# IBM General Parallel File System

GPFS to skrót od General Parallel File System, systemu plików klastra opracowanego przez IBM. Według oficjalnej strony IBM jest sprawdzonym, niezawodnym, skalowanym i wysoce wydajnym rozwiązaniem do zarządzania danymi i plikami. System oferuje wysokiej klasy narzędzia do zarządzania pamięcią masową i charakteryzuje się dużą skalowalnością. Jego mocną stroną jest wykorzystanie potencjału wydajnościowego pamięci masowej flash i automatyczne przenoszenie danych pomiędzy poziomami pamięci masowej (flash, dysk i taśma). Zastosowanie ogranicza koszty przechowywania danych nawet o 90% i przyczynia się do wzrostu poziomu bezpieczeństwa i efektywności zarządzania w środowiskach takich jak: chmury, wielkie zbiory danych i analiz. Podobnie jak w przypadku typowych systemów plików klastrowych, GPFS zapewnia równoczesny szybki dostęp do plików dla aplikacji wykonywanych na wielu węzłach klastrów. Może być używany z klastrami AIX 5L, klastrami Linux i Microsoft Windows Server lub heterogenicznymi klastrami węzłów AIX, Linux i Windows. Oprócz zapewnienia funkcji przechowywania plików, system plików GPFS zapewnia narzędzia do zarządzania klastrem GPFS i administrowania nim oraz umożliwia współdzielony dostęp do systemów plików ze zdalnych klastrów GPFS. Wyrósł on z kilku projektów badawczych równoległych systemów plików i był on dystrybuowany pod trzema nazwami: IBM General Parallel Filesystem, Elastic Storage i Spectrum Scale (który od 17 lutego 2015 jest oficjalną nazwą).

# Cechy GPFS

IBM General Parallel File System jest produktem dostępnym w trzech postaciach: jako rozwiązanie programowe, jako zintegrowane rozwiązanie IBM Elastic Storage Server (sprzęt + oprogramowanie + usługi) oraz jako usługa w chmurze. Rozwiązanie software’owe działa niemal na każdej platformie sprzętowej i obsługuje niemal każde urządzenie blokowej pamięci masowej. Działa na systemach Linux, AIX i Windows. Jako oferta IBM Elastic Storage Server obejmuje dostępność danych na wszystkich poziomach aż do użytkownika końcowego, zapewni niezawodność i integralność dzięki zastosowaniu technologii IBM Spectrum Scale RAID. Oraz jako usługa w chmurze oferuje skalowalny i wydajny system przechowywania danych oraz zintegrowane mechanizmy nadzoru nad danymi w chmurze IBM Softlayer.

GPFS został wymyślony ponieważ przyrost danych, który nie ma dziś końca jest wyzwaniem dla systemów zarządzania pamięcią masową i danymi. Nowe aplikacje generują olbrzymie ilości danych nieustrukturyzowanych: pliki tekstowe, audio, wideo. Zarządzanie nimi musi uwzględniać zarówno standardowe platformy, jak i nowe środowiska (np. chmury). Dla zrównoważenia obciążeń i eliminacji wąskich gardeł, które wpływają na pogorszenie pracy systemów zarządzania plikami wymyślono GPFS. Wg IBM-u eliminuje on wąskie gardła zapewniając równoległy dostęp do danych. Wyeliminowano pojedyncze miejsca “przytykające” zarządcę plików lub nierównowagi w intensywności dostępu. System upraszcza zarządzanie ilością danych, udostępniając jedną przestrzeń nazw, którą łatwo, szybko i bez ograniczeń można skalować poprzez proste dodawanie zasobów. Eliminuje to problem chaotycznej rozbudowy systemu zarządzania plikami.

GPFS współdziała na skalę globalną, jest świetnym rozwiązaniem dla organizacji, które są rozproszone geograficznie, ponieważ pojedyncza przestrzeń nazw ma zasięg globalny. Jednak newralgiczne dane znajdują się w pobliżu użytkowników, którzy ich potrzebują. Rozwiązanie IBM zarządza cyklem życia danych. Mechanizmy zarządzania i automatyzacji cyklu życia danych pomagają w zachowaniu kontroli nad kosztami przechowywania danych i czynią procesy obsługi kopii zapasowych oraz usuwania skutków katastrof integralnymi składnikami kompleksowego rozwiązania. Wg IBM-u produkt w sposób nieosiągalny dla innych produktów zarządza całym cyklem życia danych, przynosząc oszczędności rosnące w postępie geometrycznym dzięki automatyzacji w oparciu o strategie i zarządzanie poziomami pamięci masowej.

