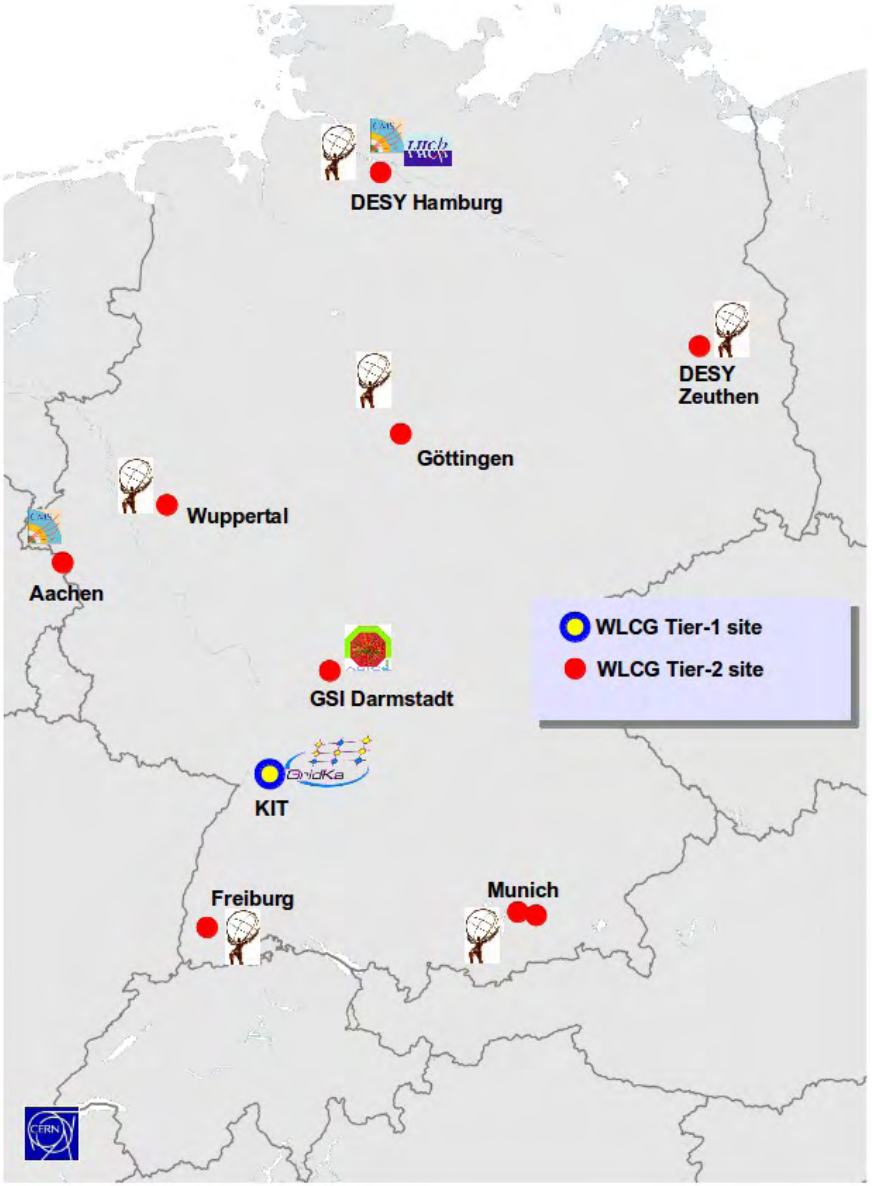
LHC Computing Status

17. Nov 2012
KET Tagung
G.Duckeck
LMU

- Überblick
- HEP Computing in DE
  - CPU Nutzung, Daten Verteilung, Storage
- T2 Ressourcen und Finanzierung
- Status Computing Stellen & Projekte

Input und Mitarbeit:
HGF Grid-PB
GridKa TAB
NAF Team



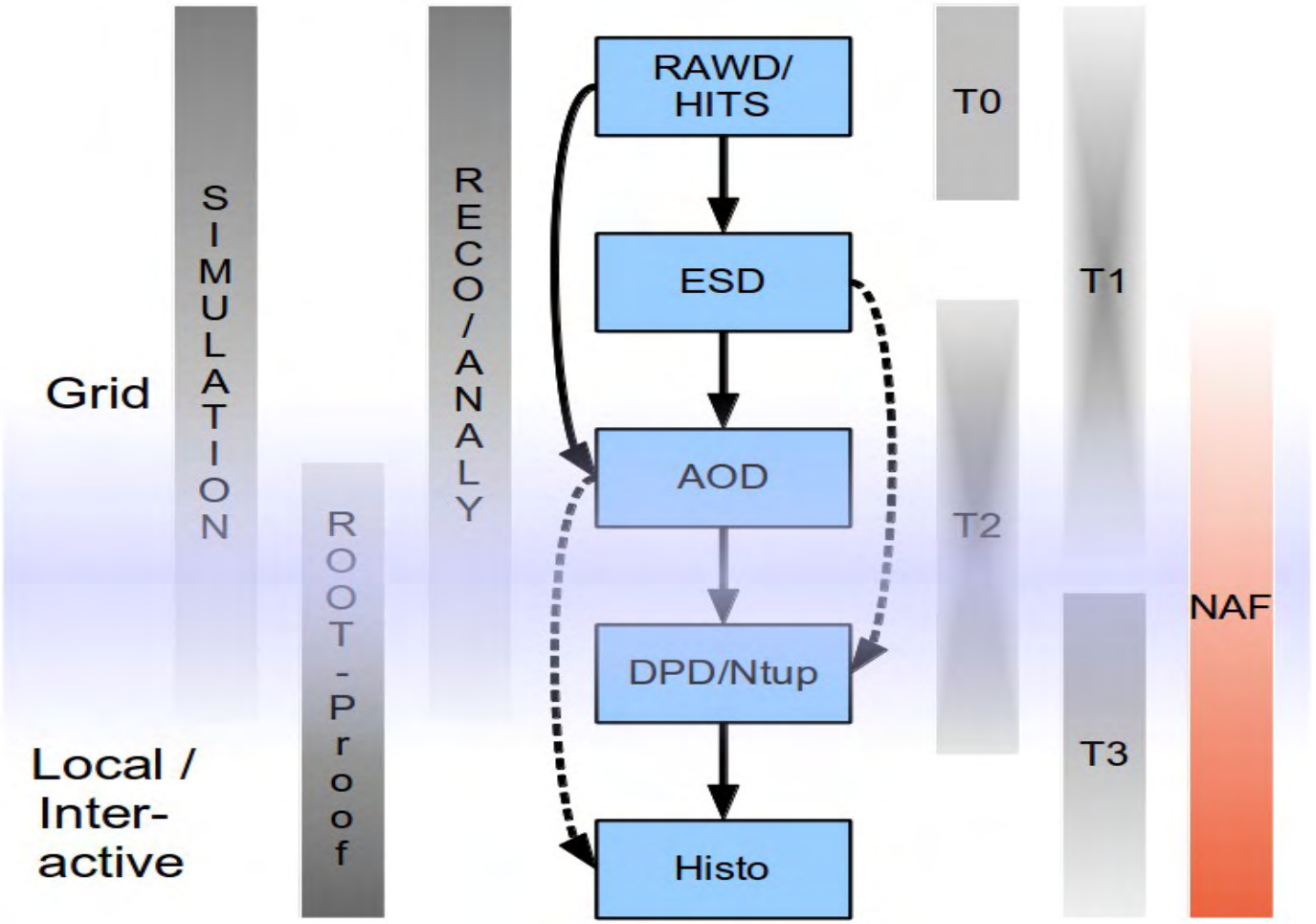
LHC Offline Computing

- Hauptteil der Ressourcen im Grid:

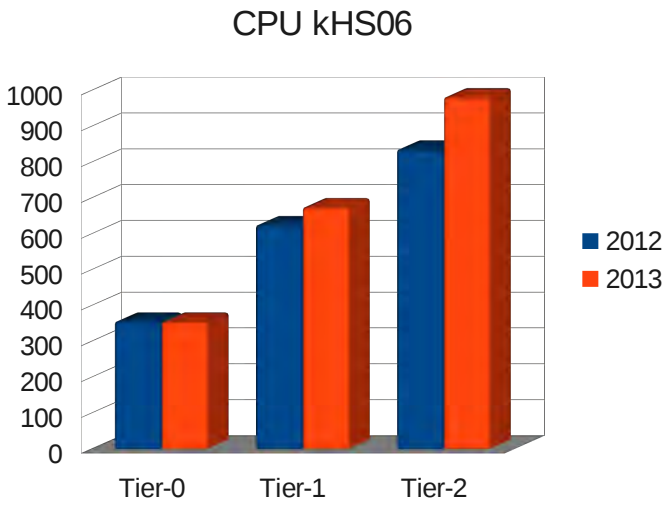
  Reconstruction, Simulation, Analyse

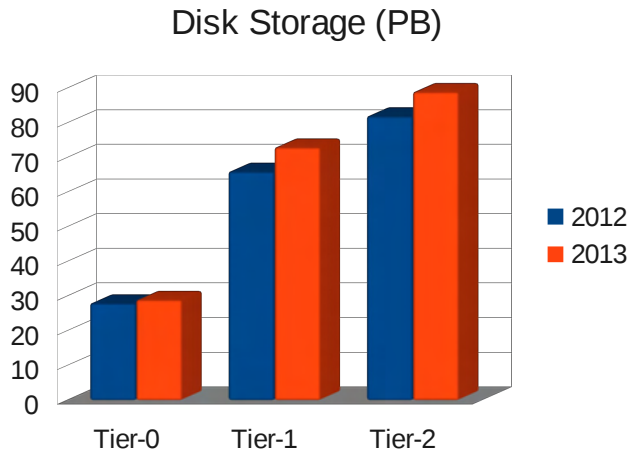
- Ergänzt durch Tier3/NAF:

