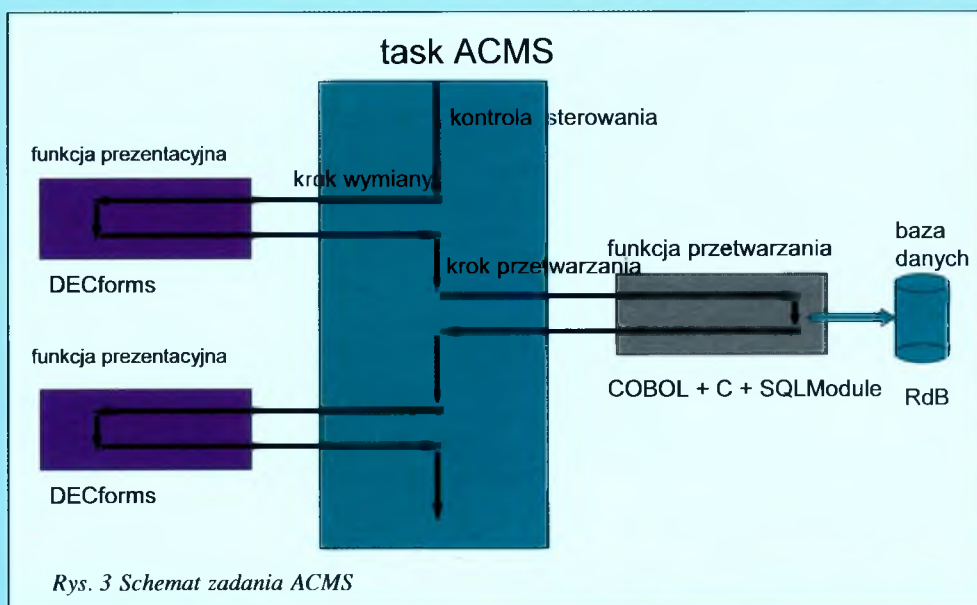
Zasadniczym problemem w osiągnięciu wymaganego czasu odpowiedzi było właściwe dostrojenie bazy danych. Zestaw atrybutów używanych do wyszukiwania był bogaty, co implikowało konieczność użycia wielu indeksów - zarówno sortowanych (o strukturze B-drzew), jak i opartych na funkcjach mieszających. Wyobrażenie o wielkości bazy danych może dać następujące zestawienie: struktura liczy ok. 70 tabel, najistotniejsze z nich (obejmujące dane osobowe) zawierają ok. 1,5 miliona rekordów, na każdej z tych kluczowych tabel utworzonych jest ok. 20 indeksów, tabele przechowujące dokumenty docelowo mają liczyć od 3 do 10 milionów rekordów. W tych warunkach baza Rdb/VMS potwierdziła swą dobrą opinię jeśli chodzi o działania optymalizatora zapytań i możliwość strojenia, co pozwoliło osiągnąć wymagany czas odpowiedzi na poziomie 1 sekundy przy formułowaniu zapytań opartych na dowolnym zestawie atrybutów na odpowiednio zaindeksowanych tabelach.

Przy optymalizacji należało uwzględnić również zapytania nieprecyzyjne, czyli przede wszystkim zawierające znaki specjalne (*wildcards*), jak również pewne mechanizmy bardzo specyficzne dla tworzonego systemu. Jednym z mechanizmów jest możliwość zadawania zapytań, w których wyszukiwanie odbywa się z uwzględnieniem imion podobnych (np. umieszczenie w polu IMIĘ tekstu "Wioletta" powoduje wyszukanie również rekordów zawierających imię "Wioleta", "Violetta" lub "Violeta"). Inną cechą systemu jest możliwość zamiany nazwisk rodo-

*Ważną rolę
pełniły
cotygodniowe
spotkania z
udziałem
menedżera
projektu
z ramienia
firmy Digital*

*Zasadniczym
problemem w
osiągnięciu
wymaganego
czasu
odpowiedzi
było właściwe
dostrojenie
bazy danych*

Do budowy
strony
zewnętrznej
systemu
użyto
DECforms



Rys. 3 Schemat zadania ACMS

wego i przybranego przy zadawaniu zapytania. Wszystkie te mechanizmy wymagały złożonego algorytmu generowania dynamicznego zapytania SQL oraz wykorzystania zaawansowanych mechanizmów optymalizacji w bazie danych.

