

Moritz Hartkopf

Proseminar Speicher- und Dateisysteme: Scientific Clouds

Agenda

1. Beispiele Scientific Clouds
2. Anforderungen an Scientific Clouds
3. Cloud Computing allgemein
4. Daten - Auffinden und Interpretieren
5. Vorteile und Risiken
6. Resümee
7. Ausblick

Max-Planck-Gesellschaft

The image shows the exterior of the Max-Planck-Gesellschaft building, a large modern glass structure with a central courtyard. The building is composed of several interconnected glass wings, creating a transparent, multi-story facade. In the center, there is a small green courtyard with a few trees. To the right, there is an outdoor seating area with white tables and chairs, and several closed white umbrellas. The sky is overcast.

Mitarbeiter: 22000

Etat/Jahr: 1.8 Milliarden €

Max-Planck-Gesellschaft



MAX-PLANCK-GESELLSCHAFT



Max-Planck-Institut für
 Neurobiologie und Biochemie

[2]


 GWDG Compute Cloud
 [Servers](#)
[Create](#)
[Connect](#)
[Monitor](#)
[Account](#)
[Logout](#)

Your servers

IPs	vCPUs	RAM	Image	Name	State	C	Actions
10.1.0.2 141.5.99.1	1	2048	OpenSuse 12.2 KDE Desktop x64	DV Treffen	ACTIVE		
10.1.0.8	1	2048	Ubuntu.12.04.LTS.Desktop.64bit	Ubuntu 12.04 Desktop 64bit	ACTIVE		
10.1.0.7	1	2048	OpenSuse 12.2 KDE Desktop x64	OpenSuse 12.2 KDE Desktop 64bit	ACTIVE		

IPs

IP	Assigned to	C	Actions
141.5.99.1	DV Treffen		
141.5.99.2			
141.5.99.3			
141.5.99.5			

Security Groups

Name	Actions
default	

GWDG Cloud

[3]



DESY

Mitarbeiter: 2300

Etat/Jahr: 230 Millionen €



DESY CLOUD



Die DESYcloud ist ein Online-Speicherdienst, um einfach und schnell an eigene Dateien zu gelangen. Der Dienst befindet sich momentan in der Pilotphase.

Bitte melden Sie sich bei Interesse bei weboffice@desy.de.



DESY Birdseye

[4]



Zugriff auf Ihre Daten »

Speichern Sie Ihre Dateien in der DESYcloud und greifen Sie bequem von überall via Smartphone, Desktop Client oder Browser zu.



Synchronisieren Sie Dateien »

Halten Sie alle Ihre Daten auf all Ihren Geräten synchron, um reibungslos zu arbeiten.

