

W jaki sposób CERN wykorzystuje OpenStack do obróbki 1 PB/s

Przez **Adam Jadczak** - 5 maja 2017

Z Timem Bellem, liderem w dziale IT w Europejskiej Organizacji Badań Jądrowych CERN rozmawiamy o jednej z największych na świecie instalacji platformy OpenStack, z której korzysta 11 tys. fizyków, narzędziach do zarządzania nią, zarządzaniem dużymi zbiorami danych i wykorzystanych systemach pamięci masowych oraz współpracy ze społecznościami open source.

CERN ma jedną z największych na świecie instalacji OpenStack. Jakie problemy rozstrzygnięto przez to open source, których nie można było rozwiązać za pomocą „tradycyjnych” technologii wirtualizacji? Dlaczego postanowiliście wprowadzić to rozwiązanie?

OpenStack pozwolił nam zautomatyzować wiele procedur związanych z zarządzaniem i utrzymaniem infrastruktury, umożliwiając np. naszym użytkownikom wykonywanie wielu działań w portalu samoobsługowym. Początek projektu wdrożenia OpenStack sięga jeszcze roku 2012, gdy zaczęliśmy przygotowywać infrastrukturę informatyczną, która miała być gotowa na drugie uruchomienie Wielkiego Zderzacza Hadronów – LHC (Large Hadron Collider) Run 2 w roku 2015.

W czasie drugiego uruchomienia akceleratora LHC mamy do czynienia z dwukrotnie wyższą energią wiązki. Wiązało się to ze znacznym wzrostem ilości przesyłanych danych i wyższym zapotrzebowaniem na moc obliczeniową. Liczba pracowników w CERN pozostała jednak na ustalonym poziomie, dlatego też musieliśmy znaleźć sposób na usprawnienie obsługi urządzeń i udostępnianie użytkownikom zasobów obliczeniowych. Jednocześnie, przy coraz szerszym upowszechnieniu chmury publicznej, oczekują oni możliwości szybkiego dostępu do samoobsługowych zasobów obliczeniowych, zamiast zgłaszać do nas zapotrzebowanie na nie.

Jak ważne są funkcje OpenStack i jakie jest znaczenie interfejsu API służącego do łączenia się z tą platformą?

Nasi użytkownicy mają zróżnicowane doświadczenie w IT. Najbardziej zaawansowani są programistami znającymi system Linux. Lubią oni używać platformy OpenStack i interfejsów API Amazon EC2 oraz wierszy poleceń, aby korzystać z chmury. Jednak wielu użytkowników korzysta również ze standardowego interfejsu webowego OpenStack, by tworzyć maszyny wirtualne do testowania i wprowadzania usług IT.

Z jakich hiperwizorów korzysta CERN? Na waszej stronie przyznajecie się do stosowania wielu różnych technologii.

Większość hiperwizorów to KVM – Kernel-based Virtual Machine. Mamy jednak również w tej samej chmurze kilkaset maszyn wirtualnych, na których pracuje Hyper-V. Korzystanie z dwóch różnych hiperwizorów pozwala nam na kierowanie obciążenia do prawidłowej technologii i porównywanie ich wydajności tak, aby wykrywać możliwości optymalizacji. Maszyny wirtualne z Hyper-V służą do uruchamiania systemów Windows, a te z KVM – do innych systemów, takich jak Linux. OpenStack kieruje obciążenie do hiperwizora właściwego typu na podstawie obrazów VM w chmurze.

Której wersji OpenStack używacie? Która jest teraz najbardziej rozpowszechniona?

Dla większości składników używamy teraz wersji Juno. Trwają jednak testy aktualizacji Kilo, do której dokonamy migracji w ciągu kilku miesięcy.

Jakie narzędzia służą do wdrażania, aktualizowania i monitorowania tak dużej instalacji OpenStack?

Oparliśmy się na dystrybucji RDO (<https://www.rdoproject.org>), którą wdrażamy za pomocą narzędzia do zarządzania konfiguracją o nazwie Puppet. Jeden z naszych zespołów odpowiedzialnych za infrastrukturę OpenStack utrzymuje zbiór szablonów konfiguracji Puppet, dzięki czemu jej wdrażanie i utrzymywanie jest łatwiejsze. Używamy także innych narzędzi open source, jak np. Elastic Search i Kibana do monitorowania oraz Rundeck i Jenkins do automatyzacji.

Tak duże środowisko i ilości danych są związane z problemem skali. Jakie są największe wyzwania, wobec których stoicie?

