This work is partially supported by projects InterExcellence (LTT17018), Research infrastructure CERN (CERN-CZ, LM2015058) and OP RDE CERN Computing (CZ.02.1.01/0.0/0.0/1 6013/0001404) from EU funds and MEYS.

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS

Exploitation of heterogeneous resources for ATLAS Computing

Jiří Chudoba on behalf of the ATLAS collaboration

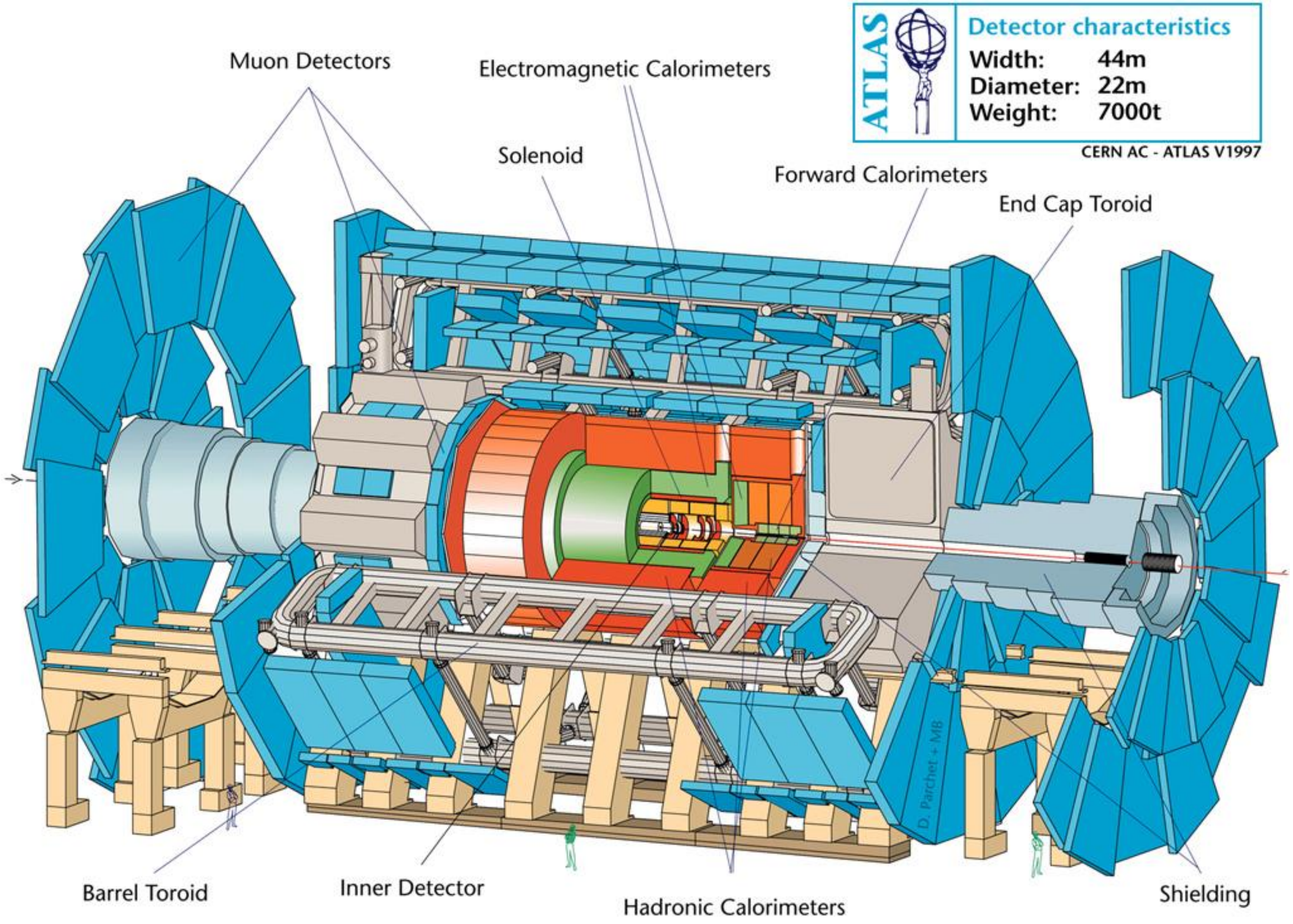
7. 7. 2018

Institute of Physics (FZU) of the Czech Academy of Sciences

Outline

- ATLAS general introduction (number of collisions, data rates and volumes, CPU requirements
- Usage of Grid as the main resource
  - geographic distribution of resources
  - heterogeneity of resources
- ATLAS Solutions for Grid
  - PanDA, JEDI, Rucio
- Newer types resources
- HPC
  - harvester, ARC cache, huge variations in number of cores
- Clouds
- ATLAS@Home