  Code Tests, Ntuple Analyse



Weltweite Ressourcen WLCG





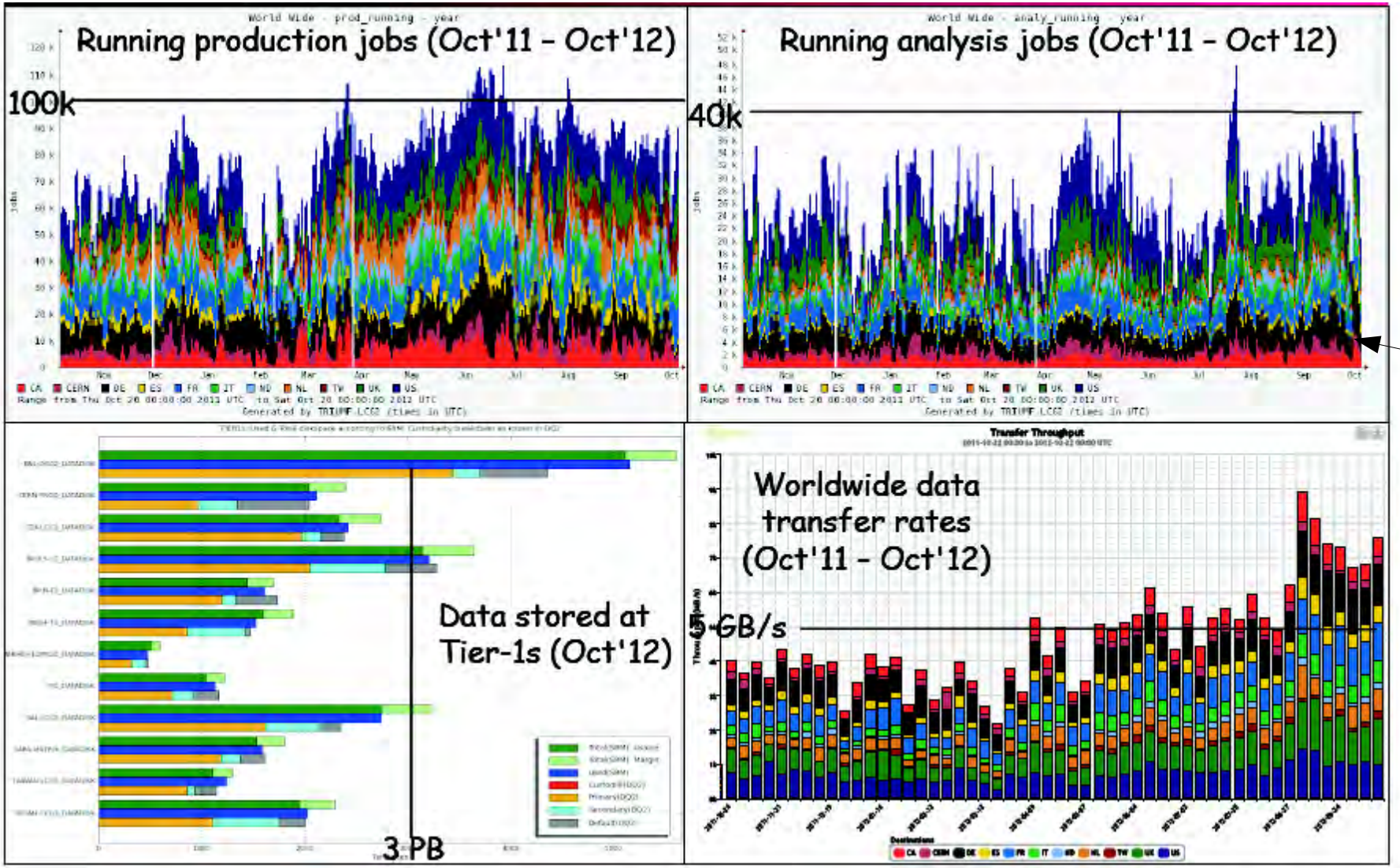
- Summe „pledged“ 2013 (150 Sites):
  - 2010 kHS06 (~200k CPU cores)
  - 190 PB Disk, 190 PB Tape
  - DE Anteil ca 9%
    - nur T1: ~14%, nur T2: ~8%
- WLCG im Vergleich:
  - LRZ SuperMuc (Platz 6(4) Top 500 Liste)
    - 155k Cores → ~3000 HS06, 10 PB Storage
  - Amazon S3: 10^{12} Objekte, ~500 PB (WLCG ~10^{9} Objekte =Files)
  - Facebook: 30 PB (2011)
  - Google DataCenter power: ~200 MW (WLCG ~6 MW)

Nutzung der Ressourcen

- Tier-1
  - Daten-Speicherung und -Verteilung, Reprozessieren, Gruppenproduktion
  - Exzellente Netzanbindung, Massenspeicher (Disk&Tape), 24x7 Betrieb
- Tier-2
  - Speicherung Analyse Daten, Analyse-Jobs, MC Simulation
  - Gute Netzanbindung, „CPU-Work-Horses“, 8x5 Betrieb (+ „best effort“)
- Tier-3/NAF/Instituts-Cluster (= „unpledged“ Ressources“)
  - „End“-Analyse, Root/Ntuples, Grid/Batch/Proof
  - oft assoziiert mit Tier-2 Site → symbiotische Nutzung von CPU & Disk
- Insgesamt sehr flexibles Modell zur optimalen Nutzung von verteilten Ressourcen für Produktion und User-Analyse
  - Computing entscheidend für die kurzen Analyse-Zyklen und Produktion von Konferenz-tauglichen Resultaten wenige Tage nach Datennahme (s. ICHEP)
- DE Zentren sehr erfolgreich beteiligt und international sichtbar

unverzichtbarer Beitrag zum LHC Programm

ATLAS Performance



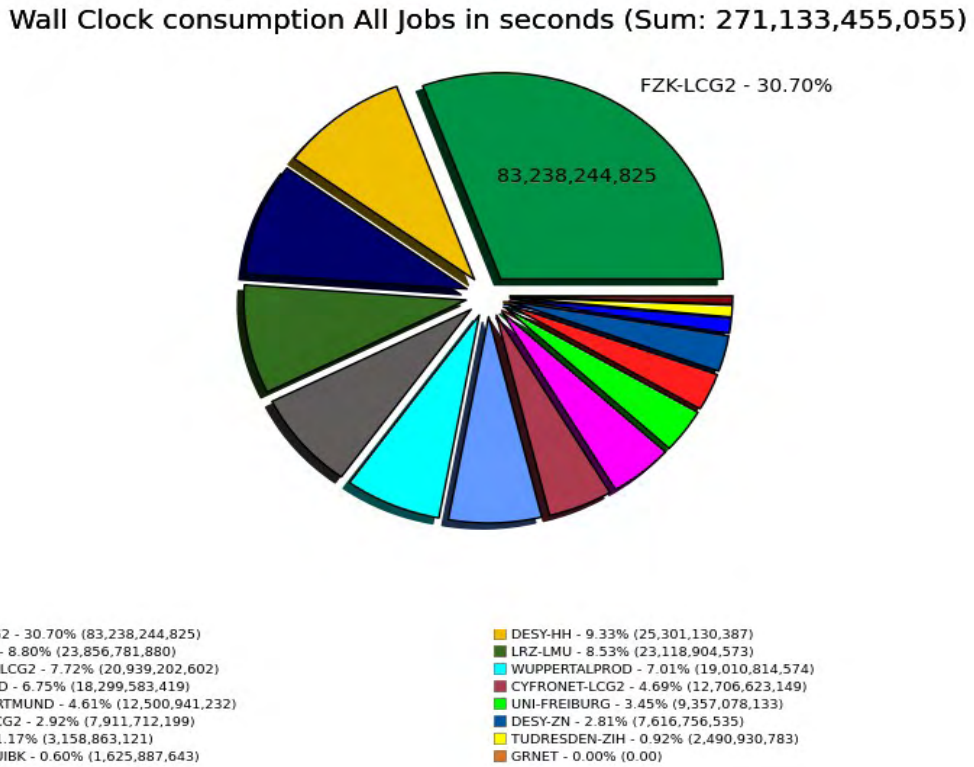
ATLAS–DE cloud per Site Jan-Sep 2012

ATLAS-DE Anteil:

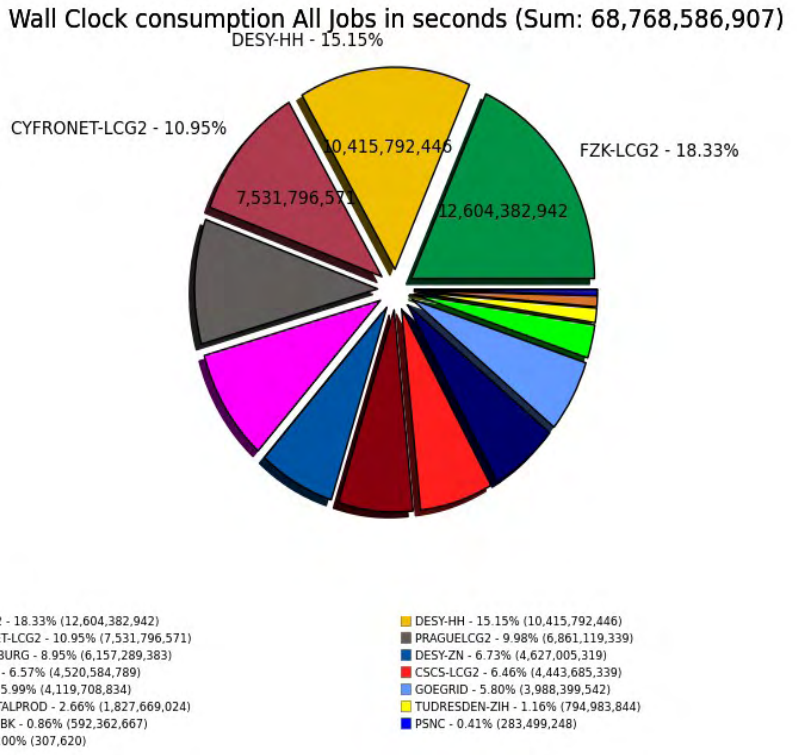
12% ATLAS gesamt

GridKa:	production 30%, analysis 18%
Desy/MPP :	production 21%, analysis 28%
University sites:	production 31%, analysis 26%
non-De sites:	production 18%, analysis 28%

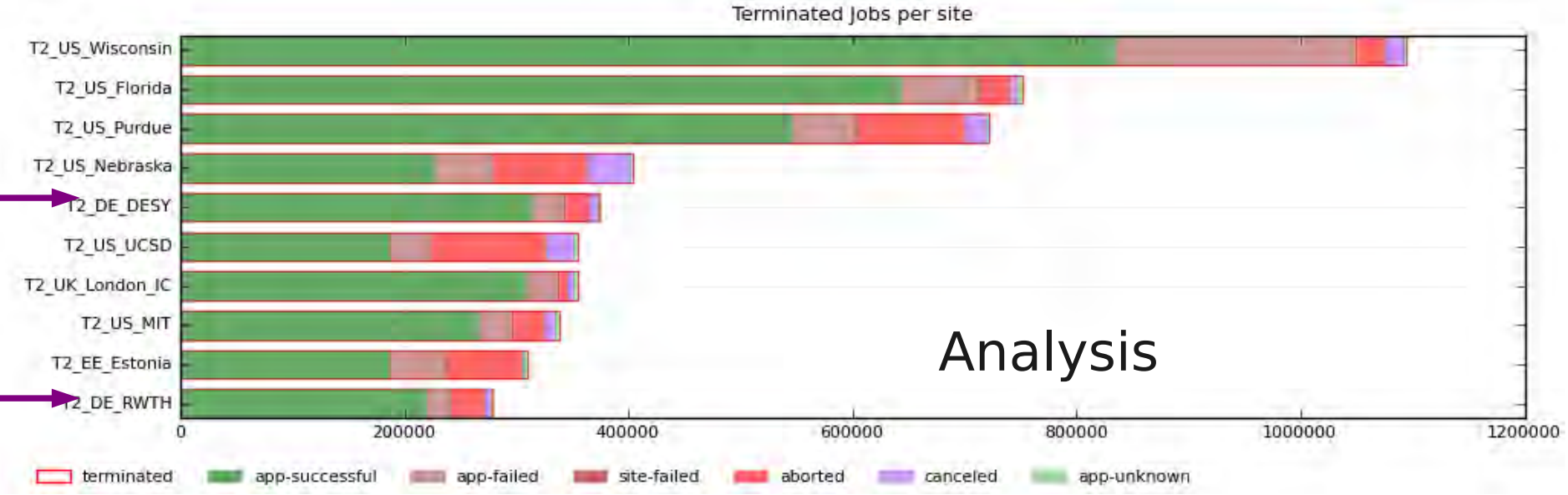
Production



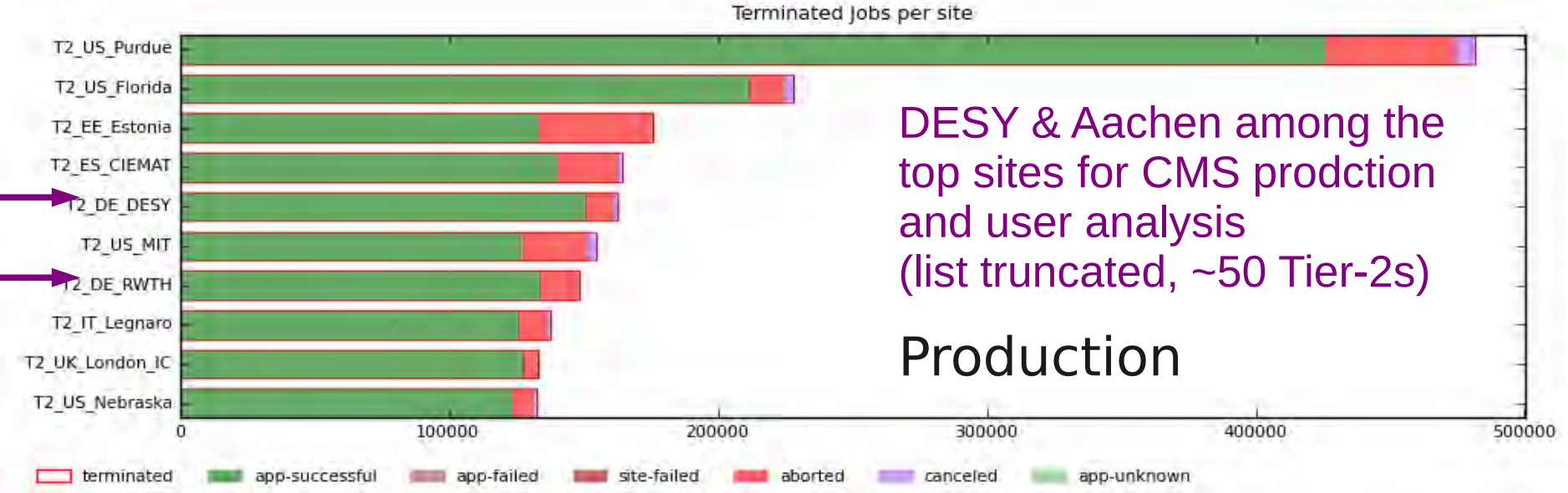
Analysis



CMS Job Statistics (October)



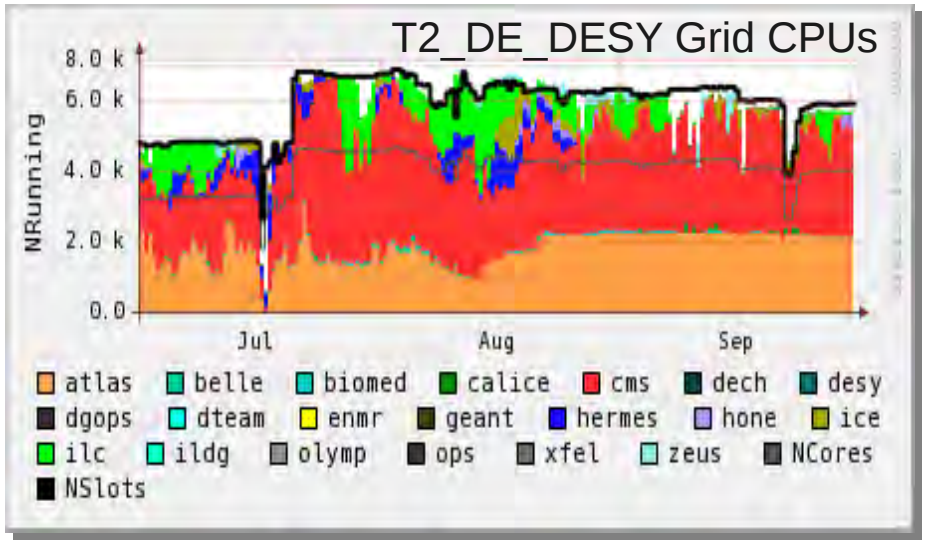
Analysis

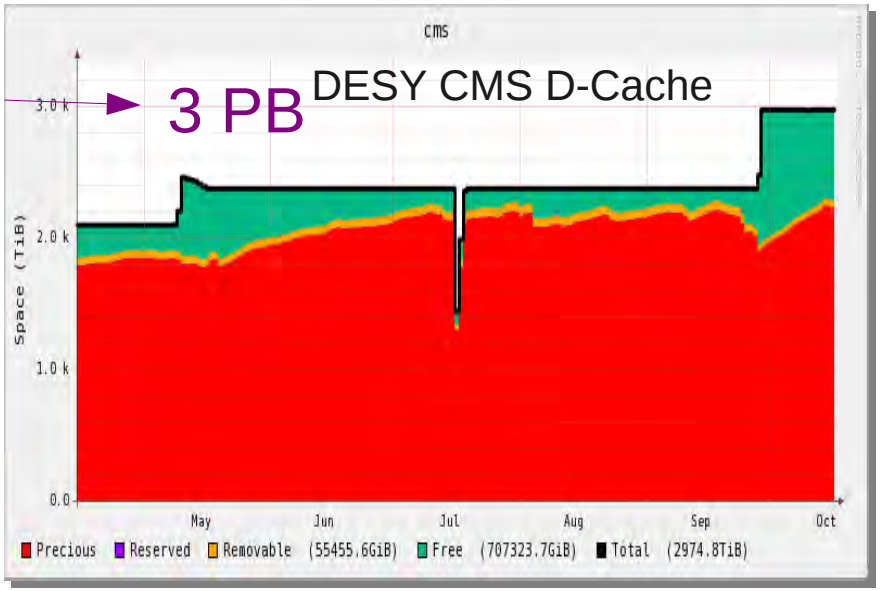


DESY & Aachen among the top sites for CMS prodction and user analysis
(list truncated, ~50 Tier-2s)

