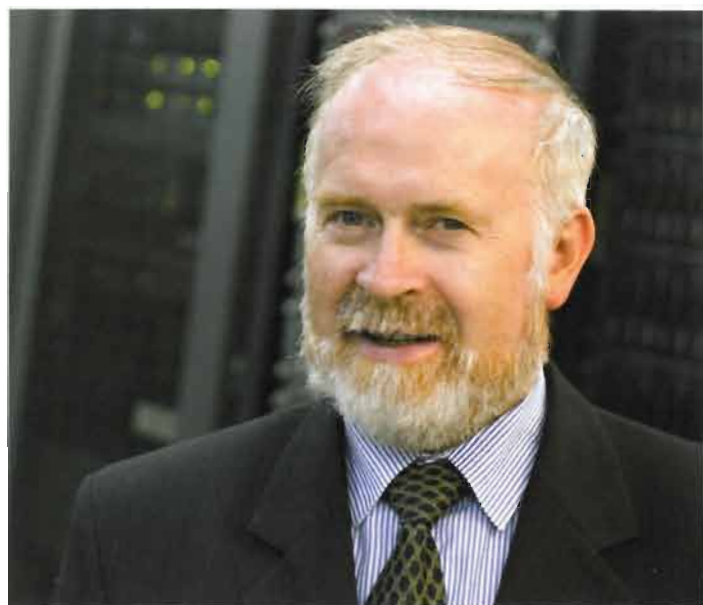


Historia Cyfronetu i EVY 8000



Cyfronet tworzy PL-Grid – rozproszoną infrastrukturę do przetwarzania i składowania dużych ilości danych na potrzeby nauki, a w przyszłości zapewne także dla znacznie szerszych zastosowań. Zdobywane przez nas obecnie doświadczenia techniczne i organizacyjne są niezwykle cenne, nie tylko dla nas, ale także dla całej polskiej nauki, która coraz lepiej i coraz pełniej będzie korzystać z możliwości, jakie będą związane z infrastrukturą PL-Grid.

Prof. dr hab. inż. Kazimierz Wiatr, Dyrektor ACK Cyfronet AGH

Środowisko naukowe w Polsce nie dysponuje nieograniczonymi funduszami na informatykę. Akademiczne Centrum Komputerowe CYFRONET AGH nie jest pod tym względem wyjątkiem.

Jakiś czas temu Cyfronet poszukiwał rozwiązania do zarządzania wielką ilością danych, związanych z bieżącymi potrzebami pracowników naukowych, badań i eksperymentów prowadzonych przez użytkowników Cyfronetu, jak również badań, w których Cyfronet współpracuje z ośrodkami naukowymi i superkomputerowymi. Okazało się, że ten ostatni przypadek stawia w praktyce największe wymagania. Cyfronet został wyznaczony jako ośrodek koordynujący w tej części Europy (7 państw) wielki program EGEE (Enabling Grids for E-sciencE), obecnie będący w drugiej fazie (EGEE II).

Aby sprostać planowanym potrzebom w dziedzinie przechowywania i wydajnego przetwarzania dużych ilości danych, Cyfronet zakupił i wdrożył macierz dyskową HP EVA 8000 o pojemności ok. 120 Terabajtów. Doskonale odpowiadała ona jego potrzebom. Zachęcony zaletami HP EVA 8000, Cyfronet zakupił ostatnio dodatkową macierz EVA z rodziny 8100, stanowiącą nowocze-

śniejszą wersję obecnie posiadanej. Charakteryzuje się ona większą wydajnością oraz znacznie mniejszym poborem mocy.

SUPER SUPERKOMPUTER

EGEE to jedna z inicjatyw związanych z przetwarzaniem danych pochodzących z wielkiego akceleratora cząstek elementarnych, budowanego pod Genewą w Europejskim Centrum Badań Jądrowych CERN. Już wstępne analizy i testy CERN wykazały, że potrzeby w dziedzinie mocy obliczeniowej i pojemności na dane będą po uruchomieniu akceleratora znacząco przewyższać możliwości jednego, a nawet kilku dużych ośrodków obliczeniowych. Szacowane potrzeby na moc obliczeniową to ok. 100 tysięcy procesorów klasy Pentium 4, zaś przewidywana pojemność na dane – ok. 20 Petabajtów rocznie.

W związku z tymi niezwykle wysokimi wymaganiami CERN zachęcił do współpracy ośrodki naukowe i centra obliczeniowe z całego świata, co doprowadziło do rozbudowy infrastruktury superkomputerowej, m.in. w ramach programów DataGrid i CrossGrid.

W skali świata wielomilionowe inwestycje pozwoliły urzeczywistnić prowadzenie wspólnych badań nad trudnymi problemami naukowymi. W ramach samego tylko programu EGEE, będącego jednym z kilku programów realizowanych równoległe, współpracuje ze sobą ponad 200 ośrodków obliczeniowych z całego świata.