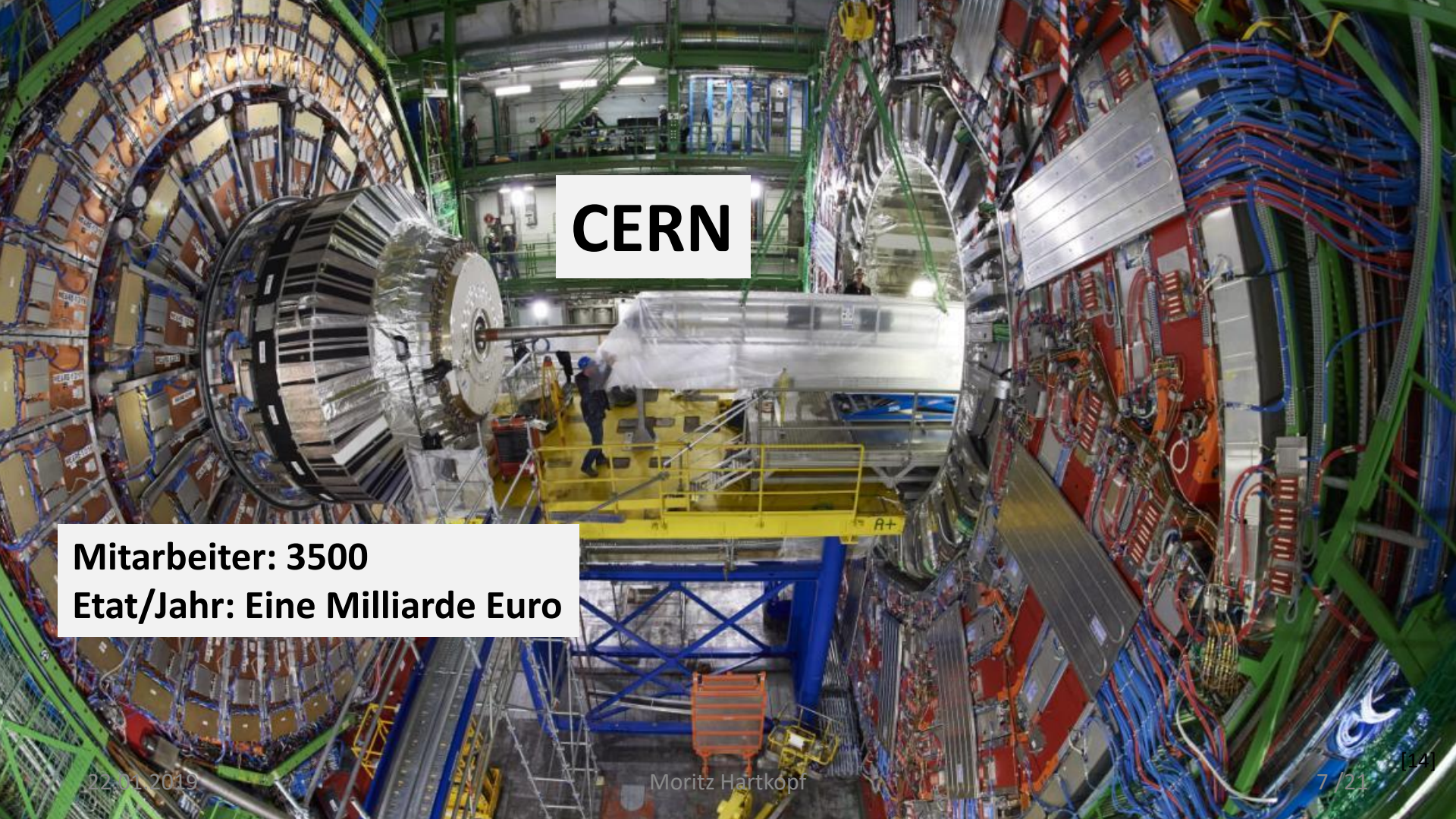


Teilen Sie mit Ihren Kollegen »

Teilen Sie Ergebnisse mit Ihren Kollegen. Dank der Speicherung in der DESY Infrastruktur sind unternehmensbezogene Daten kein Problem.

DESY CLOUD

[5]

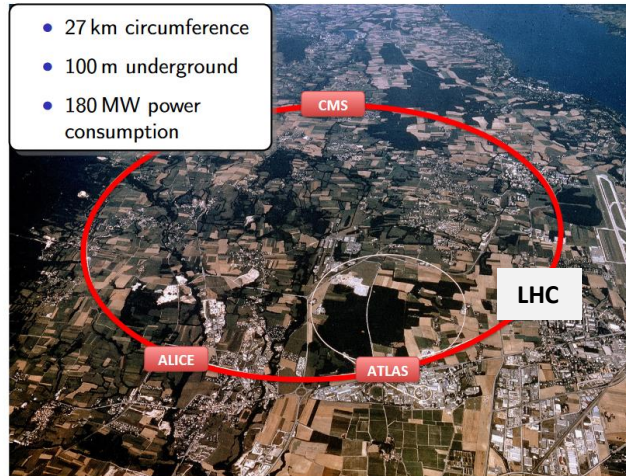


CERN

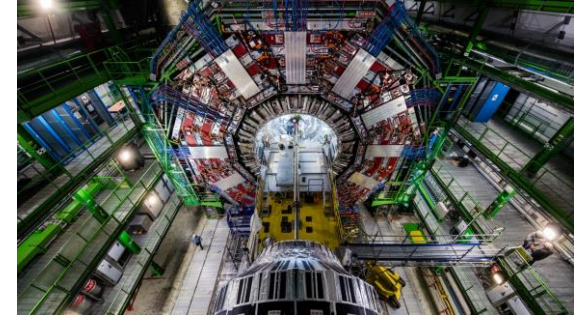
Mitarbeiter: 3500
Etat/Jahr: Eine Milliarde Euro