# Historia

GPFS wyszedł z projektów badawczych IBM „Tiger Shark File System” i „Vesta File System” i był początkowo nazywany systemem dla multimediów. Szybko okazało się, że GPFS jest szczególnie korzystny dla komputerów o dużej wydajności z powodu ich równoległej architektury. W 1998 roku GPFS pojawił się jako oficjalny produkt IBM i następca Vesta / PIOFS jako system plików zgodny z Posix .

Systemem plików był używany w superkomputerach ASCI White and Purple w Lawrence Livermore National Laboratory, jednak z czasem został przeniesiona na nowsze inne systemy operacyjne:

* AIX od 1998 roku
* Linux od 2001 roku
* Windows od 2008 roku

GPFS obsługuje również inne protokoły sieciowe, takie jak Windows CIFS. Pierwotnie system plików był wykorzystywany tylko na wielkich urządzeniach do storage’u, jednak później oddzielono go od sprzętu i zaczęto sprzedawać go jako oprogramowanie. Niedawno dodano nowe umiejętności, takie jak klastry współdzielone. W dniu 17 lutego 2015 r. Zmieniono jego nazwę na IBM w Spectrum Scale.

# Architektura

GPFS zapewnia wysoką wydajność, umożliwiając dostęp do danych na wielu komputerach jednocześnie. Większość istniejących systemów plików zaprojektowano dla środowiska pojedynczego serwera, a dodanie większej liczby serwerów plików nie poprawia wydajności. GPFS zapewnia wyższą wydajność wejścia / wyjścia poprzez usuwanie bloków danych z poszczególnych plików na wielu dyskach oraz równoległe odczytywanie i zapisywanie tych bloków. Inne funkcje oferowane przez GPFS obejmują wysoką dostępność, obsługę heterogenicznych klastrów, odzyskiwanie po awarii, bezpieczeństwo, DMAPI , HSM i ILM .

Plik zapisywany do systemu plików jest dzielony na bloki o skonfigurowanym rozmiarze, o wielkości poniżej 1 megabajta. Bloki te są rozmieszczone na wielu węzłach systemu plików, dzięki czemu pojedynczy plik jest w pełni dystrybuowany w całej macierzy dyskowej . Powoduje to dużą szybkość odczytu i zapisu dla pojedynczego pliku, ponieważ łączna przepustowość wielu fizycznych napędów jest wysoka. To sprawia, że ​​system plików jest podatny na awarie dysków - w przypadku awarii jednego dysku utracimy dane. Aby zapobiec utracie danych, węzły systemu plików mają kontrolery RAID - wiele kopii każdego bloku jest zapisywanych na dyskach fizycznych w poszczególnych węzłach. Można również zrezygnować z bloków replikowanych macierzy RAID, a zamiast tego przechowywać dwie kopie każdego bloku na różnych węzłach systemu plików.

Funkcje systemu plików obejmują:

* Rozproszone metadane, w tym drzewo katalogów. Nie ma jednego "kontrolera katalogów" ani "serwera indeksów" odpowiedzialnych za system plików.
* Wydajne indeksowanie pozycji katalogu dla bardzo dużych katalogów. Wiele systemów plików ogranicza się do niewielkiej liczby plików w jednym katalogu (często 65536 lub podobnym małym numerze binarnym). GPFS nie ma takich limitów.
* Blokowanie rozproszone - Pozwala to na pełną semantykę systemu plików Posix , w tym blokowanie dostępu do wyłącznego pliku.
* Partition Aware - Awaria sieci może podzielić system plików na dwie lub więcej grup węzłów, które mogą widzieć tylko węzły w swojej grupie. Można to wykryć za pomocą protokołu pulsu, a gdy nastąpi partycja, system plików pozostanie pod napięciem dla największej utworzonej partycji. Zapewnia to zgrabną degradację systemu plików - niektóre maszyny będą działać.
* Konserwację systemu plików można przeprowadzić online. Większość zadań związanych z utrzymaniem systemu plików (dodawanie nowych dysków, przywracanie równowagi danych na dyskach) może być wykonywane, gdy system plików działa. Zapewnia to, że system plików jest dostępny częściej, dzięki czemu sam klaster superkomputera będzie dłużej dostępny.

