cta
cherenkov telescope array

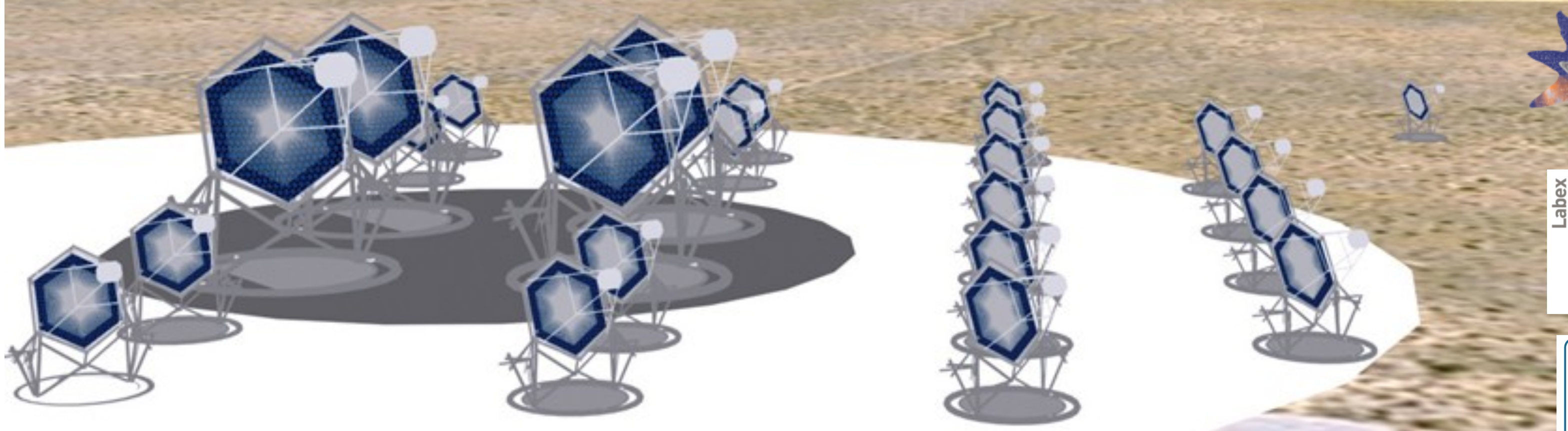
Cherenkov Telescope Array: a production system prototype

Volker Beckmann

using heavily the slides provided by

L. Arrabito[1], C. Barbier[2], J. Bregeon[1], A. Haupt[3], N. Neyroud[2]
for the CTA Consortium

[1]LUPM CNRS-IN2P3 France
[2]LAPP CNRS-IN2P3 France
[3]DESY



ACAV

Labex UnivEarthS

François Arago Centre

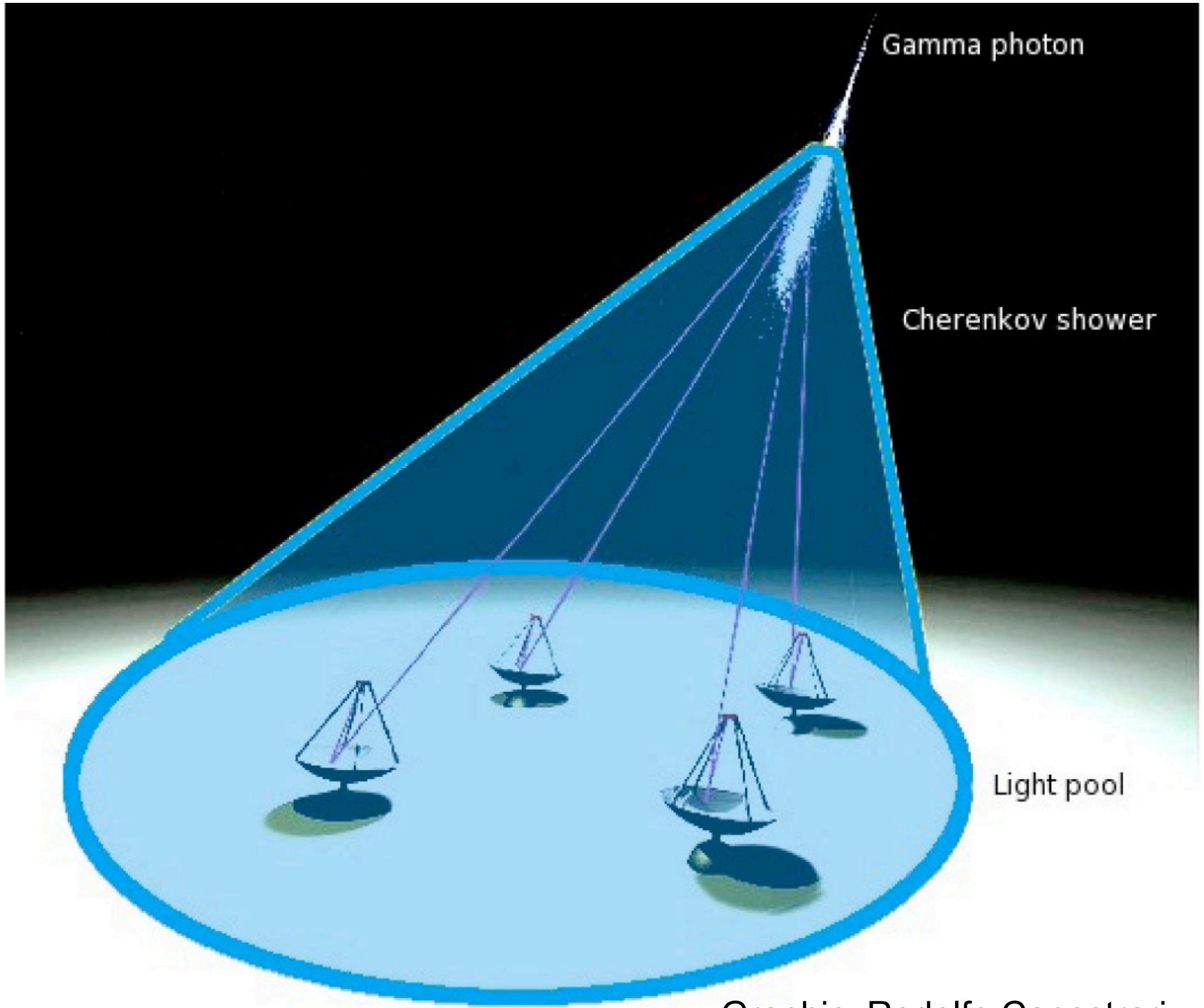
Distributed computing in astrophysics, FACe, December 10-11, 2015

Outlook:

- CTA in a nutshell
- CTA computing model:
  - Data volume
  - Data flow
  - Data processing
- CTA production system prototype:
  - Current MC simulations and Analysis
  - DIRAC for CTA
  - Resource usage: 2013-2015
- Future developments
- Conclusions