Do budowy strony zewnętrznej systemu użyto DECforms, jako specjalizowanego języka opisu interfejsu użytkownika. DECforms nadaje się bardzo dobrze do budowy aplikacji o wyraźnie wydzielonej warstwie obsługi terminala. Dostarcza bogatego zestawu poleceń, pozwalających na pełną obsługę terminala tekstowego (obecnie również graficznego). Cechą charakterystyczną pakietu DECforms, a dokładniej jego języka o nazwie IFDL jest statyczna budowa (brak jakichkolwiek struktur dynamicznych) i podział programu na stałe bloki. Dzięki tym cechom kod napisany w języku IFDL ma dobrze określoną, przejrzystą strukturę i jest łatwy w modyfikacji. Brak funkcji do operowania na ciągach znaków i liczbach można z powodzeniem zastąpić funkcjami w języku C. Wszystkie podstawowe struktury danych używane w DECforms, definiowane są na podstawie definicji we wspólnym słowniku metadanych CDD.

Moduły w DECforms zostały zintegrowane z ACMS, jako zarządcą całego systemu, pozwalając na pełne odizolowanie procesów terminalowych od procesów serwera bazy danych. Do obsługi wielu procesów terminalowych ACMS powołuje w trakcie działania systemu jeden lub kilka procesów sterują-

cych (*Control Processes*). Zadaniem każdego procesu sterującego jest efektywne zarządzanie dostępem do zadań ACMS. Synchronizacja i wymiana danych z ACMS zrealizowana została poprzez mechanizm *send-receive-transceive*, polegający na przekazywaniu wspólnie zdefiniowanych struktur danych, tzw. rekordów i oczekiwaniu na wynik działania wywołanego procesu. Ograniczenie kanału wymiany informacji DECforms-ACMS do ściśle określonego, standardowego mechanizmu pozwoliło na czytelny i strukturalny zapis dość złożonych sekwencji akcji.

Zadania algorytmiczne i dostęp do bazy danych są realizowane przez tzw. serwery proceduralne (*Procedure Servers*), zaimplementowane w językach C, COBOL i SQLModule. Procedury pogrupowane są w programy uruchamiane jako procesy systemowe. Każda z procedur w uruchomionym serwerze może być wywołana przez ACMS dla realizacji określonego przetwarzania danych.

Przy tworzeniu procedur serwerów operujących na dokumentach okazało się, że standardowe mechanizmy bazy danych nie zapewniają wystarczającej wydajności systemu. Problem polegał na tym, że z punktu widzenia użytkownika proces obsługi dokumentu musi być jedną, długotrwałą transakcją. Z punktu widzenia systemu zaś, dla minimalizacji jego obciążenia, transakcje bazodane muszą być bardzo krótkie i wykonywać się bez możliwości interwencji użytkownika. Jeśli np. średni czas obróbki dokumentu przez użytkownika wynosi 5 minut, a zapis infor-

Moduły
w DECforms
zostały
zintegrowane
z ACMS

macji do bazy danych - 15 sekund, to przy odpowiednio zaprojektowanym systemie 1 proces serwera może obsłużyć do ok. 20 równocześnie pracujących użytkowników. Aby jednak było to możliwe, dokumenty muszą być buforowane w całości w pamięci operacyjnej, co też zostało zrealizowane w systemie CRS. Wykorzystane tu zostały mechanizmy systemu operacyjnego OpenVMS pozwalające na istnienie w systemie tzw. obrazów współdzielonych (*Shareable Image*) zawierających w sekcjach globalnych dane dostępne dla różnych procesów systemowych. Mechanizm ten został zastosowany również do zwiększenia efektywności operacji na najczęściej wykorzystywanych strukturach danych - wszystkie dane stałe systemu, łącznie z całą bazą danych aktów i reguł prawnych, umieszczane są w pamięci operacyjnej w odpowiednich strukturach, dzięki czemu dostęp do nich nie wymaga odczytu informacji z bazy danych.

Oddzielnym, problemem okazał się moduł walidacji. Skomplikowana natura merytoryczna walidacji spowodowała konieczność zdefiniowania własnego języka dla zapisu reguł walidacji oraz - co za tym idzie - kompilatora i interpretera kodu pośredniego tego języka.

Język pozwala na związanie z poszczególnymi polami dokumentów, grupami pól oraz całymi dokumentami odpowiednich reguł walidacyjnych, opisujących zależności jakie powinny spełniać wartości wprowadzane w da-

nym polu, grupie pól lub dokumencie. W przypadku niespełnienia zależności wyświetlany jest komunikat, a użytkownik systemu może podjąć odpowiednie działania zaradcze.