Współpraca ta jest możliwa dzięki oprogramowaniu automatyzującemu zarządzanie przesyłaniem, składowaniem i przetwarzaniem wielkich ilości danych w ramach rozproszonej infrastruktury. Architektura tego oprogramowania zakłada silne rozproszenie przetwarzania zgodnie z koncepcją grid computing (przetwarzanie sieciowe). Również samo przetwarzanie sieciowe stanowi przedmiot badań naukowych w Cyfronecie i w wielu innych ośrodkach.

Cyfronet tworzy PL-Grid – rozproszoną infrastrukturę do przetwarzania i składowania dużych ilości danych na potrzeby nauki, a w przyszłości także innych zastosowań. Zdobywane przez nas obecnie doświadczenia techniczne i organizacyjne są niezwykle cenne dla nas, jako ośrodka obliczeniowego, ale także dla całej polskiej nauki. – mówi Dyrektor Cyfronetu.

WIELE DANYCH, WIELKIE WYZWANIA

Program EGEE II dąży do zwiększenia możliwości zarządzania danymi. Oprogramowanie grid spełnia rolę inteligentnej warstwy pośredniej (middleware), zarządzającej planowaniem obliczeń, zabezpieczaniem na ich potrzeby zasobów aplikacyjnych i sprzętowych, udostępnianiem aplikacji użytkownikom pracującym lokalnie lub w innych ośrodkach, śledzeniem postępu obliczeń, raportowaniem wyników, etc. Aplikacje modelujące, obliczeniowe i inne działające w środowisku grid middleware wykorzystują ogromne ilości danych. Ani aplikacje, ani środowisko grid, nie zarządzają jednak danymi na poziomie operacyjnym.

W Cyfronecie wiele aplikacji działa równoległe, jednocześnie zapisując i odczytując dane w macierzy dyskowej. *Zapisy i odczyty są w większości przypadków sekwencyjne, ale ponieważ równoległe odbywa się ich bardzo dużo – na rzecz wielu aplikacji i użytkowników, z punktu widzenia centrum danych, jako całości możemy mówić o obciążeniu quasi-transakcyjnym. Obsłużenie obciążeń o takiej charakterystyce wymaga bardzo wydajnej macierzy klasy korporacyjnej, na którą jednak nie mogliśmy sobie pozwolić – relacjonuje Maciej Twardy, Kierownik Działu Archiwizacji i Bezpieczeństwa Danych w Cyfronecie.*

CEL	ROZWIĄZANIE	REZULTATY
<ul style="list-style-type: none"> • Budowa zrębów infrastruktury do bezpiecznego przyjmowania, składowania, przetwarzania i udostępniania wielkich ilości danych na użytek polskich i zagranicznych środowisk naukowych 	<ul style="list-style-type: none"> • Macierz HP EVA 8000, 240 dysków FATA 500 GB, dwa kontrolery, z których każdy zawiera 4 GB pamięci buforowej i 4 interfejsy Fibre Channel • Oprogramowanie HSM – HP StorageWorks File System Extender • Oprogramowanie do tworzenia kopii bezpieczeństwa HP Data Protector • Biblioteka taśmowa HP StorageWorks ESL 712e z napędami Ultrium 460, czyli LTO3 z 698 slotami na taśmy 	<ul style="list-style-type: none"> • Bardzo wysoka wydajność zapisu i odczytu danych przez wiele aplikacji jednocześnie; stałe pasmo do/z portu na poziomie 120-150 MB/s • Bardzo dobry stosunek wydajności do ceny • Możliwość łatwego dostosowania parametrów pracy macierzy bez konieczności wzywania serwisu • Możliwość zdefiniowania wirtualnych zasobów przy jednoczesnym oszczędnym wykorzystaniu rzeczywistej pojemności



Dylemat Cyfronetu nie polegał jedynie na znalezieniu rozwiązania dostatecznie wydajnego i pojemnego. Jednym z ważniejszych czynników podczas rozważań nad wyborem architektury dla podsystemu dyskowego był niski koszt zarządzania i utrzymania. Firmy komercyjne martwią się, gdy muszą dokonywać zmian w konfiguracji grup RAID, zmieniać wielkości wolumenów czy podział pasma na interfejsach częściej niż raz na miesiąc. My nie mamy takiego komfortu – zmian należy dokonywać praktycznie codziennie, co w typowym scenariuszu oznacza wysokie rachunki za wsparcie. Tego obawialiśmy się najbardziej – mówi Marek Pogoda, Specjalista-Administrator w Dziale Archiwizacji i Bezpieczeństwa Danych w Cyfronecie.

Jednym z dawno rozpoznanych przez Cyfronet problemów był także rzeczywisty poziom wykorzystania zasobów macierzy. Sądziliśmy, że kupując pewną pojemność, po zdefiniowaniu grup RAID będziemy mogli w praktyce zappełnić je w ok. 75-85%. Po zakupie macierzy HP EVA 8000 przekonaliśmy się, że to nie zawsze musi być prawda – mówi Maciej Twardy.

