

Scientific Computing bei DESY

Yves Kemp, DESY IT
Hamburg, 19.04.2018

Was macht DESY IT im Bereich Scientific Computing?

- DESY IT entwickelt und betreibt große, zentrale Infrastrukturen für wissenschaftliches Rechnen, sowie Management und Analyse von großen wissenschaftlichen Datenmengen.
- DESY IT bietet wissenschaftliche (und ingenieur-wissenschaftliche) Software an: Bereitstellung von den Programmpaketen auf den Plattformen und Lizenzen.
- DESY IT berät Nutzergruppen bei ihren Scientific Computing Aufgaben.
- DESY IT entwickelt Software im Umfeld von Scientific Computing.



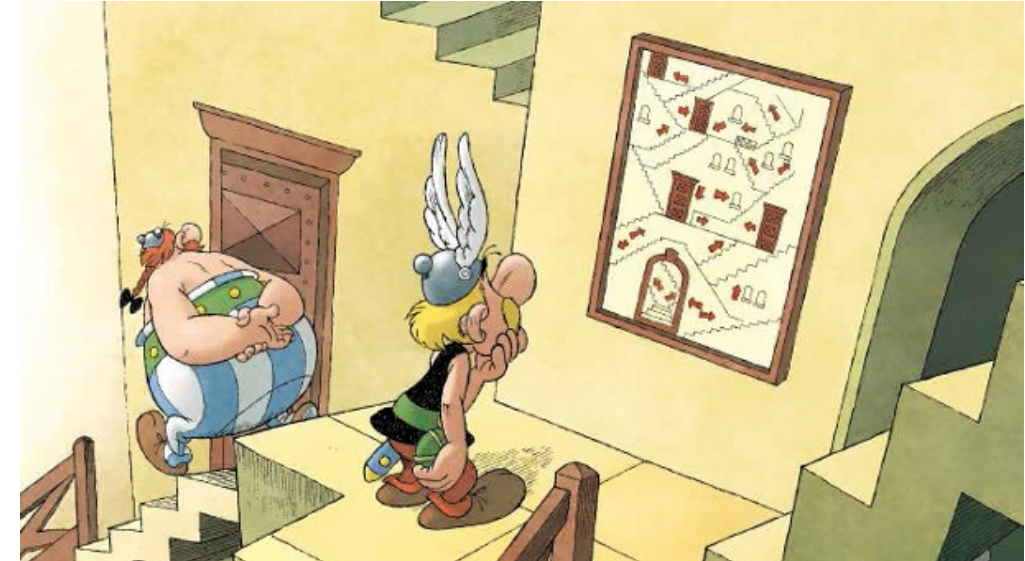
Wer macht das bei DESY IT? Das IT-Organigramm

- Fachgruppe „Scientific Computing“, geleitet von Birgit Lewendel
 - Betrieb der Grid Infrastruktur, dCache Entwicklung und Betrieb, Datennahme-Systeme von Petra-III und XFEL, Consulting, ...
- Fachgruppe „User Support“, geleitet von Martin Gloris
 - Bereitstellung der Software, und Support/Consulting
- Fachgruppe „Systeme“, geleitet von YK
 - Betrieb von NAF und Maxwell, Unterstützung beim Betrieb von anderen Systemen
- Fachgruppe „Information Fabrics“, geleitet von Peter van der Reest
 - Involviert u.a. in Collaborative Services, Datenbanken und Datamanagement (Tape)
- Fachgruppe „Network and Operations“, geleitet von Kars Ohrenberg
 - Bereitstellung und Ausbau der Datennetze für Campus- und WAN Datentransfers
- Leitung: Volker Gültzow



Hilfe: Ein Problem und 5 Ansprechspartner????