CERN Data Centre [6]

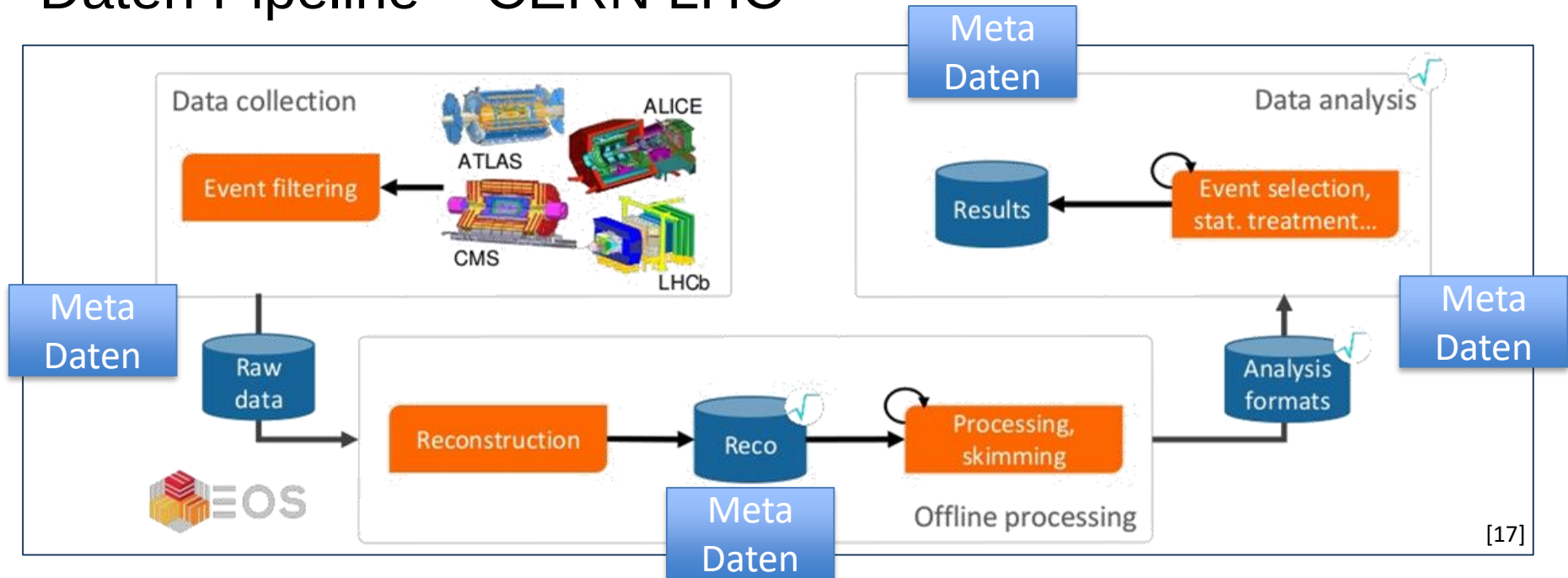


Birdseye LHC [7]



Large Hadron Collider [8]

Daten Pipeline – CERN LHC



[17]

Anforderungen an Scientific Clouds

- Langzeitspeicherung
- Finanzierung, nicht kommerziell
- Datenmengen
- Wenig externe Abhängigkeiten
- Alte Versionen müssen verfügbar bleiben
- Häufige Updates
- Schnelles Deployment

Cloud Technologie in der Wissenschaft

- Teilen von Wissen und gemeinsames Arbeiten an Projekten von unterschiedlichen Orten aus (ortsunabhängige Verfügbarkeit)
- Speicherung und Verarbeitung großer Datenmengen