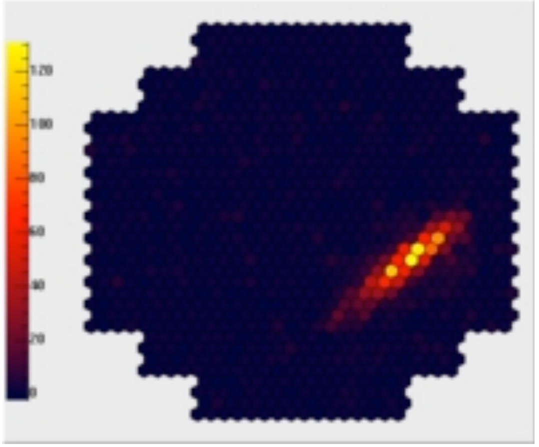
Big-Data tomorrow: CTA

Cherenkov Telescope Array: Ground based gamma-ray telescope



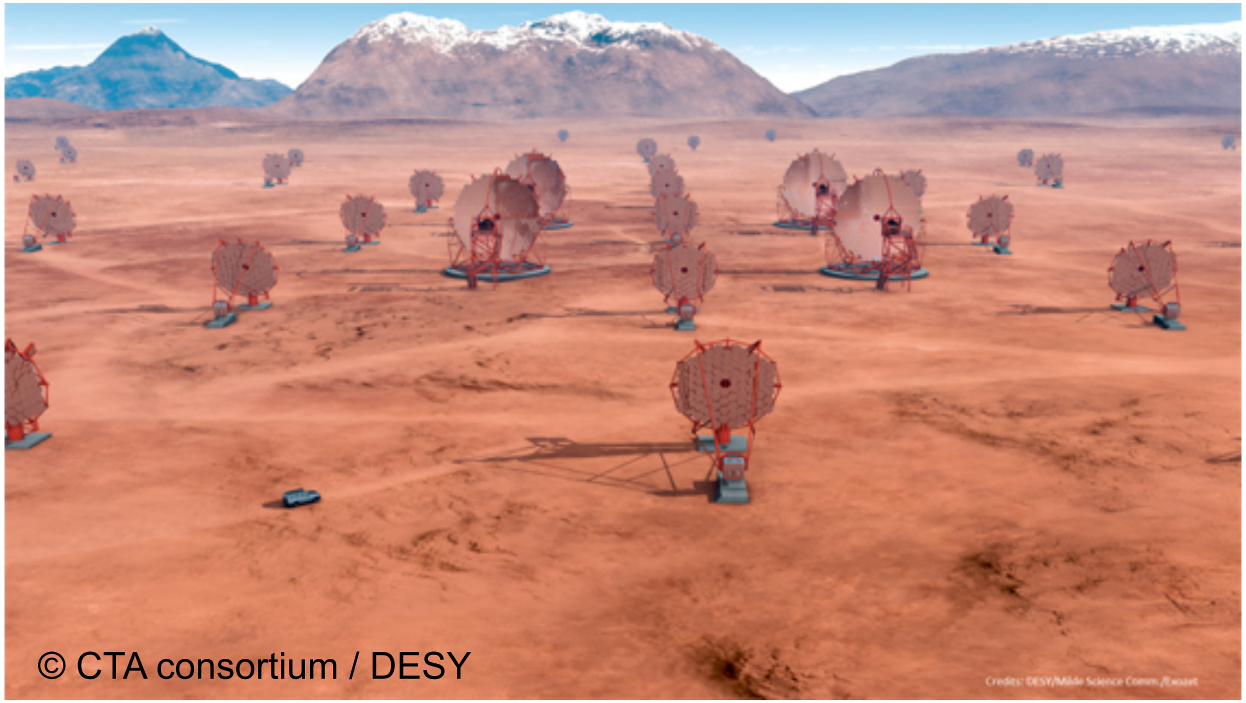
Graphic: Rodolfo Canestrari

© HESS collaboration



Big-Data tomorrow: CTA

- Ground based gamma-ray telescope
- A few 24m telescopes (20-100 GeV), Tb/s
- 10m-15m telescopes, 100 m spacing (100 GeV – 10 TeV)
- Many small size telescopes >10 TeV
- 1200 members, 200 institutes in 29 countries
- First telescopes next year, first science data 2018



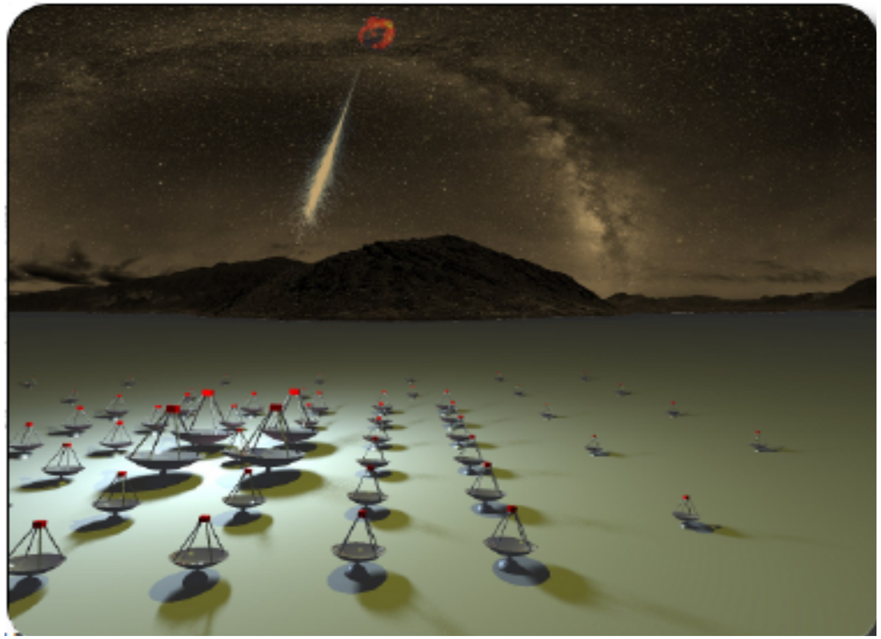
© CTA consortium / DESY

Credits: DESY/Milde Science Comm./Exozet

V. Beckmann

CTA in a nutshell

- CTA (Cherenkov Telescope Array) is the next generation instrument in VHE gamma-ray astronomy
- 2 arrays of 50-100 Cherenkov telescopes (North and South hemispheres)
- 10x sensitivity with respect to current experiments
- Consortium of ~1200 scientists in 32 countries
- Operate as an observatory



- Site locations decided for further negotiations this year:
  - North site: La Palma, Spain
  - South site: Paranal ESO, Chile
- Currently in 'Pre-construction' phase (2015-2022)
- Operations will last ~30 years

Scientific goals:

- Cosmic rays origins
- High Energy astrophysical phenomena
- Fundamental physics and cosmology