WIRTUALNA MACIERZ NA REALNE PROBLEMY

Macierz HP EVA 8000 pojawiła się w Cyfronecie w marcu 2006 r. Zanim nabyliśmy nową macierz,

staraliśmy się sprawdzić jej możliwości. Planowana konfiguracja miała 240 dysków. To niby dużo, ale każdy doświadczony informatyk wie, że liczy się wydajność urządzenia jako całości. Chcieliśmy zbadać m.in. to, jak macierz poradzi sobie z wydajnością na interfejsach, jak będzie reagować na ciągłe przepisywanie całości zawartości pamięci buforowej, jak będzie wyglądać efektywność wielodostępu, itd. – mówi Marek Pogoda.

Testy wydajnościowe z aplikacjami wykorzystywanymi na co dzień przez użytkowników Cyfronetu i próby z aplikacjami gridowymi wypadły nadzwyczaj dobrze. W testach, a także podczas późniejszej eksploatacji, udało się nam uzyskać stałą wydajność rzędu 120-150 MB/s na każdym z ośmiu interfejsów Fibre Channel 2 Gb/s. Zrównoleglenie przetwarzania na wszystkie dyski jednocześnie, realizowane przez wewnętrzne mechanizmy wirtualizujące macierzy, rzeczywiście ma fenomenalny wpływ na jej osiągi. Wydajność EVA jest tym bardziej godna uwagi, że wewnątrz działają relatywnie tanie dyski FATA, wykorzystujące technologię ATA, jednak z interfejsami komunikacyjnymi w standardzie Fibre Channel – mówi Maciej Twardy.

Przy okazji testów okazało się także, że obawy o wysokie koszty wsparcia technicznego są nieuzasadnione. Szczerze mówiąc, było to dla nas



zupelne novum. Przyzwyczailismy sie do zarzadzania macierzami z wiersza poleceń, tymczasem EVA miała elegancki, graficzny interfejs. Za pomocą konsoli graficznej można było zarządzać dowolnym aspektem działania macierzy: definiować grupy RAID, dodawać do nich nowe dyski, tworzyć pule, zwiększać ich pojemność, zarządzać uprawnieniami, pasmem, analizować statystyki... – wymienia Maciej Twardy.

Niezwykle funkcjonalną okazała się możliwość przydzielenia aplikacjom dowolnej pojemności, przy jednoczesnym wykorzystaniu tylko takiej pojemności, jaka aktualnie rzeczywiście jest potrzebna.

W celu bieżącego zabezpieczenia danych przetwarzanych w macierzy EVA HP 8000 Cyfronet stosuje sprawdzone oprogramowanie do tworzenia kopii bezpieczeństwa HP Data Protector i oprogramowaniem typu HSM – HP StorageWorks File System Extender. Do wykonywania bieżących kopii i archiwizowania danych Cyfronet wykorzystuje bibliotekę taśmową HP StorageWorks ESL 712e z napędami Ultrium 460, czyli LTO3. Backup i archiwizacja odbywają się poprzez sieć SAN zbudowaną w oparciu o rozwiązania HP StorageWorks SAN Switch: dwa 32-portowe przełączniki Fibre Channel 2 Gb/s HP StorageWorks SAN Switch oraz dwa 32-portowe przełączniki Fibre Channel 4 Gb/s, działających redundantnie (w konfiguracji lustrzanej).

W tym roku w Cyfronecie pojawiła się nowa macierz EVA, zawierająca nowsze wersje kontrolerów z portami Fibre Channel 4 Gb/s. Początkowo nie będzie wypełniona tyloma dyskami, co macierz dotychczasowa, jednak z biegiem czasu zapewne to nastąpi, ponieważ jeszcze w tym roku z CERN popłyną pierwsze „produkcyjne” porcje danych. *Architektura, funkcjonalność i niezawodność macierzy EVA zostały sprawdzone w praktyce. Obecnie mamy bardzo dobrą opinię na temat dysków FATA, do których początkowo podchodziliśmy bardzo nieufnie. Ich awaryjność okazała się mitem, a wydajność pojedynczego dysku – jeżeli wziąć pod uwagę architekturę macierzy jako całości – nie ma wielkiego znaczenia – podsumowuje Maciej Twardy.*

W SKRÓCIE

Aplikacje

- Abaqus
- Accelrys
- Ansy
- ArcInfo,
- BlackBoard
- Erdas Imagine
- Fluent
- Gaussian
- Maple
- Mathematica
- Matlab
- Opera 2D
- Oracle
- SAS
- Tripos

Rozwiązania

- Macierz HP EVA 8000, 240 dysków FATA 500 GB, dwa kontrolery, z których każdy zawiera 4 GB pamięci buforowej i 4 interfejsy Fibre Channel
- Biblioteka taśmowa HP StorageWorks ESL 712e z napędami Ultrium 460, czyli LTO3 z 698 slotami na taśmy

Oprogramowanie

- Oprogramowanie HSM HP StorageWorks File System Extender
- Oprogramowanie do tworzenia kopii bezpieczeństwa HP Data Protector