- Keine Angst: Wir reden miteinander ☺
- Manche Bereiche haben regelmässige Treffen mit IT
 - M-IT Meeting: M-Kontakt: Carsten Kluth
 - FS/Petra-III: Kontakt Thorsten Kracht, André Rothkirch
 - FH: Verschiedene Ansprechspartner, Stichwort NAF
 - ...
- Bei ganz konkreten Problemen natürlich die üblichen Support-Wege (UCO, andere Support-E-mails,...)
- Bei größeren Themen (evtl. strategisch/planischerischer Natur) gerne Kontakt mit der (Fach-)Gruppen-Leitung
- Bei Bedarf sprechen wir uns intern ab, und besprechen das jeweilige Anliegen in einer separaten Experten-Runde



Ein schneller Gang durch IT:
Welche Infrastrukturen?
Welche Zielsetzung?
Welche Software?

Rechenplattformen

Unterüberschrift, optional

Linux basierend

- Linux-Desktop: Skalierung problematisch, eher Login-Zugang zu zentralen Infrastrukturen
- PAL-Cluster: Für kleinere Aufgaben gut geeignet, Skalierung problematisch
- Große Cluster:
 - BIRD-Cluster
 - Grid-Cluster
 - HPC Cluster Maxwell

Windows basierend

- Wenig Unterstützung von Windows im Bereich des wissenschaftlichen Rechnens.
- Keine grossen Infrastrukturen, Terminal-Server wenn überhaupt



Compute Cluster: BIRD, Grid, Maxwell-HPC

BIRD & Grid:

- Mein Job läuft innerhalb eines Servers (1 oder wenige CPU-Cores)
- Netzwerk: „Moderate“ Bandbreite genügt
- Daten zB auf NetApp oder auf dCache
- Berechtigung:
 - BIRD: Batch-Ressource (über Registry oder Gruppen-Admin)
 - Grid: über VOMS (grosse Experimente)
- Ressourcen: Ca 500 Server BIRD und 500 Server Grid, insgesamt ca 20k CPU Core
- <http://bird.desy.de/htc/> und <http://grid.desy.de/>

Maxwell HPC

- Mein Job braucht viele Ressourcen (Memory, mehrere Server parallel, GPU, ...)
- Schnelles Netzwerk zur Inter-Prozess-Kommunikation oder zum Storage
- Daten zB auf GPFS (Petra-III, XFEL) oder dCache
- Berechtigung:
 - Email an maxwell.service@desy.de mit Kurzbeschreibung
- Ressourcen: Ca 400 Server, davon >50 mit GPU, ~24k CPU Core (HT)
- <https://confluence.desy.de/display/IS/Maxwell>

Storage Plattformen

Unterüberschrift, optional

dCache: DESY Eigenentwicklung

- Ca 20 PB am DESY gespeichert
- Tape-Backend möglich
- Verfügbar u.a. über NFS 4.1 Mount, über Ethernet
- Hohe aggregierte Bandbreite
- Möglichkeiten zum WAN Austausch

GPFS: Schnelle Datennahme und Analyse

- Ca 10 PB am DESY
- Spezielle Verbindung zur Datennahme
- Sehr schnellen Zugang nur über InfiniBand im Maxwell Cluster



