



# Grid – Introduction

ATLAS-D Physics Meeting Freiburg 2018

Dr. Gen Kawamura

II.Physikalisches Institut, Universität Göttingen

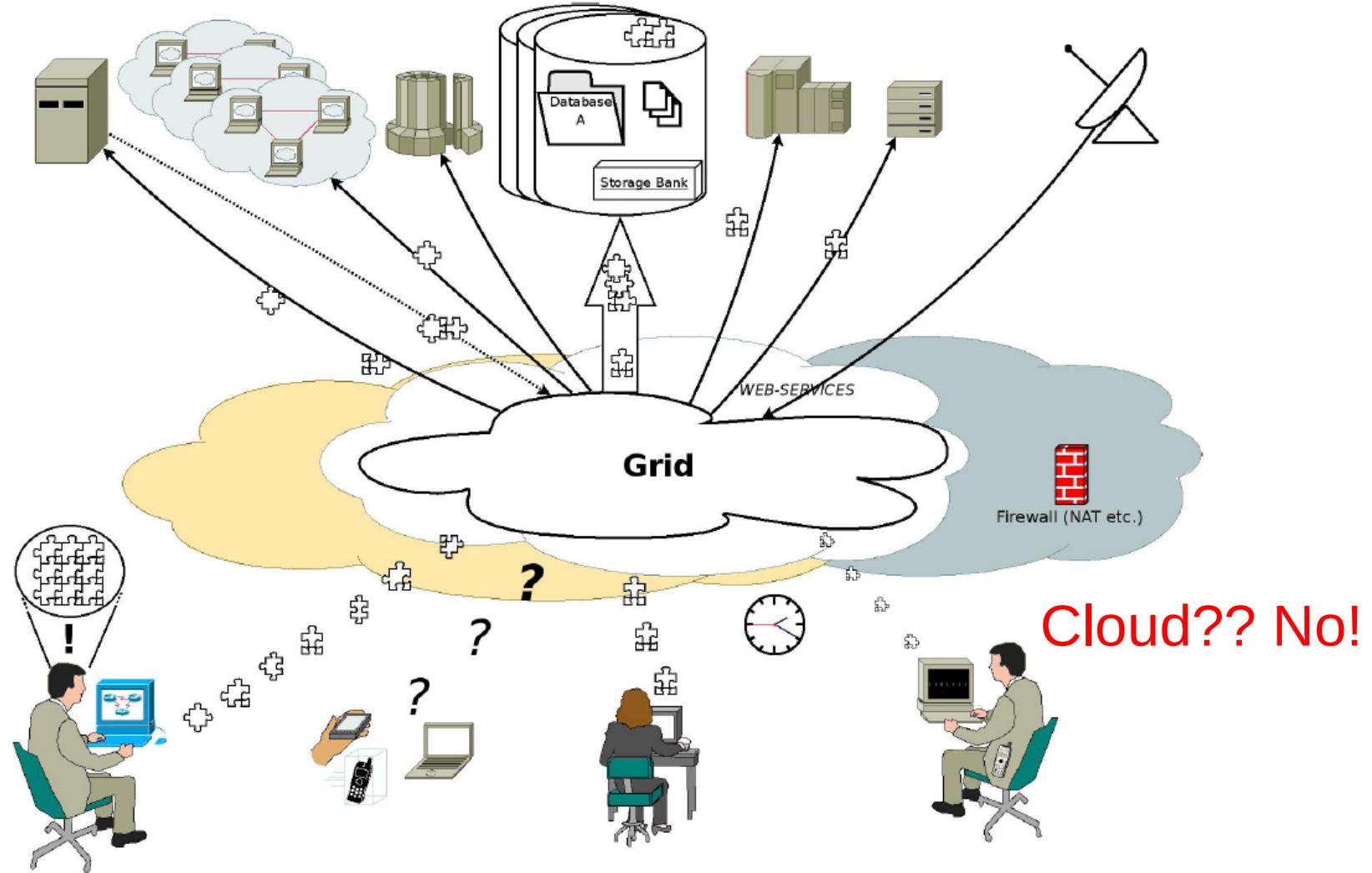
# Overview (30mins)

- Introduction to ATLAS Grid computing
  - Concepts
  - Certificate Authorities and VOMS
  - ATLAS Grid computing & WLCG Resources
  - Grid job
  - Grid user interface (CLI) and CVMFS
- PanDA (ATLAS job management system)
- ATLAS Metadata Interface (AMI)
- Rucio (ATLAS data management system)
  - Rucio basic concept
  - RSE expressions
- Links ad references
- Bakckup
  - ATLAS Resources
  - Job allocation (a tip)

# Introduction to ATLAS Grid Computing



# Introduction to Grid computing - 1



# Introduction to Grid computing - 2

- Data intensive Physical Sciences
  - High Energy & Nuclear Physics
    - Including LHC experiments at CERN
  - Gravitational wave searches
    - LIGO
  - Time-dependent 3D systems
    - Earth observation, climate modelling
    - Geophysics, earthquake modelling
    - Fluid dynamics
    - Bioinformatics, Protein simulations
  - Astronomy

# Introduction to Grid computing - 3

- A comparison

- Serial
  - Fetch/Store, Compute
- Parallel
  - Fetch/Store
  - Compute/Communicate
  - Cooperative

- *Grid*

- *Geographically independent based on internet backbone*
- *Fetch/Store (from remote data)*
- *Resource discovery*
- *Interaction with remote applications*
- *Authentication/Authorization/Standard Security*
- *Built around internet standard*
- *Compute in parallel*

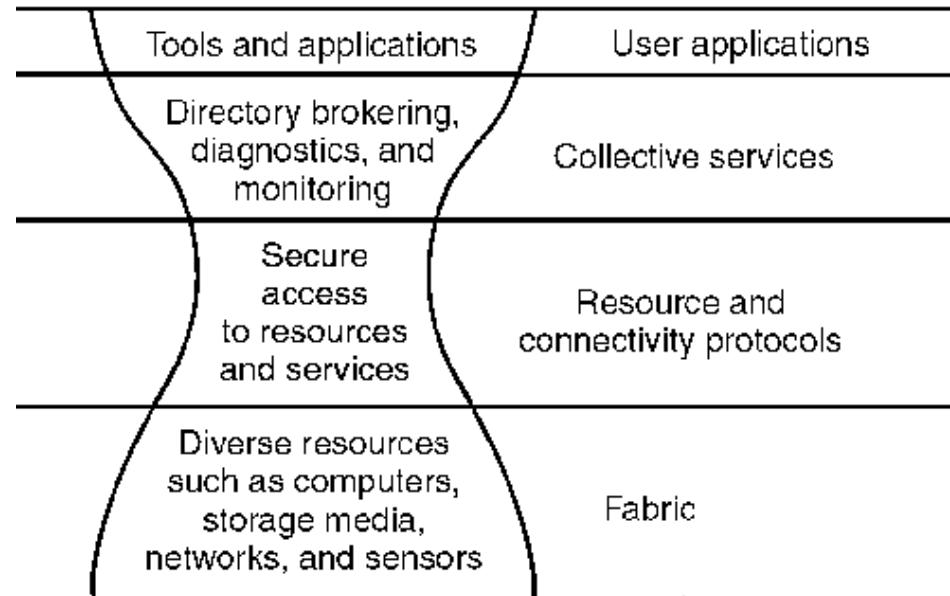


Fig. Grid Hourglass model

# Introduction to Grid computing - 4

- Grid
  - Enhancement of possibilities of the Internet to the area of high capacity computations and distributed data management
  - Ian Foster: three point checklist (2002):
    - No central administration of computing resources
    - Open standards are used
    - Non trivial quality of service
  - IBM
    - A Grid is a type of parallel and distributed system that enables the sharing, selection, and aggregation of resources distributed across multiple administrative domains based on the resources availability, capacity, performance, cost and users' quality-of-service requirements"

# Introduction to Grid computing - 5

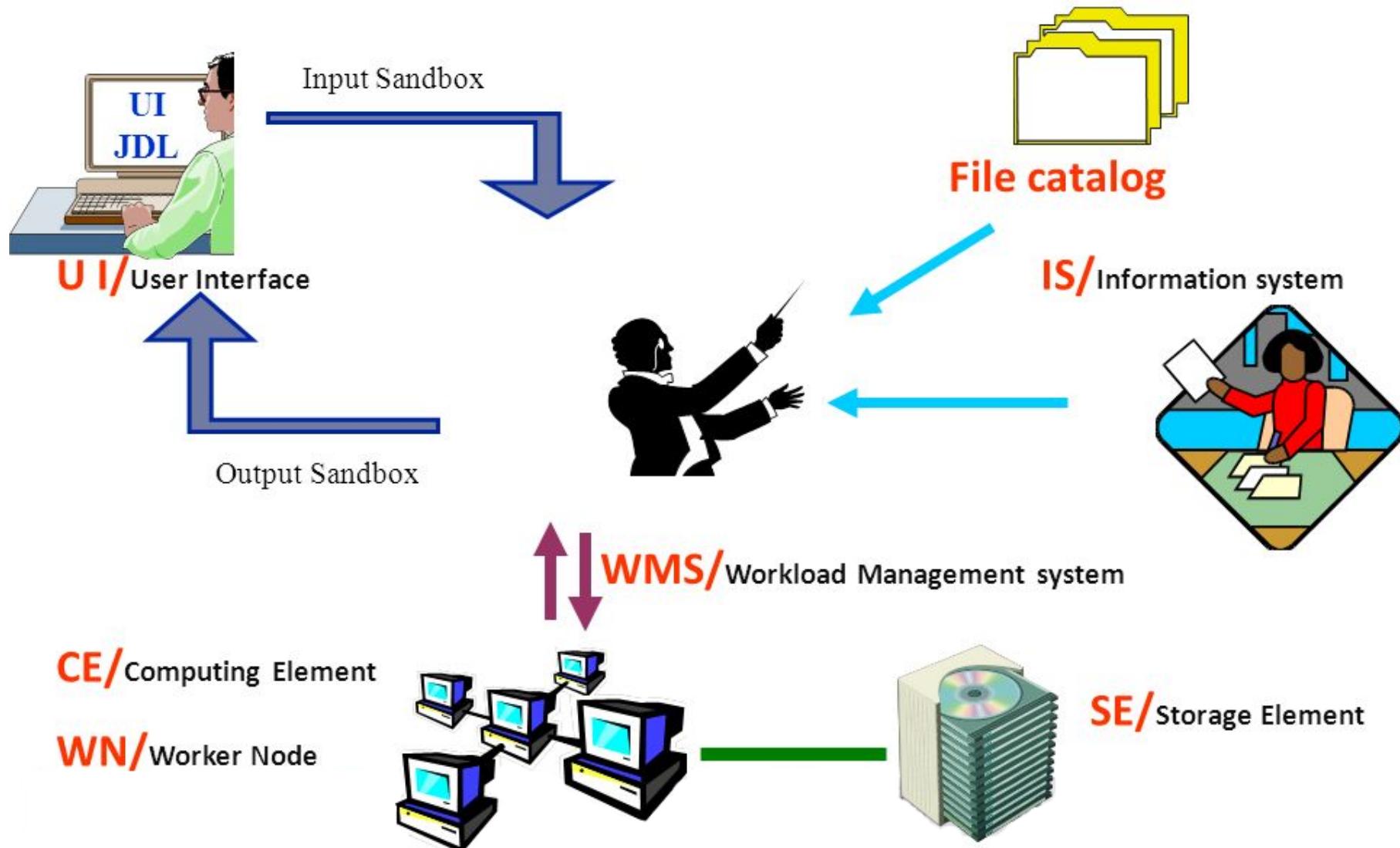
- Short history
  - 2002: Ian Foster and Carl Kesselman (eds.), The Grid: Blueprint for a New Computing Infrastructure, 2nd edition, Morgan Kaufmann Publishers,
  - 2004: middleware developent: Globus, Legion, Condor, NWS, SRB, NetSolve, *gLite*, etc
  - 2000+: start of the Global Grid
    - Global Grid Forum
    - Almost standards (Globus, Condor, ...)
  - 2006: GGF (Global Grid Forum), EGA (Enterprice Grid Alliance (launch 2004))
    - OGF (Open Grid Forum)

# Introduction to Grid computing - 6

- Classification of Grid
  - Based on Globus toolkit
    - Globus toolkit
    - gLite (developed for LCG – LHC Computing Grid)
    - NorduGrid
    - TeraGrid
    - UNICORE
  - Cloud computing
    - Amazon
    - Google
    - Yahoo
    - Microsoft
    - Etc...
- Workload balancing system
  - Sun Grid Engine (SGE)
  - LSF
  - OpenPBS
  - HT-Condor
- Community Grid (using idle resources, cycle stealing)
  - BOINC
  - SETI@HOME
  - LHC@HOME
- Application
  - Computational Grids, Data Grids

# Introduction to Grid computing - 6

## Job Workflow in gLite-WMS

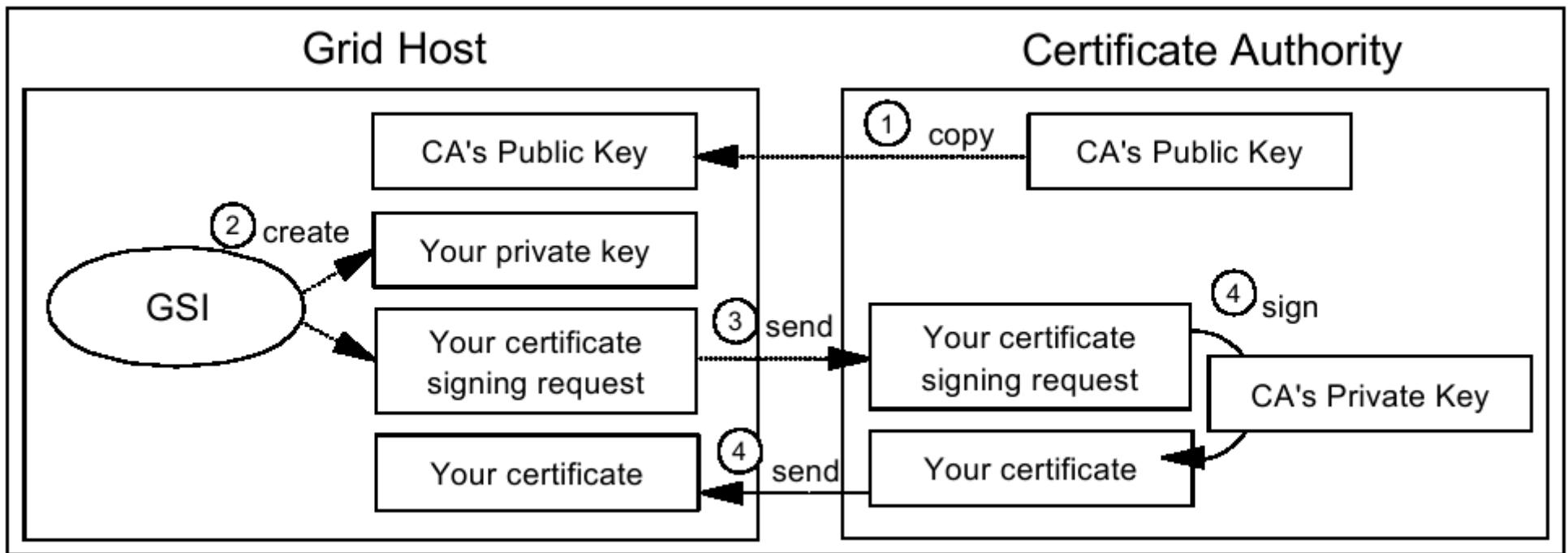


# GSI - 1

- GSI: Grid Security Infrastructure
  - Key concepts
    - PKI
    - Digital signature
    - Certificates
    - Authentication/Authorization
    - Delegation and single sign-on
  - Extra
    - Virtual Organization (VO)

