



ATLAS Grid ジョブ管理と AGIS

ATLAS ソフトウェア講習会 2016

河村 元

II.Physikalisches Institut, Universität Göttingen

Overview

- ATLAS Grid ジョブ管理
 - PanDA
 - WLCG の基本構造
 - ATLAS ジョブとデータ
 - おまけ：ジョブ実行前のデータレプリケーション
- ATLAS の計算資源情報 (AGIS)
 - ATLAS の資源とサイト
 - ATLAS の計算資源情報
 - PanDA ジョブ・キュー
 - Rucio ストレージ・エンド・ポイント
- リンクと参考文献

ATLAS Grid ジョブ管理

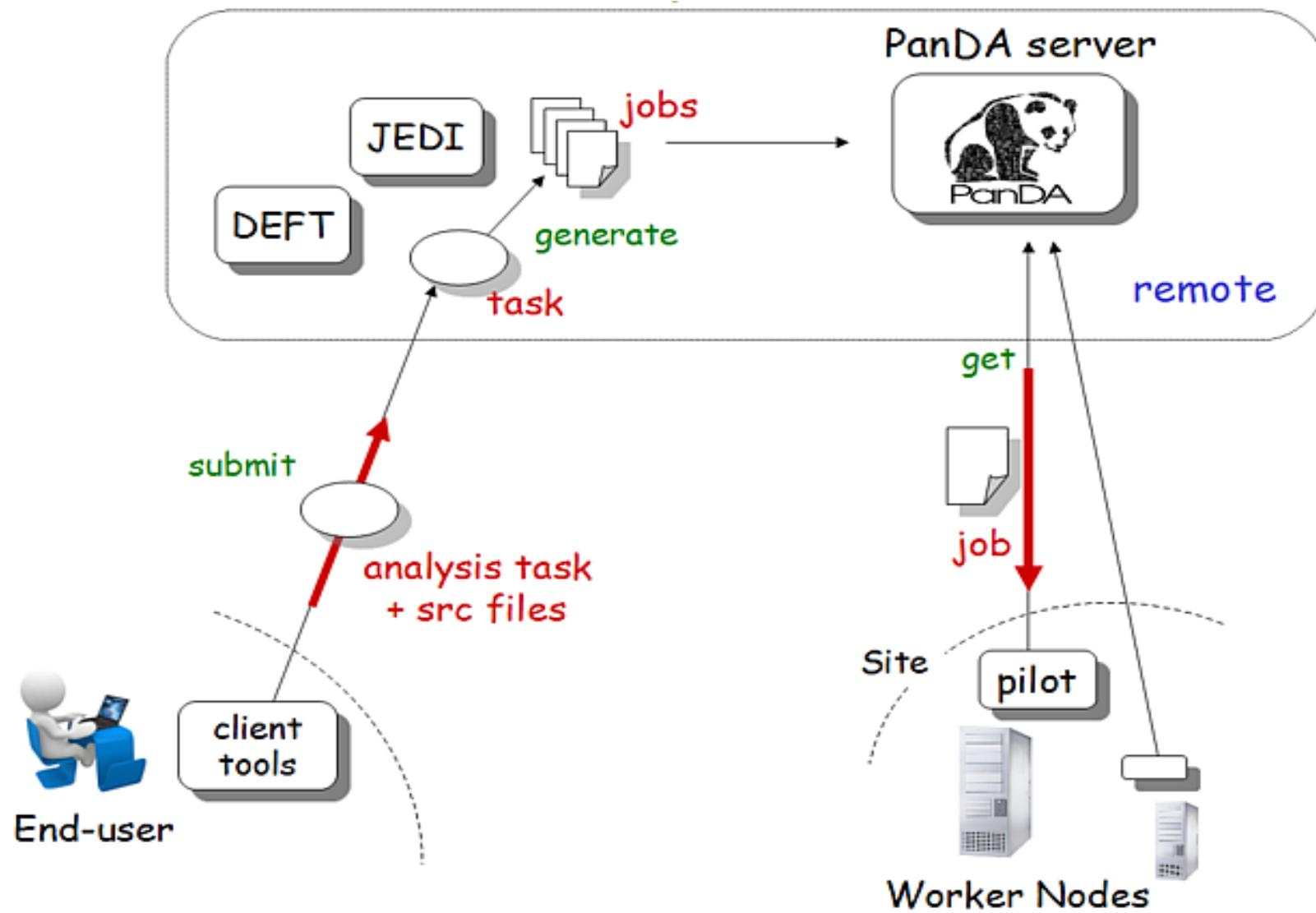


WLCG
Worldwide LHC Computing Grid

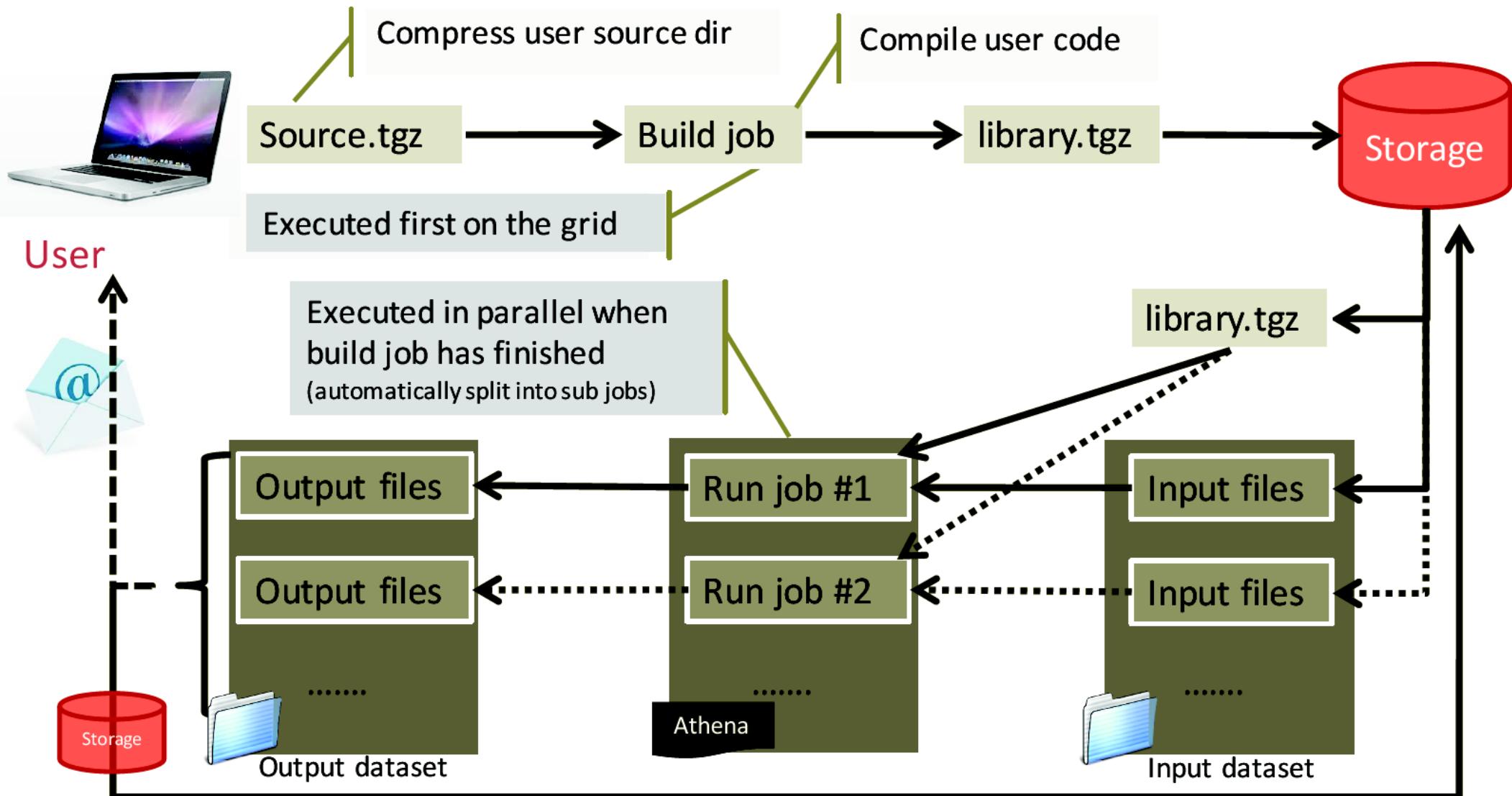
PanDA

- PanDA (Production ANd Distributed Analysis system) ATLAS 分散ジョブ管理システム
 - PanDA クライアントはジョブ管理用のクライアント・ツールセット
 - pathena - athena ジョブサブミッションツール
 - prun - python スクリプト用ジョブサブミッションツール
 - psequencer - ジョブ逐次投入用ツール。異なったシークエンスのタスク投入等のため。
 - pbook - ジョブ・ブックキーピングツール。ジョブ再投入や停止など。