Storage Plattformen _ 2

Unterüberschrift, optional

NetApp

- NFS und CIFS (teilweise gleichzeitig) möglich
- Gute Management Features

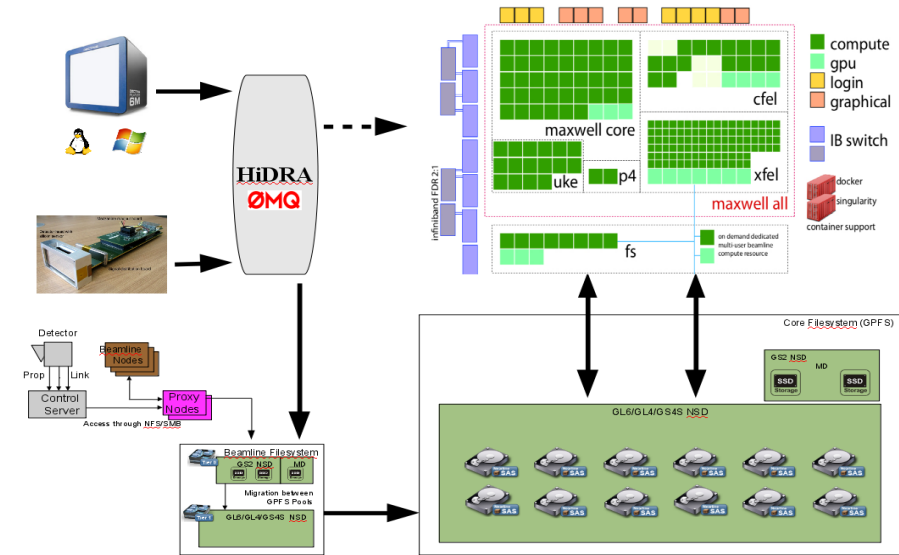
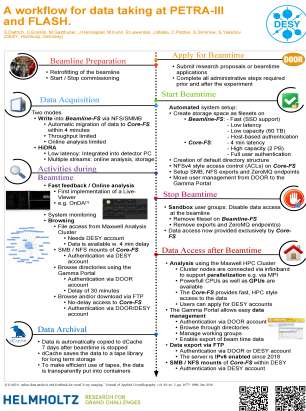
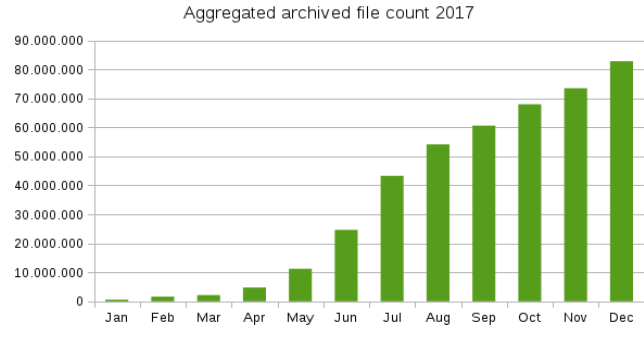
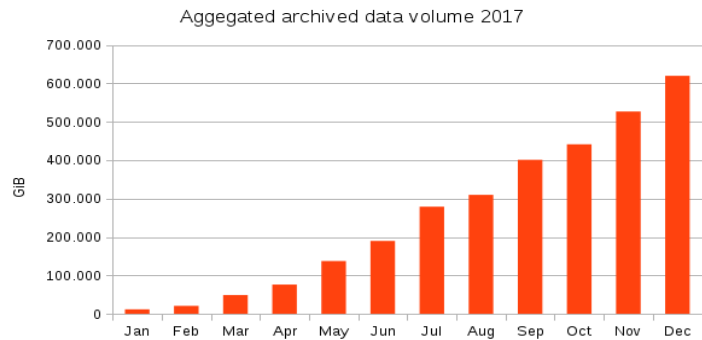
DESY Storage Cloud

- (noch) OwnCloud basierend, mit dCache als Backend
- Aktuell noch in der Pilotphase, größeres Upgrade nächste Woche