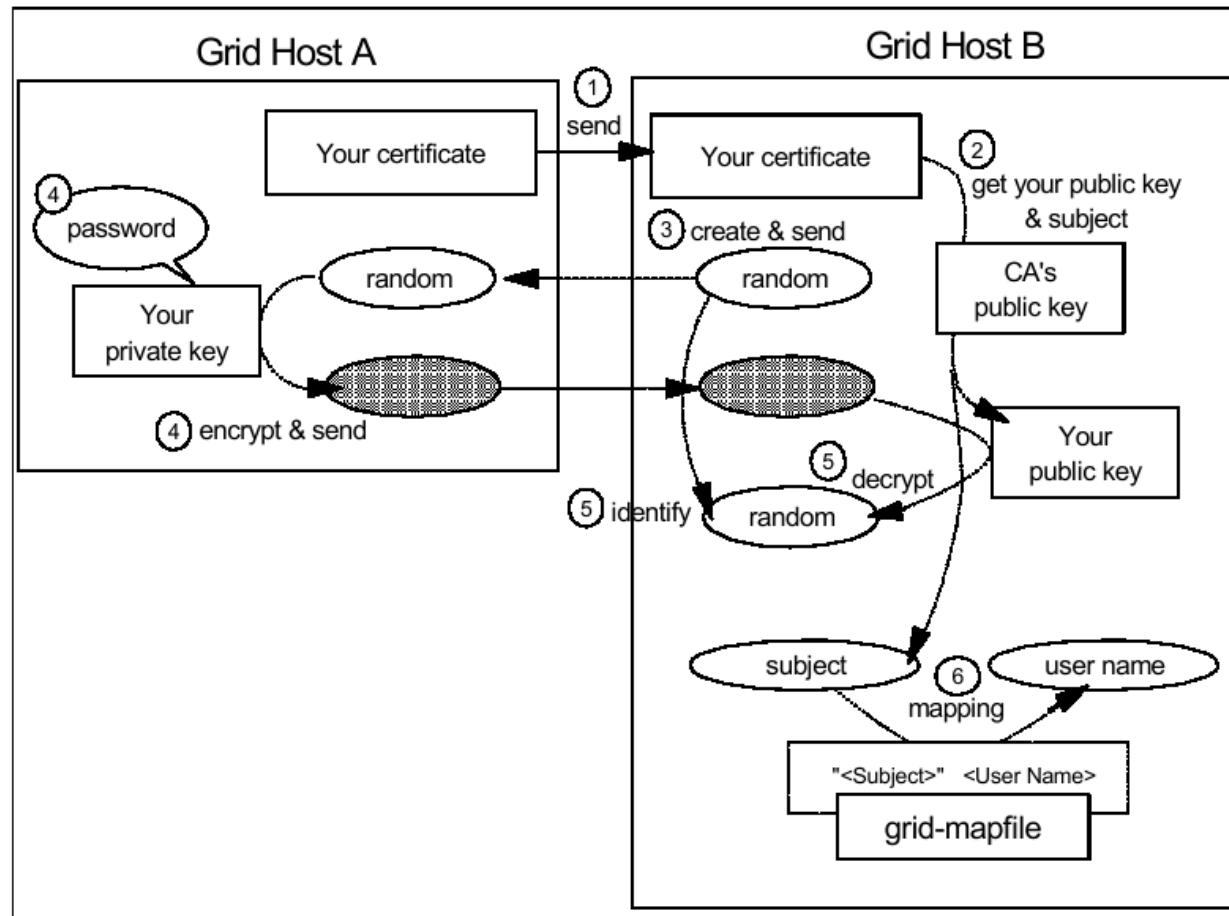
# GSI - 2

- GSI: Grid Security Infrastructure
  - Issuing → Digital signature on a PKI key



# GSI - 3

- GSI: Grid Security Infrastructure
  - Connecting to Host A from Host B



# GSI - 4

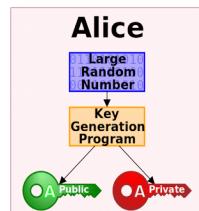
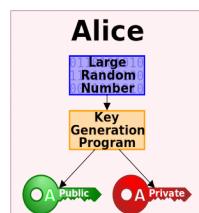
- GSI: Grid Security Infrastructure
  - Connecting to Host A from Host B
    - SSL/TLS layer from (A) to (B)
    - Forward X.509 certificate → Who I am, Which CA signed
    - Validity check through the CA public key in (B)
    - Check if (A) on the key is correct
      - (B) generates a key  $t$ , and request (A) to encrypt it
      - (A) encrypt  $t$  with its private key  $s=encrypt(t)$ , return it to (B)
      - (B) decrypt  $r=decrypt(s)$  using public key of (A)
      - If  $r == s$ , (A) is the right one
    - Vice versa (B) → (A)
    - Perform *grid user authorization* (using *grid-map*)

# GSI - 5

- GSI: Grid Security Infrastructure
  - Proxy certificate
    - Local *voms-proxy-init* command generate a PKI key pair, and Grid user certificate can sign on it

# Proxy certificate - 1

- Then, hierarchical signatures

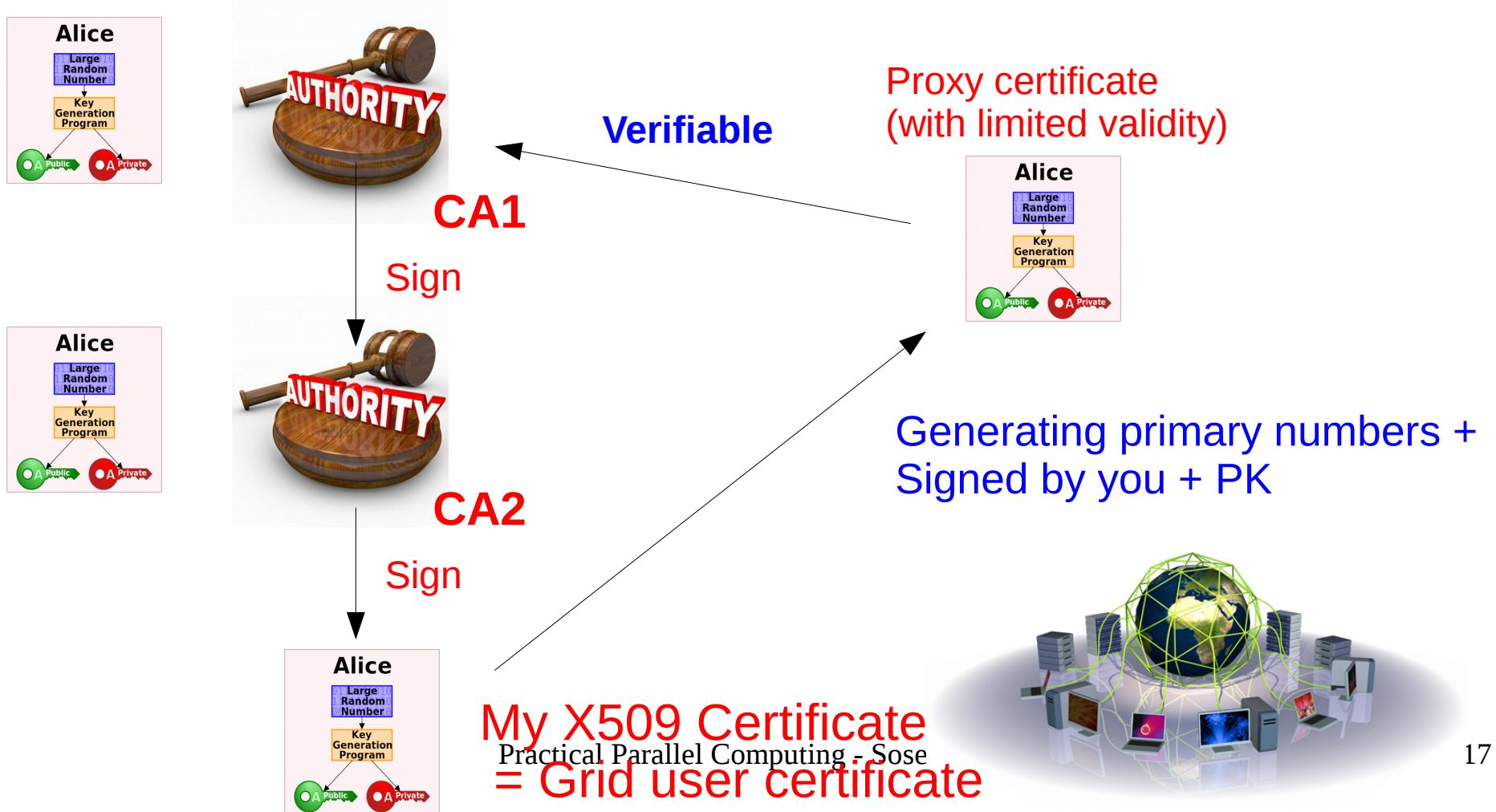


My X509 Certificate  
Practical Parallel Computing - Sose  
= Grid user certificate



# Proxy certificate - 2

- Generating a new certificate = proxy certificate



# Proxy certificate - 3

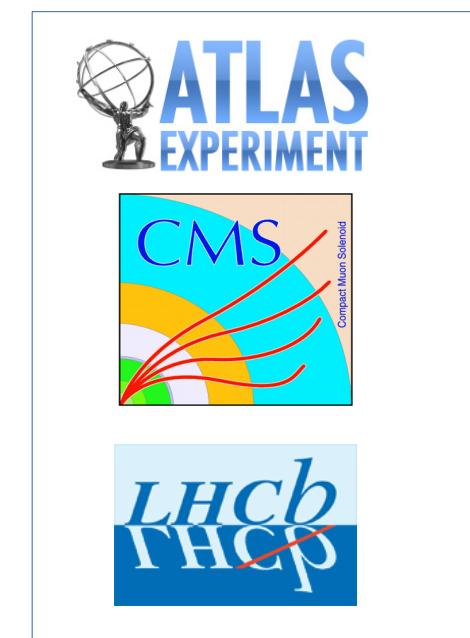
- Virtual Organization (VO), VOMS
  - VOMS extends a certificate with a VO

## Certificate Authority



VOMS server

ATLAS VOMS  
CMS VOMS  
LHCb VOMS



Tactical Parallel Computing - Sose  
= a limited copy of your certificate

# Proxy certificate - 4

- A similar analogy
  - Authenticated and authorized for your tasks

Certificate Authority



↓  
Sign



User certificate

↓  
Sign



Proxy certificate  
= a limited copy of your certificate

Controlled



Authentication  
= who you are?



Authorization  
= can enter a new land (Grid)



# Certificate Authorities and VOMS

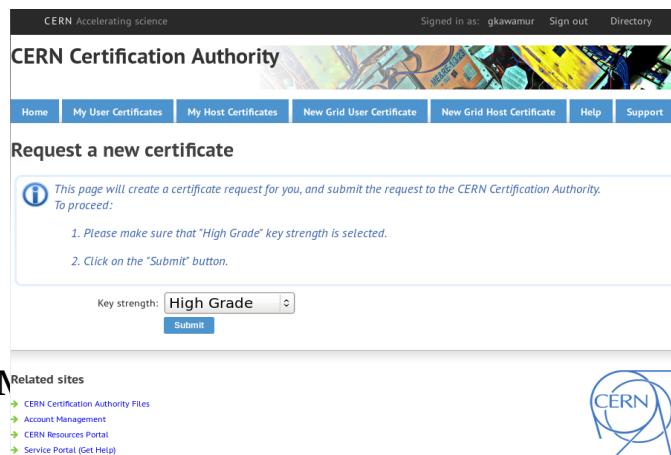
- Germany (FZK)

- <https://gridka-ca.kit.edu/>



- CERN

- <https://ca.cern.ch/ca/user/Request.aspx?template=EE2User>



# Certificate Authorities and VOMS

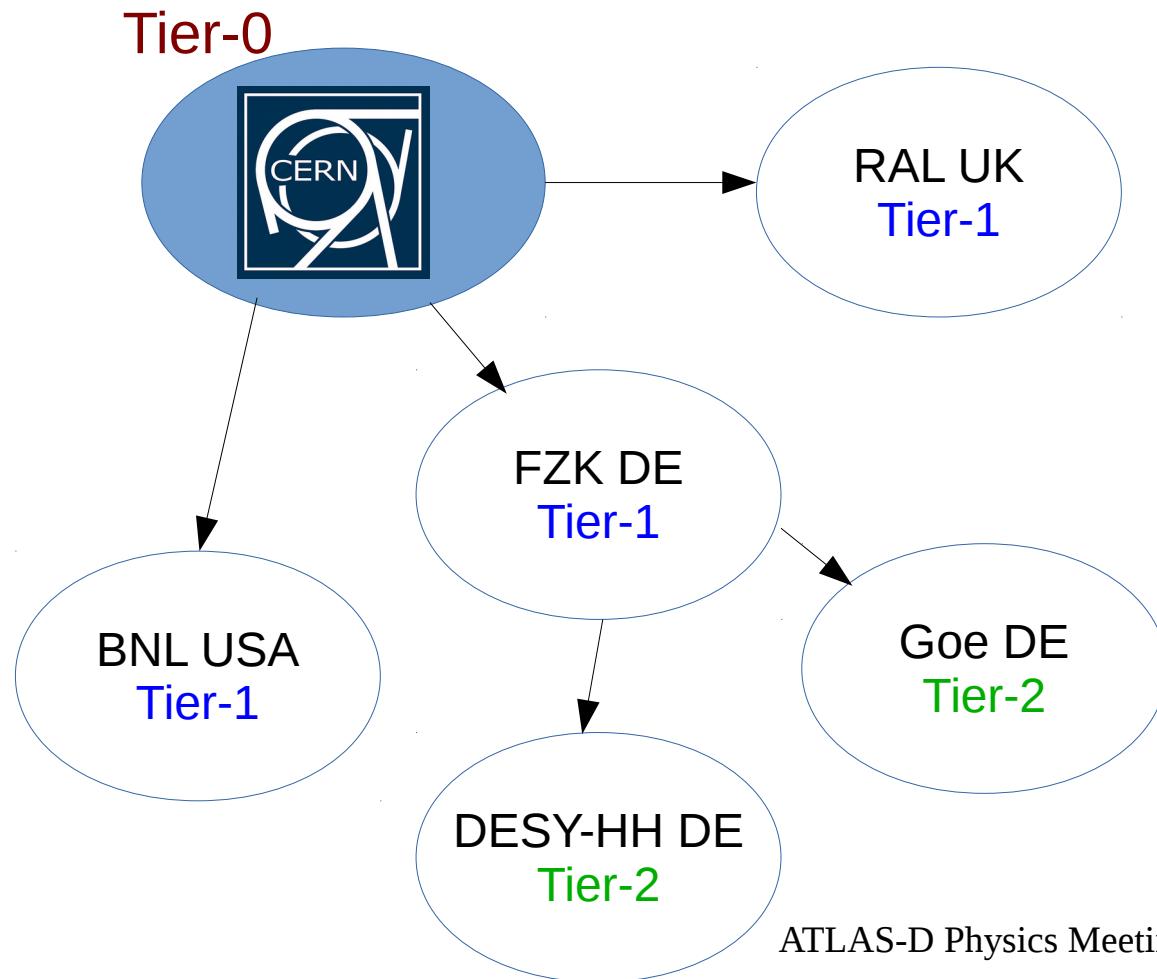
- VOMS top page
  - <https://voms2.cern.ch:8443/>
- VOMS ATLAS (request your ATLAS VO)
  - <https://voms2.cern.ch:8443/voms/atlas>
- VOMS ATLAS users in Germany
  - <https://voms2.cern.ch:8443/voms/atlas/services/VOMSCompatibility?method=getGridmapUsers&container=/atlas/de>



# ATLAS Grid computing and WLCG resources - 1

- LHC multi-tier structure

- WLCG = Worldwide LHC Computing Grid



**Tier-0:** Raw data, Data store (in tape), Pre-processing, Reconstruction

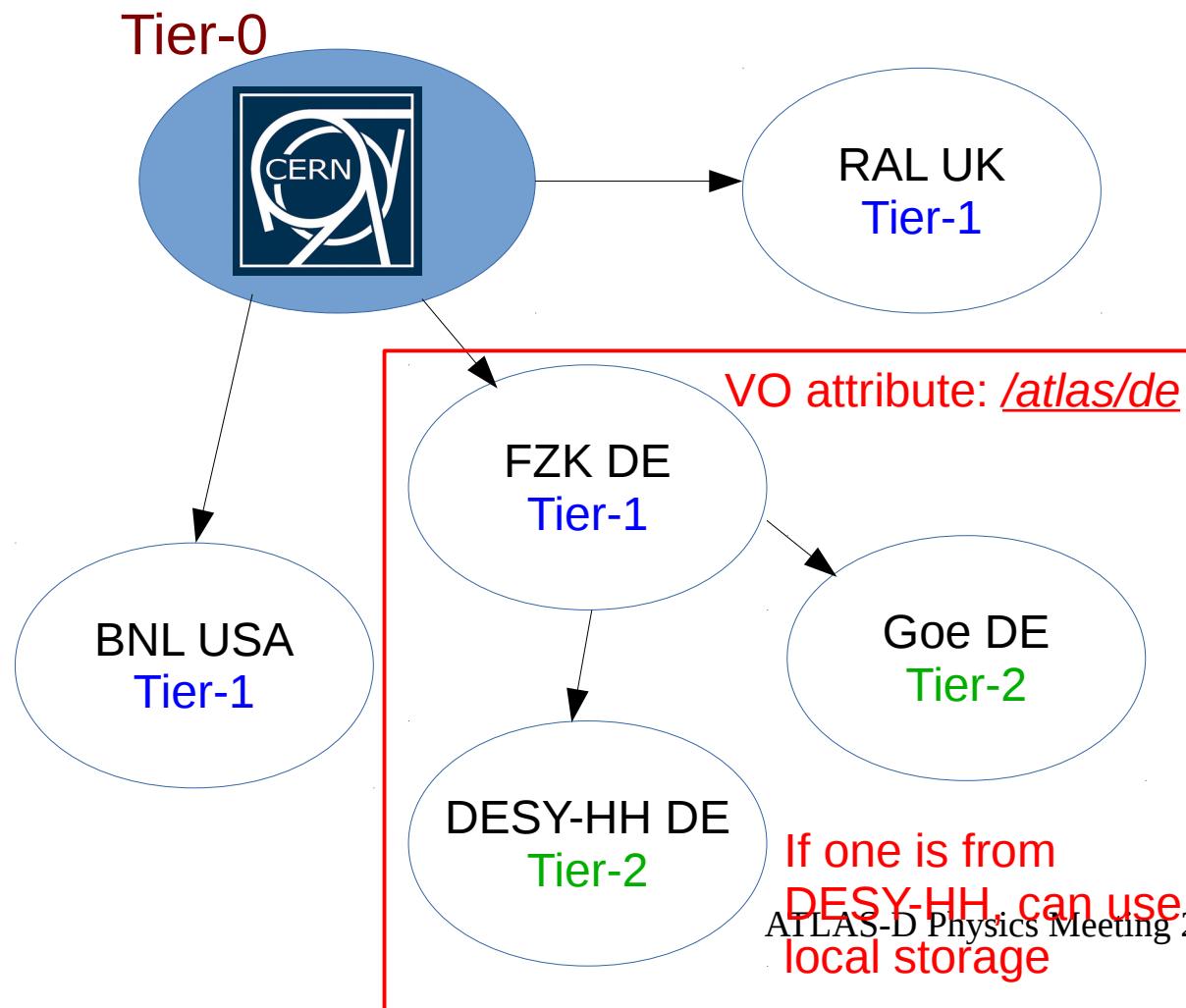
**Tier-1:** National LCG-Centre, faster network connections and larger storage spaces (e.g. Tape), MC production, user analysis, etc.