PanDA ジョブワークフロー

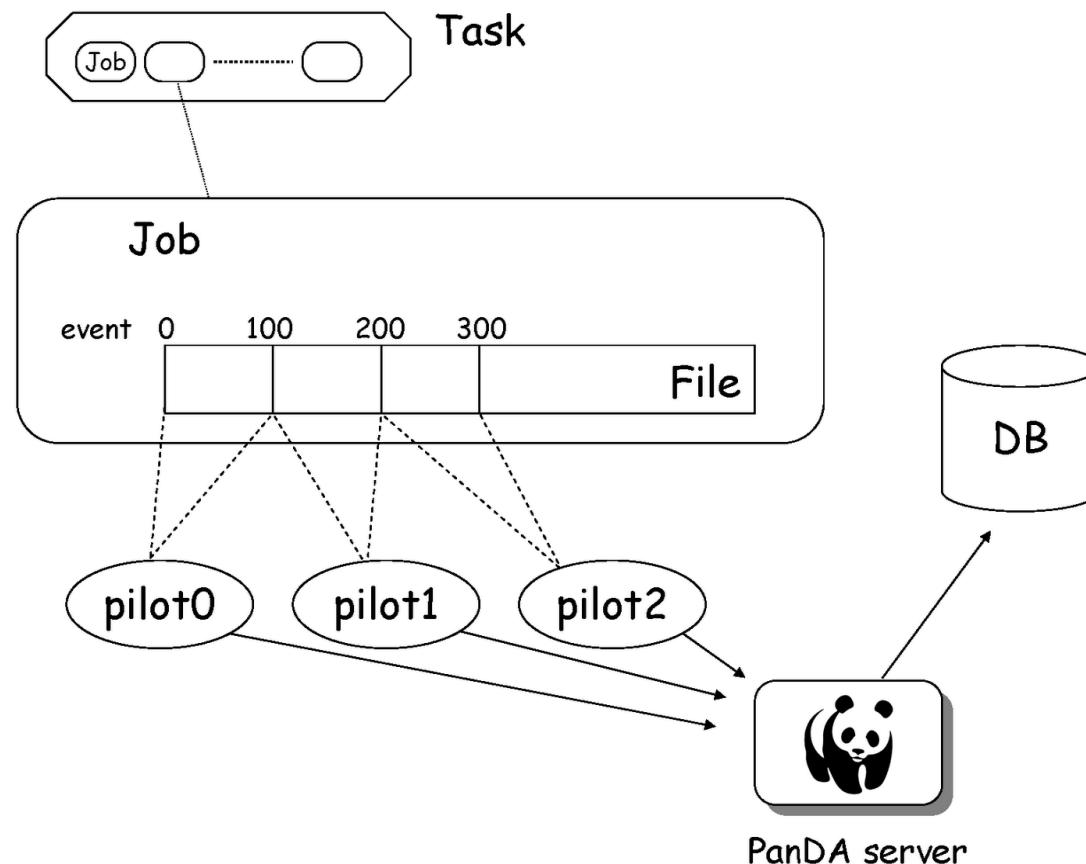


Build job & Run job



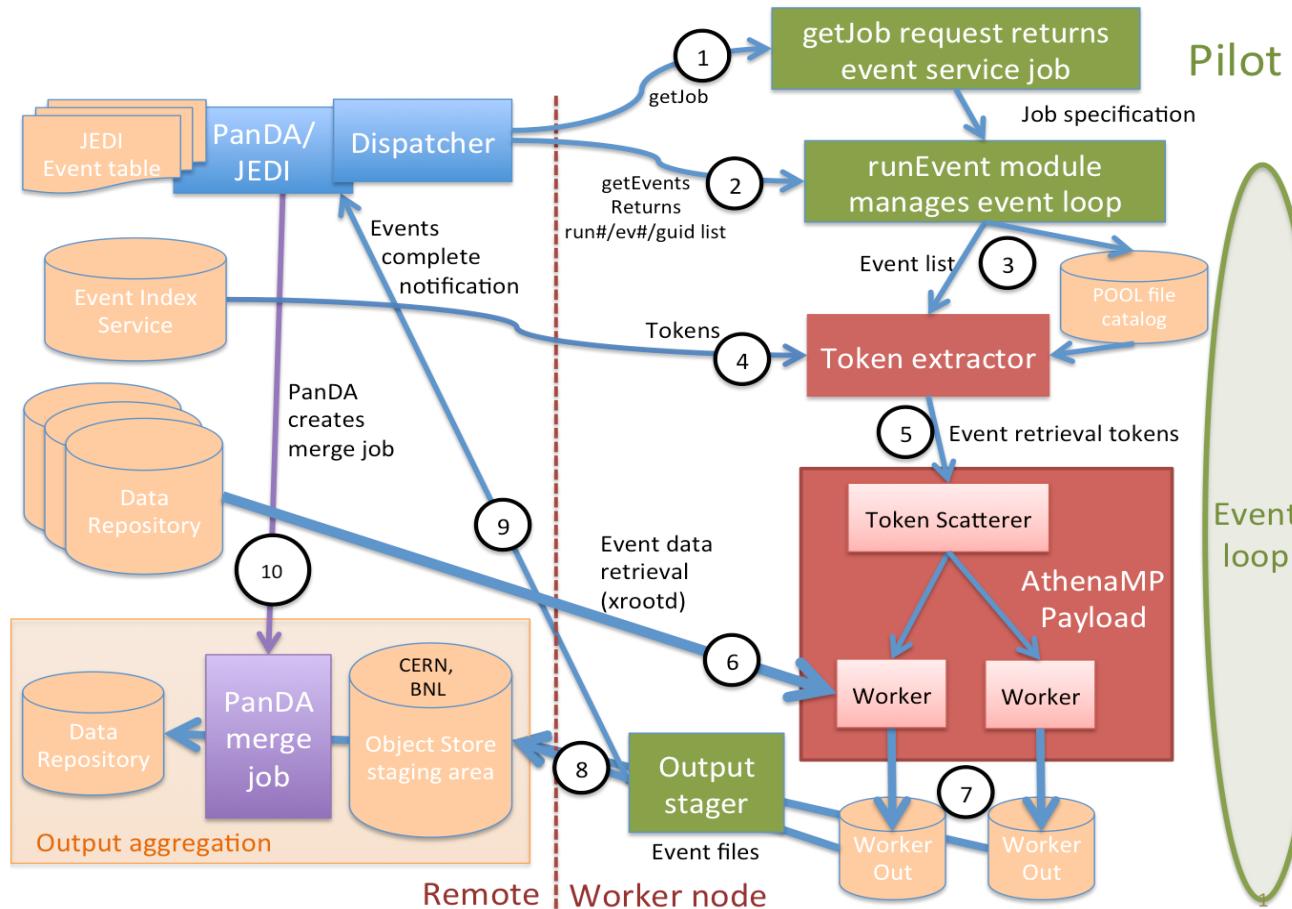
イベント・レベル処理

- PanDA は Task を物理イベント・レベルで Grid ジョブへ分割可能
 - JEDI が Task を event ベースで pilot ジョブへ



イベント・レベル処理

- PanDA は Task を物理イベント・レベルで Grid ジョブへ分割可能
 - JEDI が Task を event ベースで pilot ジョブへ



prun

- “Hello world” ジョブ

```
## PanDA client
SetupATLAS
lsetup panda

## Make a Python script
cat hello_world.py

#!/usr/bin/python
print "Hello world!"

chmod 755 hello_world.py
./hello_world.py

Hello world!

## Submitting a prun job
prun --outDS user.$USER.pruntest.$$ --exec hello_world.py

INFO : gathering files under /home/gen/tmp/for_new_comer
INFO : upload source files
INFO : submit
INFO : succeeded. new jediTaskID=5107461

## Submitting 5 prun jobs
prun --outDS user.$USER.pruntest.$$ --exec hello_world.py -nJobs=5
```

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	<input type="text"/>	<input type="button" value="Submit"/>
Batch ID	<input type="text"/>	<input type="button" value="Submit"/>
Task ID	<input type="text" value="4212437"/>	<input type="button" value="Submit"/>
Task name	<input type="text"/>	<input type="button" value="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\(1\)](#) [tmpl_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.

Task ID: 4212437 | **User:** Nurcan Ozturk | **Status:** done | **Ninputfiles:** 1 | **finished:** 1 (100%) | **failed:** 0 (0%)

States of jobs in this task: defined, waiting, pending, assigned, throttled, activated, sent, starting, running, holding, transferring, finished, failed, cancelled, merging. (The 'finished' state has a value of 2, indicated by a red arrow.)

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/ (indicated by a red arrow)

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	