Interesujące jest porównanie tego z systemem plików HDFS Hadoop , który jest przeznaczony do przechowywania podobnych lub większych ilości danych na towarowym sprzęcie - tj. W centrach danych bez dysków RAID i sieci pamięci masowej (SAN).

* GPFS obsługuje pełną semantykę systemów plików Posix. HDFS i GFS nie obsługują pełnej zgodności ze standardem Posix.
* GPFS dystrybuuje swoje indeksy katalogów i inne metadane w całym systemie plików. Z kolei Hadoop utrzymuje to na podstawowych i dodatkowych dużych serwerach, które muszą przechowywać wszystkie informacje o indeksach w pamięci RAM.
* GPFS dzieli pliki na małe bloki. Hadoop HDFS lubi bloki 64 MB lub więcej, ponieważ zmniejsza to zapotrzebowanie na pamięć Namenode. Małe bloki lub wiele małych plików szybko wypełniają indeksy systemu plików, więc ogranicz rozmiar systemu plików.

# Możliwości GPFS

Zintegrowane systemy pamięci masowej od sprzętu i oprogramowania IBM z GPFS w systemie operacyjnym Linux to:

* V7000 Unified lub IBM Storwize V7000F
* SONAS - Scale Out Network Attached Storage https://www.ibm.com/en-us/marketplace/scale-out-file-and-object-storage (Spectrum Scale)

GPFS / Spectrum Scale ma następujące właściwości funkcjonalne:

* kilka komputerów NAS może jednocześnie montować wolumin klastrowy (równolegle), a system plików jest skalowalny dla wielu klientów.
* Paski, a tym samym równoległe czytanie i pisanie są obsługiwane na poziomie pamięci masowej i pojedynczych plików. Ta równoległość pozwala osiągnąć bardzo wysokie przepustowości.
* Distributed Lock Manager: Równoległe zapisywanie do systemu plików jest możliwe, ponieważ jeden plik może być zapisany tylko w jednym procesie naraz
* Metadane i dane mogą być dystrybuowane na różnych woluminach w celu zwiększenia wydajności
* Kilka serwerów GPFS (zwanych także węzłami) działa jako klaster o wysokiej dostępności, a awarie są przechwytywane
* GPFS może wersja 4.1 też działa na zasadzie klastra shared-nothing (FPO - Umieszczenie plików Optimizer) i może w ten sposób jako HDFS pracy
* bardzo duże limity rozmiaru pliku (8 EB), wielkość katalogu, rozmiar systemu plików (8 YB), liczba plików w systemie plików (2 ^ 64)
* Obsługa HSM / Hierarchical Storage Management
* woluminy mogą być wydawane jednocześnie z CIFS i protokołem NFS , od wersji 4.1 również jako rozproszony system plików Hadoop .
* Kontrola praw dostępu działa dla systemu plików NFS (dla systemów uniksowych) z uprawnieniami do plików POSIX i dla CIFS (systemów Windows) z listami ACL . Te prawa dostępu do plików są niezależnie kontrolowane
* System plików działa zgodnie z zasadą kopiowania przy zapisie . Podobnie jak w przypadku "kopii w tle" systemu Windows, do migawek można uzyskać dostęp z dowolnego wyeksportowanego katalogu, zarówno NFS, jak i CIFS
* Możliwe jest replikacja asynchroniczna między różnymi woluminami GPFS (Active File Management)

<https://www.ibm.com/us-en/marketplace/scale-out-file-and-object-storage/details#product-header-top>

<https://www.usenix.org/legacy/events/fast02/full_papers/schmuck/schmuck.pdf>

<https://asc.llnl.gov/computing_resources/purple/>

<https://www.ibm.com/support/knowledgecenter/en/STXKQY_4.2.3/com.ibm.spectrum.scale.v4r23.doc/bl1adv_fposettings.htm>