**Tier-2:** University or Facility level computing sites. MC production, user analysis, etc.

# ATLAS Grid computing and WLCG resources - 1

- LHC multi-tier structure

- WLCG = Worldwide LHC Computing Grid



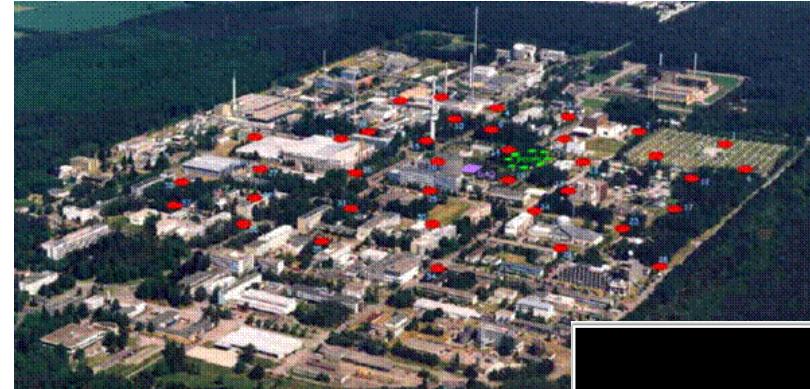
**Tier-0:** Raw data, Data store (in tape), Pre-processing, Reconstruction

**Tier-1:** National LCG-Centre, faster network connections and larger storage spaces (e.g. Tape), MC production, user analysis, etc.

**Tier-2:** University or Facility level computing sites. MC production, user analysis, etc.

# ATLAS Grid computing and WLCG resources - 2

- FZK Tier-1



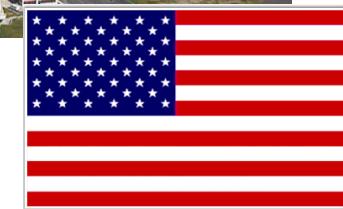
Forschungszentrum Karlsruhe  
in der Helmholtz-Gemeinschaft



- BNL Tier-1



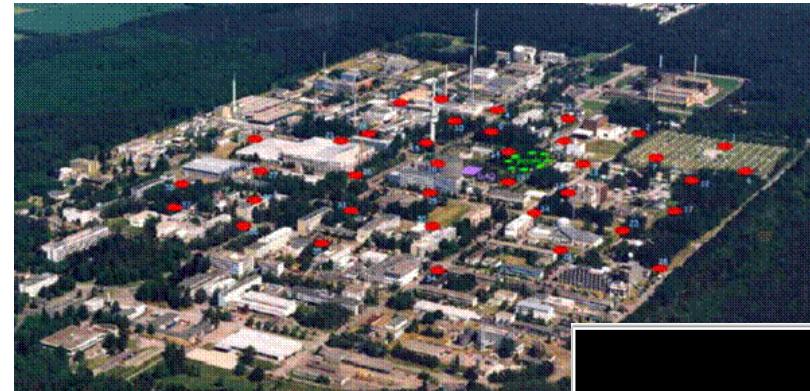
18



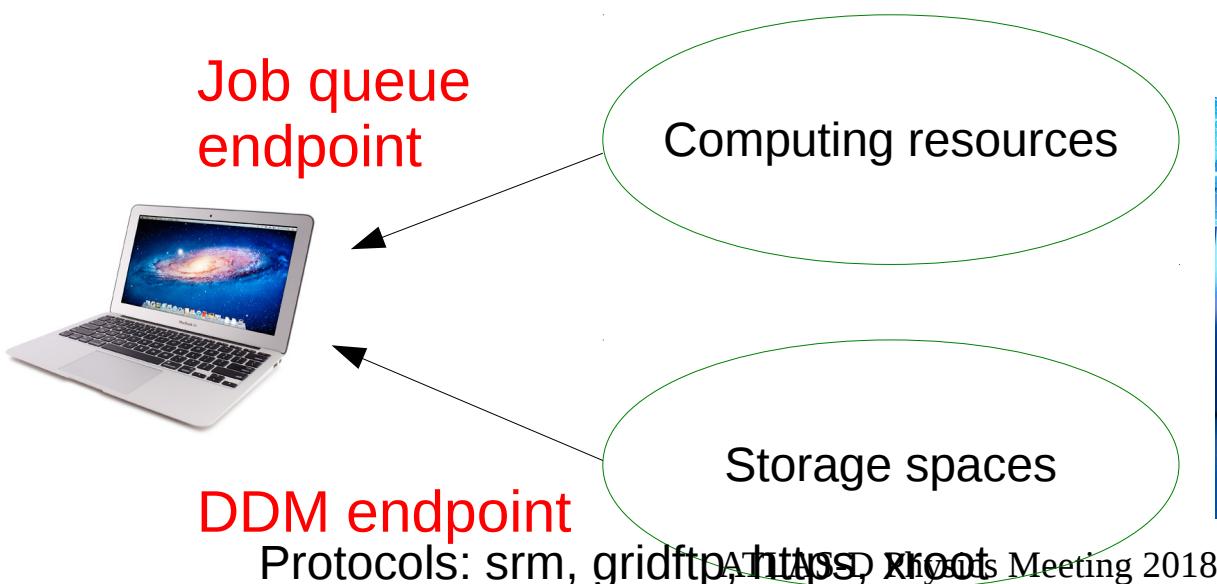
24

# ATLAS Grid computing and WLCG resources - 2

- FZK Tier-1



Forschungszentrum Karlsruhe  
in der Helmholtz-Gemeinschaft

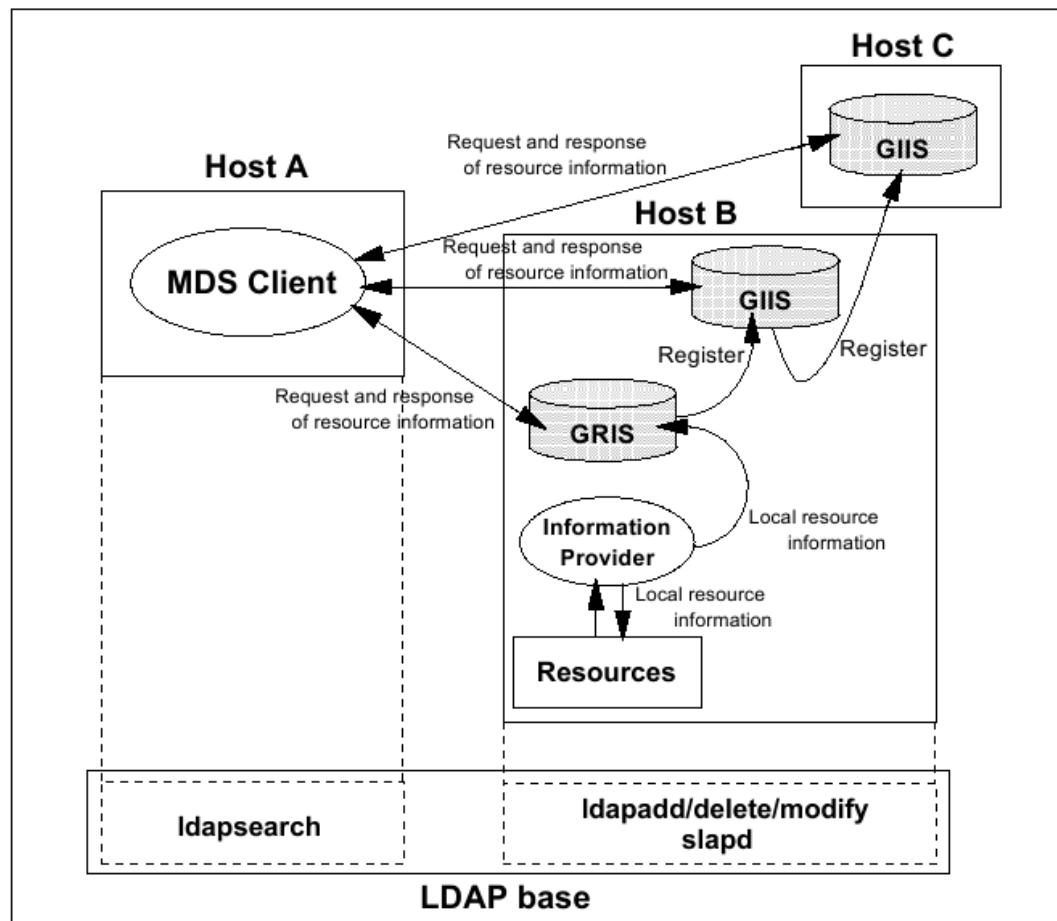


FZK Computing centre



# Information Provider

- Information Provider (MDS: Monitoring and Discovery Service)
  - Enables dynamic/static status retrieval
    - GRIS – Grid Resource Information Service
    - GIIS – Grid Index Information Service
    - has hierarchical structure similarly to DNS
    - Software based on LDAP (Lightweight Directory Access Protocol) client command *ldapsearch*
  - For example,
    - Name of services
    - Ports
    - Resources
    - Versions
    - Running/acceptable jobs



# ATLAS Grid job - 1

- Some technical terms you may often need (however not in Physics)
  - User certificate, proxy certificate, CA, Virtual Organization (VO), VOMS, authentication, authorization, Computing Element, Storage Element, Worker nodes, Workload Management System, data management system, Job, data replica, information provider, site



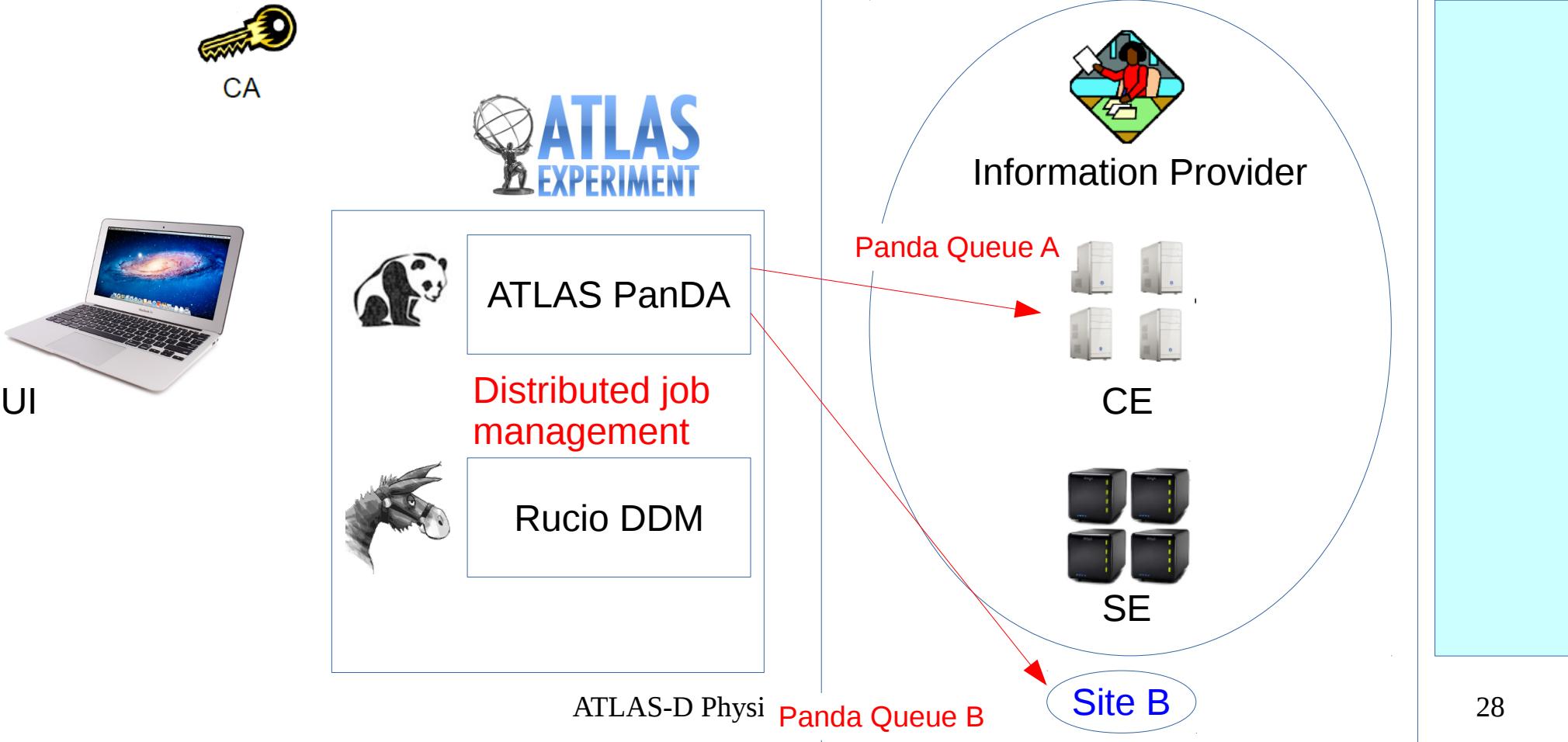
# ATLAS Grid job - 2



**WLCG**  
Worldwide LHC Computing Grid

Job  
Status

- How do they work?



