

Cloud Computing bei DESY

- Wer wir sind?
- Wieso wollen wir CloudComputing machen?
- Was sind unsere größten Herausforderungen?

Yves Kemp, DESY IT
HGF, 23.9.2020

Wer wir sind?

- DESY:
 - Beschleuniger Labor
 - Hochenergie-Physik (LHC, Belle, ILC, ...)
 - Photon Science (Petra-III, starke Links zu European-XFEL)
 - Betrieb, Bau, Entwicklung von neuen Beschleunigern und Beschleunigertechnologien (Petra-IV, Plasma-Wakefield,...)
 - Astro-Teilchen-Physik
 - Allgemein Theorie
- DESY IT Abteilung:
 - Ca 100 Mitarbeiter
 - Verantwortlich für zentrale Dienste auf dem Campus
 - Bereitstellung von Infrastrukturen für wissenschaftliche Datenhaltung und Datenanalyse
 - Grosse Grid, HTC und HPC Cluster, ~100 PB Daten, DESY, Gäste, nationale und internationale Nutzung
 - Wissenschaftliche Beratung und Beteiligung an diversen D- und EU-Projekten

Wieso Cloud Computing?

Neue Technologie um wissenschaftliche & Enterprise Workflows zu vereinfachen

- Wissenschaftliche Use-Cases
 - Nutzer wollen unabhängig sein vom lokalen Setup der Batch Knoten □ dynamische Virtuelle Maschinen / Container
 - Alternativen zu Grid & Batch Workflows für wissenschaftliche Payload, sowie Shift von Executable □ Container/VM
 - Projekte erfordern ComputeCloud (=OpenStack) als Voraussetzung
- Enterprise IT Use-Cases
 - Dynamische Skalierung von Diensten nach Last, modernes Deployment von Diensten.
 - Effizienzsteigerung im Anbieten von Diensten und Ausnutzung von Ressourcen
 - Kiosk System für IT-Nutzer / DESY-Nutzer
 - Schnellere Bereitstellung von Betriebssystemen für neue Dienste
 - Klarere Trennung zwischen Betrieb der Hardware, Bereitstellung und Betreuung Betriebssystem, Bereitstellung und Betreuung der Applikation.
 - In-house Alternative zu kommerziellen Cloud-Compute Anbietern
- Bedarf der Nutzer
 - Wenige Nutzer (ausserhalb IT) artikulieren die Notwendigkeit von Cloud Computing in-house

Umfang des Projektes

- Ganz IT muss involviert sein
 - Keine partiellen Schnellschüsse (zB Grid-Rechencluster auf OpenStack hosten)
- Tiefe Integration in IT Infrastruktur
 - Configurationsmanagement
 - Usermanagement
 - DESY Netzwerke und Management (IP, DNS, ...)
 - Security Policies
 - IT-Admins: Compute Cloud muss als normale, produktive Infrastruktur für Dienste gesehen werden
 - DESY-Nutzer: Die Infrastruktur, auf der Dienste laufen, soll egal sein.
- Manpower:
 - Keine (wenig) dedizierte Manpower
 - Zusammenarbeit von vielen Gruppen: Systeme/Virtualisierung, Netzwerk, Betrieb, Scientific Computing-Gruppen, Information Fabrics/Storage, ...

Status aktuell ... die groben Züge

- Viel Erfahrung gesammelt
- Teile der Infrastruktur werden für “early production” benutzt, insb. als Basis für Dienste für Projekte
 - → Vorträge Markus & Stefan, Kars, & Melvin
- Erste IT-Dienste in der Pilotphase auf OpenStack
 - Interesse für weitere Dienste vorhanden
 - → Vortrag Michael & Johannes
- Noch einiges an Weg vor uns
 - → Vortrag Yves
- Aktuell keine Policy in Planung à la “Alles muss auf OpenStack migriert werden” (CERN Beispiel)
 - Wir sehen Koexistenz von “klassischer” IT und OpenStack basierender IT
 - ... darüber sprechen wir in 5 Jahren aber nochmal

Cloud, Container, Orchestrierung

- In der Anfangsphase war die Richtung ganz klar auf Compute Cloud und *virtuelle Maschinen* ausgerichtet
- Mittlerweile behandeln wir im Projekt-Team auch Container, und auf Container basierende Infrastrukturen.
- ZB: Kubernetes kann, muss aber nicht auf OpenStack IaaS laufen
- Insofern: Den Blick nicht nur auf "Cloud Computing = VMs in OpenStack" verengen

**Und nun zu den
Detail-Vorträgen**

Herausforderungen & ToDos

- ... eine ungeordnete Sammlung
- ... und eher Fragen als Antworten
- ... unvollständige Liste