CTA Computing: data volume

Raw-data rate

- CTA South: 5.4 GB/s
- CTA North: 3.2 GB/s

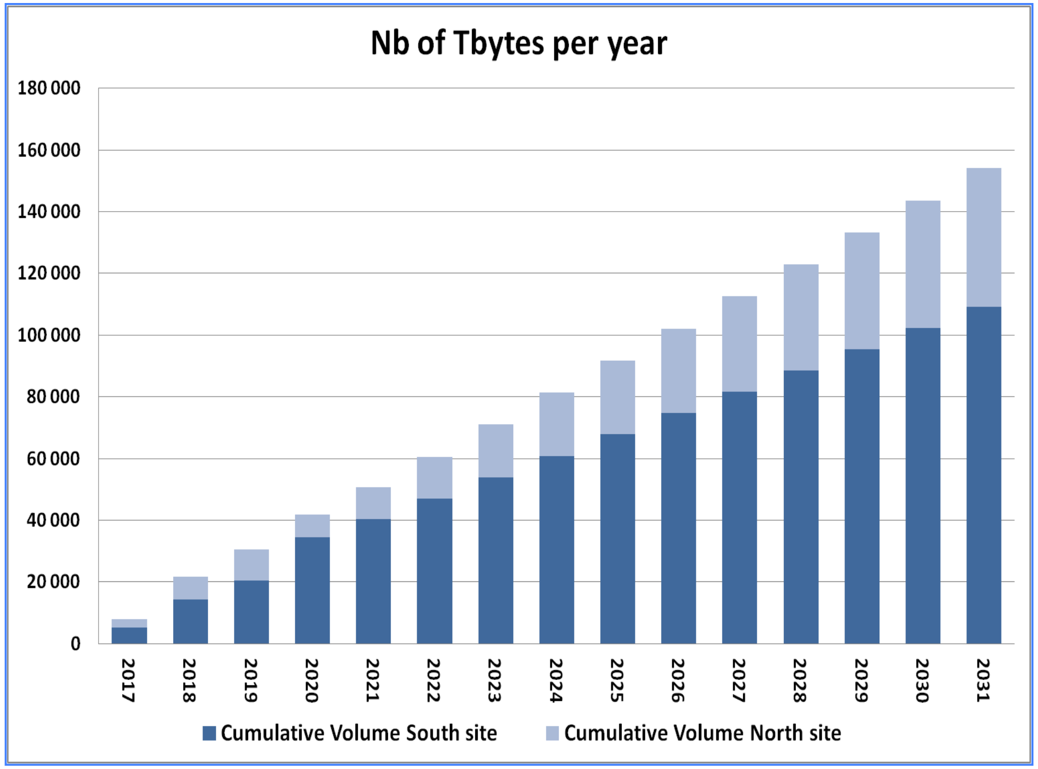
1314 hours of observation per year

Raw-data volume

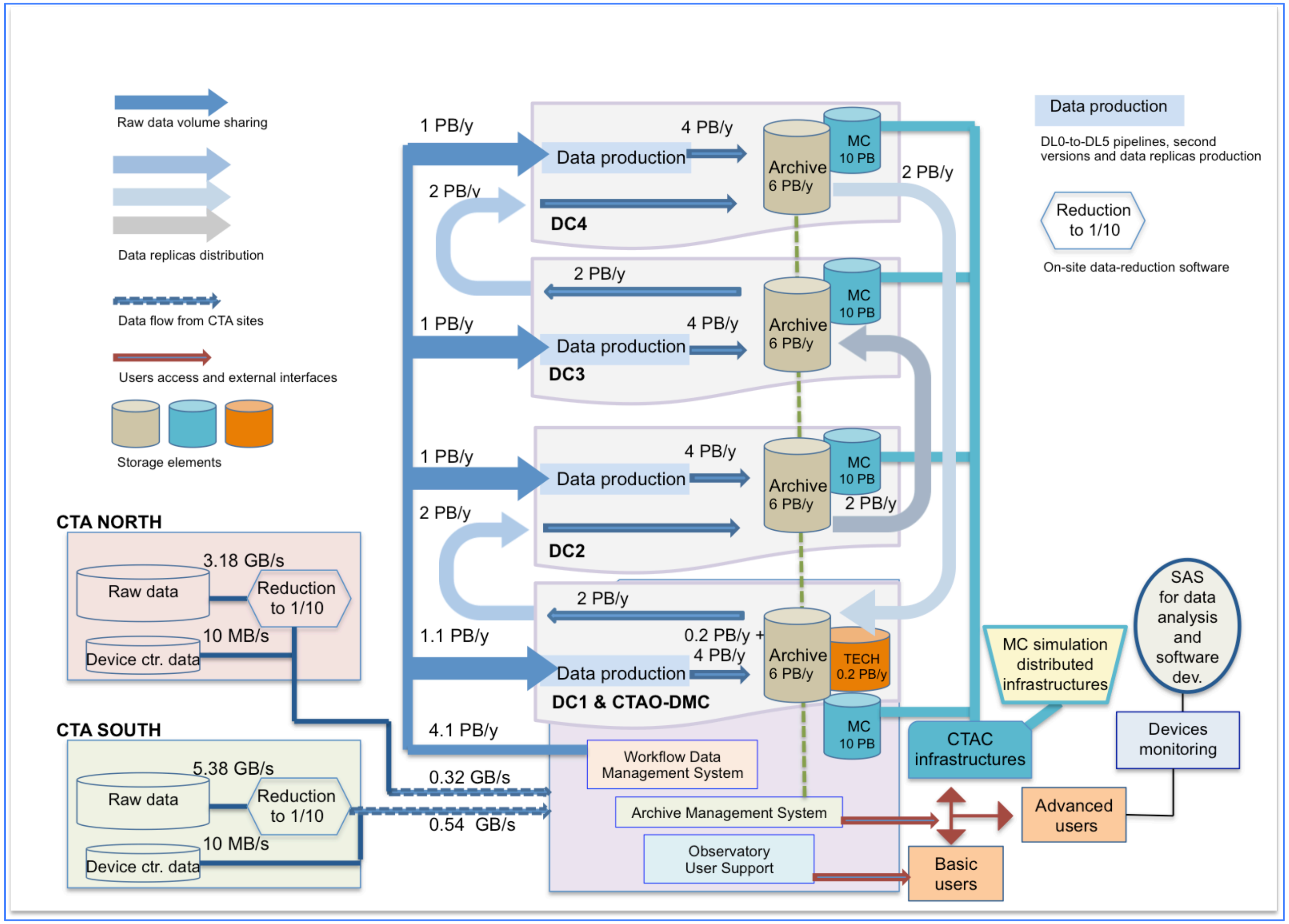
- ~40 PB/year
- ~4 PB/year after reduction