Production

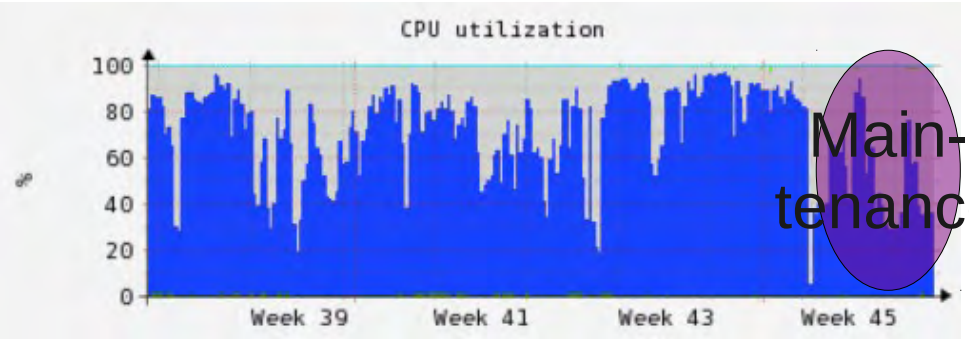
CMS Resource Usage at DESY and RWTH Aachen



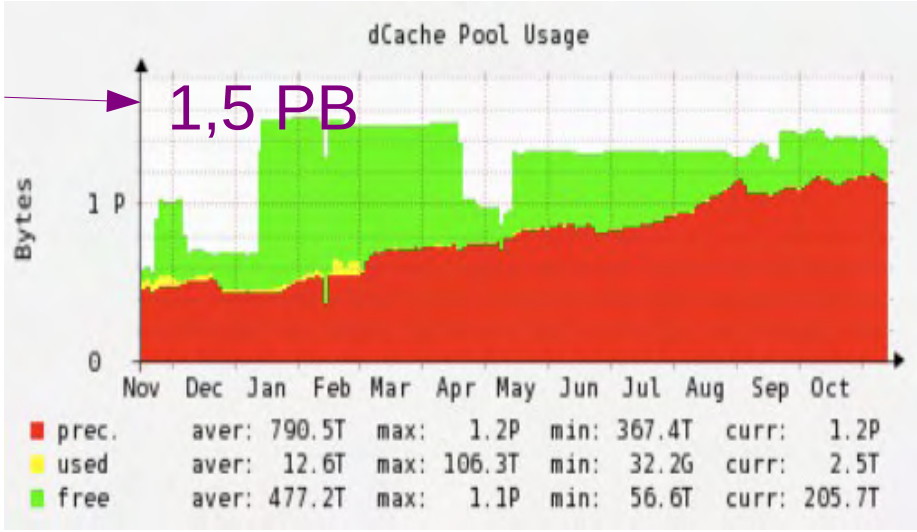


Resources are well used
Much more data will arrive !
Storage (including FSP-CMS analysis users' data) is filling up !

T2_DE_RWTH CMS CPUs

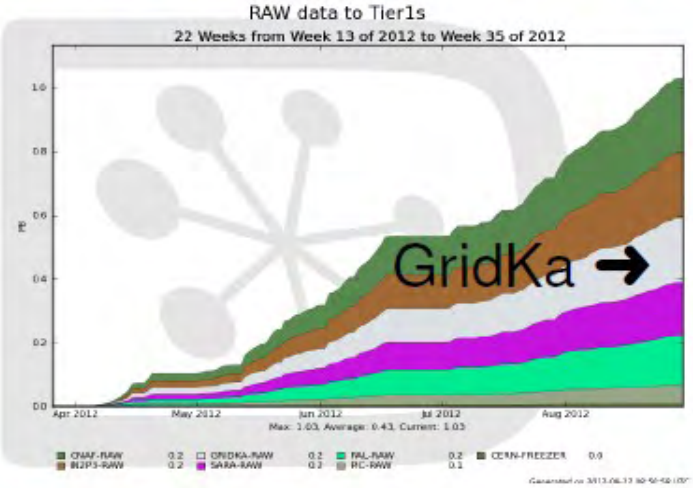


Aachen CMS storage (last year)



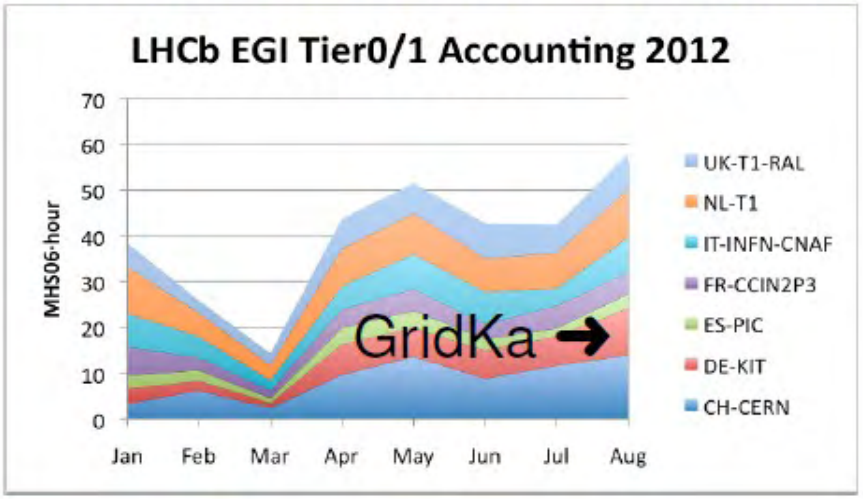
LHCb - Internationaler Vergleich

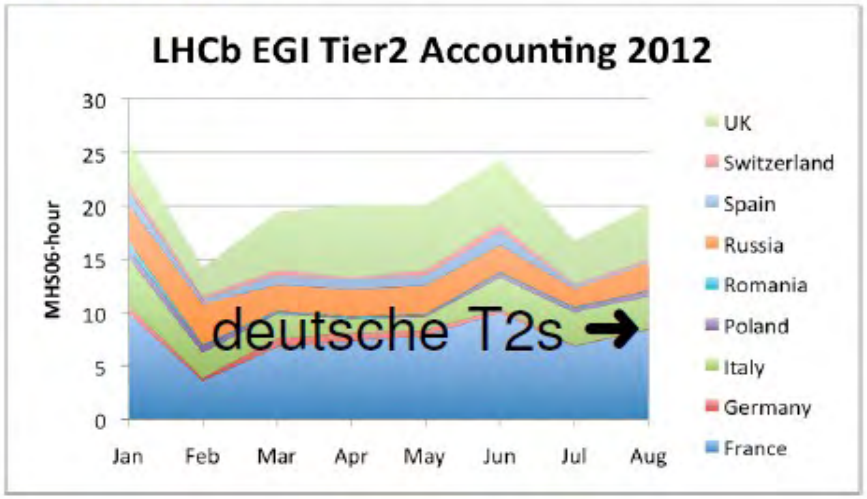
➜ Verteilung der Rohdaten an die T1-Zentren



- 200 TB Rohdatenimport zu GridKa
- gleichmäßige Verteilung zwischen T1s
- integrierte deutsche Beiträge
  - ➜ 17% der T1 Rechenleistung
  - ➜ 2% der T2 Rechenleistung (MC)