ATLAS experiment



Collaboration of more than **3000** authors from **182** institutes – **grant data access to all**

Large number of channels

+

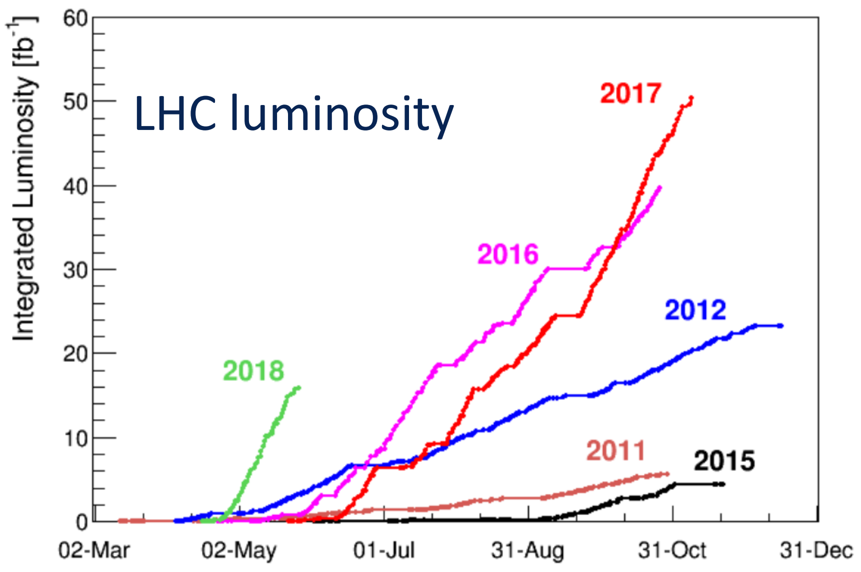
High trigger rates

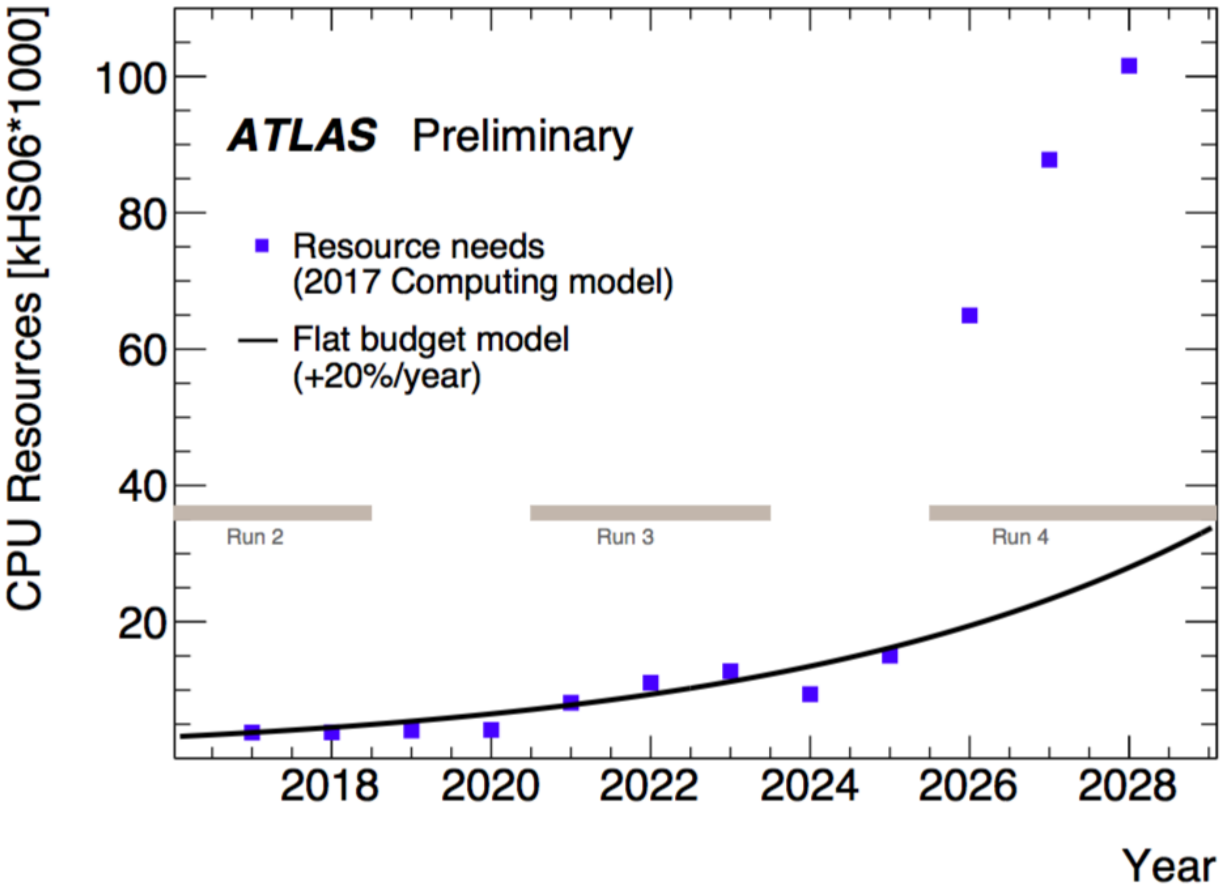
=

Huge data volumes

Computing requirements

Still increasing:





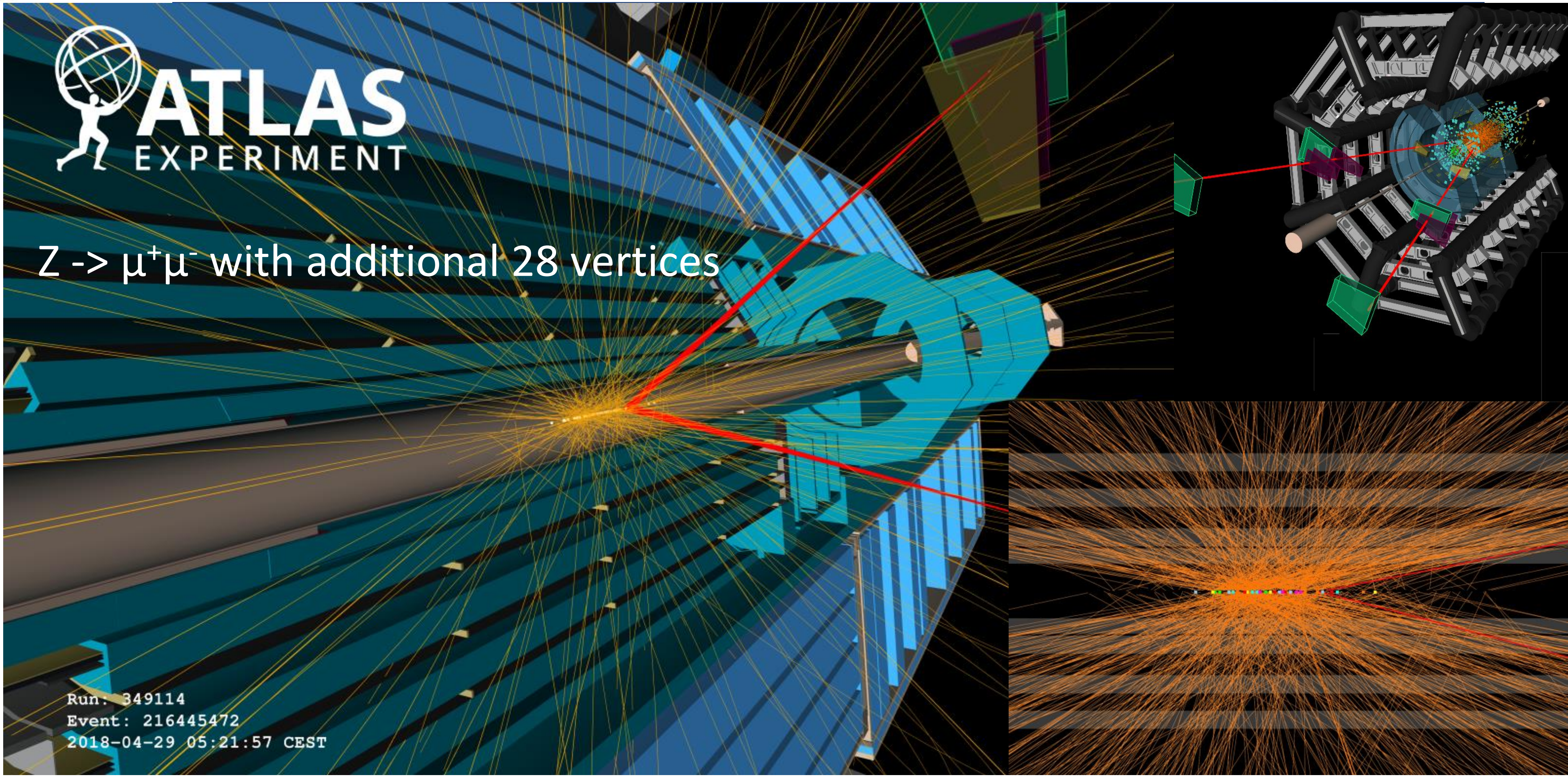
ATLAS 2018 requirements

186 PB on disk

289 PB on tape

2520 kHS06 CPU years

ATLAS experiment

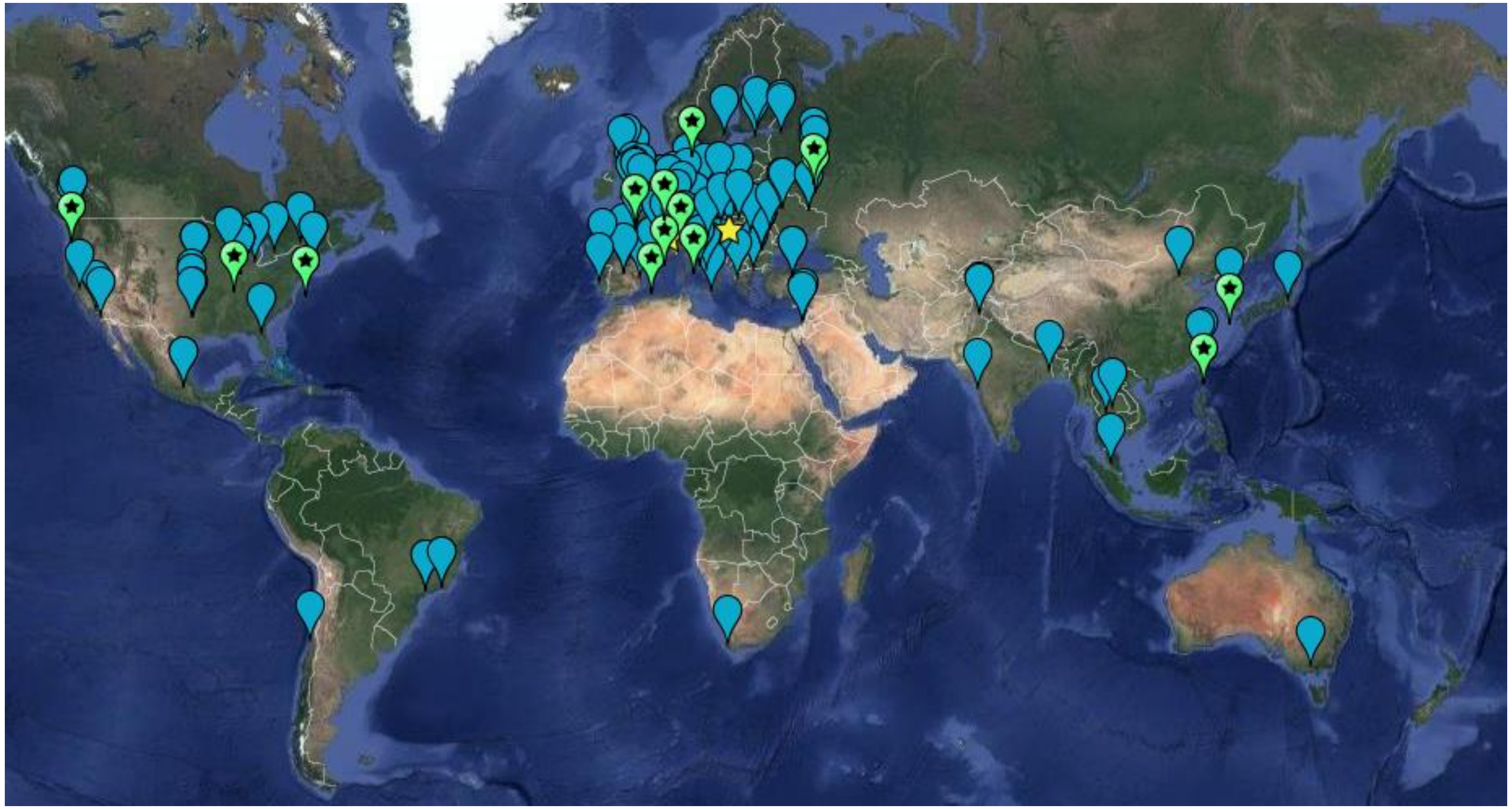


Grid - WLCG

- WLCG connects sites distributed across 5 continents
  - "EGI" sites with various implementation of CEs and SEs
  - OSG sites