Total volume

- ~27 PB/year including calibrations, reduced data and all copies



CTA Computing: data flow



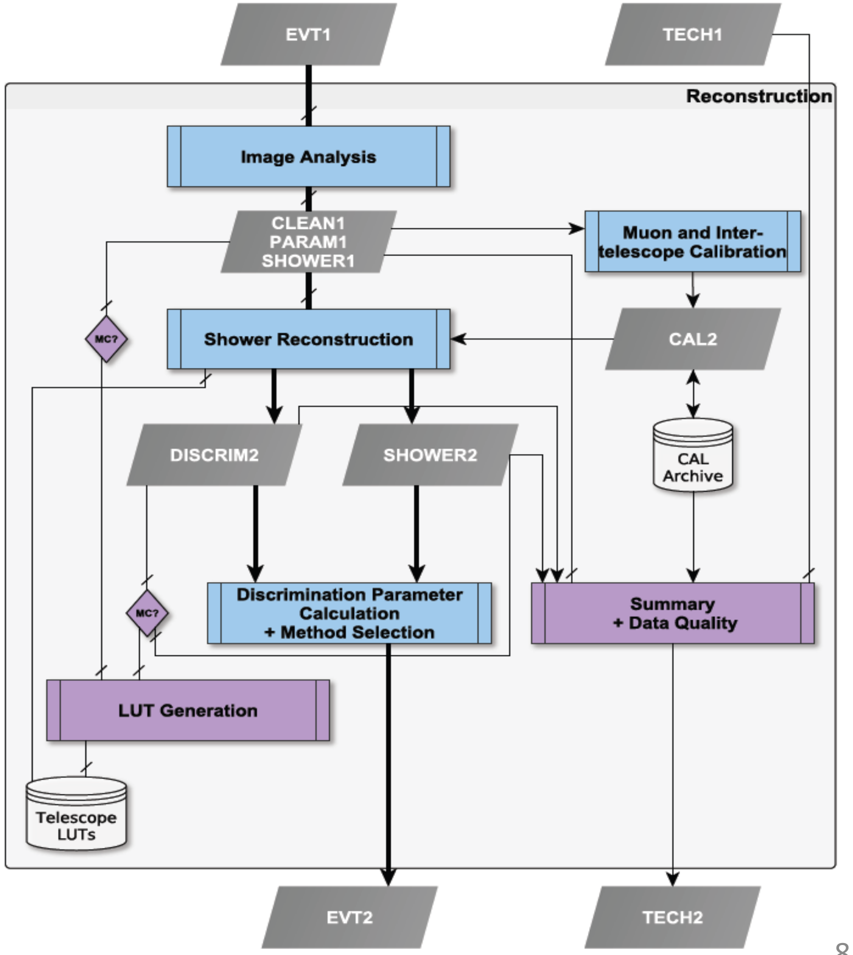
V. Beckmann

CTA Computing: data processing

From raw-data to high level science data:

- Several complex pipelines (Calibration, Reconstruction, Analysis, MC, etc.)
- Assume 1 full re-processing per year

Reconstruction Pipeline shower reconstruction step



MC simulations and Analysis

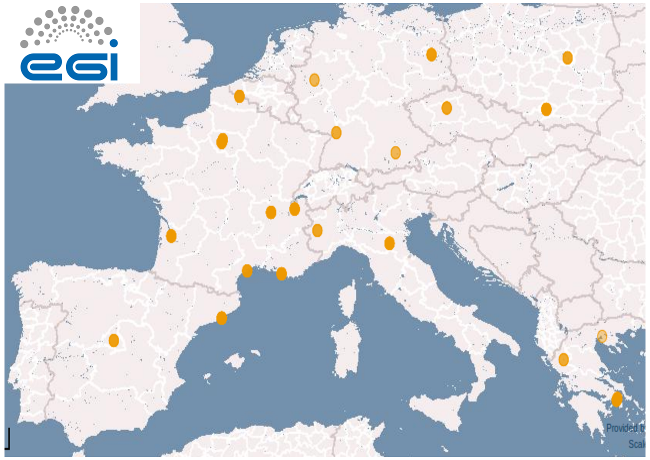
- CTA is now in 'Pre-construction' phase
- Massive MC simulations and Analysis running for 3 years
  - 'Prod2' (2013-2014)
    - Characterize all site candidates to host CTA telescopes to determine the one giving the best instrument response functions
    - 4.6 billion events generated for each site candidate, 2 different zenith angles and 3 telescope configurations
    - 8 full MC campaigns (5 sites for the South and 3 for the North)
  - 'Prod3' (2015) in progress:
    - For the 2 selected sites: study the different possible layouts of telescope arrays, pointing configurations, hardware configurations, etc.
    - 800 telescope positions, 7 telescope types, multiple possible layouts, 5 different scaling
    - Run 3 different Analysis chains on the simulated data
      - Each one processing about 500 TB and 1 M of files for 36 different configurations
- Computing is already a challenge!

MC simulations and Analysis

- Use of an existing and reliable e-Infrastructure: EGI grid
- Use of DIRAC for Workload and Data Management

CTA Virtual Organization:

- Active since 2008
- 19 EGI sites in 7 countries and 1 ARC site
- About 100 members

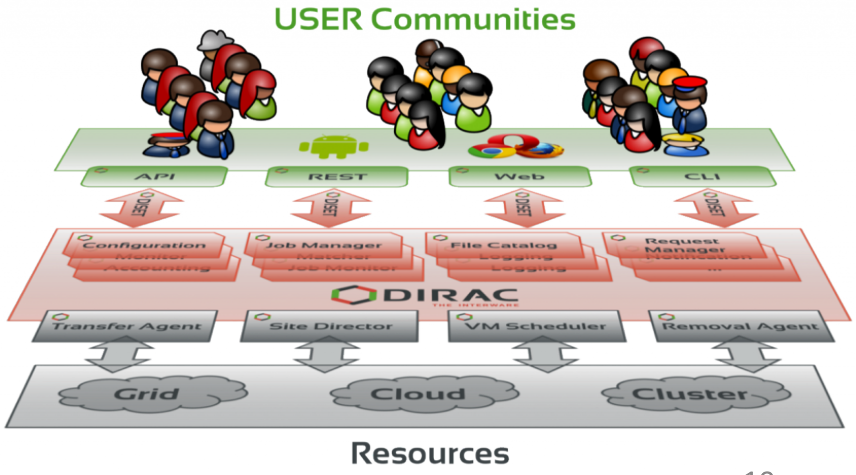


Resources:

- Dedicated storage:
  - Disk: 1.3 PB in 6 sites:
    CC-IN2P3, CNAF, CYFRONET, DESY, GRIF, LAPP
  - Tape: 400 TB in 3 sites
- CPU: 8000 cores available on average

DIRAC for CTA:

- Dedicated DIRAC instance composed of 4 main servers at CC-IN2P3 and PIC
- Several DIRAC Systems in use
- CTA-DIRAC software extension



MC simulations and Analysis

Computing Model:

- Use of CTA-DIRAC instance to access grid resources
- MC simulation uses CPU resources of all 20 grid sites supporting the CTA VO
- Output Data are stored at 6 main SE
- MC Analysis takes place at the 6 sites where MC data are stored

Computing Operations (small team of people):

- Receive production requests from MC WP (nb of events to be generated, sw to install, etc.)
- Adjust the requests according to the available resources
- Negotiate resources with grid sites on a yearly basis
- Run the productions and perform data management operations (removal of obsolete datasets, data migration, etc.)
- Support users to run their private MC productions and analysis
- DIRAC servers administration
- Development of CTA-DIRAC extension

DIRAC for CTA: main Systems in use

- Workload Management System
  - Job brokering and submission (pilot mechanism)
  - Integration of hetereogenous resources (CREAM, ARC)
  - Central management of CTA VO policies
- Data Management System
  - All data operations (download, upload, replication, removal)
  - Use of the DIRAC File Catalog (DFC) as Replica and Metadata catalog
- Transformation System
  - Used by production team to handle 'repetitive' work (many identical tasks with a varying parameter), i.e. MC productions, MC Analysis, data management operations (bulk removal, replication, etc.)