W kontrolowanych dokumentach występują złożone warunki logiczne, które zależne są również od zawartości bazy danych opisującej stan prawny. Z tego względu język do zapisu reguł walidacyjnych posiada szereg złożonych konstrukcji, taki jak możliwość dostępu do bazy danych oraz do struktur globalnych w pamięci, czy wykonywanie funkcji zewnętrznych. Operatory języka zawierają m.in. kwantyfikator ogólny ("dla każdego"), kwantyfikator szczegółowy ("istnieje") oraz złożone operatory arytmetyczne, w tym znajdujących minima i maksima w zbiorach wartości.

Zastosowane podejście do walidacji danych cechuje się dużą elastycznością oraz łatwością modyfikowania reguł walidacji w przypadku ich zmiany w trakcie eksploatacji systemu.

Zastosowanie monitora transakcyjnego miało dwa główne cele. Po pierwsze - wymaganiem systemu była spora liczba użytkowników działająca w trybie on-line i modyfikująca zawartość bazy danych. W takim przypadku zastosowanie monitora transakcyjnego, rozwiązanie powszechnie stosowane w tego typu systemach na Zachodzie, zapewnia wysoki stopień równoległości pracy użytko-

W kontrolowanych dokumentach występują złożone warunki logiczne

Na wszystkich częściach blankietu
wpisz czytelnie atramentem, długopisem
lub piśmem maszynowym jednakową
kwotę cyframi, imię i nazwisko wpłacającego
i jego adres

digital forum

PRENUMERATA

**na cztery kolejne
numery kwartalnika
DIGITALforum**

**Cena kompletu
czterech kolejnych numerów:**

**120.000,-
12,- (n. zł)**

stempel i podpis

symbol
planu kasowego

digital forum

PRENUMERATA

**na cztery kolejne
numery kwartalnika
DIGITALforum**

**Cena kompletu
czterech kolejnych numerów:**

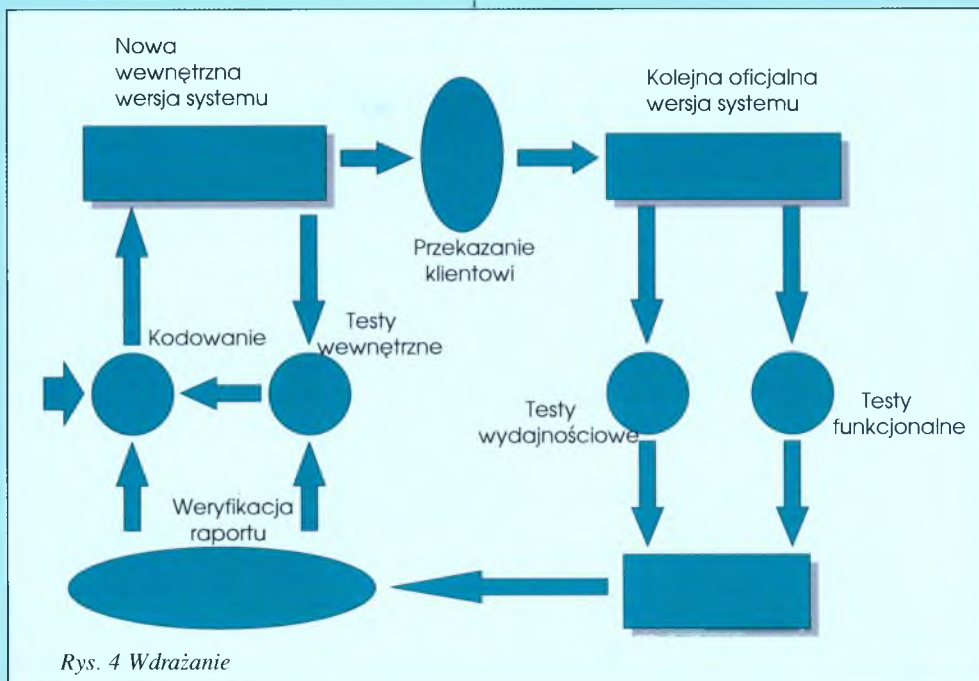
**120.000,-
12,- (n. zł)**

stempel i podpis

symbol
planu kasowego

symbol
planu kasowego

Język do zapisu reguł walidacyjnych posiada szereg złożonych konstrukcji



wników przy dużej wydajności systemu. Kluczowym elementem przy projektowaniu zadań ACMS okazało się także ich zdefiniowanie, aby zapewnić krótkie transakcje dostępu do bazy danych dla pełnego wykorzystania możliwości monitora transakcyjnego. W projekcie wykorzystano również jeszcze jeden zaawansowany mechanizm udostępniany przez system OpenVMS, bazę danych Rdb/VMS oraz ACMS, a mianowicie transakcje rozproszone, kontrolowane przez system DEC DTM (*Distributed Transaction Manager*). Umożliwiło to realizację rozwiąza-

nia, w którym jedna transakcja wysokiego poziomu (transakcja rozproszona) jest wykonywana przez wiele procesów serwerów, co pozwala na większe zrównoleglenie operacji i wzrost wydajności systemu.