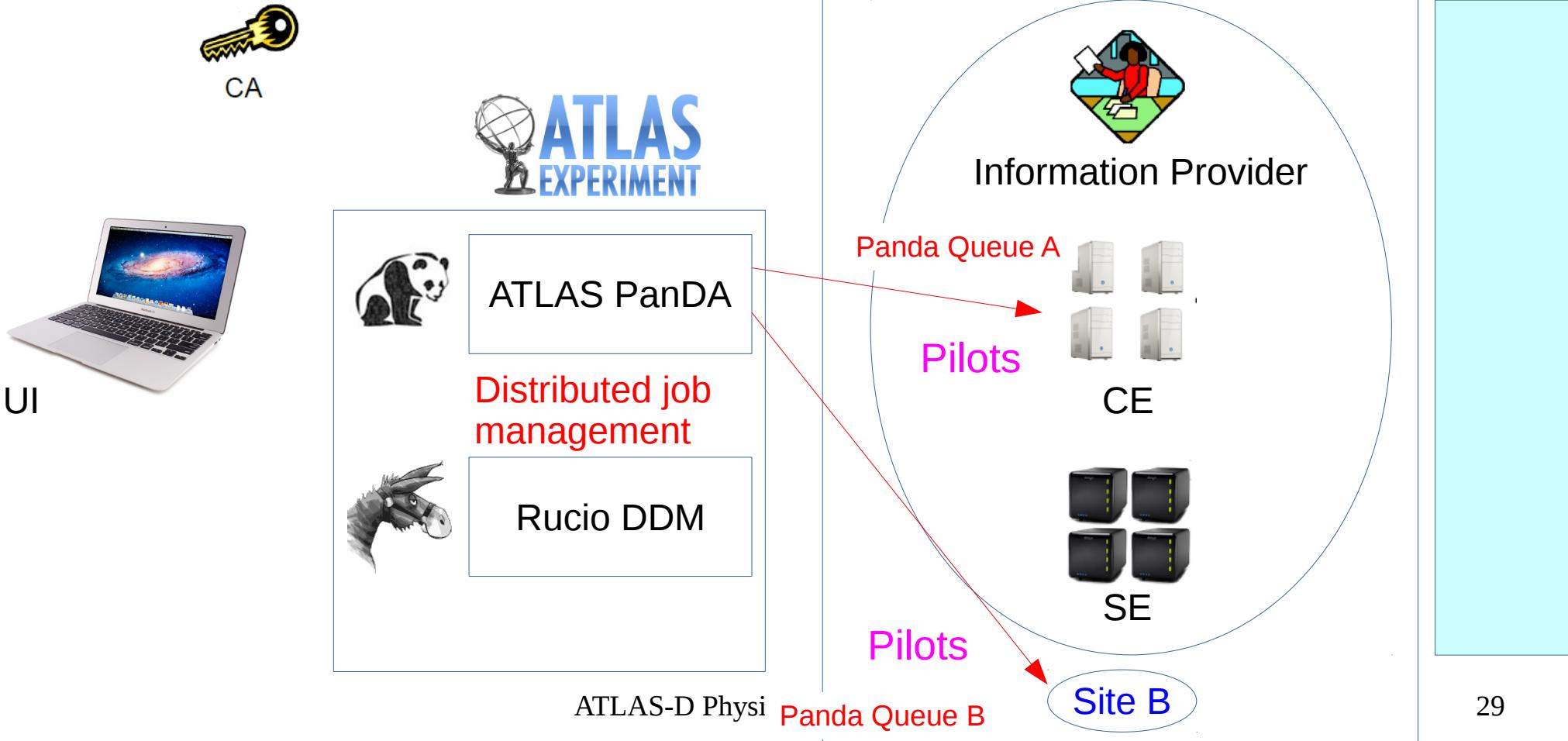
# ATLAS Grid job - 3



**WLCG**  
Worldwide LHC Computing Grid

Job  
Status

- How do they work?

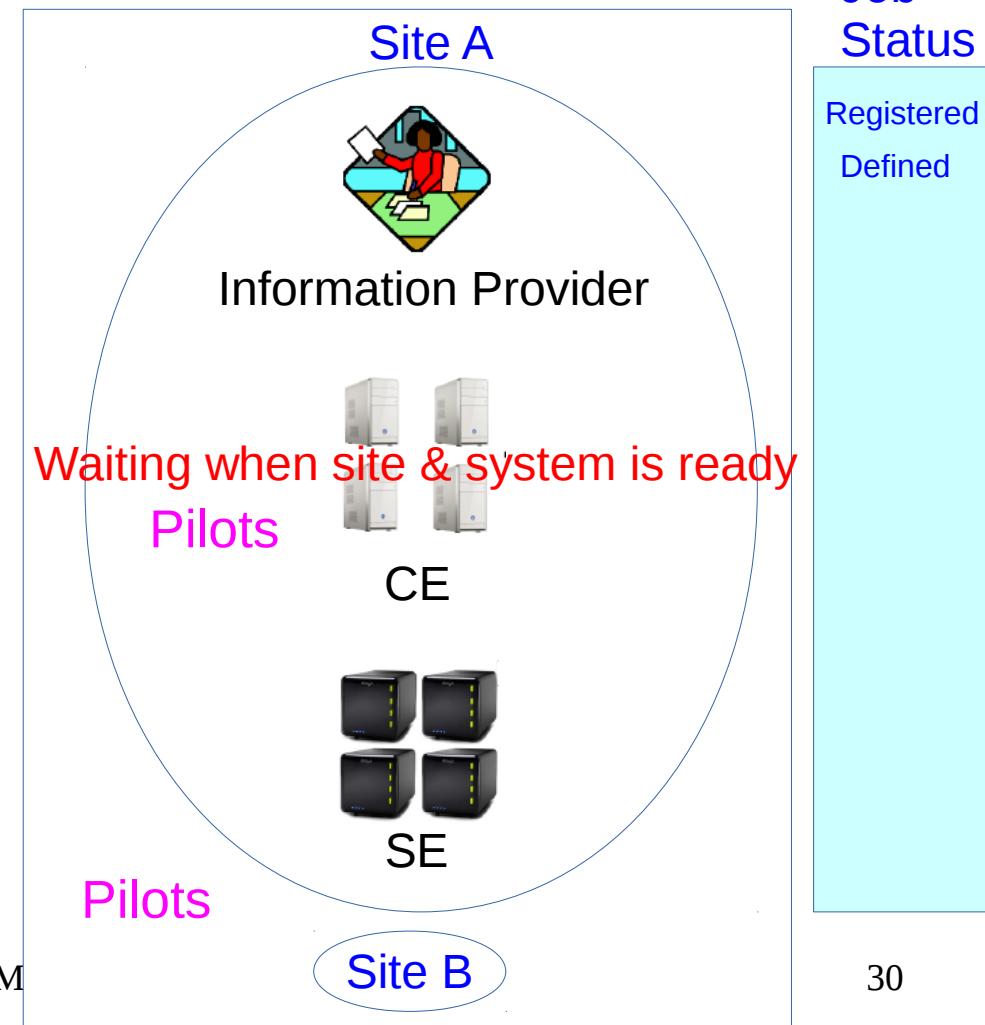
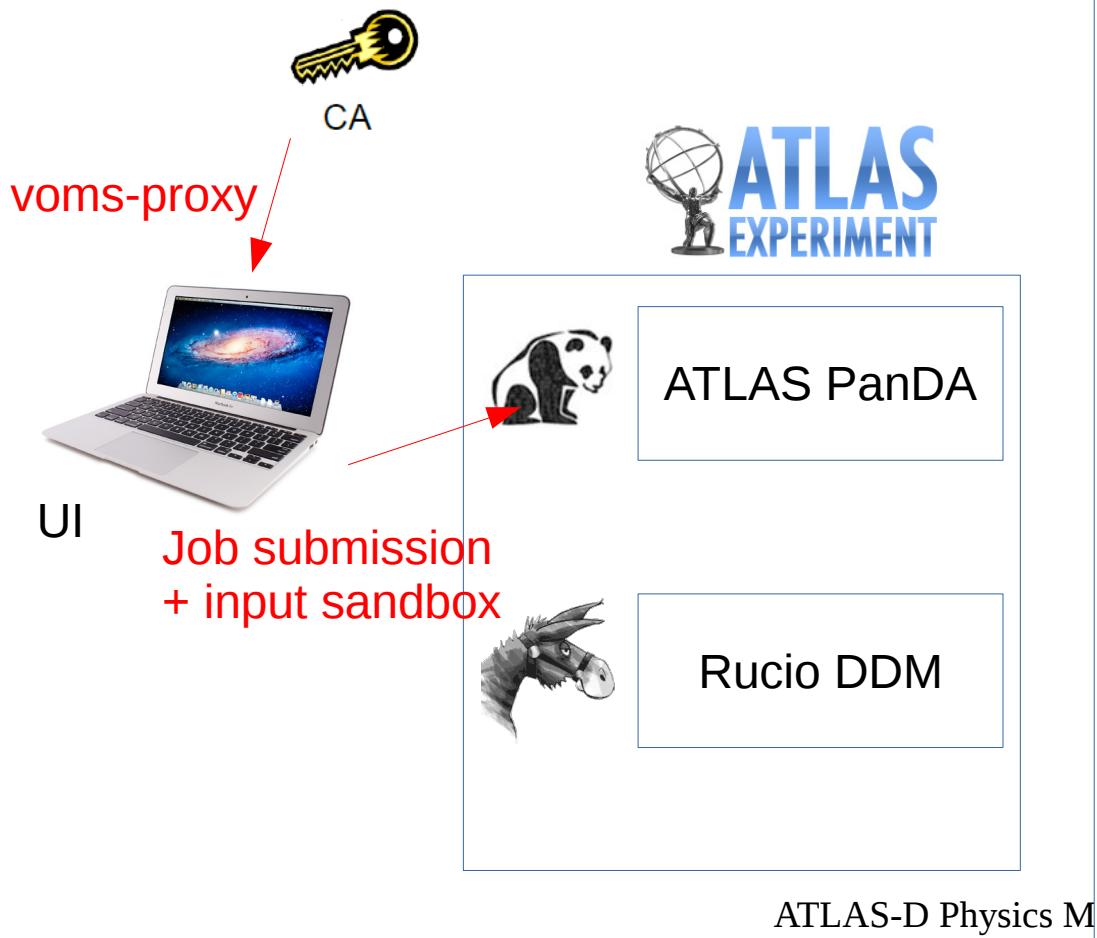


# ATLAS Grid job - 4



**WLCG**  
Worldwide LHC Computing Grid

- How do they work?

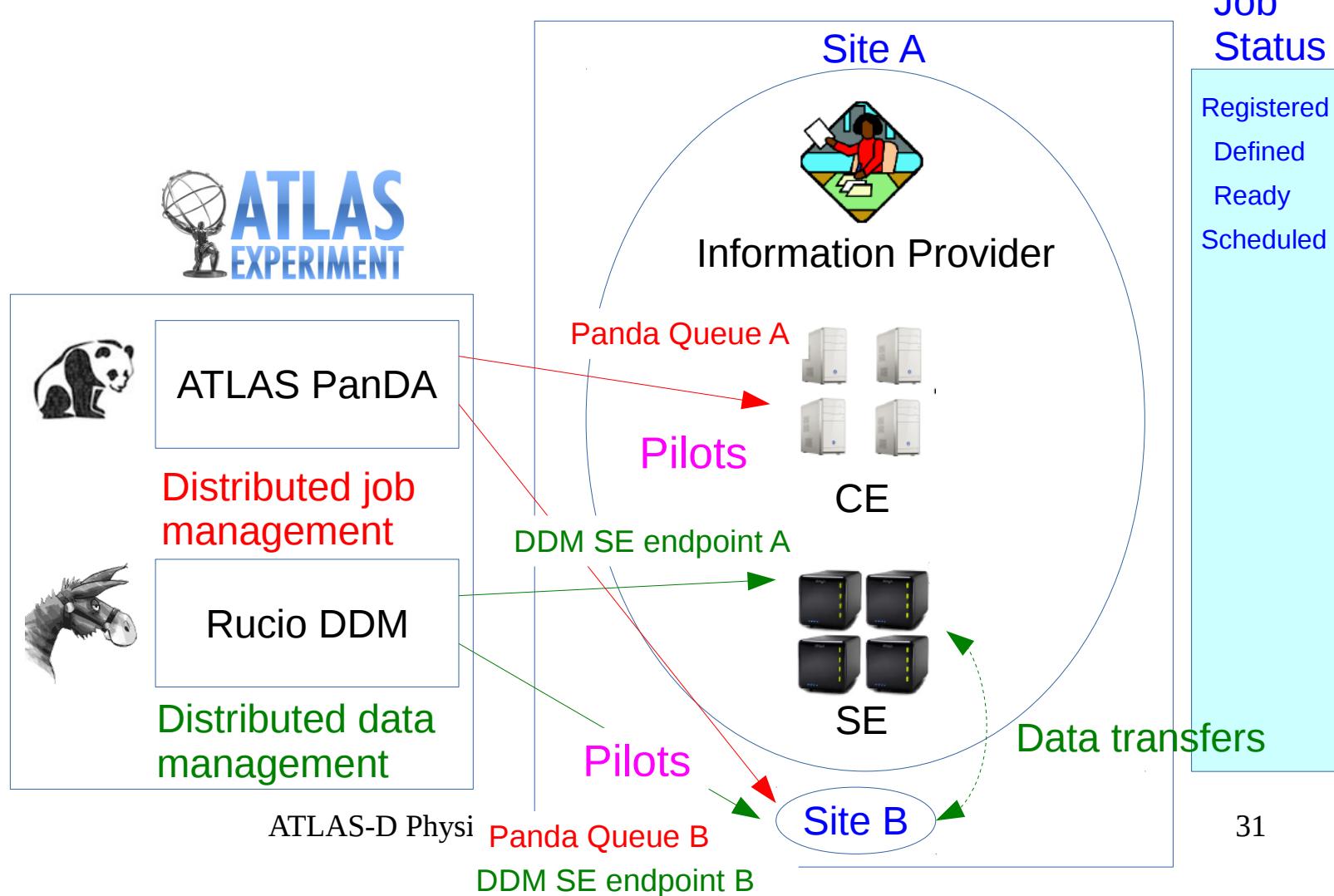


# ATLAS Grid job - 5



**WLCG**  
Worldwide LHC Computing Grid

- How do they work?

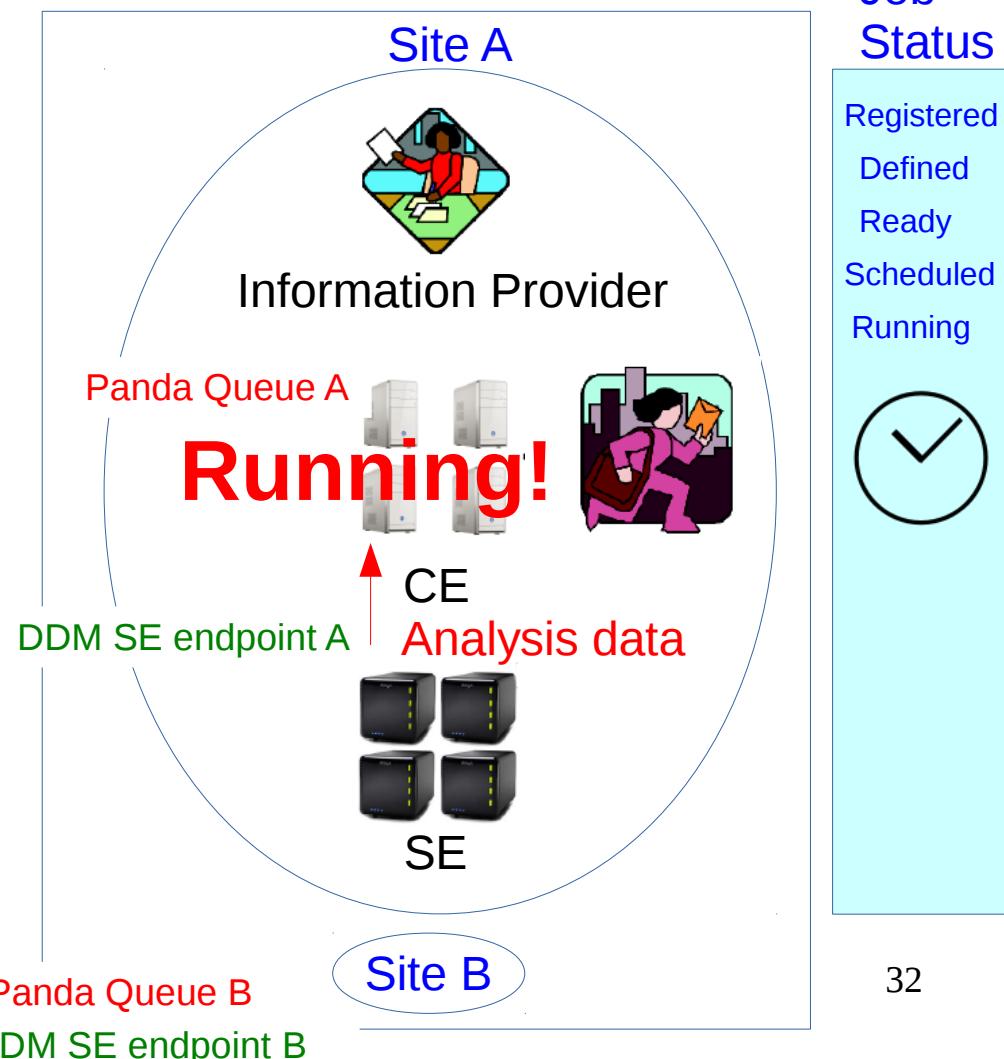


# ATLAS Grid job - 6



**WLCG**  
Worldwide LHC Computing Grid

- How do they work?