- In der Infrastruktur ähnlich wie Commercial Clouds, aber unterschiedlich in der Anwendung
→ für wissenschaftliche Berechnungen und Simulation

Cloud Anwendungsgebiete:

SaaS

- Oberste Schicht
- Anwendungssoftware
- Zugriff meist über Internet oder Web Browser

- Drop Box
- DESY Cloud

PaaS

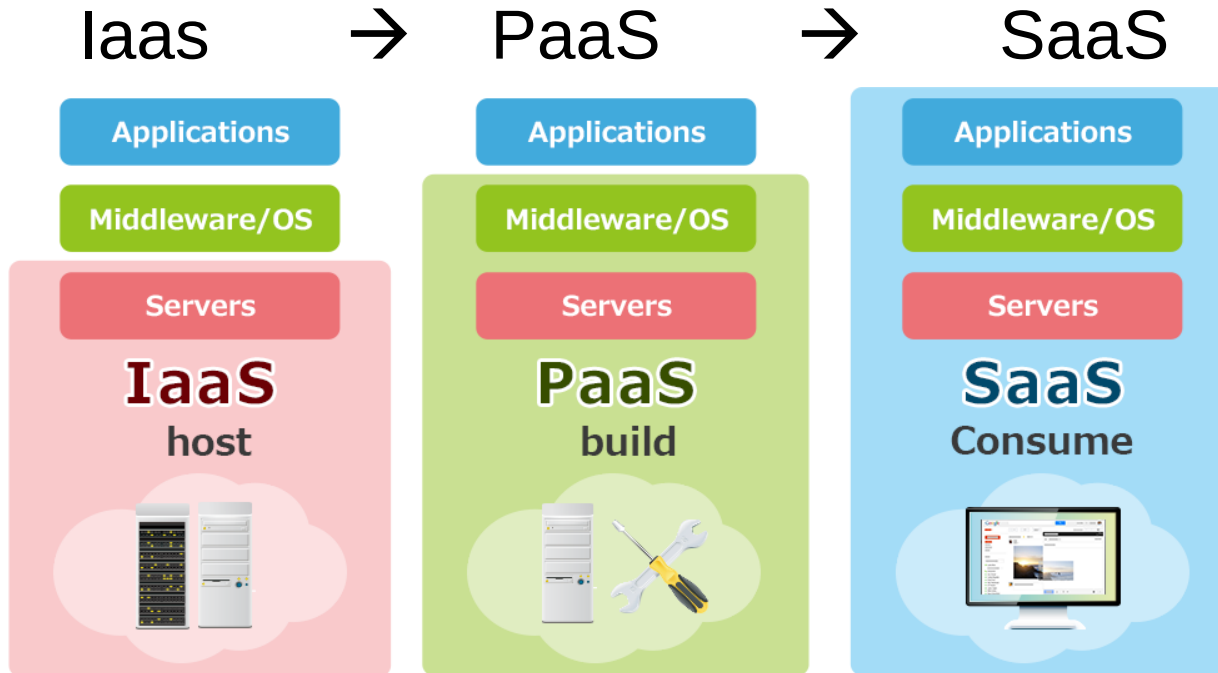
- Verbindet IaaS und SaaS
- Entwicklungsumgebungen, Datenbank Verwaltungssysteme und Web Server