panda.1007093351.25596.lib._4212437.241508507.lib.tgz (input)	panda	0	ready		panda.1007093351.25596.lib._4212437 (dispatch block: panda.1007093351.25596.lib._4212437)	
user.nozturk.pruntest.log.4212437.000001.log.tgz (log)	user.nozturk	0	finished		user.nozturk.pruntest.log.6779086	
user.nozturk.pruntest.log.4212437.000001.log.tgz (log)	user.nozturk	0	ready		user.nozturk.pruntest.log/ (destination block: sub0186192010)	
pseudo_lfn (pseudo_input)		0	finished		pseudo_dataset	
pseudo_lfn (pseudo_input)		0	unknown		pseudo_dataset	

Email notifications from PanDA

From: atlpna@cern.ch
Subject: JEDI notification for TaskID:4212437 (1/1 All Succeeded)
Date: October 7, 2014 11:50:08 AM GMT+02:00
To: Nurcan Ozturk

Summary of TaskID:4212437

Created : 2014-10-07 09:33:00 (UTC)
Ended : 2014-10-07 09:50:08.720047 (UTC)

Final Status : done

Total Number of Inputs : 1
Succeeded : 1
Failed : 0
Cancelled : 0

Error Dialog : None

Log : user.nozturk.pruntest.log/

Parameters : prun --outDS user.nozturk.pruntest --exec HelloWorld.py

PandaMonURL : <http://bigpanda.cern.ch/task/4212437/>

Report Panda problems of any sort to

the eGroup for help request
hn-atlas-dist-analysis-help@cern.ch

the JIRA portal for software bug
<https://its.cern.ch/jira/browse/ATLASPANDA>

pathena

- Grid 上での Athena ジョブ
 - pathena

```
## For example, you can seamlessly run Athena code on Grid
wget http://cern.ch/tmaeno/jobOptions.pythia16.py
pathena jobOptions.pythia16.py --outDS user.<your user name>.pythiaEventGeneration

INFO : using CMTCONFIG=x86_64-slc5-gcc43-opt
INFO : extracting run configuration
INFO : ConfigExtractor > No Input
INFO : ConfigExtractor > Output=STREAM1 pythia.pool.root
INFO : ConfigExtractor > RndmStream PYTHIA
INFO : ConfigExtractor > RndmStream PYTHIA_INIT
INFO : archiving source files
INFO : archiving InstallArea
INFO : checking symbolic links
INFO : uploading source/jobO files
INFO : submit
INFO : succeeded. new jediTaskID=4212493
```