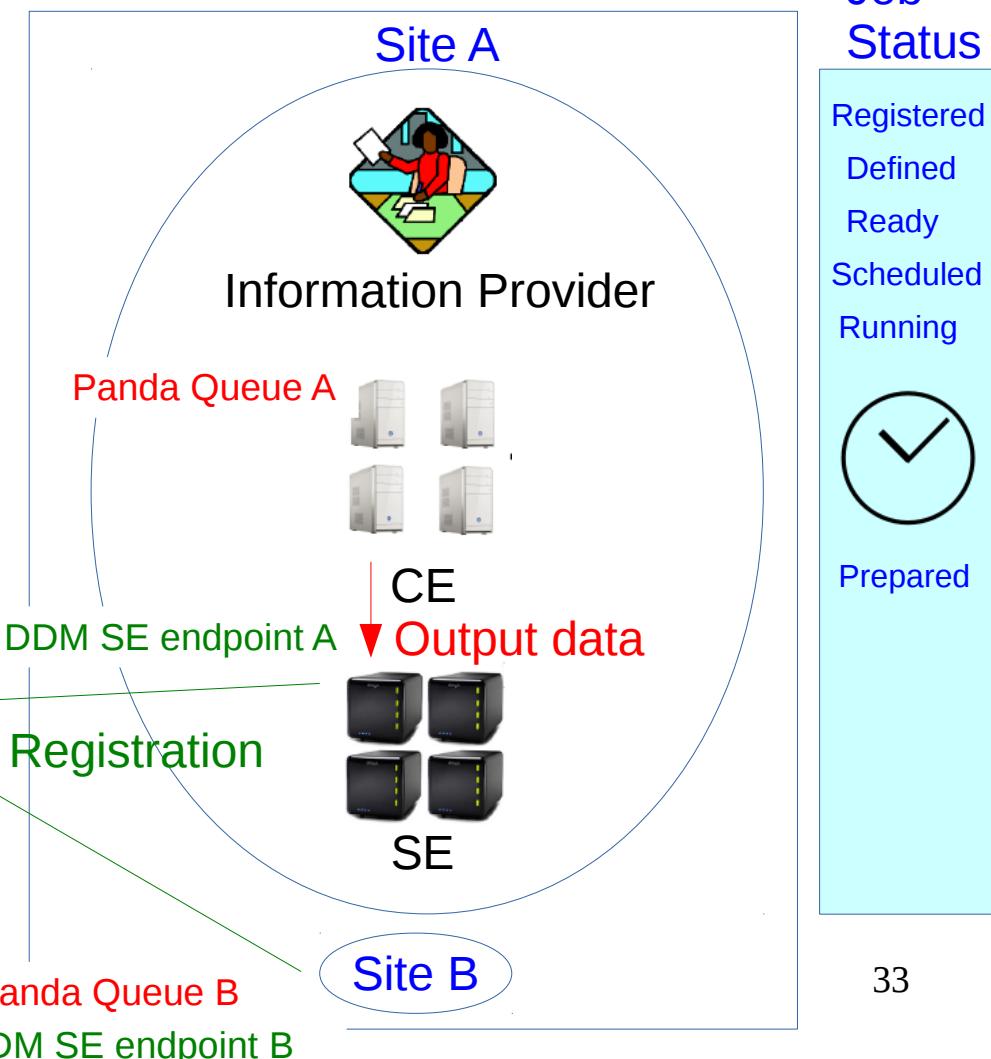
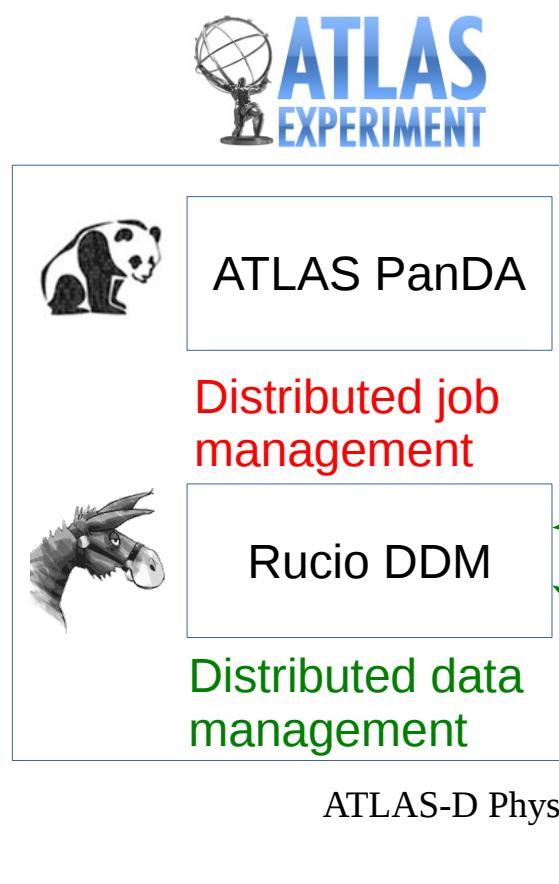


# ATLAS Grid job - 7



**WLCG**  
Worldwide LHC Computing Grid

- How do they work?



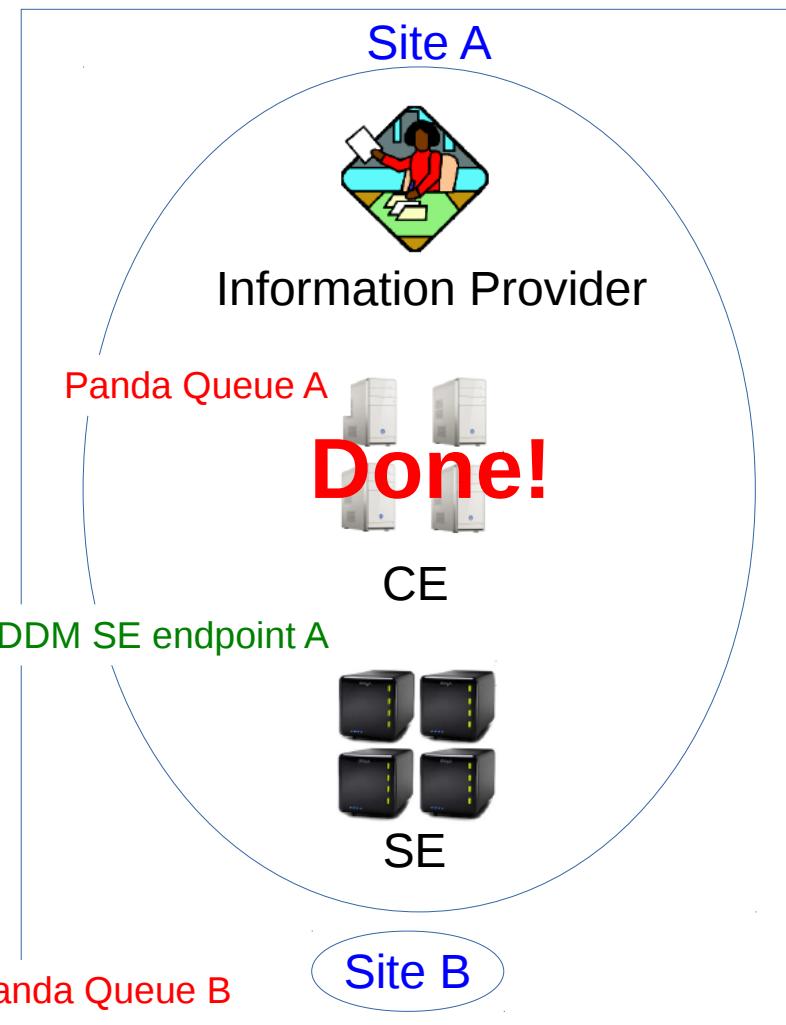
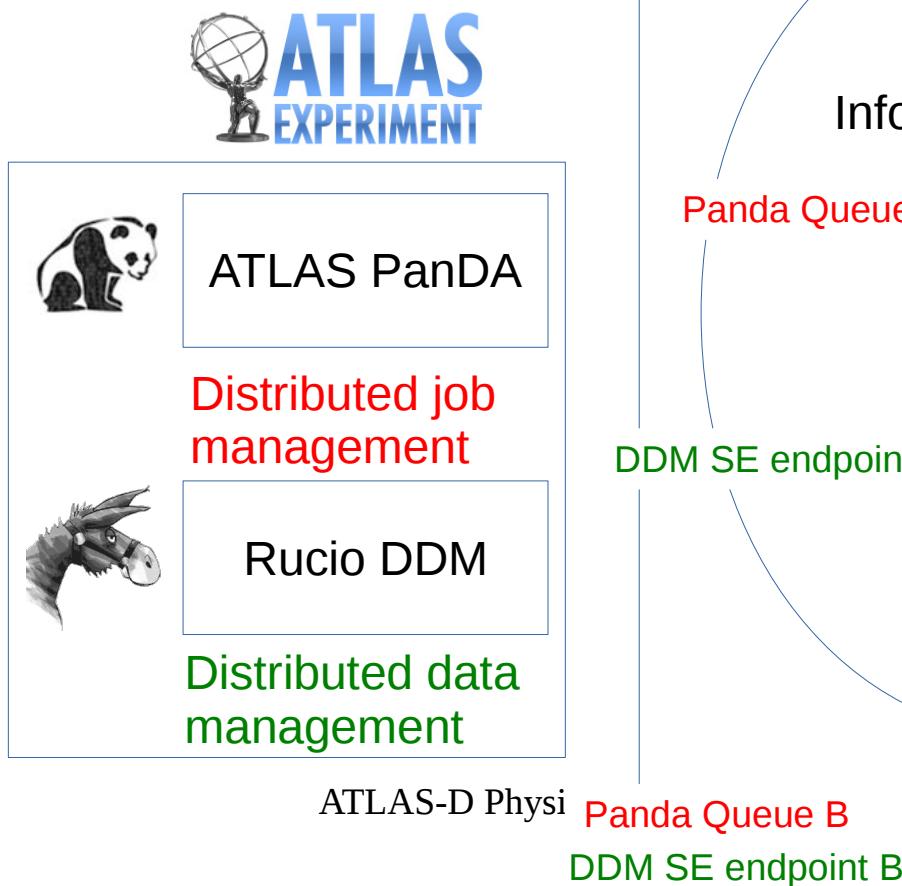
# ATLAS Grid job - 8



**WLCG**  
Worldwide LHC Computing Grid

## Job Status

- How do they work?



- Registered
- Defined
- Ready
- Scheduled
- Running


- Prepared
- Done

# Grid User Interface and CVMFS

- Grid client software originally supported by *European Middleware Initiative* (EMI)
- CVMFS is a remote repository using FUSE file system
  - e.g. /cvmfs/atlas.cern.ch → A repository of all client software we need
  - Internally using HTTP → Need of network access
- Panda client using ATLAS PanDA server is a tool to manage jobs
  - BigPanda monitoring system provides a Web interface
- Rucio client is a tool to interact with data



Just try it out!



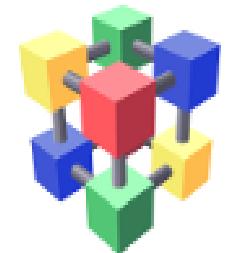
ATLAS Athena physics framework

Data management  
Rucio client

Job management  
Panda client

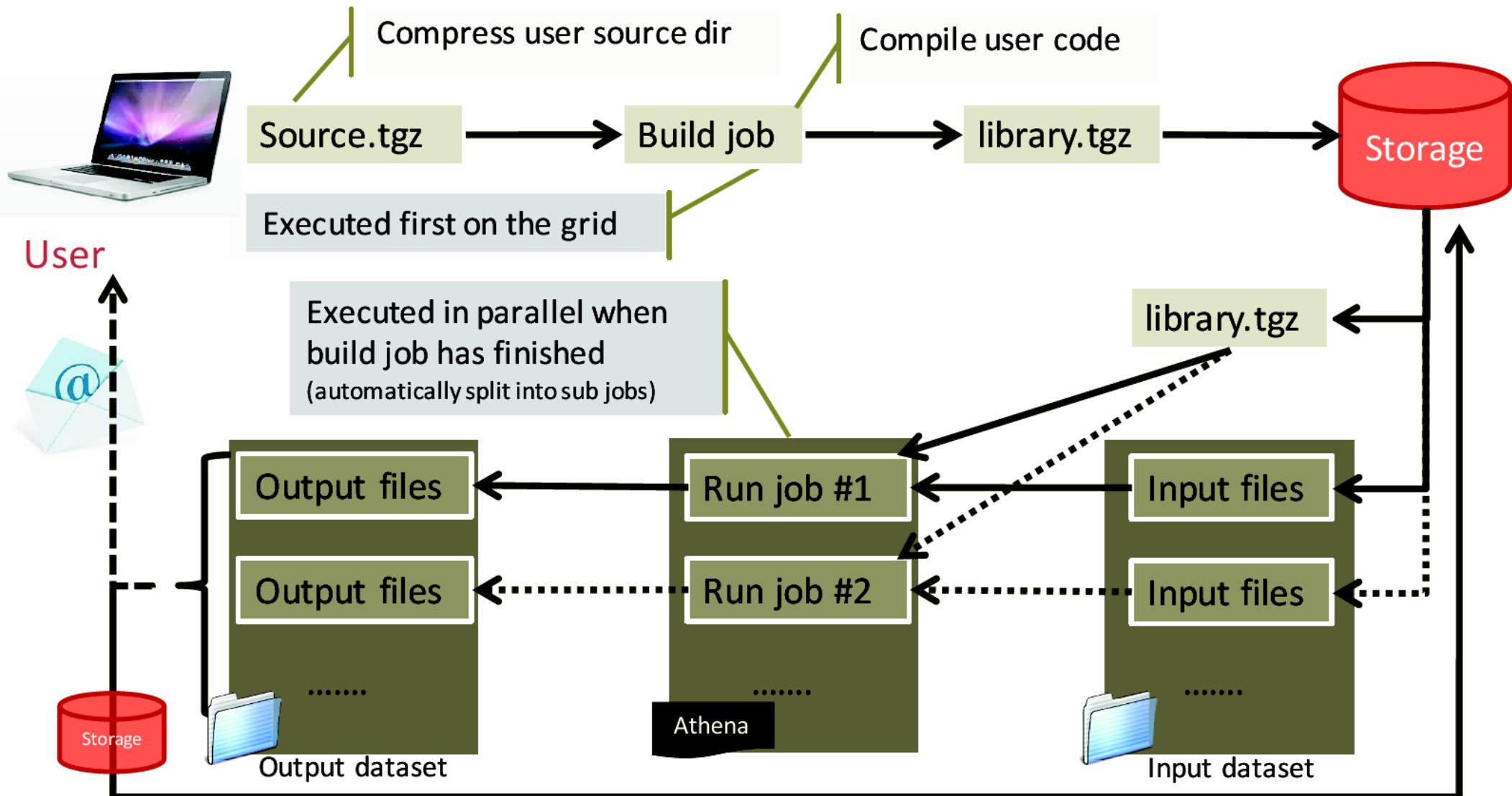
Grid client software  
EMI LCG tools

# PanDA (ATLAS Job Management System)



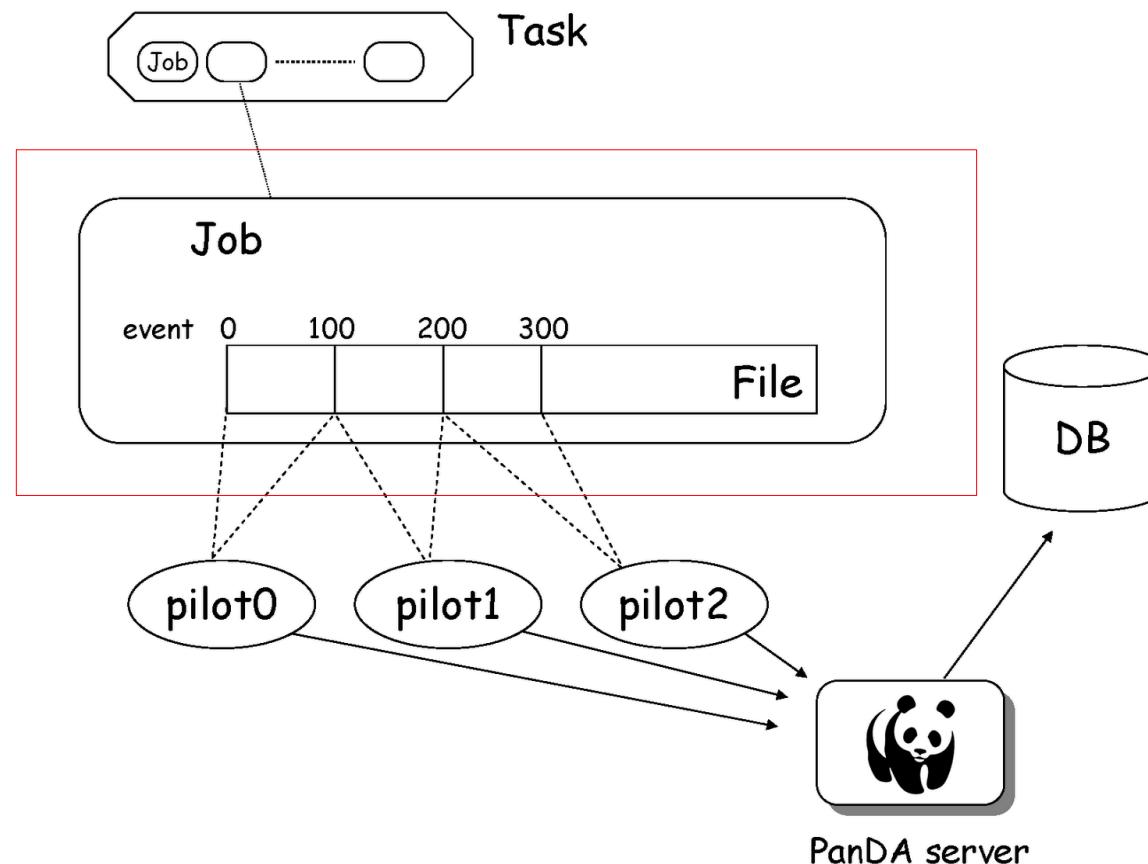
**WLCG**  
Worldwide LHC Computing Grid

# Build job & Run job



# Event level processing

- Task (calculation) can be split by evenl level handler
  - Event Level Handler can generate basic pilot jobs according to events in given file(s)



# Monitoring PanDA jobs (1/5)

Go to: <http://bigpanda.cern.ch> and enter your jediTTaskID

here →

The dashboard displays three stacked bar charts showing the number of slots of running jobs. The left chart covers 24 hours from 2014-10-06 09:00 to 2014-10-07 09:11 UTC. The middle chart covers 7 days from 2014-09-30 to 2014-10-07. The right chart covers 29 days from 2014-09-07 to 2014-10-07. The charts are color-coded by location: CERN (blue), NL (orange), EG (green), EU (red), and T0 (yellow). A legend at the bottom of each chart shows these colors.