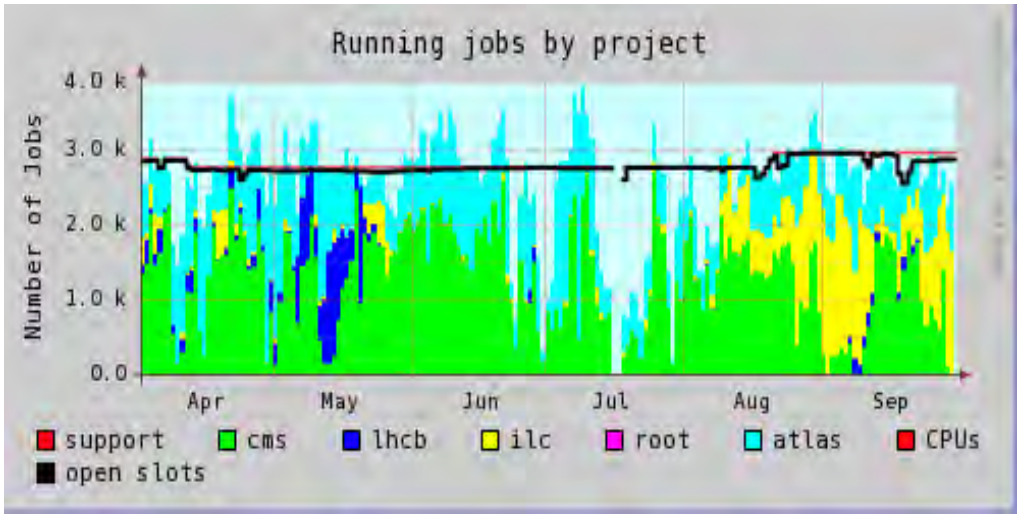
➜ Rechenzeit an T1- und T2-Zentren

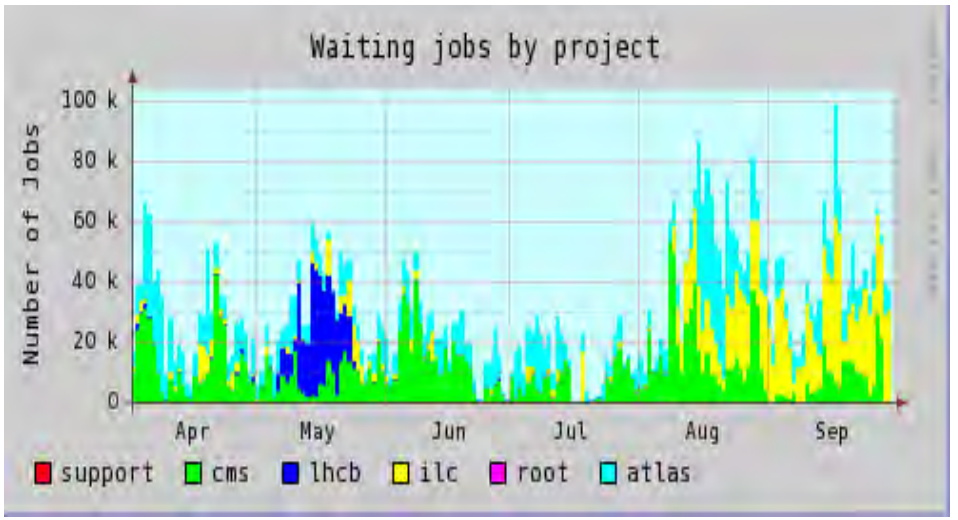




The National Analysis Facility

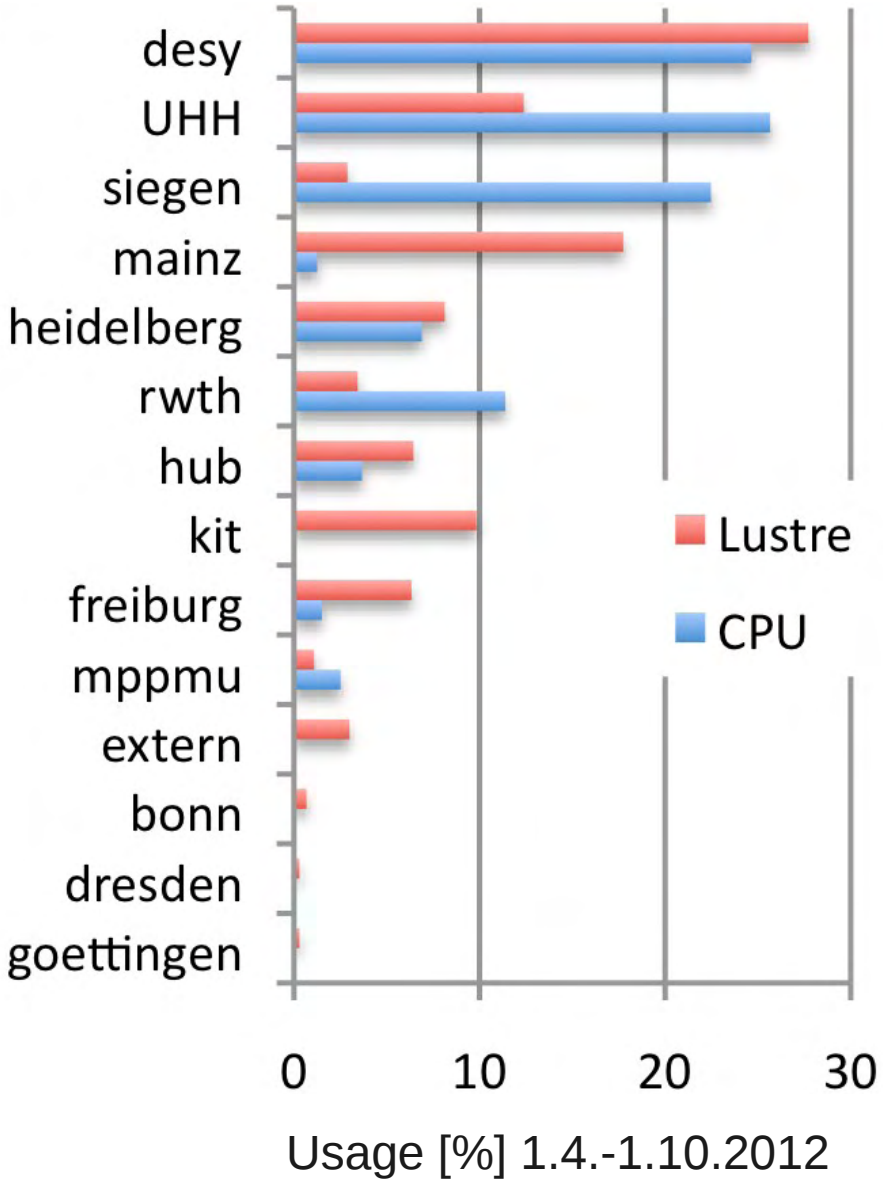
- Established 2007 in the realm of the Helmholtz Alliance at the Terascale
  - Complements the DESY Grid Infrastructure as well as the German Grid Landscape
  - Provides special resources optimized for interactive work and analysis, orthogonal ansatz w.r.t Grid – but glued together via data
- Intended for analysis of LHC data (ATLAS, CMS and LHCb), as well as ILC & CALICE data
- Resources well utilized by the experiment
  - ~3000 CPU cores, large fraction (~40%) purchased by Uni-HH/CMS Schleper Group
  - ~400 TB Lustre space, ~300 TB to be replaced by new technology soon
  - Space for NAF in dCache Grid Storage Elements, around 2.5 PB (ATLAS+CMS) in addition (beyond pledges)





The National Analysis Facility at DESY

- Usage by external institutes and groups
  - The NAF is a central DESY infrastructure for LHC analysis. DESY strong user groups – but not dominating!
  - The NAF is at the same level an infrastructure for LHC analysis to every physicists from a German institute.
  - Other institutes can contribute in HW resources – e.g. Uni-HH / CMS contributions
  - Many institutes use the NAF either as a daily workhorse, or in peak times, e.g. during PhD analyses.



DE Tier-2 Ressourcen 2013

	CPU (HS06)		Disk (PB)	
	2012	2013	2012	2013
ATLAS				
Sum 4xUni	18500	15000	3200	2800
Desy/MPP	17800	20300	2200	2400
Sum DE	36300	35300	5400	5300
ATLAS share	**11.4%**	**10.1%**	**11.3%**	**10.8%**
CMS				
Aachen	8900	7500	600	500
Desy	14700	16700	1350	1350
Sum DE	23600	24200	1950	1850
CMS share	**7.5%**	**6.9%**	**7.5%**	**7.1%**

- Computing-Share ≈ Autoren-Share:
  - ATLAS=11.5%, CMS=7.5%
- Finanzierung Uni T2s HGF Allianz 2007-2012
  - Basis für WLCG Pledges
  - Ohne weitere Finanzierung zwangsläufig Reduktion der Pledges
    - Hardware max 5 Jahre zu betreiben

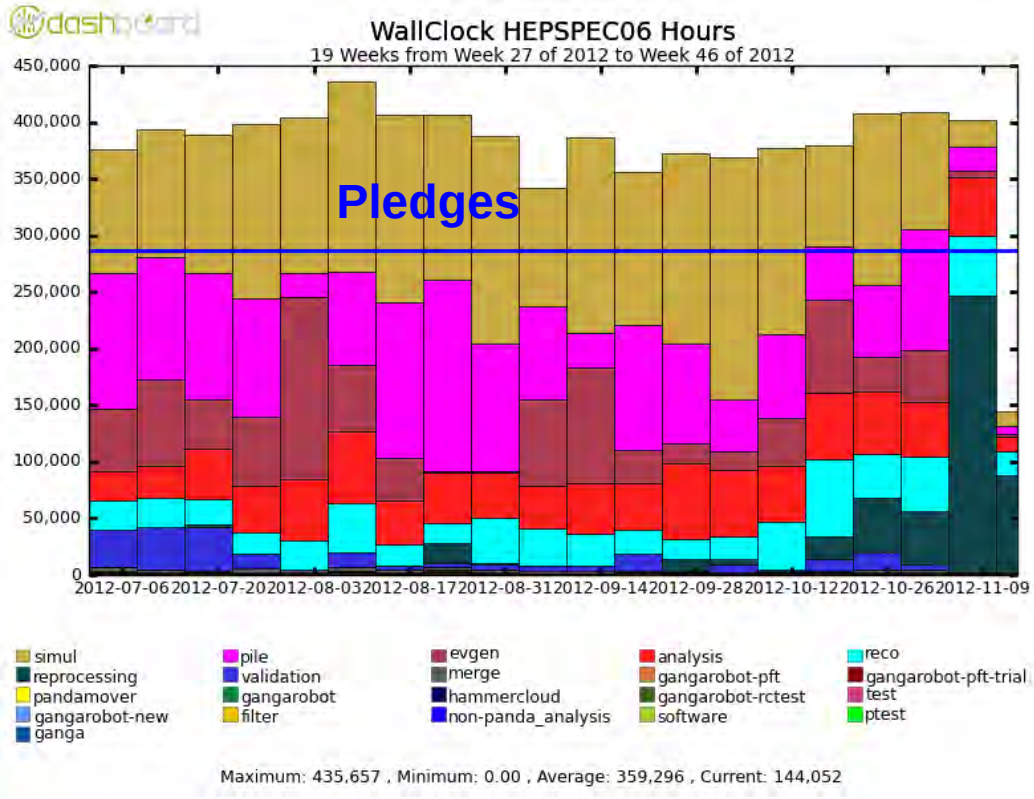
- Neu: Einmalige Sondermittel von BMBF für Tier-2 und NAF Ende 2012
  - zusätzlicher Bedarf für Computing Ressourcen gegenüber bisheriger Planung:
    - Entdeckung Higgs-ähnliches Teilchen und LHC Laufzeitverlängerung
  - aufgeteilt zwischen Uni Standorten ATLAS&CMS und NAF
  - Computing Ressourcen für 2013 damit (voraussichtlich) ausreichend abgedeckt
- Löst nicht langfristiges Problem der Finanzierung der Uni T2s !