pbook

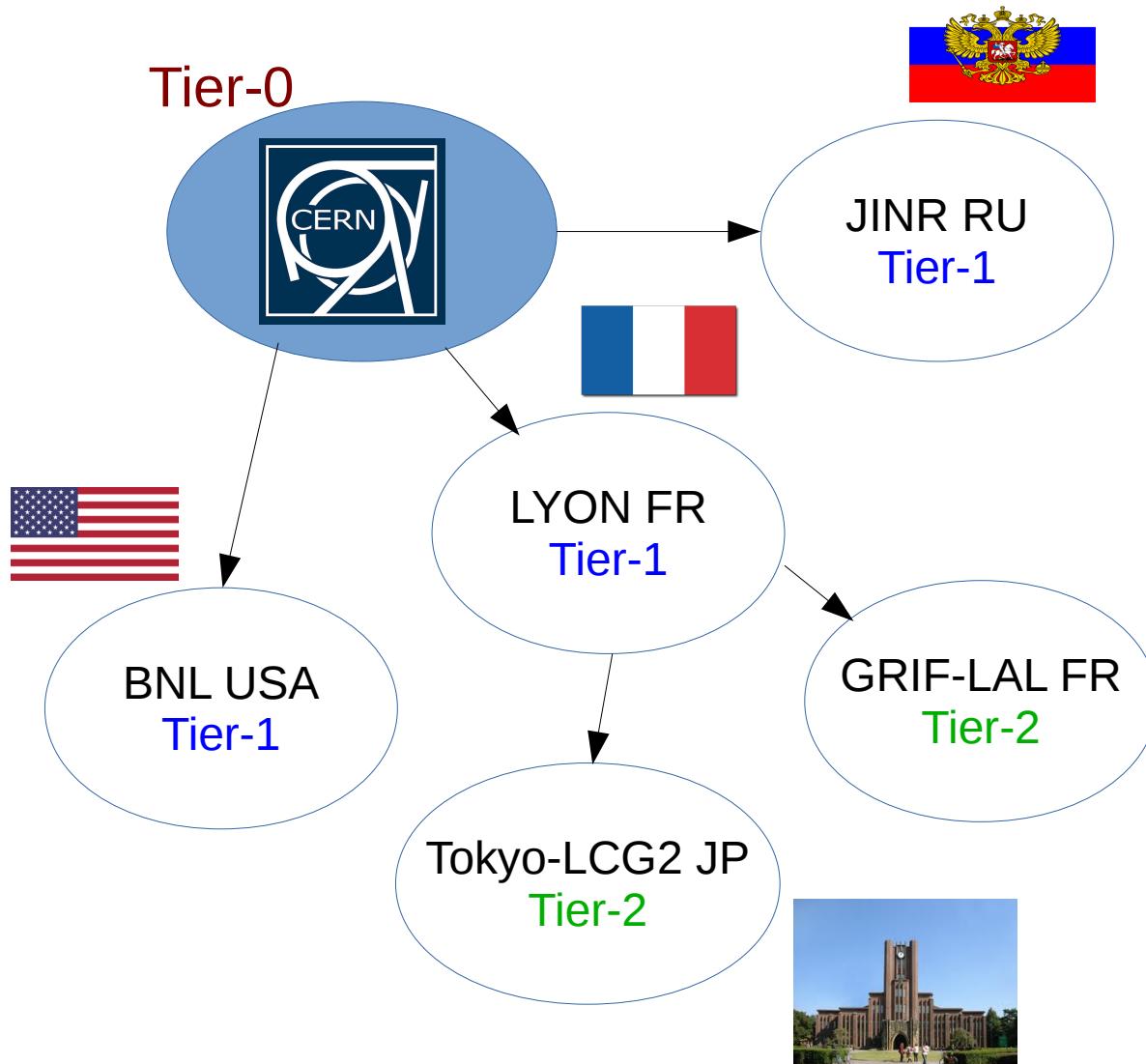
- ジョブの管理

```
## Launching PanDA book keeping interface
pbook
INFO : Synchronizing local repository ...
INFO : Got 0 jobs to be updated
INFO : Synchronization Completed
INFO : Done

Start pBook 0.5.72
>>> help()
>>> kill(4212493)
>>> retry(4212493)
>>> sync()
>>> show()
```