**Search**

PanDA job ID or name  Submit

Batch ID  Submit

Task ID  Submit

Task name  Submit

**News**

- 20140923: Job pages directly link to object store based log tarballs
- 20140913: Link DEFT request page from task list and detail pages
- 20140912: On task page highlight input containers, hide datasets by default
- 20140911: Task name search supports wildcarding
- 20140911: Show wait time, duration for jobs not yet running, completed
- 20140818: Job attempt# off for user page, JEDI jobs. Not meaningful in JEDI.
- 20140818: Task attribute summary added to user page
- 20140817: Output container list added to task detail page
- 20140817: Support clarified. Use DAST list, as ever, for dist analysis support

# Monitoring PanDA jobs (2/5)

Task is running

bigpanda.cern.ch/task/?jeditaskid=4212437&display\_limit=200

ATLAS PanDA monitor Dashboards Tasks Jobs Errors Users Sites Incidents Search Prodsys Services VO Help Generated 2014-10-07 09:34 UTC

Task 4212437: user.nozturk.pruntest/

Task ID	Jobset	Type	WorkingGroup	User	Task status	Ninputfiles   finished   failed	Created	Modified	Cores	Priority	Parent
4212437	15168	anal		Nurcan Ozturk	running	1   0 (0%)   0 (0%)	2014-10-07 09:33	10-07 09:33	1	1000	

States of jobs in this task [Show jobs](#)

defined	waiting	pending	assigned	throttled	activated	sent	starting	running	holding	transferring	finished	failed	cancelled	merging
1								1						

Jump to [job parameters](#), [task parameters](#)

View: [job list \(access to job details and logs\)](#) [child tasks](#) [prodsys task page](#) [brokerage logger](#) [JEDI action logger](#) [error summary](#)

Output containers

[user.nozturk.pruntest.log/](#)

4 datasets, show/hide by type: [all](#) [lib\(1\)](#) [log\(1\)](#) [pseudo\\_input\(1\)](#) [tmp\\_log\(1\)](#)

Dataset, container name	Type	Stream	State	Status	Nfiles	Created	Modified

Job parameters

"

# Monitoring PanDA jobs (3/5)

Task is done

Screenshot of the ATLAS PanDA monitor interface showing a completed task.

The URL in the browser bar is [bigpanda.cern.ch/task/?jeditaskid=4212437&display\\_limit=200](http://bigpanda.cern.ch/task/?jeditaskid=4212437&display_limit=200).

The main table shows the following details for Task 4212437:

Task ID	Jobset	Type	WorkingGroup	User	Task status	Ninputfiles   finished   failed	Created	Modified	Cores	Priority	Parent
4212437	15168	anal		Nurcan Ozturk	done	1   1 (100%)   0 (0%)	2014-10-07 09:33	10-07 09:50	1	1000	

Below the table, a section titled "States of jobs in this task" lists various job states:

defined	waiting	pending	assigned	throttled	activated	sent	starting	running	holding	transferring	finished	failed	cancelled	merging
											2			

A red arrow points from the text "clickable" to the "View:" button in the navigation bar.

A red arrow points from the text "link to jobs and their log files" to the "Output containers" section.

The "Output containers" section shows a link to "user.nozturk.pruntest.log/".

The bottom section displays 4 datasets with the following details:

4 datasets, show/hide by type: all lib(1) log(1) pseudo_input(1) tmpl_log(1)								
Dataset, container name	Type	Stream	State	Status	Nfiles	Created	Modified	

# Monitoring PanDA jobs (4/5)

Task details, list of jobs. Click on “PanDA ID” to go to the job details, job log files

click 

Job list Sort by Pandaid, ascending mod time, descending mod time, priority, attemptnr											
PanDA ID Attempt#	Owner Group	Task ID	Transformation	Status	Created	Time to start d:h:m:s	Duration d:h:m:s	Mod	Cloud Site	Priority	Job info
2280470784 Attempt 1	Nurcan Ozturk	4212437	runGen-00-00-02	finished	2014-10-07 09:33	0:06:28	0:0:01:04	10-07 09:49	ES ANALY_IFAE	1000	
	Job name: user.nozturk.pruntest/ #1 Datasets:										
2280470783 Attempt 0	Nurcan Ozturk	4212437	buildGen-00-00-01	finished	2014-10-07 09:33	0:00:16	0:0:01:04	10-07 09:40	ES ANALY_IFAE	2000	
	Job name: user.nozturk.pruntest/ #0 Datasets: Out: panda.1007093351.25596.lib._4212437										

 run job: runs the job at the grid site

 build job: recreates the athena environment at the grid site

# Monitoring PanDA jobs (5/5)

## How to find job log files

Screenshot of the ATLAS PanDA monitor interface showing job details for PanDA job 2280470784.

Job details for PanDA job 2280470784 (Generated 2014-10-07 10:52 UTC)

PandaID	Owner	TaskID	Status	Created	Time to start d:h:m:s	Duration d:h:m:s	Modified	Cloud Site	Priority
2280470784	Nurcan Ozturk	4212437	finished	2014-10-07 09:33	0:0:06:28	0:0:01:04	10-07 09:49	ES ANALY_IFAE	1000

Job name: [user.nozturk.pruntest/](#) type: panda-client-0.5.30-jedi-run transformation: runGen-00-00-02

Datasets:

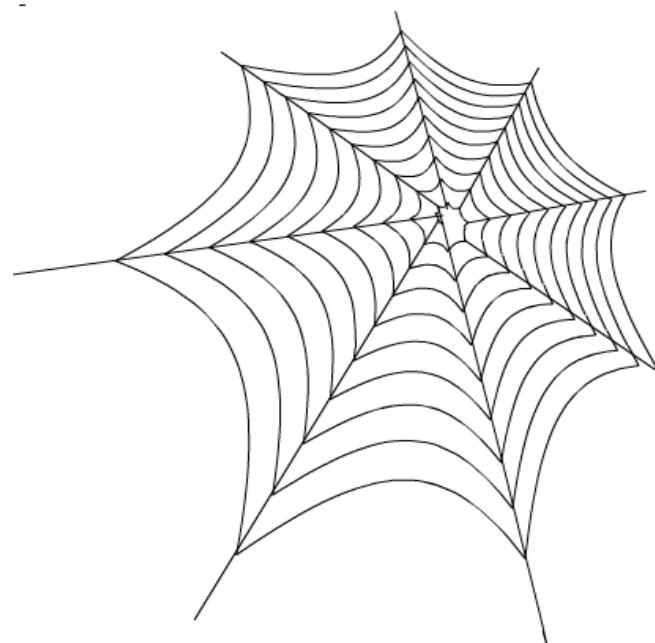
Job information logfiles pilot job stdout, stderr, batch log logger records child jobs

Jobset 15168

check athena\_stdout.txt (for pathena)/prun\_stdout.txt (for prun) and pilotlog.txt (all commands issued by the pilot) from this link

5 job files						
Filename (Type)	Scope	Size (MB)	Status	Attempt (max)	Dataset	
<a href="#">panda.1007093351.25596.lib._4212437.241508507.lib.tgz</a> (input)	panda	0	ready		<a href="#">panda.1007093351.25596.lib._4212437</a> (dispatch block: <a href="#">panda.1007093351.25596.lib._4212437</a> )	
<a href="#">user.nozturk.pruntest.log.4212437.000001.log.tgz</a> (log)	user.nozturk	0	finished		<a href="#">user.nozturk.pruntest.log.6779086</a>	
<a href="#">user.nozturk.pruntest.log.4212437.000001.log.tgz</a> (log)	user.nozturk	0	ready		<a href="#">user.nozturk.pruntest.log/</a> (destination block: <a href="#">sub0186192010</a> )	
pseudo_lfn (pseudo_input)		0	finished		<a href="#">pseudo_dataset</a>	
pseudo_lfn (pseudo_input)		0	unknown		<a href="#">pseudo_dataset</a>	

# ATLAS Metadata Interface (AMI)

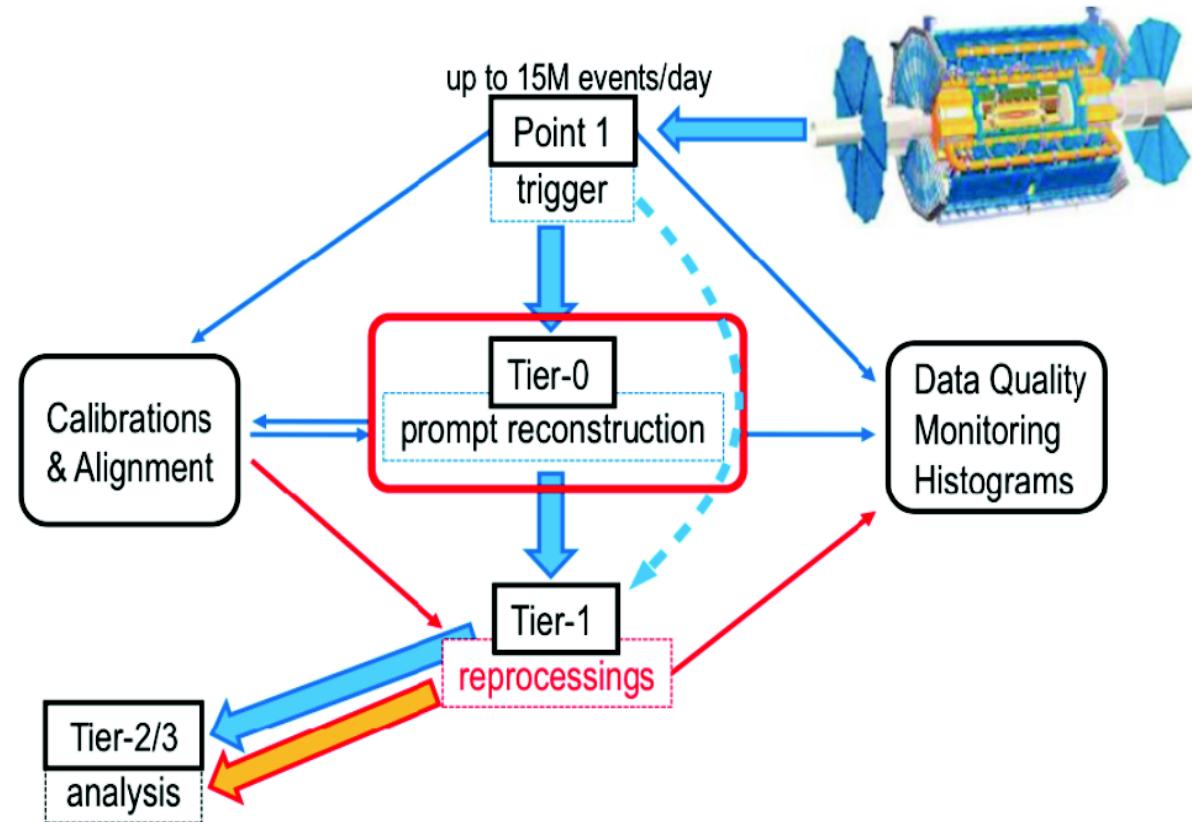


# AMI (ATLAS Metadata Interface)

- **Metadata = Data of Data (Description of data)**
- Key of ATLAS Data Life Cycle Management
- On each step of data reconstruction in ATLAS, AMI Tags are generated
  - ATLAS data set, metadata link
    - Size and origin of dataset
    - File, number of events
    - Software parameter (AMI Tags)
    - MC parameter (PDF, generator, cross section, etc)
    - Lost files and Lumi blocks
    - Link to other applications (COMA, Rucio)
    - Data period
      - Luminosity, Trigger, Data of RAW data creation etc
  - Special Interface
    - AMI-Tags
      - Software Configuration Parameters
    - AMI-Glance
      - Data + Publication
    - Definition of data periods
    - Definition of physics containers
    - Reprocessing campaigns
    - Event count comparator

# ATLAS Dataset - 2

- Tier-0
  - RAW is produced
  - Prompt data reconstruction
- ProdSys2
  - MC + Reprocessing
  - PanDA, JEDI, DEFT
- Distribution of dataset
  - Rucio



# ATLAS Dataset - 3

- Data reconstruction → Example : x353 = AMI tag
  - `Reco_tf.py —AMI=x353 --inputBSFile=tier0_RawData.data`
- Reprocessing Campaign
  - After improvement of ATLAS software or framework, often huge reconstruction jobs can run
    - The campaign require many computing resources
    - As a result, many different versions of AOD (Analysis Object Data)
      - Derivation framework: AOD → Derived AOD
      - About 1% size of original AOD
      - Selection of particular event, parameter etc
      - Different configurations, software and conditions

# ATLAS Dataset - 4

- Dataset = Collection of files
  - Collision data (data) and Monte Carlo (mc)

## Data:

project tag:  
2012 pp data  
8 TeV      Run number      stream      merged files      Data type:  
data12\_8TeV.00209980.physics\_Egamma.merge.AOD.f476\_m1223  
AMI tag describes configuration of  
each step (Tier-0 bulk reconstruction **f**,  
file merging **m**)

## Simulation:

project tag: MC DSID  
"mc12" setup      unique #  
8 TeV      for process      "human-readable" description of MC sample      merged files      Data type:  
mc12\_8TeV.119353.MadGraphPythia\_AUET2BCTEQ6L1\_ttbarW.merge.NTUP\_SMWZ.  
e1352\_s1499\_s1504\_r3658\_r3549\_p1328/  
AMI tag describes configuration of  
each step (evt generation **e**, full simulation **s**,  
reconstruction **r**, D3PD creation **p**)  
/: is a "container" (points to other datasets)

# AMI WebUI - 1

Datasets / Dataset Browser

Search Form 1: data12\_001-real\_data

1 dataset 1075 records

Query :dataset.amiStatus='VALID' AND (dataset.dataType like 'AOD') AND (dataset.streamName like 'physics\_MinBias')

more fields +	logicalDatasetName ▾ Q	nFiles ▾ Q	totalEvents ▾ Q	totalSize ▾ Q	runNumber ▾ Q	period ▾ Q
<a href="#">details</a>	data12_8TeV.00200804.physics_MinBias.merge.AOD.r4644_p1517 DQ2 - Provenance - GANGA export	18 182	585918 11645642	68.699 GB	200804 COMA Report - Periods - Run_Summary - Run_Query - DAQ_Config	A1 more Info - COMA - All Runs
<a href="#">details</a>	data12_8TeV.00200805.physics_MinBias.merge.AOD.r4644_p1517 DQ2 - Provenance - GANGA export	2	54277	1.881 GB	200805 COMA Report - Periods - Run_Summary - Run_Query - DAQ_Config	A2 more Info - COMA - All Runs
<a href="#">details</a>	data12_8TeV.00200841.physics_MinBias.merge.AOD.r4644_p1517 DQ2 - Provenance - GANGA export	4	186265	20.177 GB	200841 COMA Report - Periods - Run_Summary - Run_Query - DAQ_Config	A3 more Info - COMA - All Runs
<a href="#">details</a>	data12_8TeV.00200842.physics_MinBias.merge.AOD.r4644_p1517 DQ2 - Provenance - GANGA export	4	194266	17.213 GB	200842 COMA Report - Periods - Run_Summary - Run_Query - DAQ_Config	A3 more Info - COMA - All Runs
<a href="#">details</a>	data12_8TeV.00200863.physics_MinBias.merge.AOD.r4644_p1517 DQ2 - Provenance - GANGA export	4	99639	15.288 GB	200863 COMA Report - Periods - Run_Summary - Run_Query - DAQ_Config	A3 more Info - COMA - All Runs
<a href="#">details</a>	data12_8TeV.00200913.physics_MinBias.merge.AOD.r4644_p1517 DQ2 - Provenance - GANGA export	4	120898	16.403 GB	200913 COMA Report - Periods - Run_Summary - Run_Query - DAQ_Config	A3 more Info - COMA - All Runs
<a href="#">details</a>	data12_8TeV.00200928.physics_MinBias.merge.AOD.r4644_p1517 DQ2 - Provenance - GANGA export	2	47551	7.145 GB	200928 COMA Report - Periods - Run_Summary - Run_Query - DAQ_Config	A4 more Info - COMA - All Runs
<a href="#">details</a>	data12_8TeV.00200965.physics_MinBias.merge.AOD.r4644_p1517 DQ2 - Provenance - GANGA export	8	290279	36.652 GB	200965 COMA Report - Periods - Run_Summary - Run_Query - DAQ_Config	A4 more Info - COMA - All Runs
<a href="#">details</a>	data12_8TeV.00200982.physics_MinBias.merge.AOD.r4644_p1517 DQ2 - Provenance - GANGA export	2	40071	7.092 GB	200982	A4