Henne-Ei Problem

- Dienste laufen doch klassisch. Wieso migrieren auf eine neue, im Aufbau befindliche Plattform?
- Es sind keine Nutzer da. Wieso sollen wir Geld und Arbeit in OpenStack stecken?
- Wenig Antrieb durch IT-externe Nutzer
- Verschärftes “Problem” am DESY:
 - Hardware-Beschaffung ist recht einfach und schnell
 - Gute XEN & Vmware Infrastruktur für Server-Konsolidierung.
 - → Wir sind schon recht flexibel und agil, Vorteile von IaaS werden nicht so stark gesehen
- Lösung:
 - Enthusiasmus aller Beteiligten
 - Starker Push durch nationale und EU-Projekte, die klar auch Cloud Computing setzen

Die Daten.

- DESY hat ~100 PB an wissenschaftlichen Daten
- Die zugehörigen Rechencluster greifen aus kontrollierten IP Netzen üblicherweise über UID/GID authentifiziert auf diese zu.
 - Hinreichend sicher, effizient, einfach für Nutzer und Admins
- Wir ermöglichen wir den sicherern, einfachen, effizienten Zugriff auf diese Daten:
 - Aus user-managed OpenStack VMs?
 - Aus Containern? Kubernetes?
- Ehrlich gesagt: Wir haben einige Ansätze, aber noch keine wirklich gute Lösung
- Ohne eine gute Lösung keine tiefe Nutzung durch Wissenschaftler
- Die moderne Welt von Cloud Computing und Container etc. ist nicht gemacht für zentrale, gemountete Netzwerk-Filesysteme

Security

- In einigen Bereichen generelle Vorbehalte gegen Virtualisierungstechnologien
 - Spectre & Meltdown waren nicht wirklich hilfreich in der Argumentation
- Mitigationsstrategien im XEN&Vmware Umfeld konzeptionell nicht gut umsetzbar in OpenStack
 - Man will eigentlich *eine* grosse Infrastruktur haben, nicht viele kleine OpenStack Inseln
- Netzwerk Security wird verlagert: Von separaten Routern & Firewalls □ SecurityGroups (=Firewall im Kernel der Hypervisoren)
 - Die (unbestreitbaren) Nachteile werden stärker gesehen als die Vorteile (die sich auch erst beweisen müssen)
 - Resultat: Komplizierte Netzwerk Setups (siehe Detail-Vorträge), die das Setup erschweren, und die Akzeptanz bei ersten Nutzern verringern
- Paradigmenwechsel
 - IP Adressen → Dienste
 - Subnetze → Software definierte Netze

Dynamisches Umfeld

- OpenStack selber ist ein sehr dynamisches Umfeld
 - Viele Unter-Projekte kommen und gehen
 - Was wird wirklich gebraucht?
 - Wir gehen wir mit der Dynamik um?
- Neue Konzepte kommen
 - OpenStack und dynamische VMs haben vielleicht erst den Weg für Container und Container Orchestrierung eröffnet
 - Braucht es noch OpenStack (bei DESY) im Zeitalter von Containern und Container Orchestrierung?
- Herausforderungen an IT: Wie gehen wir mit Dynamik um? Passen Neuerungen aus der Industrie auch für uns?
- Herausforderung an Nutzer: Auch wenn die Dienste einiges abstrahieren: Die Nutzer müssen sich wohl öfter auf Neues einstellen, und Altes, Liebgewonnenes verabschieden.
- Was sind die besten Strategien aus Sicht von IT und Nutzer?

Abrechnungsmodelle & Abrechnungen

- Wir haben einmal die Nutzer aus Projekten: Wir uns zu einer bestimmten Rechenleistung verpflichtet
 - Ob diese auch abgerufen wird ist nur bedingt in unserer Hand
- Dann haben wir DESY Nutzer
 - Hier existieren typischerweise noch keine Verrechnungsverfahren. Allgemein am DESY wenig verbreitet
 - Bisheriges Modell: “Eine Gruppe kauft Hardware auf ihre Kostenstelle, IT betreibt diese im Rechenzentrum”
- Und wir haben externe Nutzer (Uni-HH, XFEL, ...)
 - Hier gibt es teilweise schon Verträge über Nutzung
- Wenn Compute Cloud als Infrastruktur breit benutzt wird, müssen für alle Nutzer Abrechnungsmodelle her
 - Kosten vorher festlegen: Bekannt bei der Spezifikation der Ressourcen
 - Genehmigungsprozesse beim Start der Systeme? Kostenfreie, eingeschränkte Minimalsysteme?
 - Genaue Abrechnung des Verbrauchs
 - In-Rechnung-Stellen des Verbrauchs
- Nicht nur technische Herausforderungen