WLCG の基本構造

- LHC Tier 構造



WLCG
Worldwide LHC Computing Grid

Tier-0: RAW データ、テープ
ドライブ、プレプロセッシング、
リコンストラクション

Tier-1: 国レベル。テープド
ライブ有。巨大計算資源と
ディスク。

Tier-2: 大学や研究機関レベ
ルのサイト。

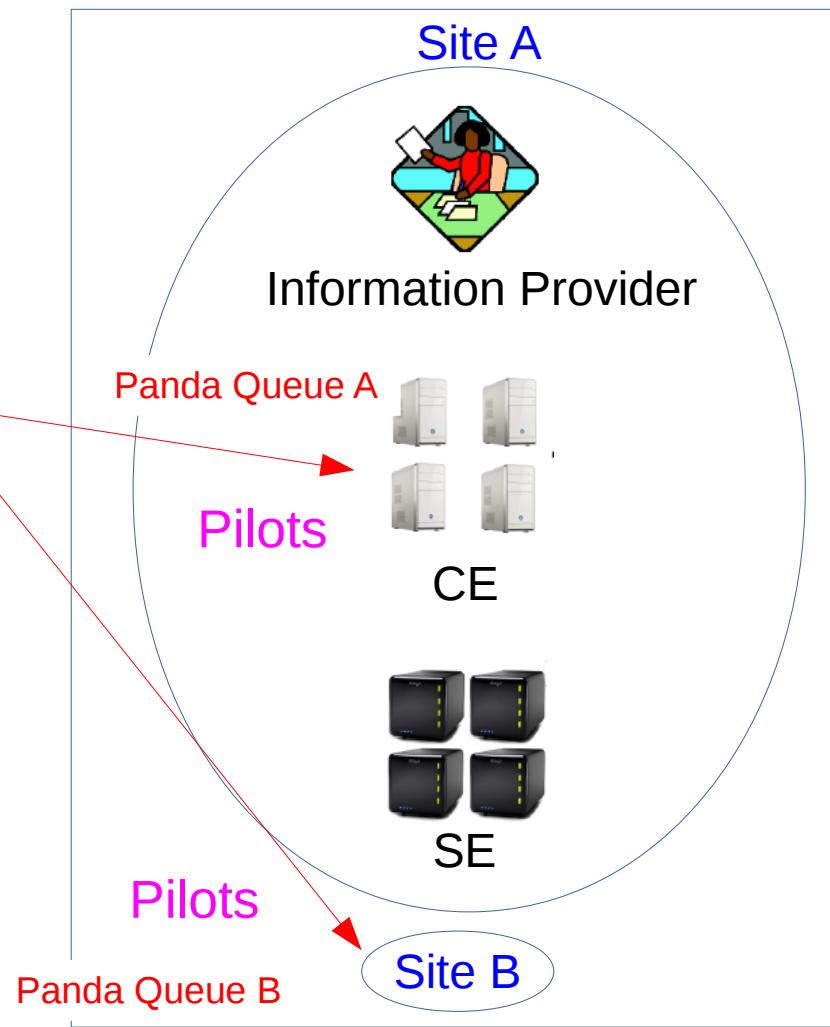
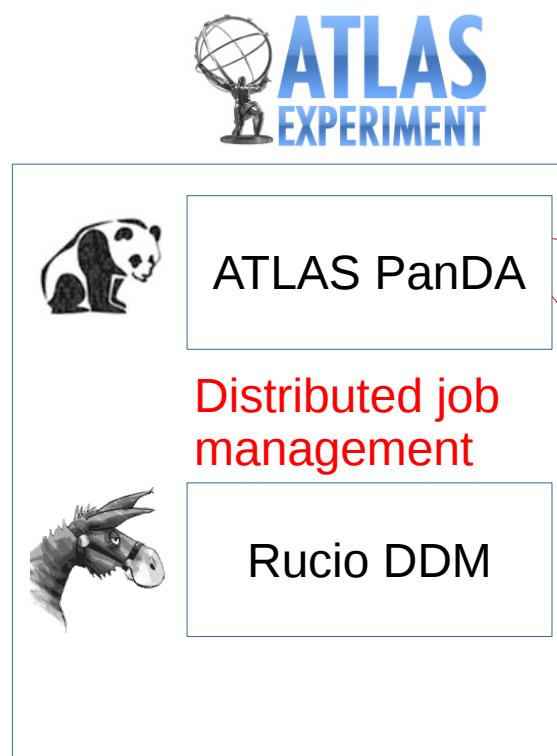
注) 階層構造自体は現在フ
レキシブルになりつつあ
り。データ転送等は各サイ
ト間のエンドポイント同士
で要求可能。