Drugim, również bardzo ważnym powodem wykorzystania monitora transakcyjnego jest możliwość łatwej strukturalizacji całej aplikacji. Warstwa pośrednia - *middleware* - pozwala na rozdzielenie obsługi interfejsu użytkownika od procedur dostępu do bazy danych, dzięki czemu moduły te mogą być

Pokwitowanie dla Wpłacającego

zł
słownie.....

wpłacający.....

adres

na rachunek
CLASSICS Sp. z o.o.
Warszawa, ul. Niemcewicz 7/9 lok. 131

IV Oddział PKO SA w Warszawie

r-k nr. 501132-40058562-2541-2-1110

Odcinek dla Posiadacza r-ku

zł
słownie.....

wpłacający.....

adres

na rachunek
CLASSICS Sp. z o.o.
Warszawa, ul. Niemcewicz 7/9 lok. 131

iv Oddział PKO SA w Warszawie

r-k nr. 501132-40058562-2541-2-1110

Odcinek dla Banku

zł
słownie.....

wpłacający.....

adres

na rachunek
CLASSICS Sp. z o.o.
Warszawa, ul. Niemcewicz 7/9 lok. 131

IV Oddział PKO SA w Warszawie

r-k nr. 501132-40058562-2541-2-1110



datownik

podpis przyjm.

Oplata

zł.....



datownik

podpis przyjm.

Oplata

zł.....



datownik

podpis przyjm.

Oplata

zł.....

realizowane niezależnie przez różne zespoły. Poszczególne zadania, składające się zwykle z naprzemiennych kroków wywołania procedur obsługi interfejsu użytkownika i procedur dostępu do bazy danych, pozwalają na naturalny opis zadań wykonywanych przez system. W ten sposób cała logika najwyższego poziomu aplikacji, logika poszczególnych podsystemów jest skupiona w jednym miejscu - w zadaniach ACMS.

Istotną zaletą zastosowanego monitora transakcyjnego jest możliwość dynamicznego reagowania na zmiany obciążenia systemu. W zależności od bieżącego stanu liczby użytkowników wykonujących poszczególne operacje powoływane są lub zwalniane odpowiednie procesy serwerów proceduralnych.

Warto też omówić problemy napotkane przy generowaniu raportów statystycznych. Celem do zrealizowania było danie użytkownikowi systemu zajmującemu się statystyką narzędzia umożliwiającego:

- uzyskiwanie raportów statystycznych zgodnych z wykonywanymi obecnie, w nie zmienionej formie graficznej;
- możliwość modyfikacji zdefiniowanych raportów statystycznych i tworzenie nowych przez użytkownika;
- uzyskiwanie zestawień statystycznych z dowolnych okresów określonych jako przedział czasu pomiędzy wskazanymi latami lub jako przedział czasu pomiędzy wskazanymi datami (parametryzacja globalna - działająca dla wszystkich raportów).
- uzyskiwanie tak zwanych raportów warstwowych, czyli dotyczących danego zagadnienia z rozbićciem na poszczególne miasta, województwa, grupy zawodowe, itp.

Podczas analizy okazało się, że żaden z dostępnych komercyjnych pakietów raportujących działających na systemie OpenVMS nie spełnia założonych wymagań. Stopień komplikacji zagadnienia wymusił skonstruowanie własnego programowanego generatora wydruków statystycznych działającego na bazie statystycznej (niezależnej od bazy Rejestru). Programowanie polega na zdefiniowaniu struktury raportu i zapisanie tej definicji w odpowiednim pliku tekstowym. Definicja taka określa źródło danych, sposób ich przetwarzania oraz formatowania wydruku.