1: number of results  
2: default order, more recent first

3: query clauses  
4: +/- fields  
5,6: filter,calculator

7: conversion of units  
8: group by, order by tools

# AMI WebUI - 2

Element's information		Children elements	
logicalDatasetName	mc14_8TeV.129173.Pythia8_AU2CTEQ6L1_gammajet_DP140.merge.AOD.e1146_s1896_s1912_r5591_r5625 RucioInfo Provenance - Campaigns - GANGA export - Series	dataset_extra	4 Records
physicistResponsible	c.gwenlan1@physics.ox.ac.uk	dataset_keywords	5 Records
nFiles	200	dataset_comment	No records found
totalEvents	999500	files	200 Records
totalSize	672.801 GB	jobOptions	No records found
dataType	AOD	prodsys_task	1 Records
prodsysStatus	ALL EVENTS AVAILABLE	field	approx_crossSection
ECMEnergy	8000	value	5 1.2217E+02
physicsComment		field	approx_GenFitEff
PDF	CTEQ6L1 - LO with LO alpha_s	value	9.6932E-04
version	e1146_s1896_s1912_r5591_r5625 Datasets - Config_Tag	field	autoConfiguration
AtlasRelease	19.0.3	value	['everything']
crossSection	122.170 nb 2 Report an error - Jira issues	field	postInclude
Trans	1: provenance & rucio 2: JIRA link for X section pbs	value	[RecJobTransforms/UseFrontier.py]
data		3: click for list of files 4: detail of prodsys task 5: cross section	

# Rucio (ATLAS data management system)

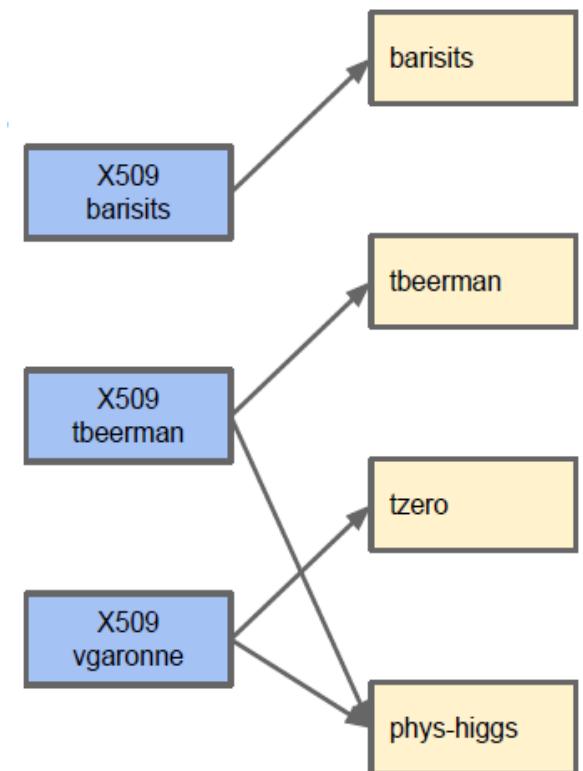


# Rucio basic concepts - 1

- Data Management System for Run-2 in ATLAS distributed computing system
  - Used to download outputs of Grid jobs, moving data and searching for them
- Rucio CLI tools from CVMFS
  - Web interface provides similar functionality

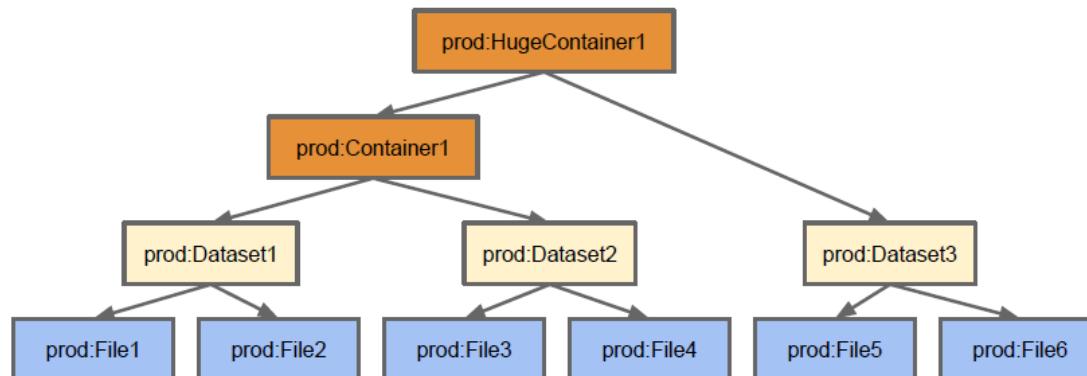
# Rucio basic concepts - 2

- Accounts
  - A Rucio Account can represent users, groups (phys-susy) or activities (panda, tzero)
  - Quota and permissions are associated to an account
  - One can connect to a Rucio account using X509 certificates/proxies, Kerberos
  - One credential can be used to map to different accounts.



# Rucio basic concepts - 3

- Rucio namespace
  - 3 types of Data Identifiers (DIDs): File, Datasets, Containers.
    - Dataset: Collection of Files
    - Containers: Collection of Datasets or Containers
  - The namespace is divided using scopes. A name is unique within a scope but can be used in other scopes. A DID is identified by a scope and a name



# Rucio basic concepts - 4

- Rucio Storage Elements (RSEs)
  - Name for a storage endpoint in Rucio, e.g.: GOEGRID\_LOCALGROUPDISK or CERN-PROD\_DATADISK
  - Can be grouped using tags (e.g. tier=2, cloud=DE)
- Replication Rules
  - Define how to replicate Data Identifiers to Rucio Storage Elements
  - E.g.: Make one replica of dataset user.gen:my.dataset on DESY-HH\_LOCALGROUPDISK
  - Using RSE expression: 2 replicas at cloud=DE&type=LOCALGROUPDISK (any German LOCALGROUPDISK)
  - Will create the minimum number of replicas to optimise storage space and minimise transfers

# RSE expressions

- Rules can be created either with an exact RSE name or by using tags that are defined on an RSE, this is then called an RSE expression
  - Examples:
    - RSEs in German cloud: cloud=DE
    - LOCALGROUPDISKs in UK: country=UK&type=LOCALGROUPDISK
    - Any T2 in Italy but not INFN-NAPOLI: cloud=FR&tier=2\INFN-NAPOLI
  - More about this can be found in the Rucio documentation:
    - [http://rucio.cern.ch/replication\\_rules\\_examples.html](http://rucio.cern.ch/replication_rules_examples.html)
- When using RSE expressions you can define a replication factor, so you can create multiple replicas for one datasets with one rule
- Also you can define the grouping of the data:
  - ALL: Rucio selects an RSE and all files will be copied to this RSE
  - DATASET: If there are multiple datasets Rucio will pick an RSE for each one and will copy all files in the same dataset to the same RSE
  - NONE: Rucio will pick a new RSE for every file, so that they spread over all available RSEs

# Links and references

- RucioUI
  - <https://rucio-ui.cern.ch/>
- Rucio Documentation
  - <http://rucio.cern.ch/index.html>
- Software twiki tutorial
  - <https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/SoftwareTutorialGettingDatasets>
- Athena Docker setup
  - <https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/AthenaMacDockerSetup>
- Docker container for CVMFS
  - <https://github.com/sbinet/docker-containers/tree/master/cvmfs-atlas>
- Binet, Sébastien, and Ben Couturier. "*docker & HEP: Containerization of applications for development, distribution and preservation.*" Journal of Physics: Conference Series. Vol. 664. No. 2. IOP Publishing, 2015.
  - <http://iopscience.iop.org/article/10.1088/1742-6596/664/2/022007/meta>
- ATLAS-D meeting 2015 Rucio Tutorial, Thomas Beermann
- Monitoring Your Grid Jobs, Andrew Washbrook University of Edinburgh, ATLAS Software & Computing Tutorials 14th January 2015 PUC, Chile
- Athena Mac Docker
  - <https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/AthenaMacDockerSetup>
- Software tutorial using Grid
  - <https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/SoftwareTutorialUsingTheGrid>

# Enjoy your hands-on exercise!

# Backup

# ATLAS Resources - 1

- AGIS (ATLAS Grid Information System)
  - <http://atlas-agis.cern.ch/agis/>

ATLAS Grid Information System			
RC Site	ATLASSite	DDMEndpoint	PANDA Queue
Service	Central Services	DDM Groups	Docs
<ul style="list-style-type: none"><li>▪ Define RC site</li><li>▪ Define Experiment site</li><li>▪ Define DDM endpoint</li><li>▪ <b>Define OS RSE endpoint (new implementation)</b></li><li>▪ Define PANDA site</li><li>▪ Define PANDA queue</li><li>▪ RC pledges</li><li>▪ Find DDM endpoints links</li><li>▪ Find TransferMatrix links</li></ul>	<ul style="list-style-type: none"><li>▪ <b>Define OS service</b></li><li>▪ Define LFC service</li><li>▪ Define SE service</li><li>▪ Define CE service</li><li>▪ Define Redirector service</li><li>▪ Define PerfSonar service</li><li>▪ Define Frontier service</li><li>▪ Define Squid service</li><li>▪ Define Central service</li><li>▪ <b>SE protocols (DDM/Panda activities)</b></li></ul>	<ul style="list-style-type: none"><li>▪ Crons list</li><li>▪ ADMINs list</li><li>▪ Changes log</li><li>▪ <b>Request ADMIN privileges</b></li></ul>	<ul style="list-style-type: none"><li>▪ Main TWiki</li><li>▪ TWiki WEBUI instructions</li><li>▪ API Docs</li></ul>
<b>DOWNTIMES</b>	<b>TOACACHE EXPORT</b>	<b>COMPARISON &amp; VALIDATION TOOLS</b>	
<ul style="list-style-type: none"><li>▪ Downtime calendar</li><li>▪ DDM Blacklisting data</li><li>▪ PANDA Blacklisting data</li></ul>	<ul style="list-style-type: none"><li>▪ <b>dynamic ToACache (changes are immediately propagated):</b> <a href="http://atlas-agis-api.cern.ch/request/toacache/TiersOfATLASCache.py">http://atlas-agis-api.cern.ch/request/toacache/TiersOfATLASCache.py</a></li><li>▪ <b>static ToACache:</b> <a href="http://atlas-agis-api.cern.ch/ToACache/TiersOfATLASCache.py">http://atlas-agis-api.cern.ch/ToACache/TiersOfATLASCache.py</a></li><li>▪ <b>previous caches:</b> <a href="http://atlas-agis-api.cern.ch/ToACache/cache/">http://atlas-agis-api.cern.ch/ToACache/cache/</a><ul style="list-style-type: none"><li>▪ ToACache with Extra data</li></ul></li><li>▪ View/Modify ToACache ExtraData (RSE integration)</li></ul>	<ul style="list-style-type: none"><li>▪ Consistency checker</li><li>▪ ToAComparator</li><li>▪ AGIS-BDII CE comparison</li><li>▪ AGIS-Schedconf-PF mon CE comparison</li><li>▪ AGIS-DIMGOCDB sites+services comparison</li><li>▪ AGIS-PANDA PandaResource+SWReleases comparison</li><li>▪ AGIS-Schedconfig (topology) comparison</li><li>▪ AGIS-Schedconfig JSON comparison</li><li>▪ AGIS-GSR services comparison</li></ul>	

# ATLAS Resources - 2

- PanDA queue end points

ATLAS Grid Information System

RC Site	ATLASSite	DDMEndpoint	PANDA Queue	Service	Central Services	DDM Groups	PandaQueue combined resources						Docs	TWiki	OLD	JSON						
Show 200 entries				First	Previous	1	Next	Last														
<a href="#">give me url of this page</a> <a href="#">hold shift click column for Multi-column ordering</a>														Status	Manual	HC	Switcher	Panda Integration	CLOUD	Final		
atlas			FZK						ACTIVE													
VO name	▲	ATLAS Site	▲	PanDA Site	▲	Template object	▲	PanDA Resource	▲	PanDA Queue	▲	state	▲	(current) status	▲	type	▲	capability	▲	CLOUD	▲	TIER
atlas		FZK-LCG2		FZK-LCG2		FZK-LCG2_VIRTUAL		ANALY_FZK		ANALY_FZK		ACTIVE		online		analysis		score		DE		T1
atlas		FZK-LCG2		FZK-LCG2		FZK-LCG2_VIRTUAL		ANALY_FZK_HI		ANALY_FZK_HI		ACTIVE		online		analysis		score		DE		T1
atlas		FZK-LCG2		FZK-LCG2		FZK-LCG2_VIRTUAL		ANALY_FZK_SHORT		ANALY_FZK_SHORT		ACTIVE		online		analysis		score		DE		T1
atlas		FZK-LCG2		FZK-LCG2		FZK-LCG2_VIRTUAL		FZK-LCG2		FZK-LCG2-all-prod-CEs		ACTIVE		online		production		score		DE		T1
atlas		FZK-LCG2		FZK-LCG2		FZK-LCG2_VIRTUAL		FZK-LCG2_HIMEM		FZK-LCG2_HIMEM		ACTIVE		online		production		himem		DE		T1
atlas		FZK-LCG2		FZK-LCG2		FZK-LCG2_VIRTUAL		FZK-LCG2_MCORE		FZK-LCG2_MCORE		ACTIVE		online		production		mcose		DE		T1
atlas		FZK-LCG2		FZK-LCG2		FZK-LCG2_VIRTUAL		FZK-LCG2_MCORE_HI		FZK-LCG2_MCORE_HI		ACTIVE		online		production		mcose		DE		T1
atlas		FZK-LCG2		FZK-LCG2		FZK-LCG2_VIRTUAL		FZK-LCG2_MCORE_LO		FZK-LCG2_MCORE_LO		ACTIVE		online		production		mcose		DE		T1

Active PanDA (job) queues  
ANALY\_ .... = Analysis queue  
ATLAS-D Physics Meeting 2018

# ATLAS Resources - 3

- DDM end points

**ATLAS Grid Information System**

RC Site ATLASTSite **DDMEndpoint** PANDA Queue Service Central Services DDM Groups **DDM Endpoints** Docs TWiki OLD JSON