ATLAS ジョブとデータ



WLCG
Worldwide LHC Computing Grid

- あくまで概略

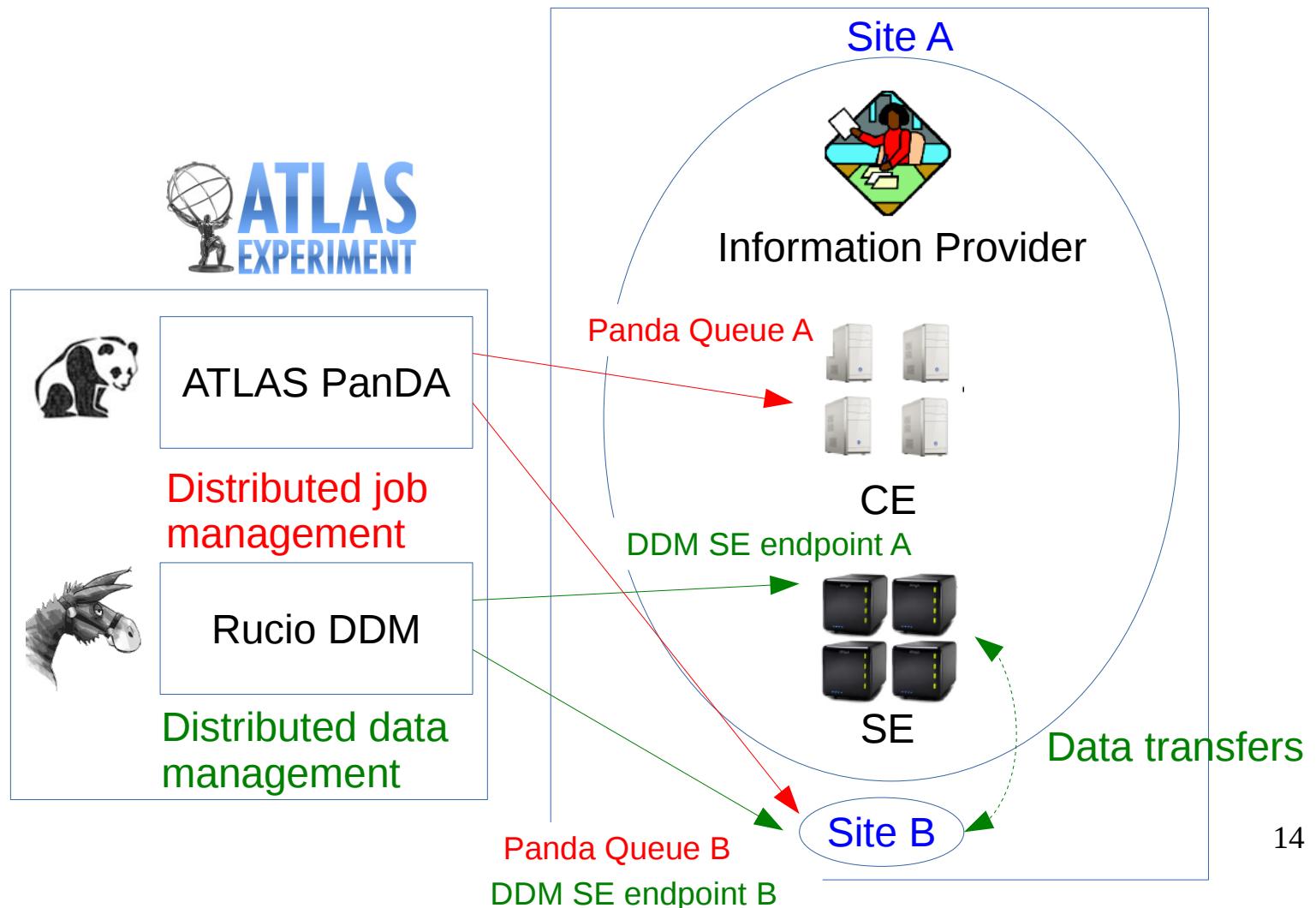


ATLAS ジョブとデータ



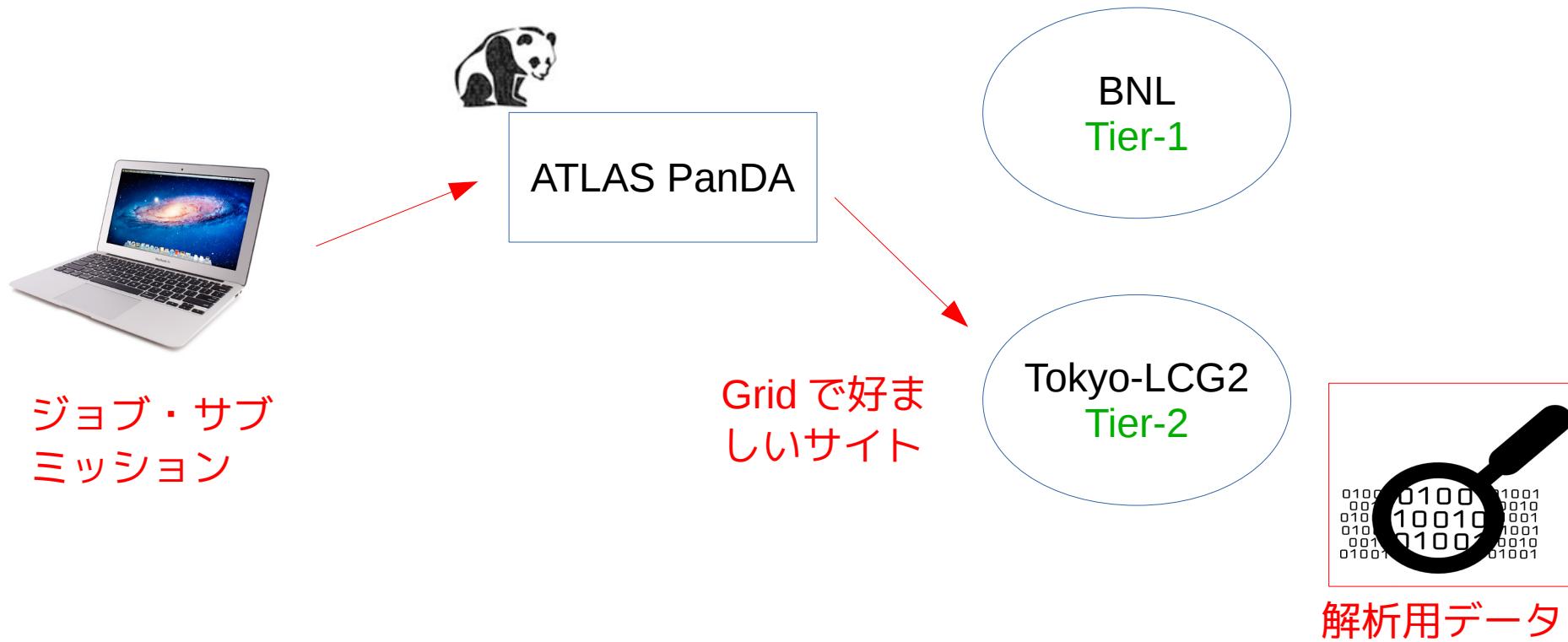
WLCG
Worldwide LHC Computing Grid

- あくまで概略



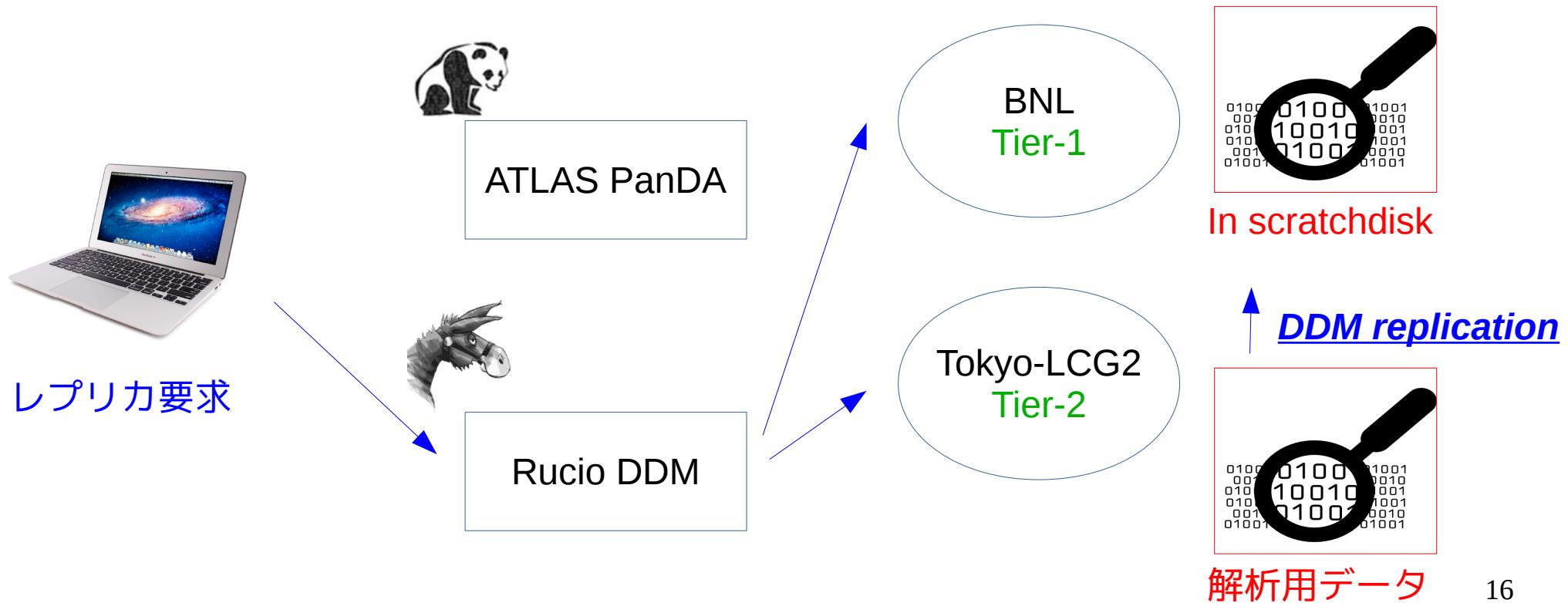
おまけ：ジョブ実行前のデータレプリケーション

- データエラー やクラッシュ 等理解不能な状況にはまって抜け出せない場合に試す価値あり
 - 大量のデータ処理を行うジョブにありがち