- Linux
- MySQL

IaaS

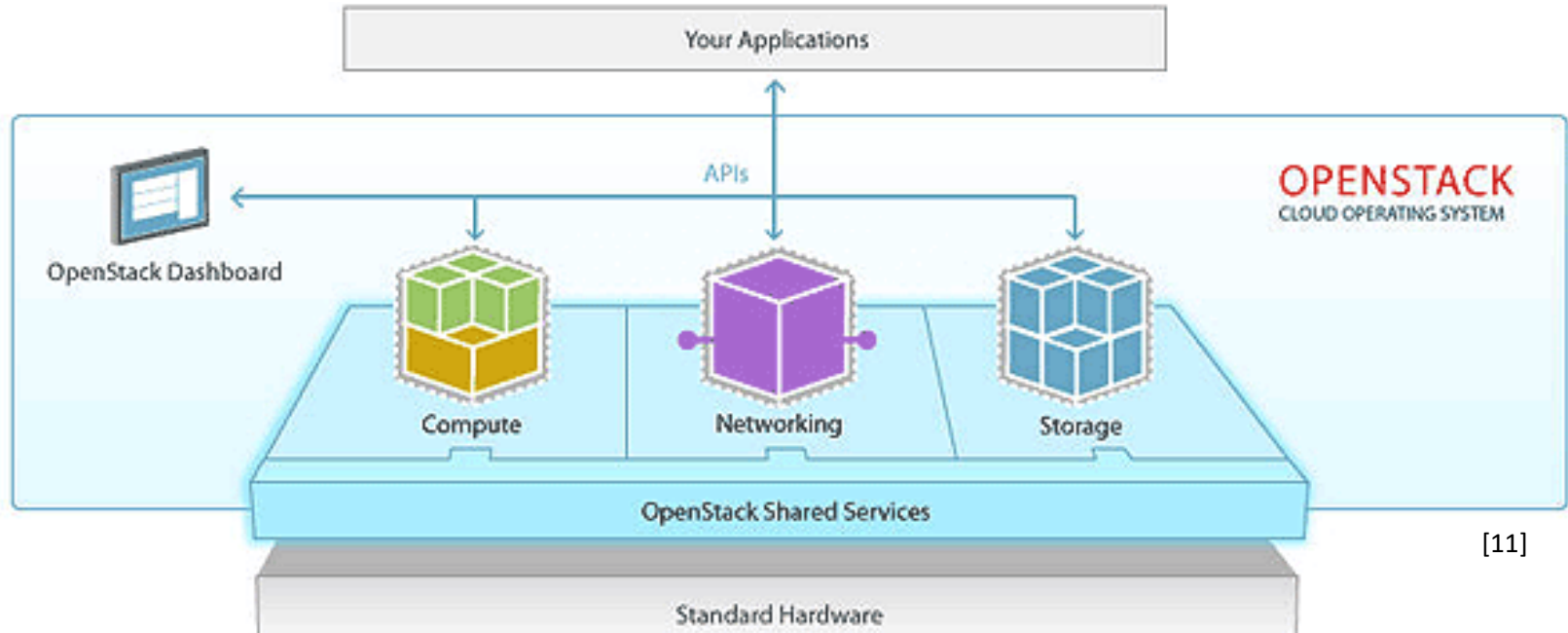
- Unterste Ebene des Cloud Computing
- Grundlegende Ressourcen: CPU, Storage, Netzwerk
- Zugang zu virtueller Hardware

- AWS S3
Storage oder
EC2 CPU



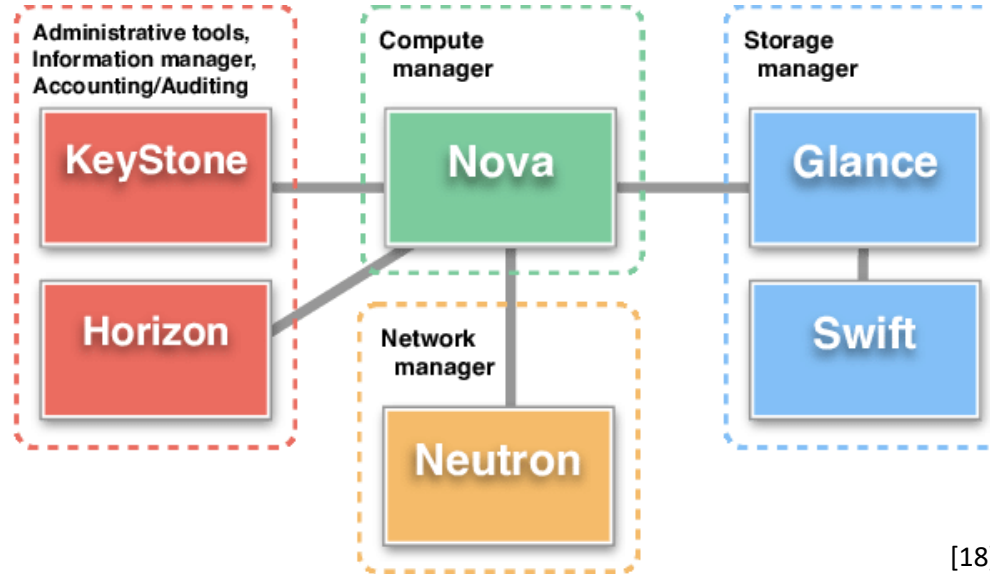
[9]