Brauchen wir Tier-2 an 9 Standorten in D?

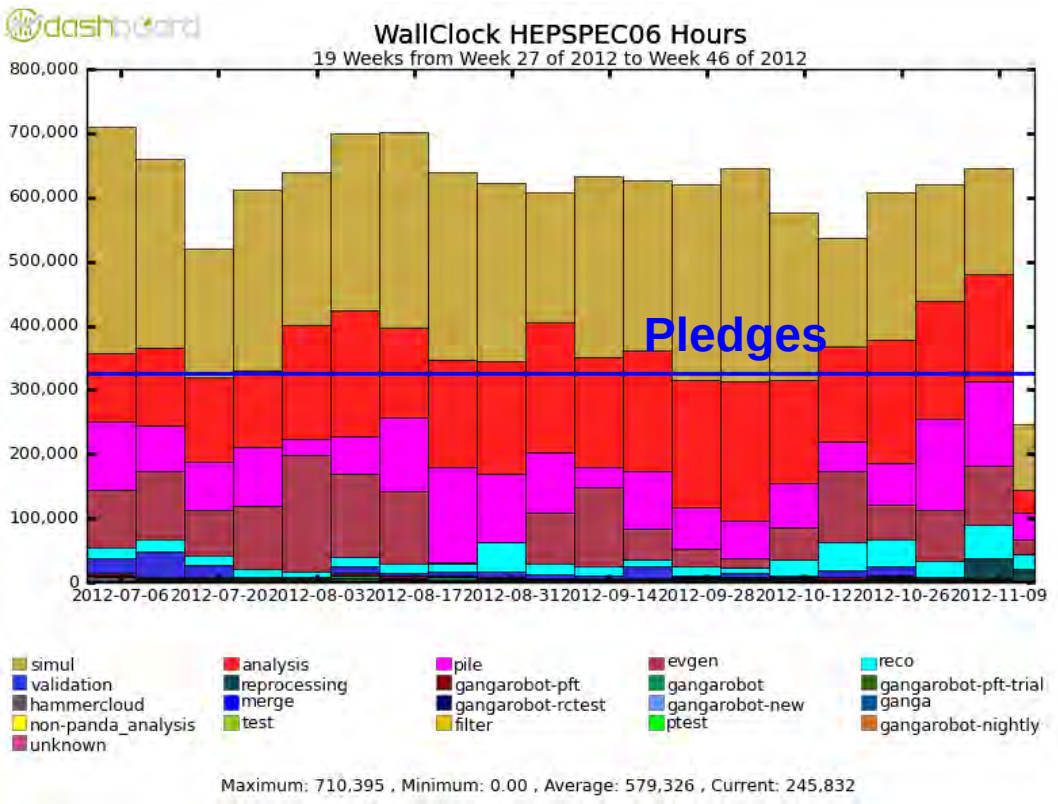
- Offensichtlich mehr Personal nötig um Ressourcen auf viele Standorte verteilt zu betreiben
  - 9 Sites: ca. 2 Admin * 9 Sites + 2-5 Ops * 3 Expt ≈ 25-30
  - 3 Sites: ca. 3-4 Admin * 3 Sites + 2-3 Ops * 3 Expt ≈ 15-20
- Aber Grundidee WLCG Modell:
  - lokale Ressourcen möglichst flexibel nutzen
- International „viele“ T2 Sites üblich, DE eher wenig:
  - US: 25, UK: 19, FR 8, IT: 10, ...
- Erhöhter Personal-Einsatz mehr als ausgeglichen durch
  - Nutzung lokaler Infrastruktur&Personal (Gebäude, Strom, lokale Admin, …)
  - Opportunistische Nutzung von T3 Kapazität → signifikante extra Ressourcen
  - Ausbildung Computing-Experten an Uni-Gruppen

Beispiel ATLAS Tier-1 und Tier-2 CPU Nutzung vs Pledge

ATLAS all Tier-1:
Nutzung – Pledge: +15%



ATLAS all Tier-2:
Nutzung – Pledge: +85%



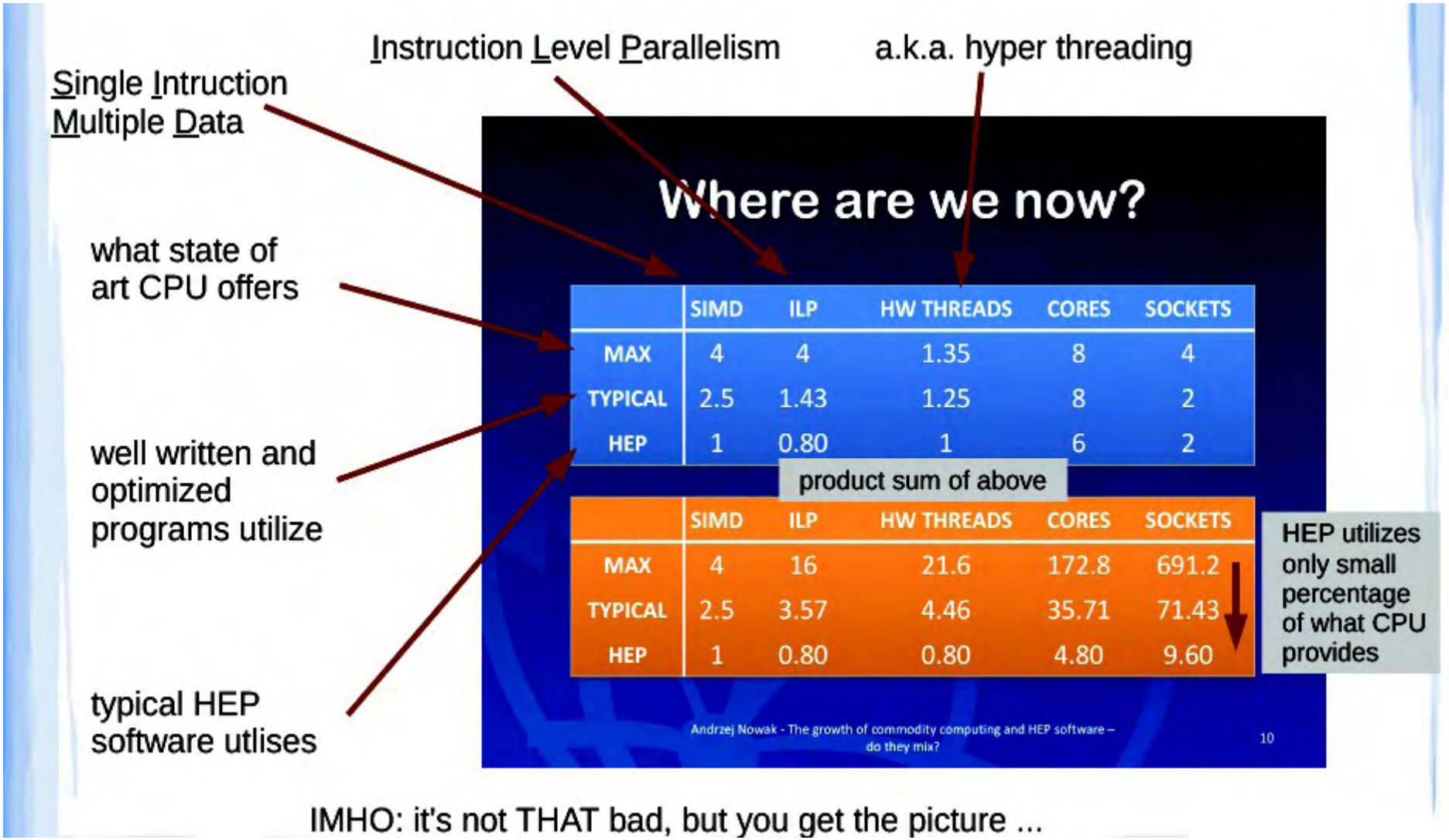
Stellen Situation DE Computing

- Standbeine:
  - BMBF Verbundforschung
    - ATLAS&CMS Stellen für experiment-spezifischen Betrieb der Zentren
      - Halbierung der Stellen für FP 2012-15
    - Core Computing Aufgaben und Entwicklung im Experiment:
      - Stellen beibehalten (I.W. in-kind M&O A)
  - HGF Allianz
    - dCache Support, Entwicklung Grid Computing Tools ~3 FTE 7/2007 bis 6/2012
    - HGF Computing Project 2013/14: 135 kEuro/a (≈2.5 FTE) genehmigt
      - High performance data access & storage, monitoring tools, flexible Job Submission (cloud, HPC), WAN Networks (LHCONE), Training
  - Wichtige Beiträge von Desy und KIT (neben T1&T2 Betrieb)
    - NAF-Support, dCache Entwicklung und Support, Schulungen
- In Summe Manpower für Grid Computing Betrieb deutlich reduziert
  HGF Computing Projekt entscheidend um mit Weiterentwicklungen Schritt zu halten

