the 451 group

451 Market Insight Service | User Deployment Report

CERN virtualizes scheduling to launch a private cloud for rocket scientists

Analyst: William Fellows

February 24, 2010

The **European Organization for Nuclear Research** (CERN) is extending its physical and virtualized server into a private cloud in order to address its biggest challenge – scheduling the growing number of job requests across its datacenters.

Early Adopter Snapshot

Key areas of innovation

CERN is virtualizing its job scheduler environment to support the volume of requests for compute and data processing across its systems as Large Hadron Collider experiments multiply. This is a private cloud for scientists – at scale.

Challenges to adoption

Local disk I/O – virtualizing to hard disk – is the key performance bottleneck CERN faces. Supporting the kind of data-intensive requirements it has (size, I/O) across an on-demand infrastructure makes the use of public cloud a nonstarter for it now. The data management problem hasn't been readily solved on its private cloud yet either. It reports that virtualization vendor lock-in hurts.

The 451 take

Rocket-science-scale processing will always bring a unique set of challenges, but in virtualizing its scheduling function with Platform ISF, CERN is taking a similar path to some of the largest investment banks and life science companies. All are seeking to make better use of internal resources to accommodate demand by using dynamic scheduling and virtualization. Operational and consumption models deliver this as a private cloud.

Physical

The Fabric Infrastructure and Operations group at CERN has 4,000 machines running batch tasks supporting the Large Hadron Collider (LHC). Some 10,000 physicists around the world submit jobs to process snapshots of LHC data using CERN's servers and storage systems. CERN's challenge is to manage and schedule those jobs as the volume of data collected and the number of experiments performed on it increase. As well as LHC data, CERN is also responsible for producing, collecting and analyzing data from other projects such as Compass and Opera. Most of its servers are 1U and 2U racks with two motherboards. Each motherboard has two CPUs with four cores – some older systems have two cores. It has 30,000-35,000 cores in total, and could theoretically run the same number of jobs, since its workloads are single-core and single-threaded. It doesn't use traditional HPC mechanisms such as message passing interface or parallelism, and the resources are not configured as a single pool.

Company name
CERN - European Organization for Nuclear Research

Activities
Finding out what the universe is made of and how it works

Head office
Geneva

Project
Private cloud

Key suppliers
Platform Computing, OpenNebula, Xen, KVM

Although a longtime user of **Platform Computing** LSF since the 1990s, CERN didn't really touch grid, even if that's how it has been characterized. Storage is different. Storage is treated as a single whole active cache (pool) so that data doesn't have to be moved onto the individual nodes themselves. Data can be backed up to tape and restored to the cache as needed in the batch mode. CERN has 50PB disk and 60PB tape. While **IBM** and **Hewlett-Packard** are associated with CERN in various capacities (and **StorageTek** for storage), the server estate is tendered from midrange suppliers.

Email, Web, admin, database and other supporting IT infrastructure and activities are handled by other groups. CERN has some 200 staff in IT, 50 of whom work on the infrastructure to support its physicists.

Virtual

CERN already has 50,000-60,000 VMs (Xen and KVM) deployed in the batch environment that connect to LSF – but going forward, it wants to be able to connect to any scheduler. For this reason, it is starting to use the open source **OpenNebula** meta-scheduler.

CERN is now also using Platform's ISF adaptive cluster management tools alongside LSF. Between them, ISF and LSF manage both physical and virtual machines (short-lived VMs for batch processing and longer-life VMs for more traditional server loads). After considering a number of suppliers, it chose ISF because of, it says, its superior policy-driven capability. In effect, CERN is virtualizing its scheduling function and using ISF to apply and manage policy across resources.

Cloud

Scheduling through its batch cluster is CERN's key challenge as more users submit more jobs to use more data on its systems from around the world. It becomes more difficult, and therefore harder to

schedule centrally, when science experiments generate their own momentum and, therefore, additional processing needs. Binaries generated in one place may not work in another. It requires greater agility for this. It also needs to be able to create policies that can determine how much compute capacity could be allocated within LHC to different projects, such as Higgs Boson or Supersymmetry, on the fly, according to different priorities. For this reason, the private cloud model, where VMs can be generated as requested and on-demand, is the next step in its journey. CERN may also use ISF to separate analytics and simulation tasks that have only small data outputs (or tasks with small data inputs) from other CPU cluster work to get better utility and run them in a cloud.

Challenges and opportunities

CERN estimates that CPU overhead for using virtualization is now less than 5%, an acceptable overhead, and it now pushes as much work as it can through the VMs. While there is little impact on network I/O, the key challenge is accessing data on local disks as I/O is distributed around the network. The problem with virtualization as a whole is that vendors tend to tie organizations into using other virtualization elements in their stack – running and managing heterogeneous environments is tough, in its experience.

A key challenge will be the ability to support intensive data requirements in the cloud. Creating additional capacity on-demand in a cloud, rather than using hard-wired resources, will require new mechanisms to get data in and out of the cloud to support this demand. Moreover, it's this data management challenge that would be the inhibitor to using external clouds. The problem, as far as CERN is concerned, is that third-party clouds are, for the most part, opaque – you don't know where the resources are or what they are.