Dla osiągnięcia założonej funkcjonalności, konieczne było utworzenie w bazie danych parametryzowanych perspektyw, tak aby móc uzyskiwać różne strumienie danych wejściowych dla programu generatora wydruków, zmieniając jedynie odpowiednie parametry, a nie zmieniając samych perspektyw. Osiągnięte to zostało dzięki bardzo wygodnemu mechanizmowi bazy Rdb/VMS, polegającemu na możliwości definiowania funkcji zewnętrznych napisanych w dowolnym języku programowania, które można wywoływać z poziomu języka SQL.

Wdrażanie i testowanie systemu

Wdrażanie systemu odbywało się etapami z zasadniczym podziałem na dwie, wspomniane wcześniej, fazy projektu, kończone szczegółowym odbiorem systemu. Etapy pośrednie - kamienie milowe - zakładały instalowanie kolejnych wersji aplikacji z coraz bardziej rozbudowaną funkcjonalnością. Typowy cykl wdrożenia kolejnego modułu aplikacji przebiegał następująco: implementacja - wewnętrzne testy - przekazanie klientowi kolejnej wersji systemu - testy modułu systemu wykonywane przez pracowników klienta kończone raportem ewentualnych usterek i niezgodności z wizją klienta - akceptacja (lub odrzucenie) zmian zależnie od wagi przez szefa projektu bądź przez menedżera ze strony firmy Digital - poprawianie usterek przez zespół implementujący - wewnętrzne testowanie - kolejna wersja aplikacji. Dla modułów o najbardziej rozbudowanej funkcjonalności proces wdrażania miał wiele takich cykli.

Ważnym elementem w procedurze wdrażania systemu było intensywne testowanie systemu przez klienta i strojenie systemu przez pracowników firmy PentaComp. Strojenie przebiegało w kilku etapach. Po wdrożeniu aplikacji zdalnej nastąpiło podstawowe strojenie bazy danych w celu optymalizacji zapytań. Wdrożenie modułów aplikacji obsługujących dokumenty spowodowało konieczność strojenia całego systemu, tzn. bazy danych Rdb/VMS, systemu operacyjnego OpenVMS i pakietu ACMS. Proces strojenia systemu, w którym mieściły się przeprowadzane kilkakrotnie próby obciążenia systemu, trwał nieustannie aż do czasu ostatecznego oddania systemu. Obejmował on modyfikację struktury fizycznej bazy danych, zmiany parametrów bazy, systemu operacyjnego

*Etapy
pośrednie
- kamienie
milowe
- zakładały
instalowanie
kolejnych
wersji
aplikacji
z coraz
bardziej
rozbudowaną
funkcjonalno-
ścią*

*System,
po dwóch
latach pracy,
został
wdrożony.
Można więc
już mówić
o sukcesie.
System
osiągnął przy
tym założoną
funkcjonal-
ność i w tym
zakresie
spełnia on
wymagania
klienta*

i systemu ACMS. Pozwoliło to osiągnąć wydajność o ok. 30 procent większą w stosunku do pierwszych testów obciążenia. Wykorzystywane były narzędzia monitorowania i analizy obciążenia bazy danych, systemu operacyjnego i ACMS.

Bardzo istotną cechą wykorzystywanego oprogramowania systemowego jest możliwość zmiany pewnych istotnych parametrów w czasie działania aplikacji, co umożliwia bieżące dostrajanie systemu do aktualnych wymagań. Zadanie to wykonywane jest z powodzeniem przez administratorów systemu, dzięki czemu możliwości systemu wykorzystywane są w sposób optymalny.

Oczywiście warunkiem niezbędnym wdrożenia nowego systemu jest intensywne przeszkolenie jego przyszłych użytkowników, przy czym firma przygotowująca system może przeprowadzić takie szkolenia tylko od strony technicznej. Klient natomiast, we własnym zakresie (choć najlepiej we współpracy z firmą przygotowującą system), powinien przeprowadzać szkolenia merytoryczne pracowników dotyczące tego, jak poprawnie wprowadzać do systemu informacje, jak przy wprowadzaniu danych interpretować różnorodne sytuacje, itp. Pracownik musi bowiem nie tylko umieć obsłużyć aplikację, ale mieć również dostateczną wiedzę merytoryczną na temat przedmiotu działania systemu. Wa-

żne jest też, by użytkownik miał możliwość odpowiednio długi czas pracować na danych testowych, pod kontrolą osób znających techniczną i merytoryczną stronę systemu.