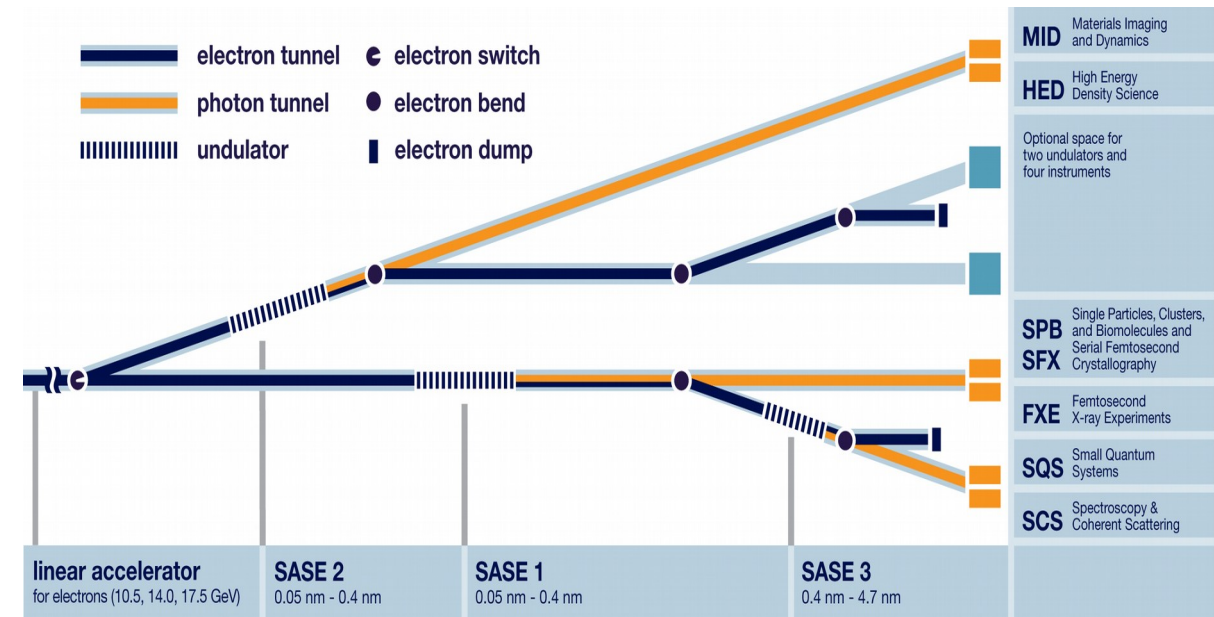
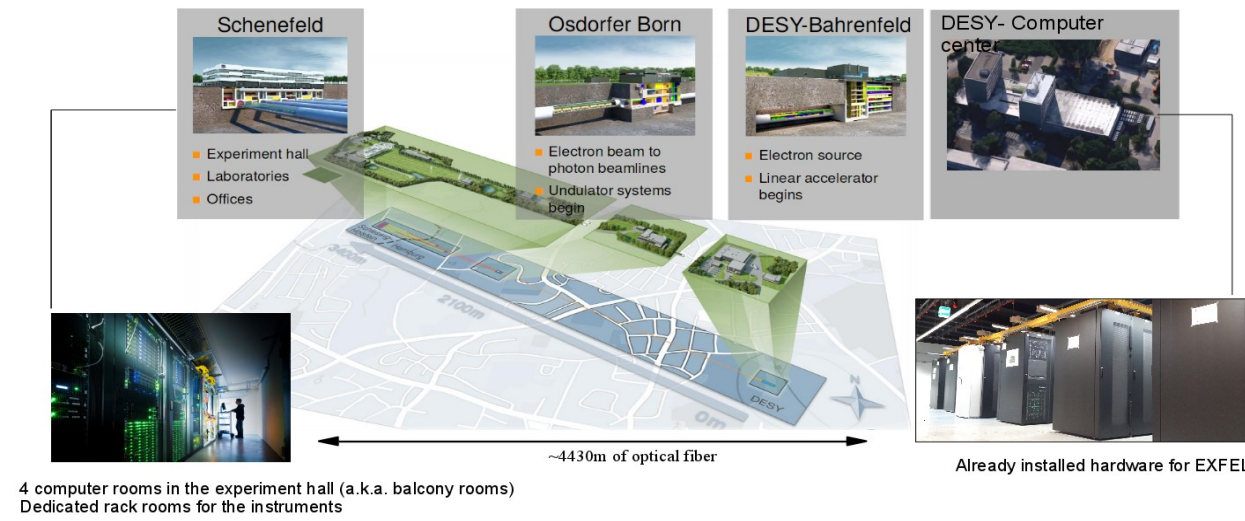
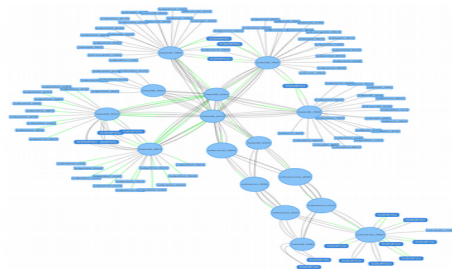
PETRA-III and FLASH

- ASAP³ data taking and management workflow is well established and running smoothly
- 2 new beam lines added, more to come
- Focus is moving to data analysis
 - Maxwell HPC cluster, GPUs ...
- New data sources are being added to ASAP³
 - Detector development, on-site laser and microscope experiments, long term XFEL accelerator monitoring, ...