DIRAC for CTA: DIRAC File Catalog

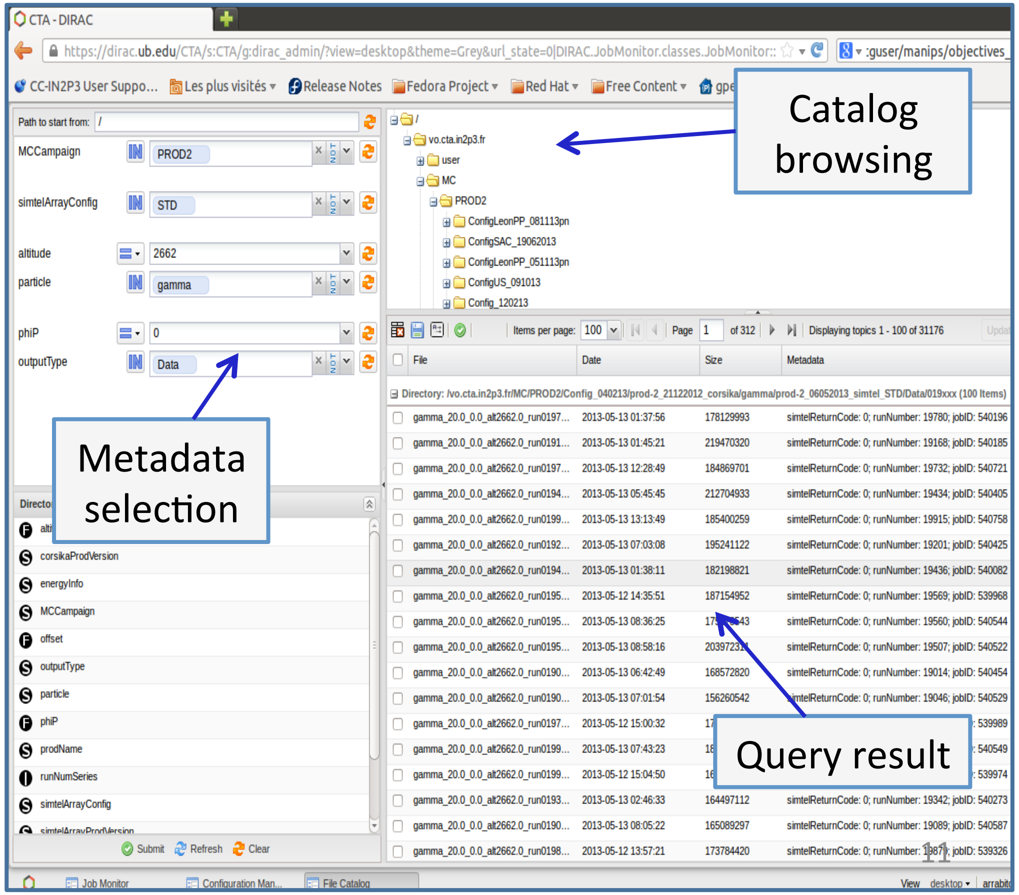
- In use since 2012 in parallel with LFC. Full migration to DFC in summer 2015
- More than 21 M of replicas registered
- About 10 meta-data defined to characterize MC datasets

Query example:

cta-prod3-query --site=Paranal --particle=gamma --tel_sim_prog=simtel --array_layout=hex --phiP=180 --thetaP=20 --outputType=Data

Typical queries return several hundreds of thousands of files

DFC web interface



CTA - DIRAC

https://dirac.ub.edu/CTA/s:CTA/g:dirac_admin/?view=desktop&theme=Grey&url_state=0|DIRAC.JobMonitor.classes.JobMonitor::

Path to start from: /

MCCampaign	IN	PROD2
simtelArrayConfig	IN	STD
altitude	=	2662
particle	IN	gamma
phiP	=	0
outputType	IN	Data

Catalog browsing

```
/
  vo.cta.in2p3.fr
    user
    MC
      PROD2
        ConfigLeonPP_081113pn
        ConfigSAC_19062013
        ConfigLeonPP_051113pn
        ConfigUS_091013
        Config_120213
```

Metadata selection

Directo…
- F alti…
- S corsikaProdVersion
- S energyInfo
- S MCCampaign
- F offset
- S outputType
- S particle
- F phiP
- S prodName
- I runNumSeries
- S simtelArrayConfig
- S simtelArrayProdVersion

Items per page: 100 Page 1 of 312 Displaying topics 1 - 100 of 31176

Directory: /vo.cta.in2p3.fr/MC/PROD2/Config_040213/prod-2_21122012_corsika/gamma/prod-2_06052013_simtel_STD/Data/019xxx (100 Items)