Współpracowaliśmy z grupą roboczą ds. dużych wdrożeń w OpenStack, aby zidentyfikować potencjalne zatory w tworzonej przez nas infrastrukturze. Mamy także kontakt z zespołami programistycznymi OpenStack, by wdrażać i sprawdzać tworzone przez nas optymalizacje. Wiemy jednak, że witryny takie jak eBay, Yahoo i Rackspace mają znacznie większe instalacje OpenStack niż nasza chmura, a to oznacza, że korzystamy również z ich doświadczeń.

Interesujące nas zmiany związane ze skalą naszej instalacji OpenStack – takie jak tworzenie i zapisywanie w określonej puli połączeń do baz danych czy wykorzystanie w systemie uwierzytelniania rozwiązania do buforowania pamięci podręcznej ogólnego przeznaczenia

Memcached – wdrożono już w poprzednich wydaniach. Memcached stworzony został przez Danga Interactive dla serwisu LiveJournal. Jest udostępniany teraz na licencji BSD i często używany do przyspieszenia dynamiki baz danych, poprzez buforowanie danych oraz obiektów w pamięci RAM. w celu zredukowania czasu odczytu zewnętrznego źródła danych.

Jak w CERN jest zrealizowany system pamięci masowych? Wiem, że korzystacie z CEPH. Jakie są zalety i wady tej technologii?

CEPH służy do tworzenia blokowej pamięci masowej w naszej chmurze prywatnej. Dzięki temu użytkownik może uzyskiwać dodatkowe miejsce na dysku, odpowiednio do swoich potrzeb i w odpowiednim dla siebie czasie. Rozwiązanie to działa bez problemu od 18 miesięcy. Z kolei ogromne ilości danych z eksperymentów przeprowadzanych w czasie działania Wielkiego Zderzacza Hadronów przechowywane są w wyspecjalizowanych pamięciach masowych, takich jak opracowane u nas rozwiązania dedykowane – CASTOR i EOS. W tym bowiem przypadku potrzebny jest dostęp do plików, a nie do bloków pamięci. CASTOR – czyli Cern Advanced STORage manager – to hierarchiczna baza danych, bazująca na pamięciach taśmowych o pojemności ponad 100 PB. EOS zapewnia znacznie szybszy dostęp do danych, wykorzystując dyski twarde. Natomiast pamięć Flash jest używana do zapisywania plików dziennika CEPH i buforów pamięci podręcznej w warstwach baz danych.

Jaka technologia służy do dystrybuowania ruchu sieciowego w CERN?

CERN wykorzystuje Ethernet do łączności z większością serwerów obliczeniowych i ogólnosiwiatową sieć TCP/IP, którą stosujemy do łączności z innymi laboratoriami sieci obliczeniowej LHC. Aplikacje analizujące wyniki doświadczeń z zakresu fizyki wysokich energii nie potrzebują bowiem sieci o niskich opóźnieniach, takich jak InfiniBand.

Mówi się o 1 PB informacji powstającym co sekundę w czasie działania Wielkiego Zderzacza Hadronów. W jaki sposób zapewniana jest wysoka dostępność, kopie zapasowe i przywracanie działania po awarii? Czy możliwe jest wykonanie kopii tak dużej ilości danych?

Dane eksperymentalne kopiowane są do różnych lokalizacji na świecie. Każda z nich ma również możliwość przechowywania dużych zbiorów danych. Gwarantuje to istnienie co najmniej jednej kopii danych poza CERN. Chociaż eksperymenty wytwarzają 1 PB/s danych, są one zredukowane przez odfiltrowywanie do kilku GB/s. Dzieje się tak poprzez wybór i przekazanie najbardziej interesujących zdarzeń do innych lokalizacji. Mamy również drugie centrum danych w Budapeszcie, obsługujące niektóre usługi i umożliwiające ich przywrócenie w razie awarii w centrum danych CERN w Genewie. Chmura CERN została uruchomiona

produkcyjnie w lipcu 2013 r. Obecnie wykorzystuje 90% mocy obliczeniowej w naszych centrach danych w Genewie i Budapeszcie.

Czy OpenStack jest idealnym rozwiązaniem? Czego w nim najbardziej brakuje?

Prowadzone są intensywne prace rozwojowe nad OpenStack, a w każdym wydaniu pojawia się wiele ulepszeń. Jedno z najważniejszych wyzwań jest właśnie z tym związane. Dotyczy ono nadążania za regularnymi aktualizacjami, które pojawiają się co sześć miesięcy. Pomagają w tym narzędzia, takie jak Puppet. Z kolei wymiana doświadczeń na konferencjach OpenStack umożliwia nam przygotowanie się na potencjalne problemy z funkcjonowaniem naszej infrastruktury. Są też nowe, interesujące dla naszej społeczności projekty rozwijające ekosystem OpenStack, m.in. Docker i Kubernetes tworzony w ramach projektu Magnum.