European XFEL

- Data taking off-campus in Schenefeld, storage & analysis on the DESY campus
- User runs in November 2017, March & May 2018
 - More continuous operation starting in summer
- About 1 PiB of data stored already, up to 50 TiB per day
- Calibration and analysis using the Maxwell HPC cluster at DESY
- One SASE active, two more starting this year
- Infiniband monitoring developed by IT



Software auf *Maxwell*: Engineering und andere wichtige Produkte

• Kommerzielle Produkte

- ansys, avizo, comsol, jetbrains
- maple, mathematica, matlab
- vgstudio, intel, portland, u.v.a. ...
- In der Regel über Netzwerk-Lizenzen im DESY-Netz verfügbar
- In der Regel ausschließlich für akademische Anwendungen



• Frei verfügbare Produkte

- openFoam, Orca, Paraview, VTK/ITK
- QuantumEspresso, Salome
- diverse Anwendungen für X-ray Simulationen
- diverse Anwendungen aus der Strukturbioogie
- Deep Learning: Tensorflow, Keras, u.v.a. ...
- In der Regel ausschließlich für akademische Anwendungen



Und die Zukunft?

**Folien von Volker Gülzow aus
der POF Evaluierung Februar
2018**

Scientific Computing Development @ DESY

Scientific Computing is a key feature for LK I research activities and is an integral part of the LK II facilities

Scientific Computing

Simulation

Automation and
Control Systems

Data Analytics
Machine Learning
in the program MT

Hardware
Acceleration

Software Engineering

Data Management
and Storage

Scientific computing development @ DESY

A key feature for LK I research activities and also part of the LK II facilities

Scientific computing is considered as one of the key topics for research at DESY.

The following key activities are currently being prepared or intensified:

- Building an internal **virtual group** for scientific computing across divisions.
- Preparation of the topic „**Data Management and Analytics**“ in the program “Matter & Technologies”.
- Scientific computing will significantly contribute to the planned **Center for Data and Computing Science** (CDCS) on a cross disciplinary level.
- Scientific computing will consolidate and strengthen the effort to acquire **National, European & International Projects**.