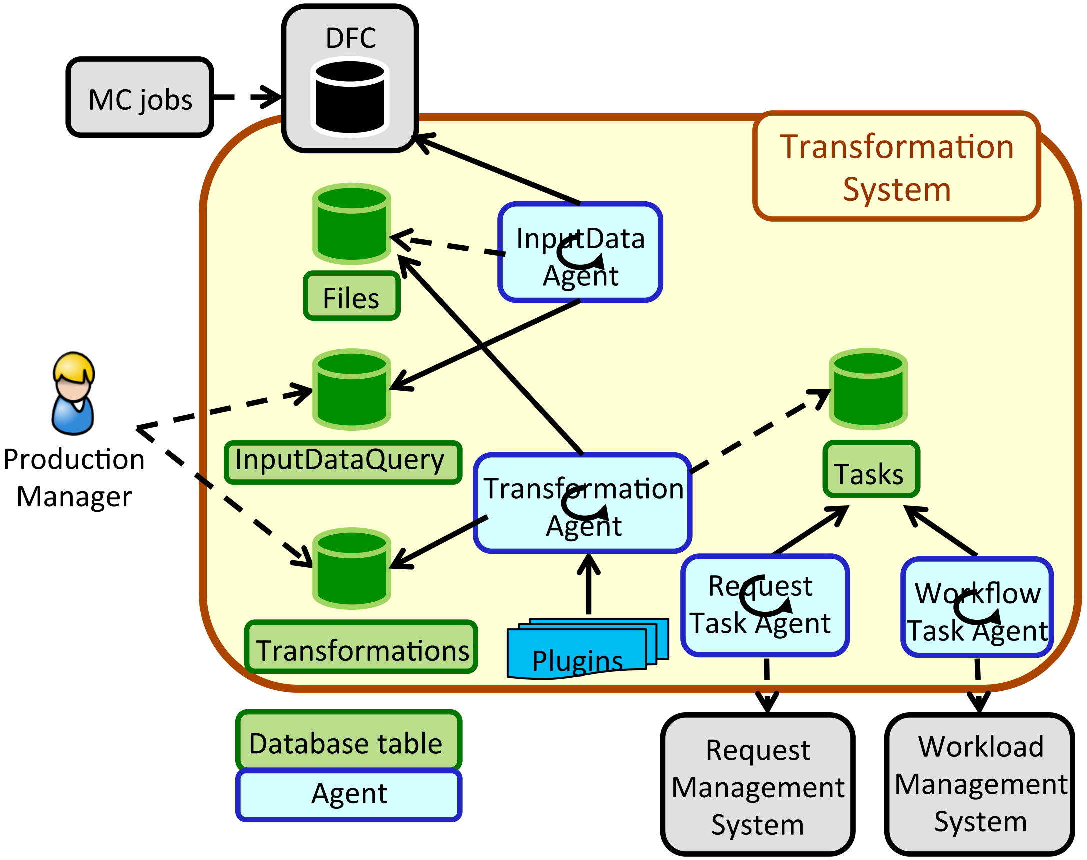
File	Date	Size	Metadata
gamma_20.0_0.0_alt2662.0_run0197...	2013-05-13 01:37:56	178129993	simtelReturnCode: 0; runNumber: 19780; jobID: 540196
gamma_20.0_0.0_alt2662.0_run0191...	2013-05-13 01:45:21	219470320	simtelReturnCode: 0; runNumber: 19168; jobID: 540185
gamma_20.0_0.0_alt2662.0_run0197...	2013-05-13 12:28:49	184869701	simtelReturnCode: 0; runNumber: 19732; jobID: 540721
gamma_20.0_0.0_alt2662.0_run0194...	2013-05-13 05:45:45	212704933	simtelReturnCode: 0; runNumber: 19434; jobID: 540405
gamma_20.0_0.0_alt2662.0_run0199...	2013-05-13 13:13:49	185400259	simtelReturnCode: 0; runNumber: 19915; jobID: 540758
gamma_20.0_0.0_alt2662.0_run0192...	2013-05-13 07:03:08	195241122	simtelReturnCode: 0; runNumber: 19201; jobID: 540425
gamma_20.0_0.0_alt2662.0_run0194...	2013-05-13 01:38:11	182198821	simtelReturnCode: 0; runNumber: 19436; jobID: 540082
gamma_20.0_0.0_alt2662.0_run0195...	2013-05-12 14:35:51	187154952	simtelReturnCode: 0; runNumber: 19569; jobID: 539968
gamma_20.0_0.0_alt2662.0_run0195...	2013-05-13 08:36:25	175[illegible]543	simtelReturnCode: 0; runNumber: 19560; jobID: 540544
gamma_20.0_0.0_alt2662.0_run0195...	2013-05-13 08:58:16	2039723[illegible]	simtelReturnCode: 0; runNumber: 19507; jobID: 540522
gamma_20.0_0.0_alt2662.0_run0190...	2013-05-13 06:42:49	168572820	simtelReturnCode: 0; runNumber: 19014; jobID: 540454
gamma_20.0_0.0_alt2662.0_run0190...	2013-05-13 07:01:54	156260542	simtelReturnCode: 0; runNumber: 19046; jobID: 540529
gamma_20.0_0.0_alt2662.0_run0197...	2013-05-12 15:00:32	17[illegible]	[illegible]: 539989
gamma_20.0_0.0_alt2662.0_run0199...	2013-05-13 07:43:23	18[illegible]	[illegible]: 540549
gamma_20.0_0.0_alt2662.0_run0199...	2013-05-12 15:04:50	16[illegible]	[illegible]: 539974
gamma_20.0_0.0_alt2662.0_run0193...	2013-05-13 02:46:33	164497112	simtelReturnCode: 0; runNumber: 19342; jobID: 540273
gamma_20.0_0.0_alt2662.0_run0190...	2013-05-13 08:05:22	165089297	simtelReturnCode: 0; runNumber: 19089; jobID: 540587
gamma_20.0_0.0_alt2662.0_run0198...	2013-05-12 13:57:21	173784420	simtelReturnCode: 0; runNumber: 1987[illegible]; jobID: 539326

Query result

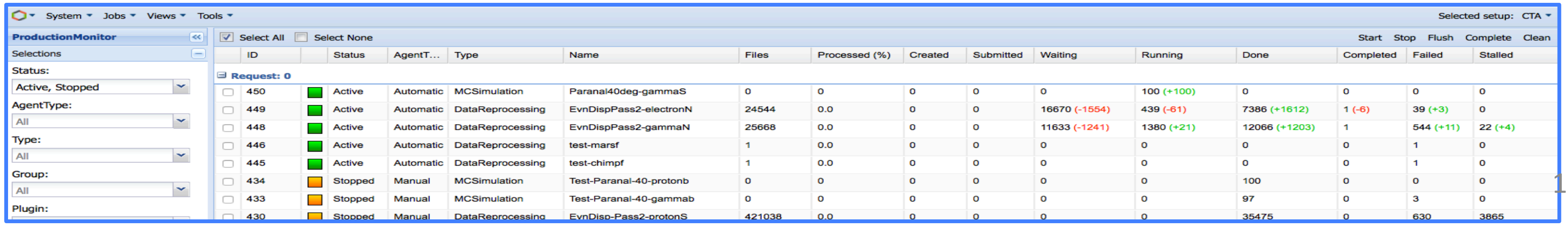
DIRAC for CTA: Transformation System

Transformation System Architecture

- The Production Manager defines the transformations with meta-data conditions and ‘plugins’
- InputData Agent queries the DFC to obtain files to be ‘transformed’
- Plugins group files into tasks according to desired criteria
- Tasks are created and submitted to the Workload or Request Management System



Transformation Monitoring



Status: Active, Stopped

AgentType: All

Type: All

Group: All

Plugin:

Request: 0

	ID		Status	AgentT...	Type	Name	Files	Processed (%)	Created	Submitted	Waiting	Running	Done	Completed	Failed	Stalled
□	450	🟩	Active	Automatic	MCSimulation	Paranal40deg-gammaS	0	0	0	0	0	100 (+100)	0	0	0	0
□	449	🟩	Active	Automatic	DataReprocessing	EvnDispPass2-electronN	24544	0.0	0	0	16670 (-1554)	439 (-61)	7386 (+1612)	1 (-6)	39 (+3)	0
□	448	🟩	Active	Automatic	DataReprocessing	EvnDispPass2-gammaN	25668	0.0	0	0	11633 (-1241)	1380 (+21)	12066 (+1203)	1	544 (+11)	22 (+4)
□	446	🟩	Active	Automatic	DataReprocessing	test-marsf	1	0.0	0	0	0	0	0	0	1	0
□	445	🟩	Active	Automatic	DataReprocessing	test-chimpf	1	0.0	0	0	0	0	0	0	1	0
□	434	🟧	Stopped	Manual	MCSimulation	Test-Paranal-40-protonb	0	0	0	0	0	0	100	0	0	0
□	433	🟧	Stopped	Manual	MCSimulation	Test-Paranal-40-gammab	0	0	0	0	0	0	97	0	3	0
□	430	🟧	Stopped	Manual	DataReprocessing	EvnDisp-Pass2-protonS	421038	0.0	0	0	0	0	35475	0	630	3865

Resource usage: 2013-2015

- Use of about 20 grid sites
- 5000-8000 concurrent jobs for several weeks
- Users analysis also running in parallel (private simulations, analysis)
- More than 7.7 M jobs executed

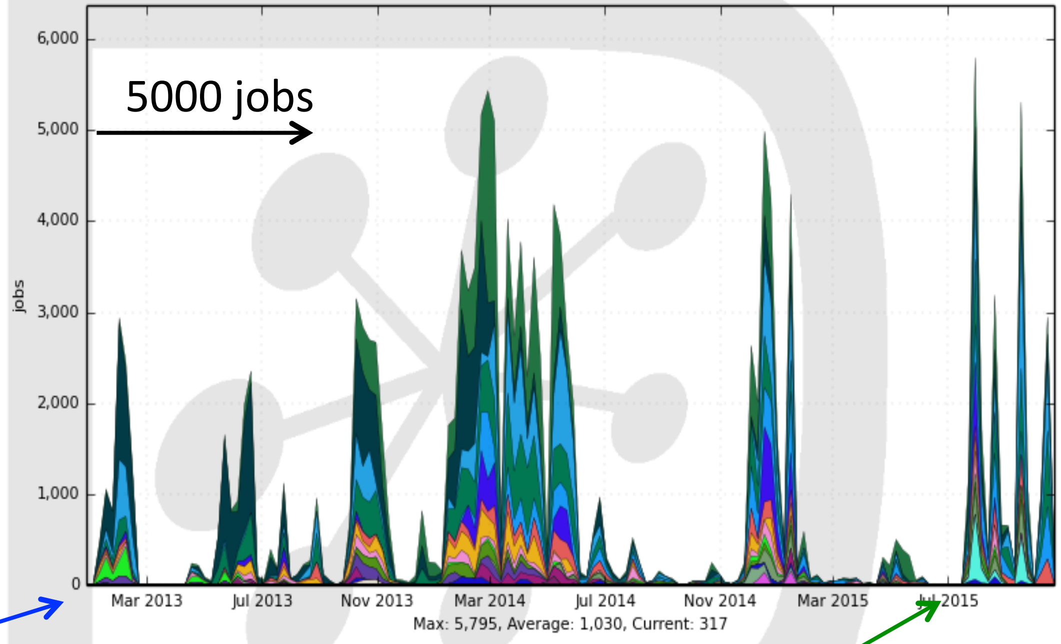
Prod2

- 148.56 M HS06 hours
- 640 TB

Prod3

- 28.8 M HS06 hours
- 785 TB

Running jobs by site



Max: 5,795, Average: 1,030, Current: 317

Start prod2

El Leoncito

SAC

Aar

US

SPM

Tenerife

Paranal

Start prod3

MC Paranal

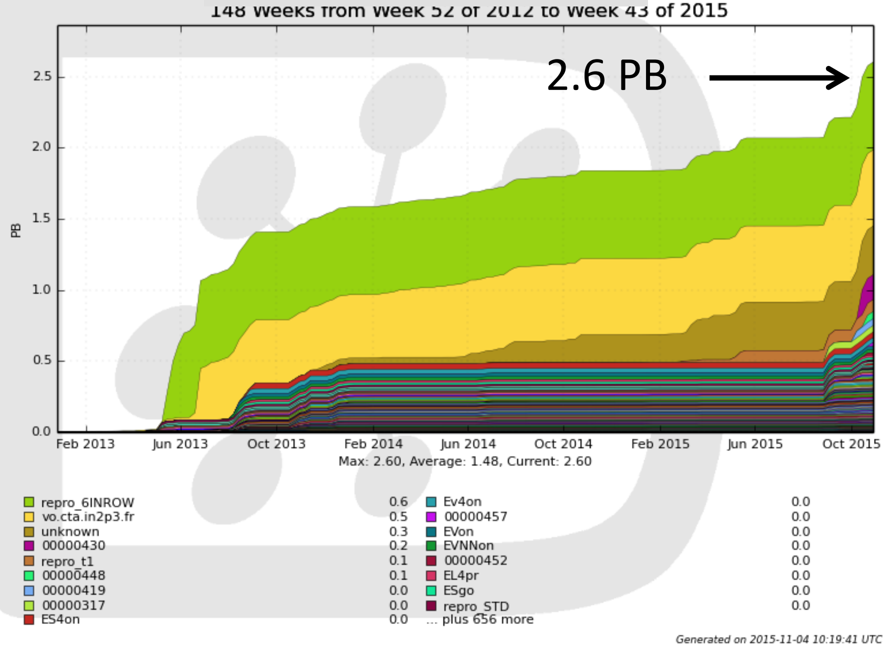
MC Analysis

Users Analysis

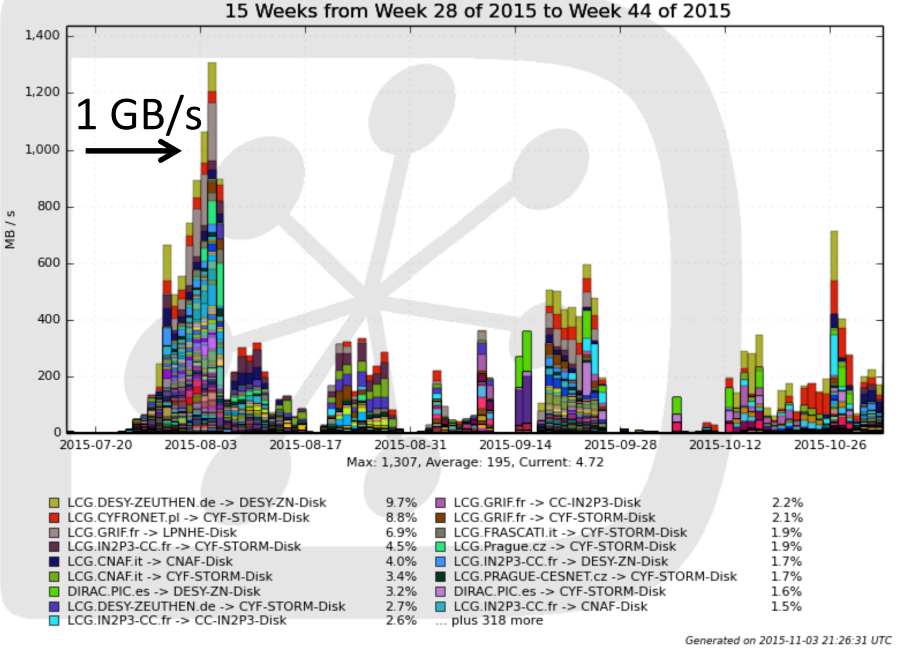
Resource usage: 2013-2015

- About 2.6 PB processed (MC Analysis)
- Throughput of 400-800 MB/s during ‘prod3’