Jakie są główne problemy przy korzystaniu z OpenStack w tak dużym środowisku?

Modularna architektura OpenStack sprawia, że zarządzanie tak dużą – jak nasza – chmurą prywatną jest dużo łatwiejsze dzięki możliwości zarządzania mniejszymi jednostkami, a następnie łączenia ich w większe struktury. Ta funkcjonalność OpenStack jest coraz powszechniej wykorzystywana przez takie firmy, jak Rackspace i GoDaddy. W celu dalszego ulepszania funkcjonalności pojedynczych modułów współpracujemy także z takimi organizacjami, jak centrum BARC w Bombaju.

CERN aktywnie uczestniczy w rozwoju projektów open source. Czy udział w tej społeczności umożliwi tworzenie nowych funkcji albo realizację celów biznesowych?

To prawda, mamy długą tradycję współpracy przy projektach open source i wnoszenia do nich rozwiązań powstałych w CERN. W wielu obszarach – takich jak zintegrowane zarządzanie dostępem do zasobów czy, wspomniana już, rozbudowa funkcjonalności poszczególnych modułów, tzw. cells – dokonaliśmy wielu ulepszeń. Współpraca ze społecznością open source umożliwia nam jednak korzystanie z pracy innych osób i organizacji, gdy potrzebujemy dodatkowych funkcji, np. wsparcia uwierzytelniania użytkowników w jednej chmurze i umożliwiania im dostępu do zasobów w innej chmurze.

Czy taki model rozwoju oprogramowania jest odpowiedni do tworzenia profesjonalnych rozwiązań dla systemów o krytycznym znaczeniu? Jakie jest ryzyko związane z krytycznością systemu zbudowanego na technologii otwartej i wsparcie techniczne otrzymywane od tej społeczności?

Istnieje wiele modeli korzystania z OpenStack. Od korzystania z dystrybucji takiej jak RDO, przez chmurę prywatną jako usługę, aż po chmurę publiczną. Firmy mogą wybierać to, co im najbardziej odpowiada. Komercyjna pomoc techniczna jest przydatną opcją do rozważenia, a otwarty charakter OpenStack zapewnia różne możliwości w tym zakresie, włącznie ze szkoleniem dostępnym na witrynie OpenStack Marketplace.

OpenStack jest używany z aplikacjami tzw. trzeciej generacji opartymi na architekturze mikroserwisów. Czy aplikacje tego typu są używane w CERN?

Zapowiadane wsparcie OpenStack dla Dockera jest interesujące i obecnie oceniamy je w ramach unijnego projektu Indigo-DataCloud poświęconego stworzeniu tzw. chmury danych.

Czy udostępnianie infrastruktury OpenStack naukowcom poza instytutem? A jeśli tak, to jak to robicie?

Około 11 000 fizyków pracujących nad projektami w CERN może się logować do usług w naszej chmurze prywatnej. Wdrożyliśmy zintegrowany system zarządzania dostępem do zasobów EduGain, aby umożliwić użytkownikom – zalogowanym w domach czy macierzystych instytutach – uzyskanie dostępu do naszej chmury w CERN. Niektóre funkcje, takie jak EduGain, zostały opracowane we współpracy z Rackspace w ramach OpenLab, który służy do współpracy pomiędzy CERN i branżą IT.

Jak w ogólnym zarysie przedstawia się infrastruktura IT w CERN? Czy wasze centrum danych ma jakieś szczególne cechy, które różnią je od typowego centrum danych? Gdy pracuje LHC, musicie np. radzić sobie ze skokowym wzrostem obciążenia, a potem nic się nie dzieje do następnego eksperymentu...

Obecnie wykorzystujemy ponad 4000 maszyn wirtualnych, a nasza fizyczna infrastruktura korzysta z ponad 120 000 rdzeni procesorów. Jednak nasze podstawowe centrum danych, Meyrin, zbudowano w 1970 roku XX wieku, dlatego obsługa wymagań dotyczących zasilania i chłodzenia dla współczesnych usług stanowi dla nas duże wyzwanie. Zwiększyliśmy nasze możliwości obliczeniowe dzięki uruchomieniu drugiego ośrodka w Budapeszcie.

Czy w CERN macie jakieś szczególne wyzwania związane z bezpieczeństwem?

CERN jest dużym laboratorium, z którego codziennie korzysta wiele tysięcy fizyków. W celu szybkiego identyfikowania wszelkich problemów z bezpieczeństwem współpracujemy ze społecznościami dostarczającymi rozwiązania open source, takie jak OpenStack. Natomiast za pomocą takich narzędzi jak Puppet możemy w razie potrzeby szybko wprowadzać nowe wersje systemu Linux i innych aplikacji, z których korzystamy. Regularnie spotykamy się z