Open Stack



[11]

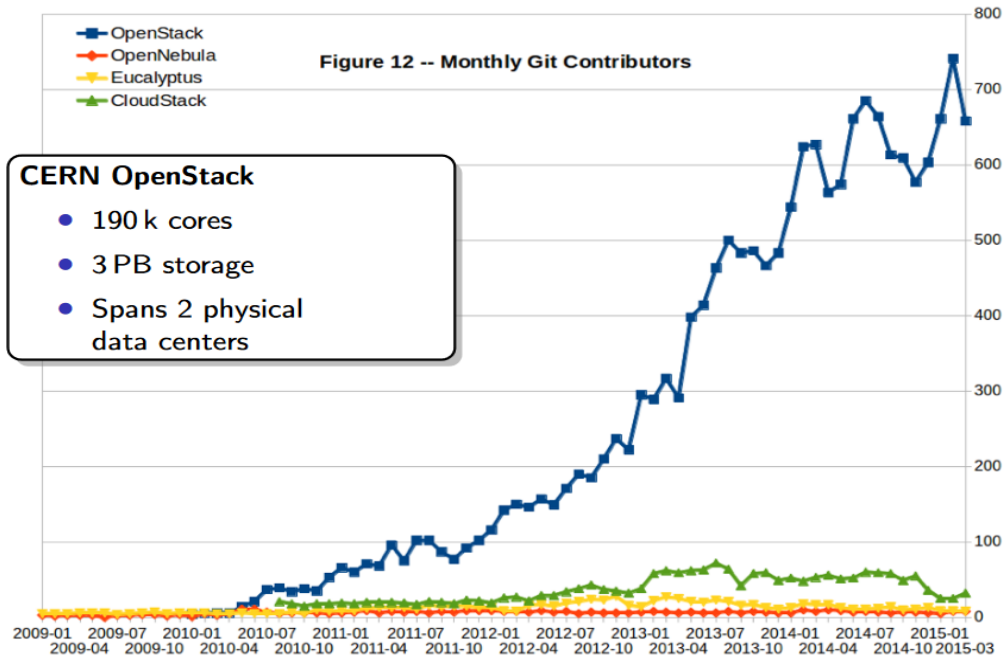
Open Stack – Kernservices



[18]

Open Source Cloud Software

- Open Stack CERN wächst
- Git Contributors
> 700/Monat
- Verteilte Umgebung



[1]

CERN Cloud Controller

Flexible Integration von Cloud Stacks

- Commercial + Open Source
- Experimentelle Lösungen möglich
- Keine Sackgasse/Einbahnstraße
- Offen für viele Anforderungen

The success of CernVM is largely based on the fact that it runs in practically **all** cloud environments.

Hypervisor / Cloud Controller	Status
VirtualBox	✓
VMware	✓
KVM	✓
Xen	✓
Microsoft Hyper-V	✓
Vagrant	✓
OpenStack	✓
OpenNebula	✓
CloudStack	✓
Amazon EC2	✓
Google Compute Engine	✓
Microsoft Azure	✓
Docker	✓ ¹

¹ Release Candidate

26 / [16]

Metadaten – Daten finden und verwenden in der Cloud

Metadaten – Hintergrund

- Standardisierung
- Auffindbarkeit
- Wiederverwendbarkeit
- Nachvollziehbarkeit

Metadaten – Beispiel DESY

Metadata common to all DESY experiments.