Virtual Group „Scientific Computing“ at DESY

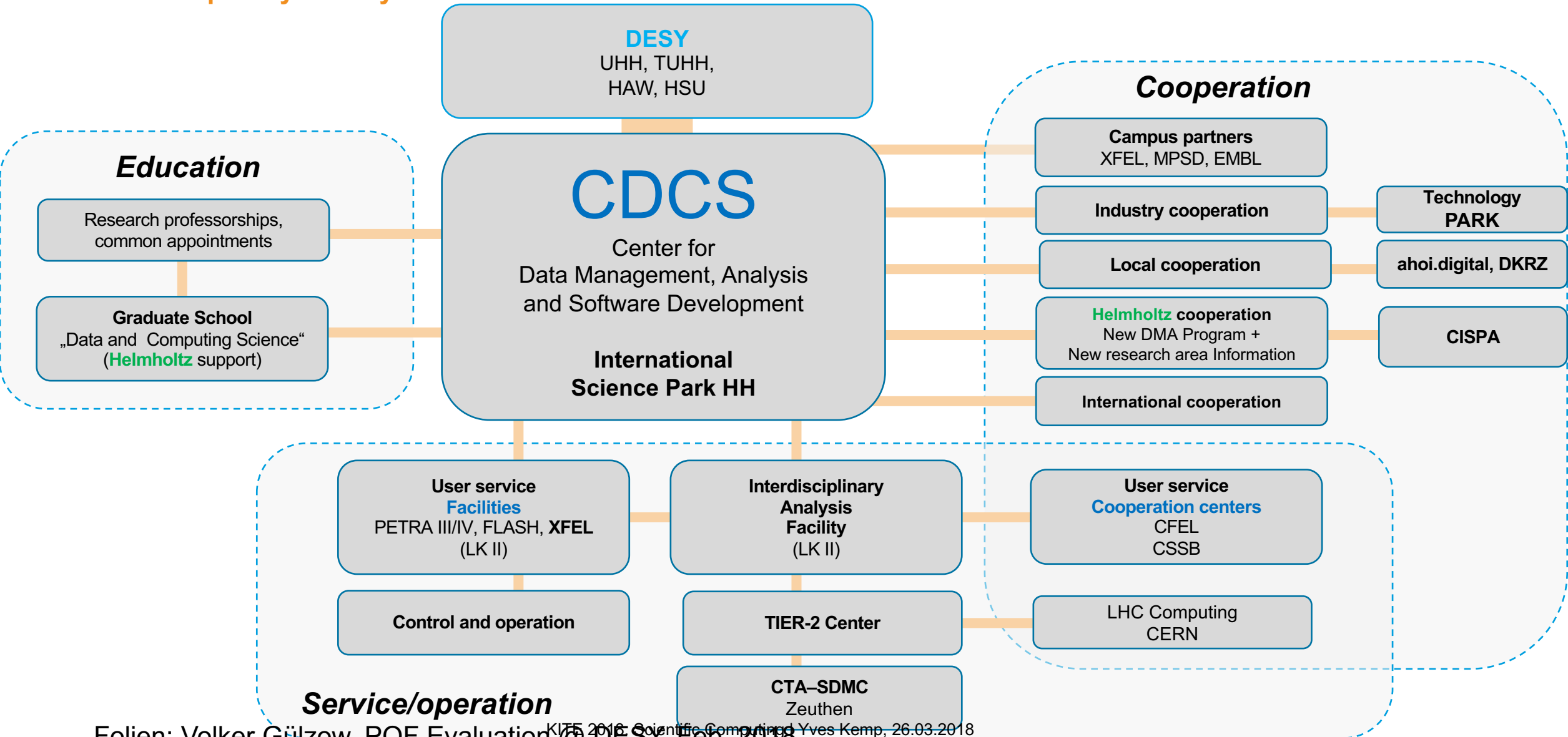
Across DESY divisions

Common Topics	FH Division	FS Division	M Division	ZN
Technologies & Methods	<i>LK 1</i>	<i>LK 1</i>	<i>LK 1</i>	<i>LK1</i>
Algorithms	SW development Analysis	SW development Analysis	Modelling and Simulation	Simulations Methods
HPC, GPU Optimization	Methods Algorithms HPC	Methods Algorithms	Analysis methods Algorithms	Algorithms Technology
Parallelization	Optimization Big Data	<i>LK 2</i> <i>PERFORMANCE</i>	<i>LK 2</i> <i>PERFORMANCE & FLASH</i>	HPC Optimization Analysis
Clouds and Virtualization	<i>LK 2 (TIER-2)</i>	Support for scientific & industrial users Near Real time	Accelerator control Data Processing Data Acquisition	SW Development <i>LK II (TIER-2)</i>
Policies and Coordination	Grid computing Big Data Management	Analysis Data Management	Data Analysis Data Visualization	Grid computing Big Data
Visualization	Cloud			Management

The virtual group spanning divisions

Proposal: Center for Data and Computing Science (CDCS).

Cross disciplinary activity



Scientific Computing @ DESY:

- Bereits jetzt viele Aktivitäten bei IT (und sicherlich anderen Gruppen)
- Im Zuge der DESY Strategie 2030 ist eine Verstärkung des Scientific Computing bei DESY zu erwarten