Processed data



Throughput



V. Beckmann

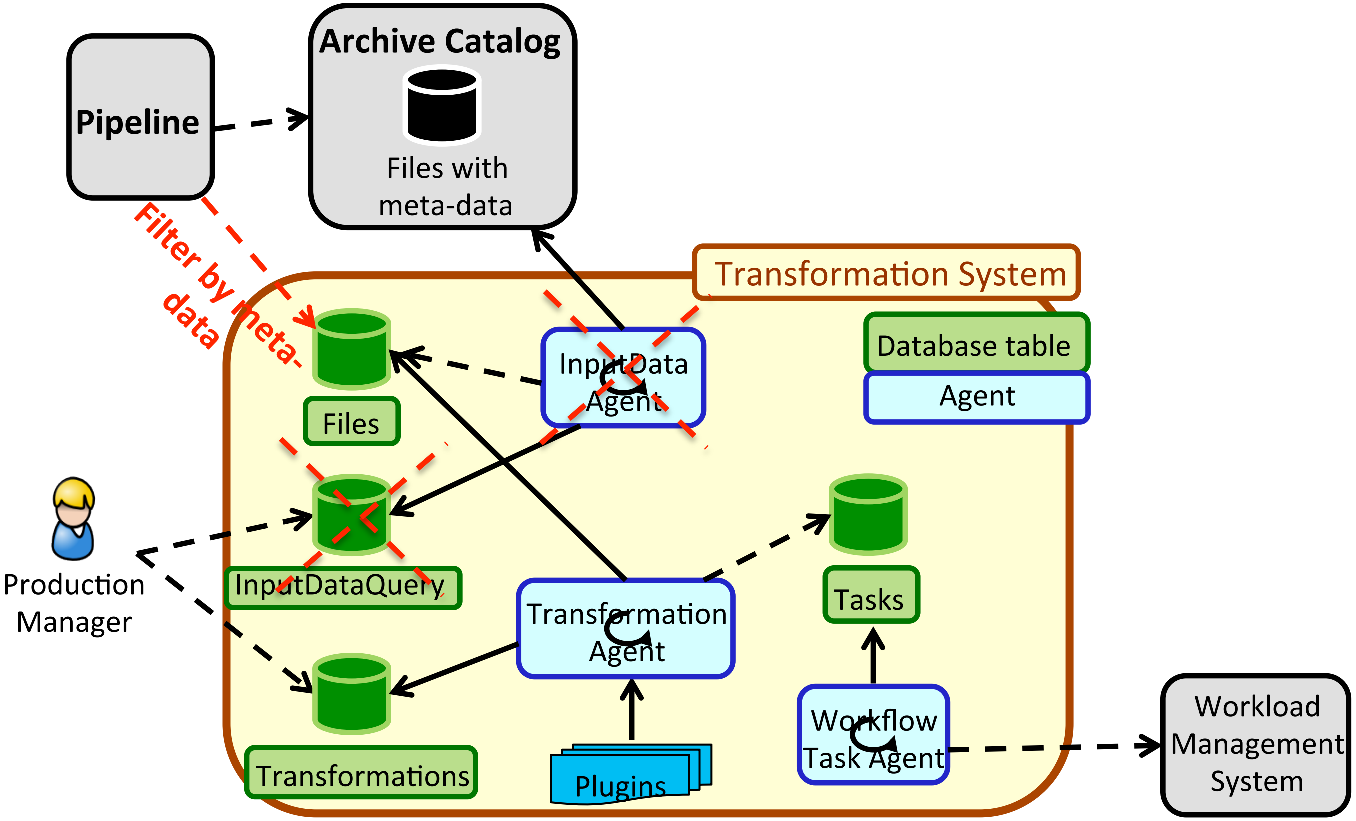
Future developments

- Until now, we have been using almost all DIRAC functionalities

- For long term operations CTA is developing several systems:
  - Archive system to store and manage CTA data
  - Pipeline framework to handle all CTA applications
  - Data model to define the whole meta-data structure of CTA data
  - Science gateway for end-users data access
  - DIRAC will be used for the Workload and Production Management

- Future developments:
  - Develop interfaces between DIRAC and the other CTA systems
  - Improvement of the DIRAC Transformation System toward a fully data-driven system (next slide)
  - Improve CTA-DIRAC hardware setup

Toward a fully data-driven Transformation System



Developments in progress:

- When new files are registered, a filter based on meta-data is applied to send them on the fly to their matching transformation
- No need anymore to perform heavy queries of the Archive

Improve CTA-DIRAC hardware setup

- DIRAC is based on a Service Oriented Architecture
- Each DIRAC System is composed of Services, Agents and DBs
- CTA-DIRAC instance is a rather modest installation, composed of:
  - 3 core servers:
    - 1 server running all Services (except DM) and a few Agents (4 cores)
    - 1 server running all Agents except TS (2 cores)
    - 1 server running the DataManagement System and Transformation Agents (16 cores)
  - 2 MySQL servers:
    - 1 server hosting all DBs except FileCatalogDB
    - 1 server hosting the FileCatalogDB
  - 1 web server
- Observed high load on the core servers when running several 'productions' in parallel
- Need to add more servers, but also optimize the component distribution on the servers

Conclusions

- CTA will produce about 27 PB/year and it will operate for ~30 years
- A production system prototype based on DIRAC has been developed for MC simulation productions:
  - Required minimal development of a CTA-DIRAC extension
  - Minimal setup and maintenance of the instance running all the components, mostly the work of 1 person with help of DIRAC team and sys admins upon request
- Successfully used during last 3 years for massive MC simulations and analysis:
  - > 1.6 PB produced
  - ~ 2.6 PB processed
  - > 20 M files registered in the Catalog
- Future developments:
  - Improve production system for real data processing pipeline
    - Build interfaces between DIRAC and other CTA systems (Archive, Pipeline, etc.)
    - Further develop DIRAC Transformation System toward a fully data-driven system

V. Beckmann