おまけ：ジョブ実行前のデータレプリケーション

- レプリケーションを作ったあとにジョブ実行
 - 理由: Grid ジョブはデータロケーションを好む。
ローカルなデータアクセスは圧倒的に速い



ATLAS の計算資源情報（AGIS）



WLCG
Worldwide LHC Computing Grid

ATLAS の資源とサイト

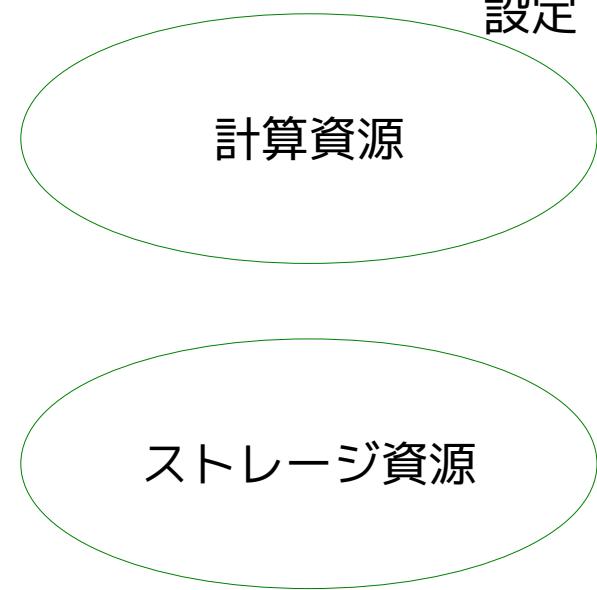
- FZK Tier-1



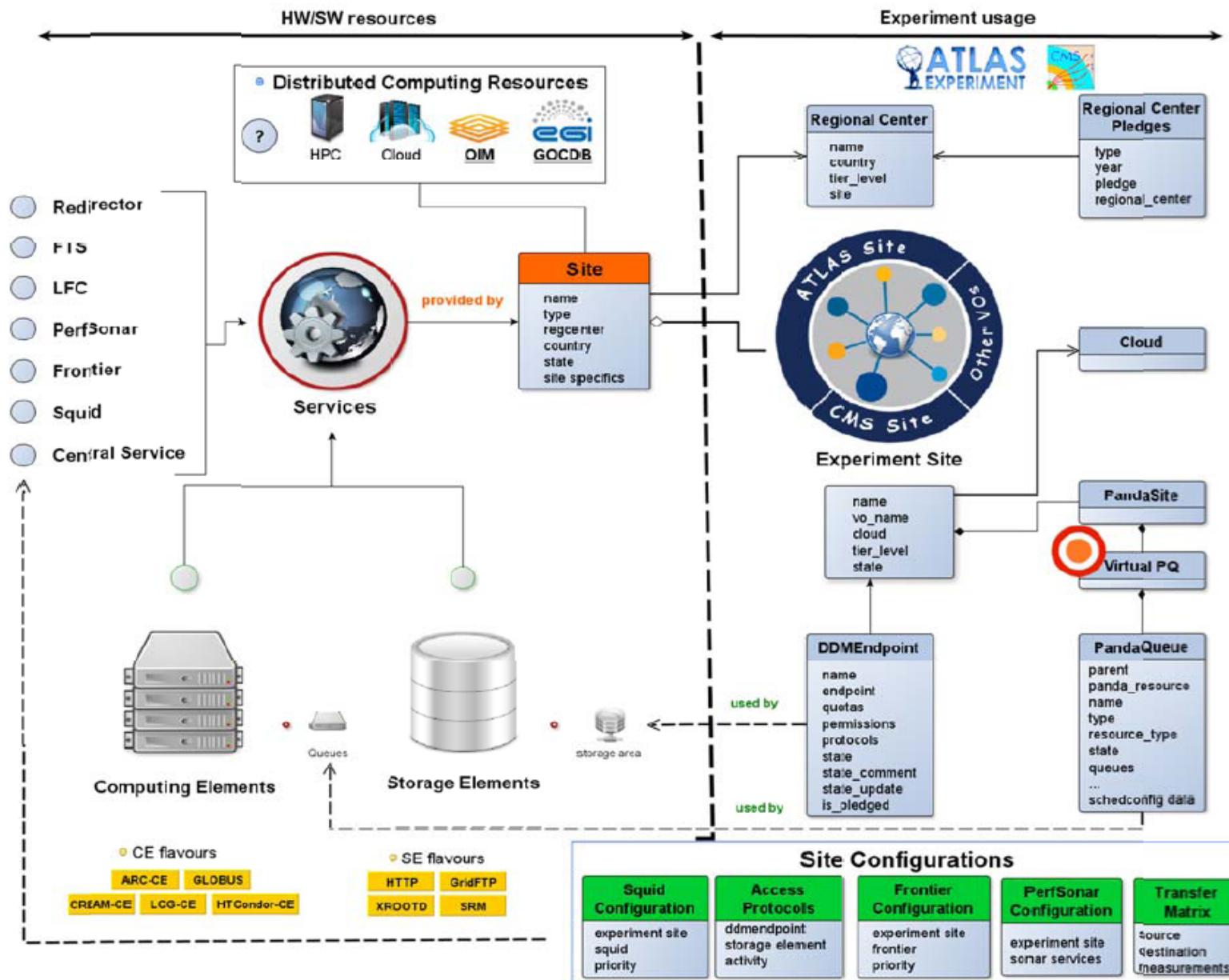
Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft



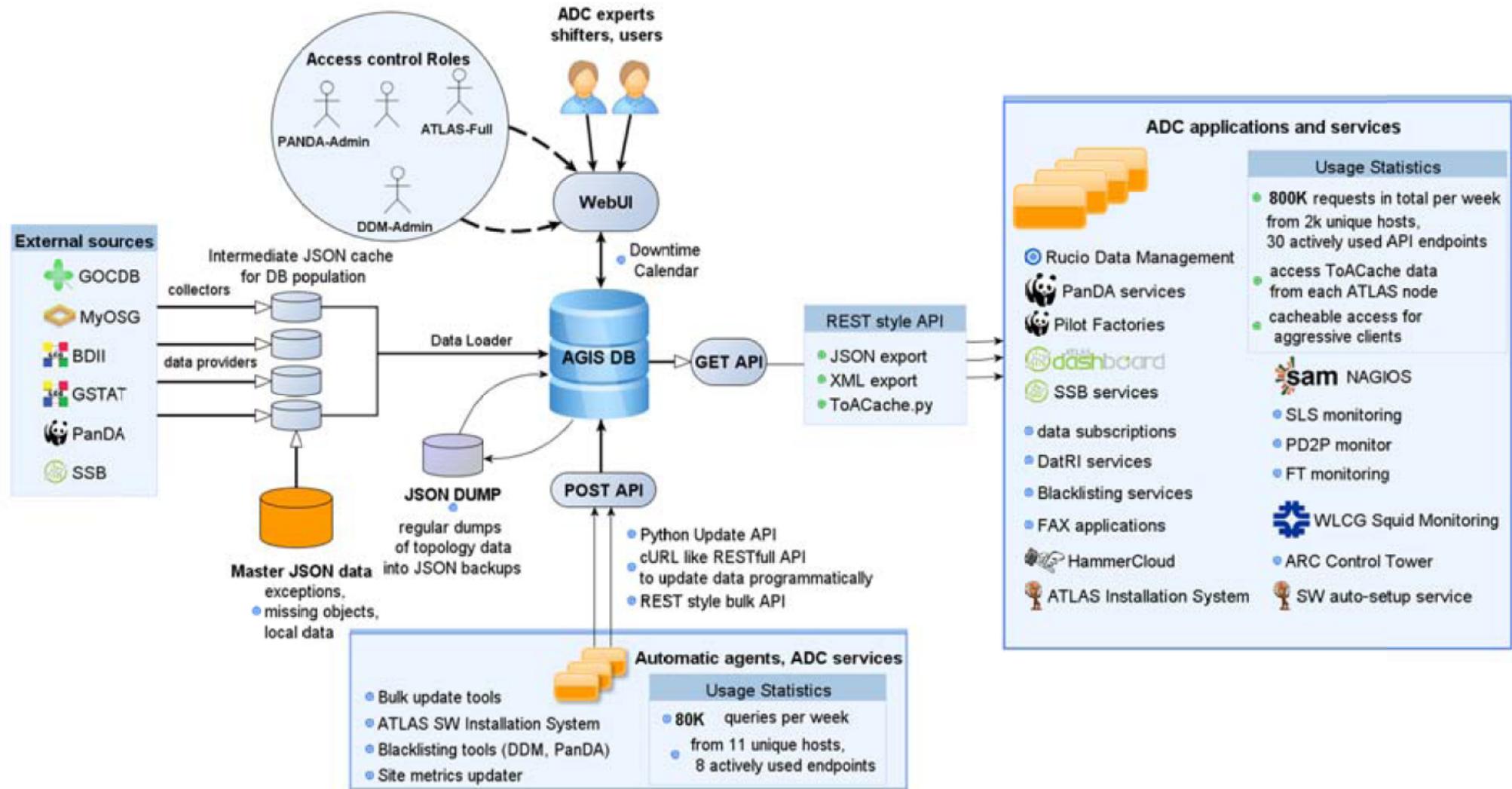
FZK Computing centre



ATLAS の資源とサイト - 抽象化（計算資源と実験）



ATLAS の資源とサイト - 実装



ATLAS の計算資源情報

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

PanDA ジョブ・キュー

- PanDA ジョブ・キュー・エンドポイント

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																		
													give me url of this page hold shift click column for Multi-column ordering						VO name	ATLAS Site	PanDA Site	Template object	PanDA Resource	PandaQueue	state	(current) status	type	capability	Final	Status	Manual	HC	Switcher	Panda Integration	CLOUD	TIER
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					Clone ANALY_FZK	ACTIVE	online	analysis	score	DE	T1																				
atlas		FZK-LCG2	FZK-LCG2	FZK-LCG2_VIRTUAL	ANALY_FZK_HI					Clone ANALY_FZK_HI	ACTIVE	online	analysis	score	DE	T1																				
atlas		FZK-LCG2	FZK-LCG2	FZK-LCG2_VIRTUAL	ANALY_FZK_SHORT					Clone ANALY_FZK_SHORT	ACTIVE	online	analysis	score	DE	T1																				
atlas		FZK-LCG2	FZK-LCG2	FZK-LCG2_VIRTUAL	FZK-LCG2					Clone FZK-LCG2-all-prod-CEs	ACTIVE	online	production	score	DE	T1																				
atlas		FZK-LCG2	FZK-LCG2	FZK-LCG2_VIRTUAL	FZK-LCG2_HIMEM					Clone FZK-LCG2_HIMEM	ACTIVE	online	production	himem	DE	T1																				
atlas		FZK-LCG2	FZK-LCG2	FZK-LCG2_VIRTUAL	FZK-LCG2_MCORE					Clone FZK-LCG2_MCORE	ACTIVE	online	production	mcose	DE	T1																				
atlas		FZK-LCG2	FZK-LCG2	FZK-LCG2_VIRTUAL	FZK-LCG2_MCORE_HI					Clone FZK-LCG2_MCORE_HI	ACTIVE	online	production	mcose	DE	T1																				
atlas		FZK-LCG2	FZK-LCG2	FZK-LCG2_VIRTUAL	FZK-LCG2_MCORE_LO					Clone FZK-LCG2_MCORE_LO	ACTIVE	online	production	mcose	DE	T1																				

アクティブな PanDA (job) キュー
ANALY_ = Analysis キュー

Rucio ストレージ・エンド・ポイント

- DDM Rucio Storage End points

ATLAS Grid Information System											Docs	TWiki	OLD	JSON							
RC Site	ATLASSite	DDMEndpoint	PANDA Queue	Service	Central Services	DDM Groups	DDM Endpoints														
FZK											First	Previous	1	Next	Last						
							DDM Endpoint	State	DDM Site	ATLAS Site	Site	ATLAS TIER	CLOUD	type	Full Endpoint	Token	Domain	LFC Name	FTS Master	FTS Master2	FTS Test
							Master3														
DDM Endpoint	State	ACTIVE	FZK											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/managerv2?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/managerv2?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/managerv2?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/managerv2?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/managerv2?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/managerv2?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/managerv2?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/managerv2?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/managerv2?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/managerv2?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/managerv2?SFN=/pnfs/gridka.de/atlas/disk-only/atlasscratchdisk/	CERNFTS3_REST https://fts3.cern.ch:8446	CERNFTS3PILOT_REST https://fts3-pilot.cern.ch:8446					

アクティブな DDM storage end points

リンクと参考文献

- Software twiki tutorial
 - <https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/SoftwareTutorialGettingDatasets>
- ATLAS-D meeting 2016 Grid/Rucio Tutorial, Gen Kawamura
- Monitoring Your Grid Jobs, Andrew Washbrook University of Edinburgh, ATLAS Software & Computing Tutorials 14th January 2015 PUC, Chile
- Software tutorial using Grid
 - <https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/SoftwareTutorialUsingTheGrid>
- ATLAS Software Tutorial, Feb 2016
 - <https://indico.cern.ch/event/465378/>
- Calafiura, Paolo, et al. "The ATLAS Event Service: A new approach to event processing." Journal of Physics: Conference Series. Vol. 664. No. 6. IOP Publishing, 2015.
- Anisenkov, Alexey, et al. "AGIS: Evolution of Distributed Computing information system for ATLAS." Journal of Physics: Conference Series. Vol. 664. No. 6. IOP Publishing, 2015.