NeXus field or attribute path	ICAT field	Remark
/NXentry/experiment_identifier	Investigation	a unique beamtime ID
/NXentry/title	Investigation.Summary	description of the experiment
/NXentry/NXinstrument/name	Instrument.Fullname	name of the beamline
/NXentry/NXinstrument/name@short_name	Instrument.Name	short name of the beamline
/NXentry/NXinstrument/NXsource/name	Facility.Fullname	accelerator name
/NXentry/NXinstrument/NXsource/name@short_name	Facility.Name	short name for the accelerator
/NXentry/NXsample/name	Sample.Name	name of the sample
/NXentry/NXsample/chemical_formula	Sampletype. Molecularformula	the sample chemical formula
/NXentry/start_time	Instrument.Start_Date	start time of the scan
/NXentry/end_time	Instrument.End_Date	end time of the scan

[10]

Vor- und Nachteile



- Kosten ergeben sich je nach Verbrauch/Benutzung
- gemeinsame Projekte ohne Ortsgebundenheit
- mobiler Zugriff von überall
- Nachhaltigkeit



- Sicherheit
- Datenschutz
- keine garantierte Langzeitspeicherung bei Commercial Clouds

Resümee

- **Ähnliche** technologische **Infrastruktur** wie bei Commercial Clouds (Server, Storage, Netzwerk) und **Ziel** der Virtualisierung
- Aber **unterschiedliche Anforderungen** und **Anwendungsgebiete**: Langzeitspeicherung, extrem schnelle Datenverarbeitung und Speicherung

Ausblick

Aussichtsreiche Lösung im Umgang mit der wachsenden Datenflut, durch größerer Experimente und voranschreitende Technologie, z.B. leistungstärkere Teilchenbeschleuniger

→ **Hypothese:** Bald Standard bei großen Forschungsprojekten und in Großforschungseinrichtungen

Quellen

Textquellen:

1. <https://www.bitkom.org/Presse/Presseinformation/Nutzung-von-Cloud-Computing-in-Unternehmen-boomt.html#>
2. <https://cernvm.cern.ch/sites/cernvm.web.cern.ch/files/hep-cloud.pdf>
3. <https://home.cern/news/opinion/computing/end-computings-steam-age>
4. <https://home.cern/science/computing>
5. <https://home.cern/science/computing/volunteer-computing>
6. <https://home.cern/news/news/cern/european-science-cloud-open-day>
7. <https://home.cern/science/computing/data-centre>
8. <https://home.cern/science/computing/storage>
9. <http://eos.web.cern.ch/>
10. https://www.mpg.de/6957705/JB_2013

Quellen

Textquellen:

11. <https://www.youtube.com/watch?v=qIn0homyLm4>
12. <https://www.bmbf.de/de/auf-dem-weg-zur-european-open-science-cloud-5216.html>
13. <https://sureshemre.wordpress.com/2015/02/22/cern-lhc-large-hadron-collider-is-waking-up/>
14. http://www.desy.de/forschung/beschleuniger/index_ger.html
15. https://it.desy.de/dienste/speicherdienste/desycloud/index_ger.html
16. <https://www.golem.de/news/wissenschaft-10-000-cpus-und-1-petabyte-fuer-die-cloud-von-helix-nebula-1802-132802.html>
17. <https://www.cloudcomputing-insider.de/metadaten-sind-das-wertvollste-gut-in-einer-multicloud-welt-a-782375/>
18. <https://de.ryte.com/wiki/Metadaten>
19. https://de.wikipedia.org/wiki/Cloud_Computing#Cloud_Computing_in_der_Wirtschaft
20. <https://phys.org/news/2018-06-cern-major-reap-atom-smasher.html>

Quellen

Textquellen:

21. <https://en.wikipedia.org/wiki/OpenNebula>
22. <https://de.wikipedia.org/wiki/CERN>
23. http://www.desy.de/femto/sites2009/site_www-desy/content/e187923/e187955/e198864/femto_2015_1_dt_ger.pdf
24. http://www.desy.de/aktuelles/news_suche/index_ger.html?openDirectAnchor=1213&two_columns=1
25. <https://www.abendblatt.de/hamburg/article213327465/Roentgenlaser-700-Terabyte-Daten-fuer-340-Forscher.html>
26. <https://cloud.google.com/blog/topics/inside-google-cloud/subatomic-particles-and-big-data-google-joins-cern-openlab>
27. <https://www.youtube.com/watch?v=N4eT9Lfvuro>
28. <http://www.desy.de/~jkotan/nexusdesyintegration.pdf>
29. <https://www.slideshare.net/databricks/cerns-next-generation-data-analysis-platform-with-apache-spark-with-enric-tejedor>
30. <https://www.datenschutzbeauftragter-info.de/die-cloud-saas-paas-und-iaas-einfach-erklaert/>

Quellen

Textquellen:

31. <https://softwareengineering.stackexchange.com/questions/336763/is-cloud-storage-such-as-dropbox-saas-paas-iaas-or-something-else-entirely>
32. <https://www.computerwoche.de/a/was-sie-ueber-die-cloud-wissen-muessen,2504589,2>

Quellen

Bildquellen:

1. <https://cernvm.cern.ch/sites/cernvm.web.cern.ch/files/hep-cloud.pdf> S.22
2. https://fritsch-tschaids.de/de/assets/galleries/225/_thumbs/large_fritsch_tschaids_architekten_MR_01_Luftbild_601.jpg
3. <https://tchadcarriere.com/wp-content/uploads/2017/10/The-MPH-Center-for-Quantum-Optics-max-planck-gesellschaft.jpg>
4. http://innovatives-brandenburg.de/sites/default/files/styles/news_detailed/public/2015-09-29_luftbild_mit_beschleunigern_rs-0045_a.jpg?itok=GE9_HHoG
5. https://it.desy.de/dienste/speicherdienste/desycloud/index_ger.html
6. http://mediaarchive.cern.ch/MediaArchive/Photo/Public/2012/1207154/1207154_03/1207154_03-A5-at-72-dpi.jpg
7. <https://cernvm.cern.ch/sites/cernvm.web.cern.ch/files/hep-cloud.pdf> S.22
8. <https://www.scitecheuropa.eu/wp-content/uploads/2018/06/%C2%A9-2017-2018-CERN.jpg>
9. https://pbxl.co.jp/wordpress/wp-content/uploads/2015/05/cloud_en.png
10. <http://www.desy.de/~jkotan/nexusdesyintegration.pdf>

Quellen

Bildquellen:

11. <http://www.sdnstack.com/wp-content/uploads/2015/07/openstack-sm.png>
12. <https://www.doranth-post-architekten.de/wp-content/uploads/CIMG2395-1980x1213.jpg>
13. <http://www.haz.de/var/storage/images/haz/nachrichten/der-norden/uebersicht/elektriker-stirbt-auf-desy>
14. [_gelaende-in-hamburg/619388475-1-ger-DE/Elektriker-stirbt-auf-Desy-Gelaende_big_teaser_article.jpg](http://www.haz.de/var/storage/images/haz/nachrichten/der-norden/uebersicht/elektriker-stirbt-auf-desy-gelaende-in-hamburg/619388475-1-ger-DE/Elektriker-stirbt-auf-Desy-Gelaende_big_teaser_article.jpg)
15. <https://s1.ibtimes.com/sites/www.ibtimes.com/files/styles/full/public/2017/03/04/cern-lhc-cms.jpg>
16. <https://cernvm.cern.ch/sites/cernvm.web.cern.ch/files/hep-cloud.pdf> S.26
17. <https://image.slidesharecdn.com/01enrictejedor-181010193440/95/cerns-next-generation-data-analysis-platform-with-apache-spark-with-enric-tejedor-9-638.jpg?cb=1539200135>
18. https://www.researchgate.net/profile/Anthony_Simonet/publication/316903190/figure/fig2/AS:568953308626944@1512660447768/Core-Services-of-OpenStack.png