Show 200 entries **FZK** First Previous 1 Next Last

give me url of this page hold shift + click column for Multi-column ordering

DDM Endpoint	State	DDM Site	ATLAS Site	ATLAS TIER	CLOUD	type	Full Endpoint	FTS Master	FTS Test
FZK-LCG2_DATADISK	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	DATADISK	token:ATLASDATADISK:srm://atlassrm-fzk.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/atlas/disk-only/atlasdatadisk/	CERNFTS3_REST   https://fts3.cern.ch:8446	CERNFTS3PILOT_REST   https://fts3-pilot.cern.ch:8446
FZK-LCG2_DATATAPE	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	DATATAPE	token:ATLASDATATAPE:srm://atlassrm-fzk.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/atlas/atlasdatatape/	CERNFTS3_REST   https://fts3.cern.ch:8446	CERNFTS3PILOT_REST   https://fts3-pilot.cern.ch:8446
FZK-LCG2_GROUPTAPE_PERF-EGAMMA	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	GROUPTAPE	token:ATLASMCTAPE:srm://atlassrm-fzk.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/atlas/atlasgrouptape/perf-egamma/	CERNFTS3_REST   https://fts3.cern.ch:8446	CERNFTS3PILOT_REST   https://fts3-pilot.cern.ch:8446
FZK-LCG2_LOCALGROUPDISK	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	LOCALGROUPDISK	token:ATLASLOCALGROUPDISK:srm://dgridsrm-fzk.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/datas/atlaslocalgroupdisk/	CERNFTS3_REST   https://fts3.cern.ch:8446	CERNFTS3PILOT_REST   https://fts3-pilot.cern.ch:8446
FZK-LCG2_LOCALGROUPTAPE	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	LOCALGROUPTAPE	token:ATLASLOCALGROUPTAPE:srm://dgridsrm-fzk.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/datas/atlaslocalgrouptape/	CERNFTS3_REST   https://fts3.cern.ch:8446	CERNFTS3PILOT_REST   https://fts3-pilot.cern.ch:8446
FZK-LCG2_MCTAPE	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	MCTAPE	token:ATLASMCTAPE:srm://atlassrm-fzk.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/atlas/atlasmctape/	CERNFTS3_REST   https://fts3.cern.ch:8446	CERNFTS3PILOT_REST   https://fts3-pilot.cern.ch:8446
FZK-LCG2_PERF-EGAMMA	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	GROUPDISK	token:ATLASDATADISK:srm://atlassrm-fzk.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/atlas/disk-only/atlasgroupdisk/perf-egamma/	CERNFTS3_REST   https://fts3.cern.ch:8446	CERNFTS3PILOT_REST   https://fts3-pilot.cern.ch:8446
FZK-LCG2_PERF-IDTRACKING	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	GROUPDISK	token:ATLASDATADISK:srm://atlassrm-fzk.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/atlas/disk-only/atlasgroupdisk/perf-idtracking/	CERNFTS3_REST   https://fts3.cern.ch:8446	CERNFTS3PILOT_REST   https://fts3-pilot.cern.ch:8446
FZK-LCG2_PERF-TAU	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	GROUPDISK	token:ATLASDATADISK:srm://atlassrm-fzk.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/atlas/disk-only/atlasgroupdisk/perf-tau/	CERNFTS3_REST   https://fts3.cern.ch:8446	CERNFTS3PILOT_REST   https://fts3-pilot.cern.ch:8446
FZK-LCG2_PPSSCRATCHDISK	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	SPECIAL	token:ATLASPPSSCRATCHDISK:srm://ppssrm-kit.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/atlas/atlasppsscratchdisk/	CERNFTS3_REST   https://fts3.cern.ch:8446	
FZK-LCG2_SCRATCHDISK	ACTIVE	FZK-LCG2	FZK-LCG2	T1	DE	SCRATCHDISK	token:ATLASSCRATCHDISK:srm://atlassrm-fzk.gridka.de:8443/srm/managerv?SFN=/pnfs/gridka.de/atlas/disk-only/atlasscratchdisk/	CERNFTS3_REST   https://fts3.cern.ch:8446	CERNFTS3PILOT_REST   https://fts3-pilot.cern.ch:8446

Showing 1 to 11 of 11 entries

**Active DDM storage end points**

# ATLAS Resources - 4

- **SCRATCHDISK** (Tier1 + Tier2s in Germany)
  - FZK-LCG2\_SCRATCHDISK
  - DESY-HH\_SCRATCHDISK
  - DESY-ZN\_SCRATCHDISK
  - LRZ-LMU\_SCRATCHDISK
  - WUPPERTALPROD\_SCRATCHDISK
  - UNI-FREIBURG\_SCRATCHDISK
  - GOEGRID\_SCRATCHDISK
- **LOCALGROUPDISK** (e.g. DESY-HH and UniGoettingen)
  - DESY-HH\_LOCALGROUPDISK
  - GOEGRID\_LOCALGROUPDISK
  - ..... \_LOCALGROUPDISK

# ATLAS Resources - 4

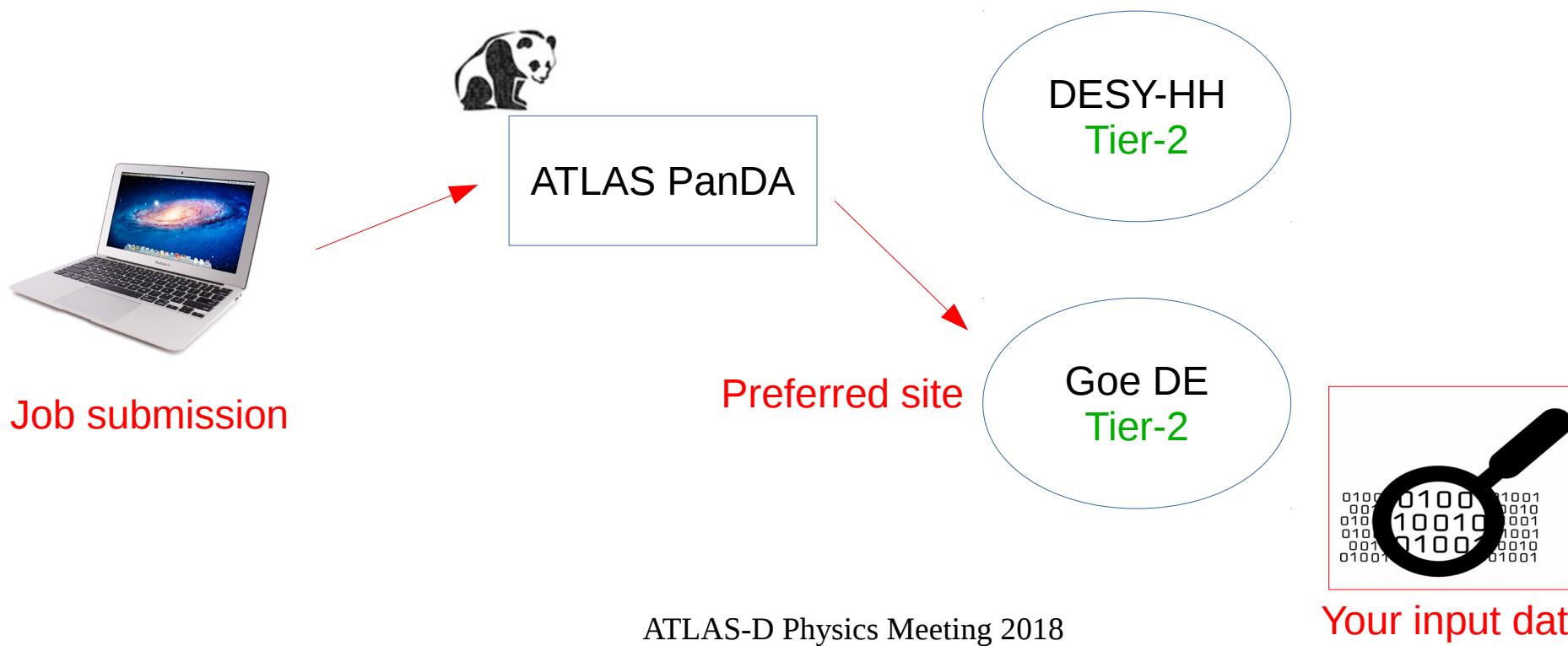
- **SCRATCHDISK** (Tier1 + Tier2s in Germany)
  - FZK-LCG2\_SCRATCHDISK
  - DESY-HH\_SCRATCHDISK
  - DESY-ZN\_SCRATCHDISK
  - LRZ-LMU\_SCRATCHDISK
  - WUPPERTALPROD\_SCRATCHDISK
  - UNI-FREIBURG\_SCRATCHDISK
  - GOEGRID\_SCRATCHDISK

Storages for temporary data of PanDA jobs. Would be automatically ***REMOVED!***
- **LOCALGROUPDISK** (e.g. DESY-HH and UniGoettingen)
  - DESY-HH\_LOCALGROUPDISK
  - GOEGRID\_LOCALGROUPDISK
  - ..... \_LOCALGROUPDISK

Permanently ***KEPT.*** Generally speaking, in total a few hundred TB in each site

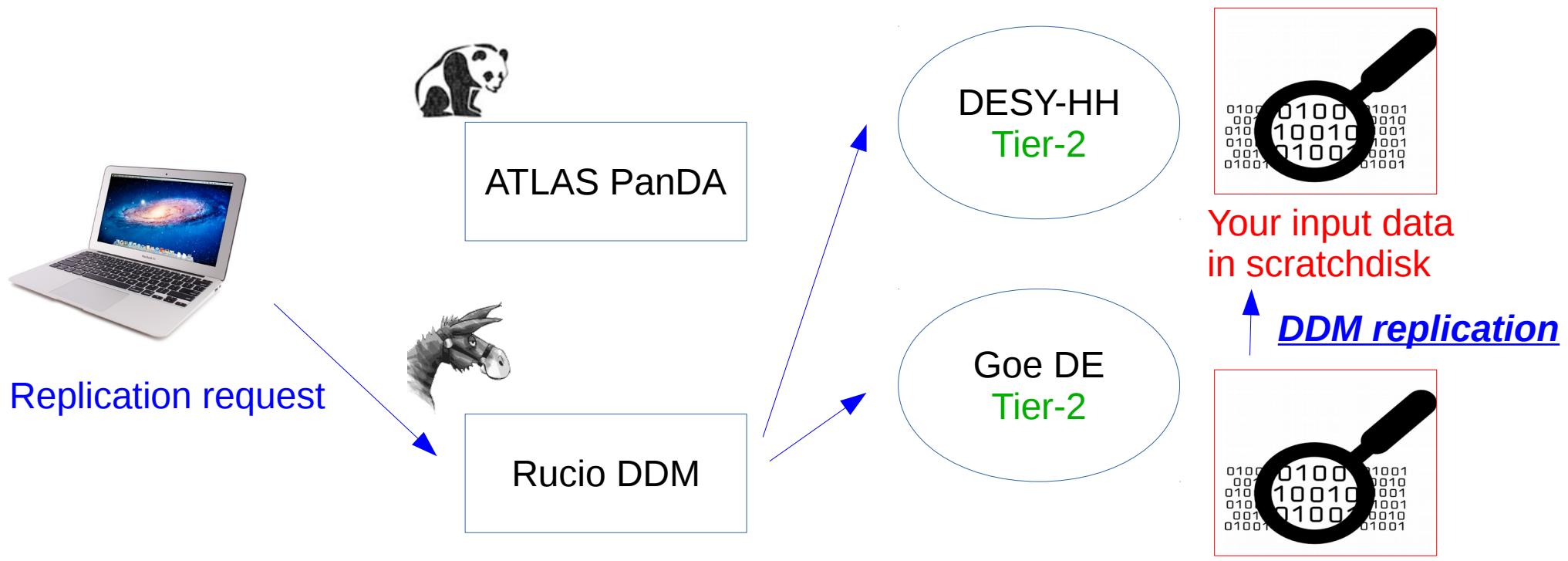
# Job allocation (a tip) - 1

- A tip: User job allocation policy among sites
  - Rule: Grid jobs (should) go to their data locations



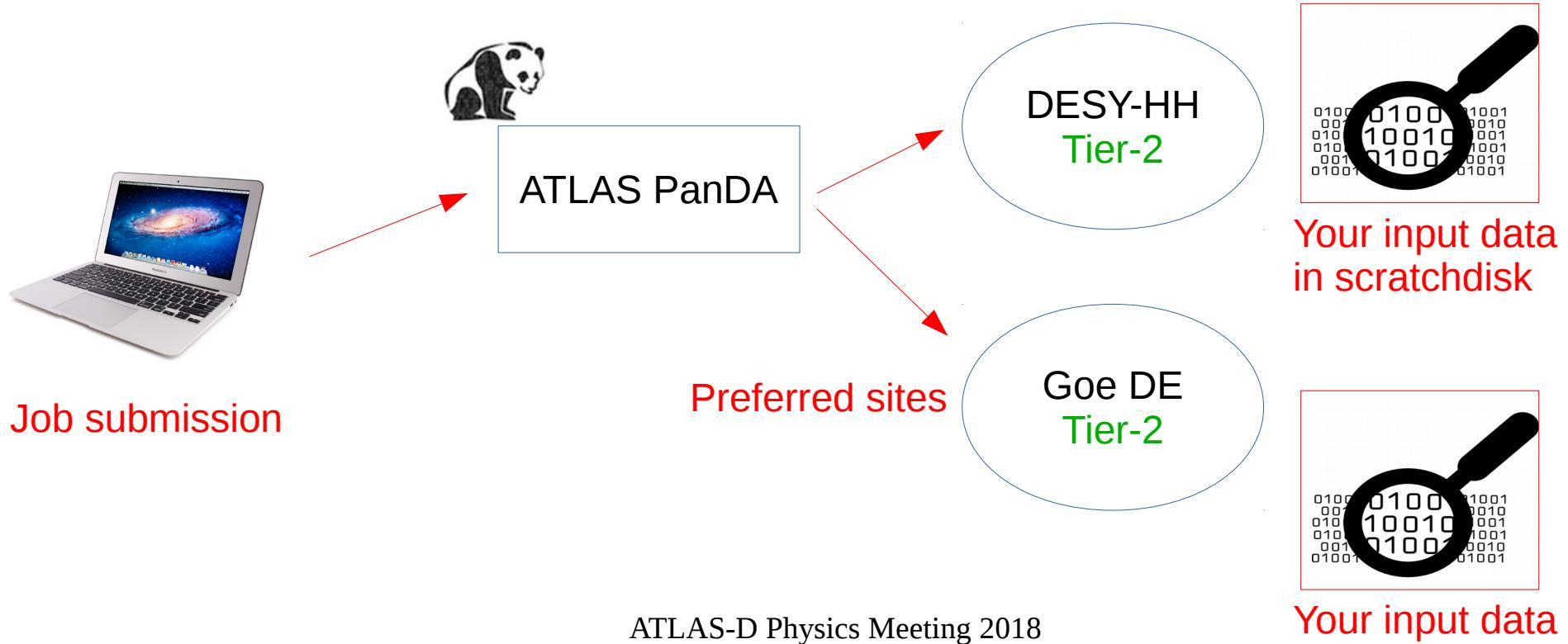
# Job allocation (a tip) - 2

- A tip: User job allocation policy among sites
  - Rule: Grid jobs (should) go to their data locations



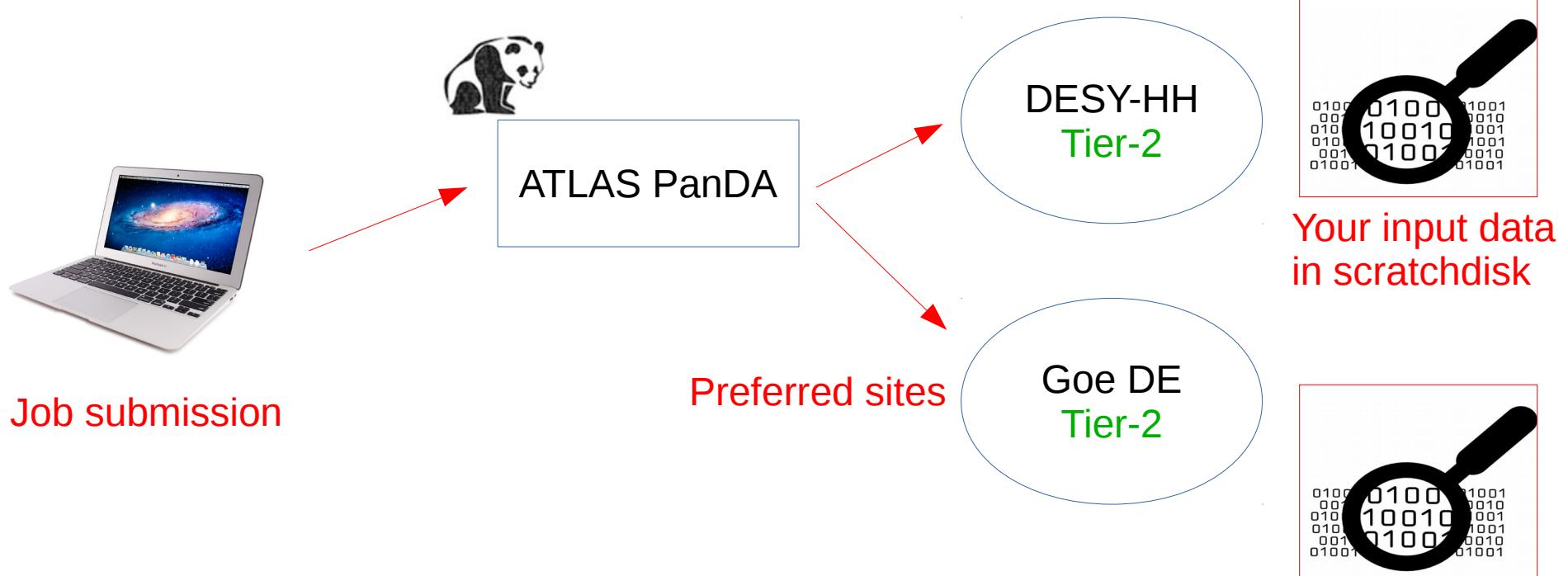
# Job allocation (a tip) - 3

- A tip: User job allocation policy among sites
  - Rule: Grid jobs (should) go to their data locations



# Job allocation (a tip) - 4

- A tip: User job allocation policy among sites
  - Rule: Grid jobs (should) go to their data locations



Local access is better than remote access  
(or to avoid an error in PanDA, “local data does not exist”)

ATLAS-D Physics Meeting 2016