Najistotniejsze spostrzeżenie dotyczące wdrożenia systemu związane jest ze współpracą z klientem. Prawidłowe wdrożenie systemu nie może się obyć bez dużego zaangażowania klienta we współpracę z zespołem projektującym system. Realizacja i wdrożenie nowego systemu w firmie wymaga wręcz pewnej determinacji ze strony klienta. Niezbędna jest tutaj też dobra wola obu stron, dobra atmosfera i wzajemne zaufanie, albowiem w tak dużym przedsięwzięciu nie wszystkie sprawy muszą być rozwiązywane w sposób ściśle formalny. Dzięki temu, że założenia te były w opisywanym projekcie spełnione, osiągnięcie sukcesu w realizacji systemu było możliwe.

Podsumowanie

System, po ponad dwu latach pracy, został wdrożony. Można więc już mówić o sukcesie. System osiągnął przy tym założoną funkcjonalność i w tym zakresie spełnia on wymagania klienta. Zastosowana technologia okazała się skuteczna w realizacji projektu o dużym stopniu komplikacji i wyśrubowanych założeniach wydajnościowych.

Podstawowy zespół programistyczny przy projekcie "Centralny Rejestr Skazanych" tworzyli:

*mgr inż. Piotr Salata - szef zespołu
mgr inż. Wojciech Chybowski
mgr inż. Dariusz Figura
mgr inż. Wojciech Kamiński
mgr inż. Michał Kirpluk*

PentaComp s.c ul. Kleczewska 99, 01-826 Warszawa, tel./fax: (22) 346464
email: pentaco@ikp.atm.com.pl

Firma PentaComp specjalizuje się w:

- tworzeniu dużych systemów informatycznych dla przedsiębiorstw i instytucji państwowych w oparciu o technologię middleware
- integracji pracy biurowej w oparciu o pakiet LinkWorks firmy Digital Equipment Corp.
- doradztwie w zakresie projektów software'owych,
- auditingu projektów software'owych
- szkoleniach w zakresie baz danych i narzędzi CASE

CO ROBIĆ, JEŚLI NADLUDZKIM WYSIŁKIEM UPCHNĘLIŚMY WSZYSTKIE NIEZBĘDNE DANE...

... A WCIĄŻ PRZYCHODZĄ NOWE?

Właśnie wtedy, gdy myślisz, że Twój system informatyczny nie jest w stanie przerobić więcej danych, przychodzi kolejna porcja. Nieważne, dlaczego tak się dzieje. To się zdarza. Próbujesz więc temu zaradzić. Rozbudowujesz system. Zmieniasz platformę na architekturę klient/serwer. Idealne rozwiązanie masz w zasięgu ręki. Jest przecież wysoce wydajna rodzina komputerów AlphaServer firmy Digital, która z łatwością rozwiąże problemy każdego przedsiębiorstwa - od linii lotniczych po instytucje finansowe. Przykładowo, komputery AlphaServer 2100

5/300 mogą wykonywać ponad 3000 transakcji na sekundę. Systemy AlphaServer 8200 i 8400 są jedynymi serwerami, które mogą w pełni wykorzystać 64-bitową bazę danych Oracle7™ pracując 100 razy szybciej niż konwencjonalne systemy 32-bitowe. Taka jest potęga 64-bitowej Alphy.

Umożliwia ona znacznie

szyb-

sze i tańsze

realizowanie zadań,

często niewykonywalnych dla

tradycyjnych systemów 32-bitowych. Dla komputerów AlphaServer

przygotowano tysiące aplikacji, wśród nich są te, których potrzebujesz. Budując cały system informatyczny z Digitalem, wkraczasz w przyszłość. Nie ma co czekać. Zrób to. Zadzwoń jeszcze dziś, aby uzyskać

digital™

dodatkowe informacje.

Digital Equipment Poland,

ul. Wołoska 18, 02-672 Warszawa, tel.: (48 22) 64 00 123, fax: (48 22) 64 00 111.

SYSTEMY ALPHA SERVER

oferują średnim i dużym przedsiębiorstwom szeroki zakres wydajności:

Ilość transakcji na sekundę
od 130 do 11000+

Maksymalna pamięć operacyjna
od 192 MB do 14 GB

Przepustowość magistrali I/O
od 132 do 1200 MB/sek.

SPECint92 - 116.8 do 337.8

SPECfp92 - 139.7 do 503.2



DIGITAL EQUIPMENT POLSKA

ul. WOŁOSKA 18 (d. KOMAROWA)

02-672 WARSZAWA

tel. 640-01-23

fax 640-01-11

Biuro w Gliwicach

ul. Akademicka 16

44-100 Gliwice

tel./fax 832. 37-20-44

