

System Składowania Danych ACK Cyfronet AGH

Obserwowane obecnie zjawisko lawinowo rosnącej ilości wytwarzanych informacji w postaci cyfrowej dotyczy także środowiska naukowego. Dostęp do potężnych superkomputerów umożliwia podejmowanie analiz wielkoskalowych problemów badawczych, w efekcie czego powstają olbrzymie zbiory danych. Wymagają one całkowicie nowego podejścia do procesów przetwarzania i przechowywania informacji. Problem ten, stanowiący obecnie jedno z jedno z najważniejszych wyzwań współczesnego świata cyfrowego, określany jest pojęciem BigData. Również i w ACK Cyfronet AGH wyraźnie widoczna jest korelacja pomiędzy wzrostem oczekiwań w zakresie dostępnej pojemności, szybkości oraz dodatkowych funkcjonalności zasobów pamięci masowej, spowodowanym oferowaniem coraz wydajniejszych systemów obliczeniowych. Niezwykle zróżnicowana charakterystyka tematyczna realizowanych zadań, wykonywanych w infrastrukturze udostępnianej przez Centrum, znajduje swoje odbicie w stopniu złożoności architektury Systemu Składowania Danych – głównej platformy pamięci masowej dla Komputerów Dużej Mocy Obliczeniowej. System ten złożony jest z kilku głównych elementów:

- sieci SAN – dedykowanej, wydajnej i wysokodostępnej sieci przeznaczonej do komunikacji urządzeń wchodzących w skład Systemu Składowania Danych oraz klientów wykorzystujących udostępniane zasoby lub usługi;
- macierzy oraz serwerów dyskowych oferujących przestrzeń przeznaczoną na składowanie danych użytkowników o dużej rozpiętości parametrycznej – począwszy od szybkich lecz równocześnie drogich i mało pojemnych rozwiązań, a skończywszy na urządzeniach bardzo pojemnych i relatywnie tanich, ale zarazem o ograniczonej wydajności;
- serwerów usługowych wraz ze specjalistycznym oprogramowaniem usługowym i wirtualizacyjnym, udostępniających dla użytkowników funkcjonalności takie jak: automatyczne kopie zapasowe i archiwalne, hierarchiczne systemy składowania danych, wydajne sprzętowe platformy plików bądź rozproszone sieciowe systemy plików;
- bibliotek taśmowych oraz specjalizowanego oprogramowania służącego zapisywaniu krytycznych danych użytkowników na nośnikach magnetycznych;
- infrastruktury dodatkowej, np. sieci Ethernet, Infiniband, rozwiązań wspomagających zarządzanie infrastrukturą IT oraz umożliwiających bezpieczne przechowywanie nośników magnetycznych.

Zrównoważony rozwój tak złożonego systemu wymaga zachowania właściwych proporcji pomiędzy jego poszczególnymi elementami składowymi, przy równoczesnej korelacji z planowanymi działaniami w obszarze Komputerów Dużej Mocy Obliczeniowej. Długoletnie doświadczenia ACK Cyfronet AGH na tym polu pokazują jednoznacznie, że tylko taki sposób kształtowania infrastruktury w efekcie końcowym umożliwi maksymalizację efektów podejmowanych prac modernizacyjnych.