Perspektive

- Wir brauchen „kritische Masse“ für Computing Expertise in DE
  - interessante Projekte und Perspektive für Post-docs & Mitarbeiter
- Computing Vorhersagen immer unsicher aber unwahrscheinlich dass HEP Computing in 10 Jahren genauso funktioniert wie jetzt
  - in etlichen Bereichen nicht nur „mehr vom Gleichen“ sondern „disruptive“ Technologie Trends
    - many-core CPUs, GPUs, SSD vs HDD, commercial clouds & HPC
  - Umstellung & Nutzung potentiell sehr aufwändig
- Viele Optimierungen / Aufräumarbeiten geplant für LHC LS1 ...
  - insbesondere Software – Parallelisieren, many core, GPU
  - nicht-trivial neue Anforderungen zu erfüllen
    - high Pileup, trigger rate 1 kHz
- … und noch viel mehr nötig für Phase-2/2020

Beispiel: Software Optimierung (R. Seuster)



IMHO: it's not THAT bad, but you get the picture ...

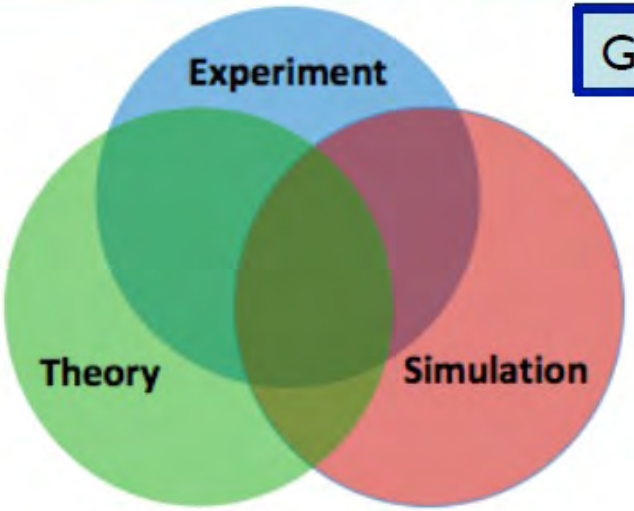
Zusammenfassung

- LHC Computing weltweit und in DE mit guter Stabilität und Performance
  - Maximierung verfügbarer Ressourcen ermöglicht durch die flexible Nutzung von vielen verteilten Zentren
  - Gute Funktion von Sites&Services Voraussetzung für kurze Analyse-Zyklen und schnelles Produzieren von Konferenz-tauglichen Resultaten
- Entscheidend ausreichend Computing Ressourcen (auch in D) zu haben
  - Computing sollte nicht Möglichkeiten der LHC Analyse beschränken
  - BMBF Sondermittel für Tier-2/NAF gewährleisten zusätzliche Ressourcen für die besonderen Anforderungen in 2013
  - Langfristige Tier-2 Finanzierung nach wie vor ungelöst !
- Computing Stellensituation
  - schwierig für Experiment-Betrieb durch Halbierung BMBF Stellen
  - HGF Computing Projekt ermöglicht Kontinuität für Entwicklungen 2013/14

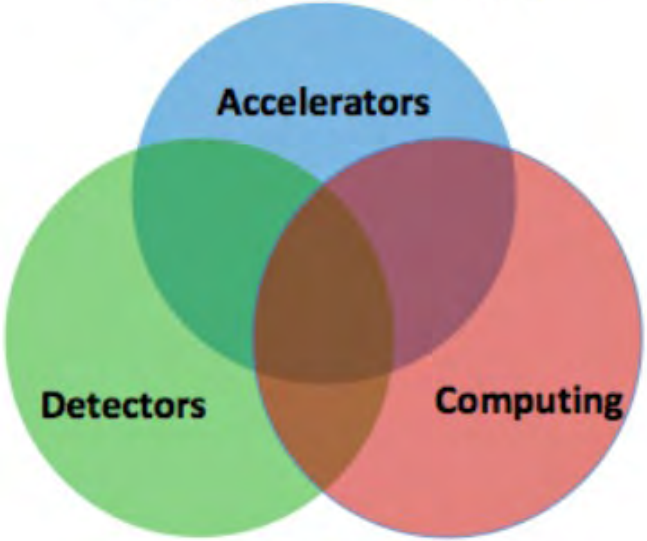
Physics and Technology

Glen Crawford / DOE

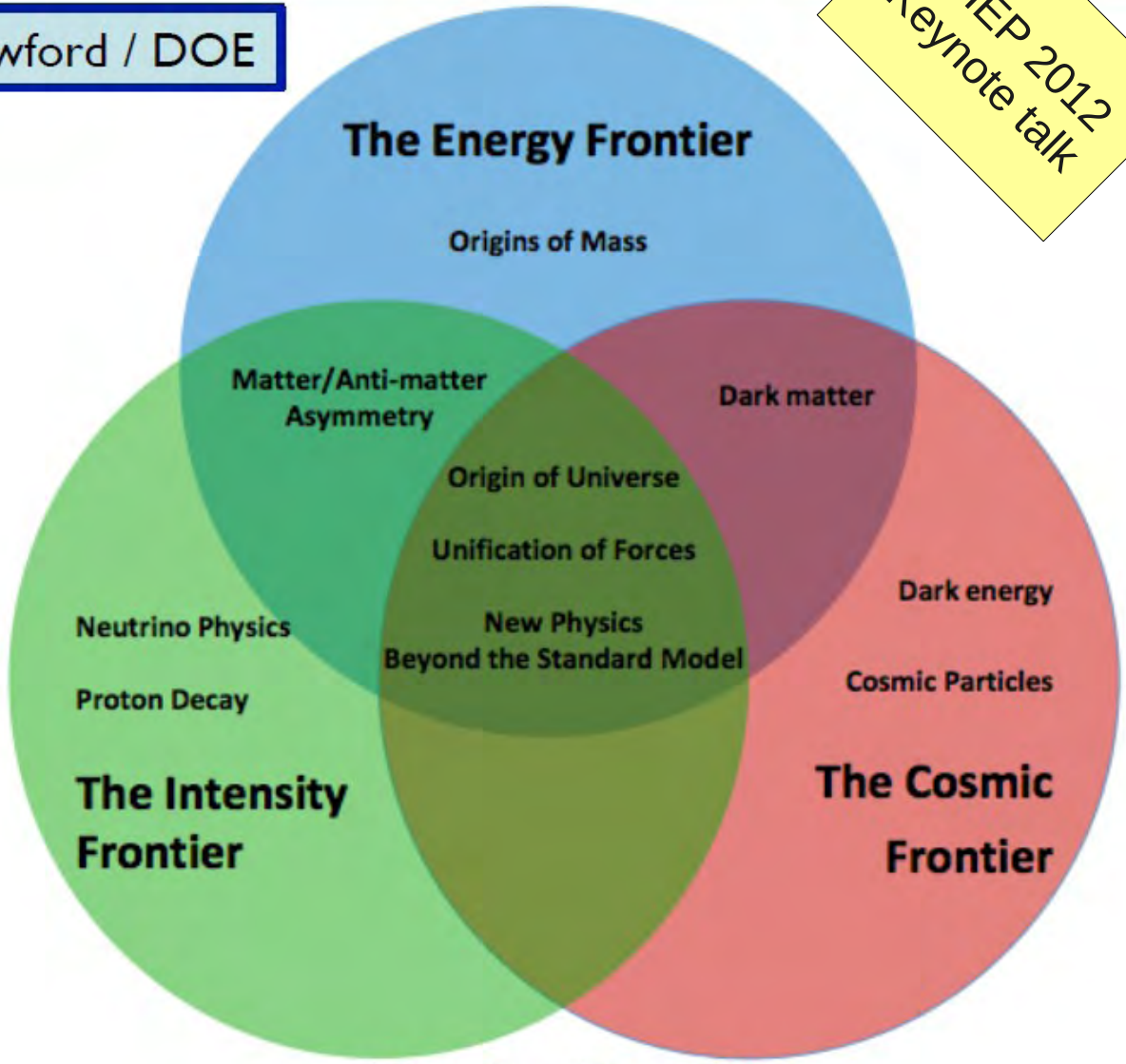
CHEP 2012 Keynote talk



Along Three Paths



Enabled by Advanced Technologies in:



Physics Frontiers

Zusätzliches Material

GridKa-Ressourcen 2013

Exp.		VO-Requirements/ GridKa Anteil 2013	Pledge 2012	Vorläufiger Pledge 2013 (30.09.12)	Mögliches Szenario: Leichte Umverteilung bei CPU, konstante Speicher-Ressourcen je VO. (ca % der Anforderungen / der Stand 12.11.12 tatsächlich gepledgten Ressourcen)
Alice		**25%**			
	CPU	120000 / 30000	40000	30000	30000 (25% / 29.7%)
	Disk	10800 / 2700	2700	2225	2700 (25% / 35.3%)
	Tape	21000 / 5250	5250	5250	5250 (25% / 37.2%)
Atlas		**12,5%**			
	CPU	319000 / 39875	32380	39875	39875 (12.5% / 12.0%)
	Disk	33000 / 4125	3375	3375	3375 (10.2% / 9.6%)
	Tape	40000 / 5000	4500	4500	4500 (11.3% / 11.0%)
CMS		**10%**			
	CPU	165000 / 16500	15000	17500	17500 (10.6% / 11.7%)
	Disk	26000 / 2600	2200	2200	2200 (8.5% / 9.3%)
	Tape	50000 / 5000	5100	5100	5100 (10.2% / 10.6%)
LHCb		**16.67%**			
	CPU	110000 / 18337	19200	19200	19200 (17.5% / 20.8%)
	Disk	8600 / 1433	1610	1450	1610 (18.7% / 22.5%)
	Tape	10800 / 1800	1050	1050	1050 (9.7% / 12.1 %)

CPU: [HEPSPEC06] Disk: [TB]

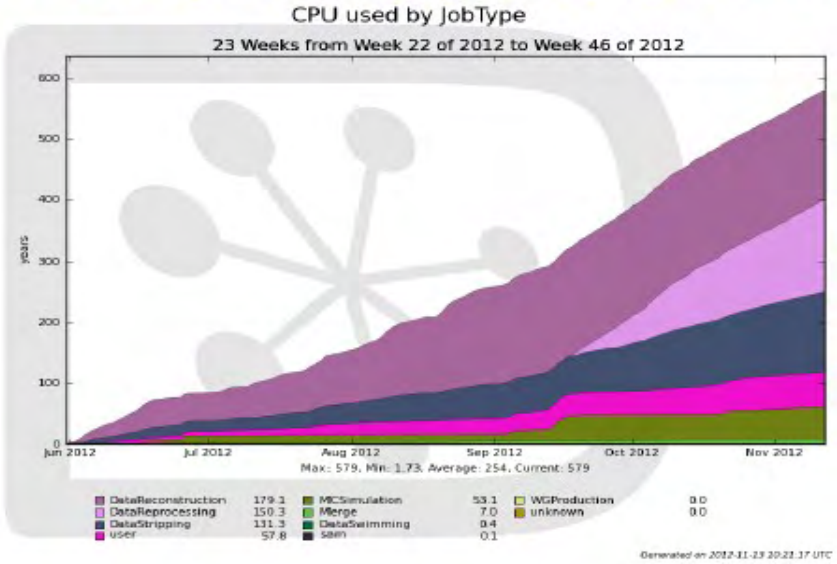
ATLAS DE Storage Überblick

	dCache/DPM	SPACE TOKENS (TB)				
Site	Version	DATADISK	GROUPDISK	PRODDISK	SCRATCHDISK	LOCALGROUPDISK
CSCS-LCG2	1.9.5-27	403/445	17/49	2/10	12/27	1/10
CYFRONET-LCG2	1.8.2	204/236	51/54	1/11	14/27	9/11
DESY-HH	1.9.12-12	694/772	296/375	8/27	32/68	672/838
DESY-ZN	1.9.12-21	505/565	117/118	7/18	26/37	496/557
FZK-LCG2	1.9.12-11	2634/2676	405/439	0/24	55/98	-1/-1
GoeGrid	1.9.12-20	474/830	54/98	16/39	14/29	68/78
HEPHY-UIBK	unset	94/107	0/1	0/5	1/3	0/1
LRZ-LMU	1.9.12-21	373/429	74/121	7/27	19/44	196/218
MPPMU	1.9.12-20	383/439	55/146	1/15	17/39	90/117
praguelcg2	unset	645/737	119/150	4/20	19/40	124/147
PSNC	1.8.4	47/54	-1/-1	0/5	3/21	-1/-1
UNI-FREIBURG	1.9.12-21	499/560	32/146	9/13	20/43	159/161
wuppertalprod	2.0.3-1	459/518	120/195	8/20	22/44	104/112

>1.9 PB in LOCALGROUPDISK → User-spezifische Analyse Daten an DE T2&NAF

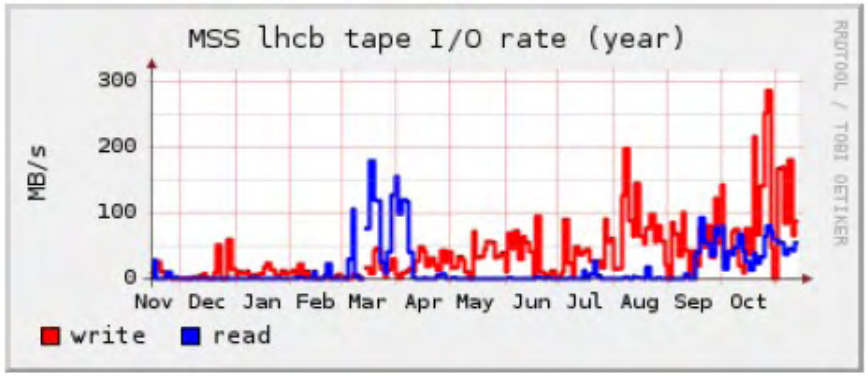
LHCb – Nutzung von GridKa in 2012

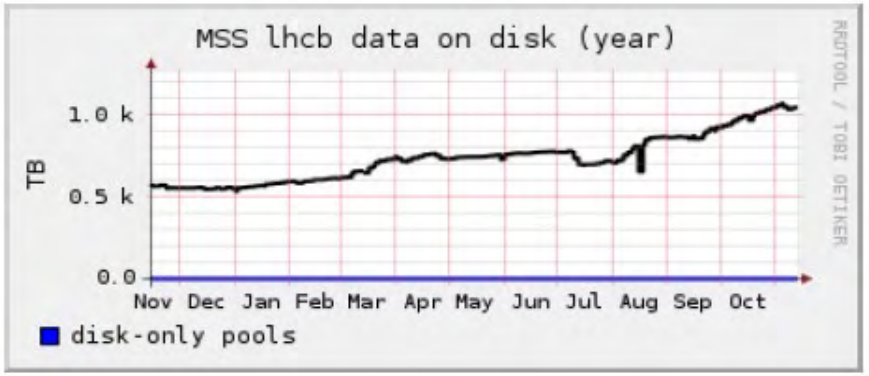
➜ Rechenzeit seit Juni 2012



- 580 CPU-Jahre seit Juni 2012
- 80% zentrales Data-Processing
- 10% User-Jobs
- 10% Monte Carlo
- Tape-I/O bis zu to 300 MB/s
- Plattenbedarf: Verdopplung in 2012
- derzeitige Belegung > 1 PB

➜ Disk und Tape





“NAF evolution” or “NAF 2.0”: Idea and status

- The current NAF successful and important
- Needs changes to be successfully continue its mission in the future
  - Changes in the Helmholtz Alliance “Physics at the Terascale”
  - Current design is 5 years – new technologies and user expectations
  - User requested changes e.g. via GridCenter Review Taskforce or via NUC
  - Want to focus more on data centricity for even better analysis performance
- Status
  - Several discussions in NAF provider group, with NUC and with experiments
  - Blueprint for a “NAF evolution” (or “NAF 2.0”) sent to Alliance Grid Project Board and NAF computing experts as well as PRC (DESY review committee) – general positive feedback
  - First prototype systems set up and being used – first plan until end 2012
  - Further directions and plans depend on outcome of the tests
  - Proposal to have independent NAF service for ATLAS split at two sites in discussion (disfavored by ATLAS)

Ressourcen Überblick

DE ATLAS Resources				
	CPU (HS06)		Disk (PB)	
	2012	2013	2012	2013
ATLAS				
FR	4430	3252	783	654
GOE	3853	3853	1000	1000
LMU	5780	4620	670	540
WUP	4430	3252	783	654
MPP	5780	5917	670	883
DESY	12000	14400	1500	1560
Sum DE T2	36273	35294	5406	5291
ATLAS share	11.4%	10.1%	11.3%	10.8%
KIT T1	32380	39875	3375	3375
ATLAS share	12.5%	12.5%	12.5%	10.2%

DE CMS Resources				
	CPU (HS06)		Disk (PB)	
	2012	2013	2012	2013
CMS				
Aachen	8875	7500	600	500
Desy	14750	16690	1350	1350
Sum DE T2	23625	24190	1950	1850
CMS share	7.5%	6.9%	7.5%	7.1%
KIT T1	15000	17500	2200	2200
CMS share	10.3%	10.6%	10.0%	8.5%

DE LHCb Resources				
	CPU (HS06)		Disk (PB)	
	2012	2013	2012	2013
Desy T2	3200	3200		
LHCB share	7.4%	7.0%		
KIT T1	19200	19200	1610	1610
LHCb share	17.0%	17.5%	16.9%	18.7%

WLCG Betrieb 365d x 24h