WLCG Capacities in 2018	
cores	694000
CPU capacity	8768 kHS06
disk	382 PB
tape	361 PB

US capacities not included

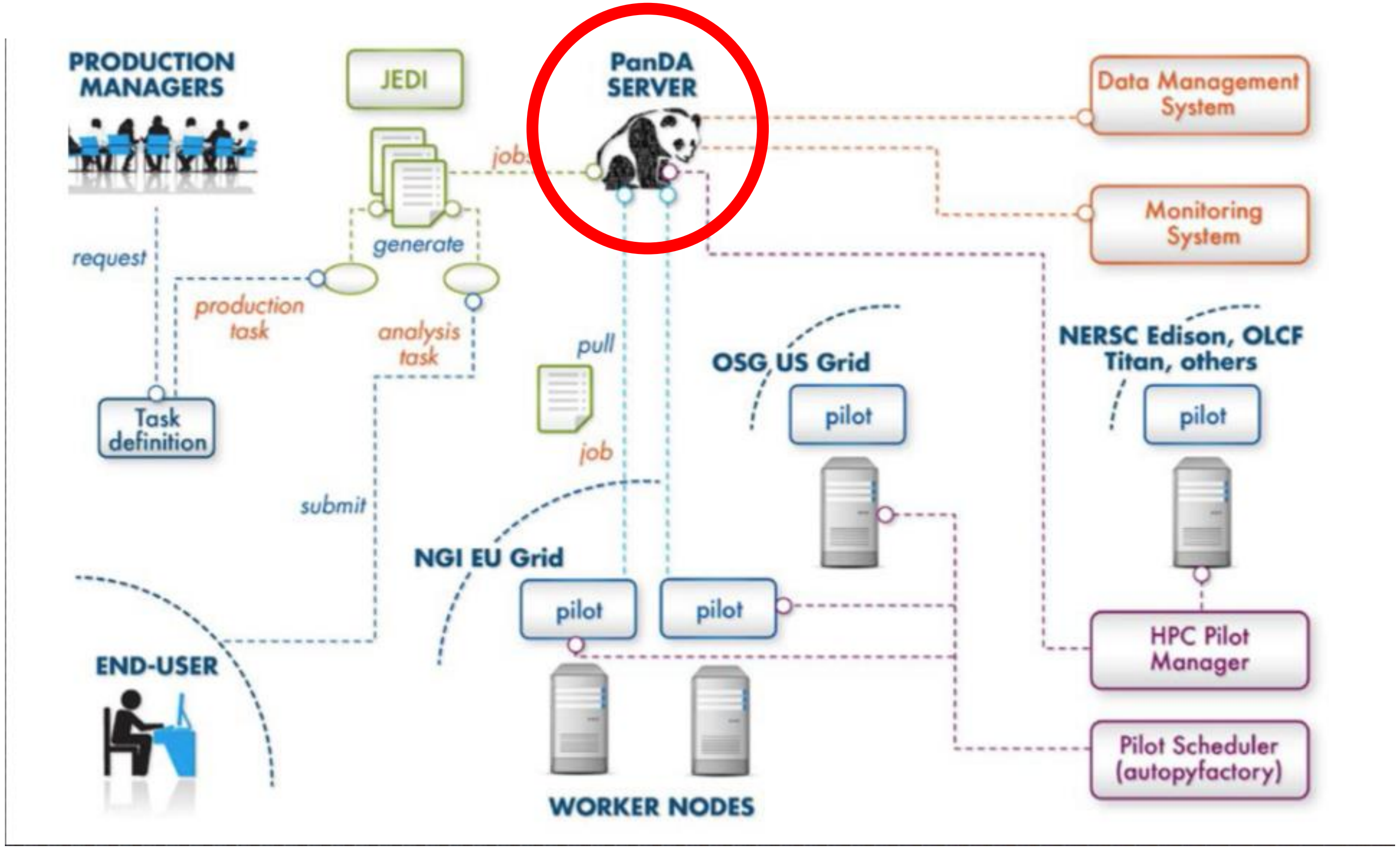


ATLAS Distributed Computing Solutions

Big effort to maximize resource usage and to automatize workflows

PanDA workload management system

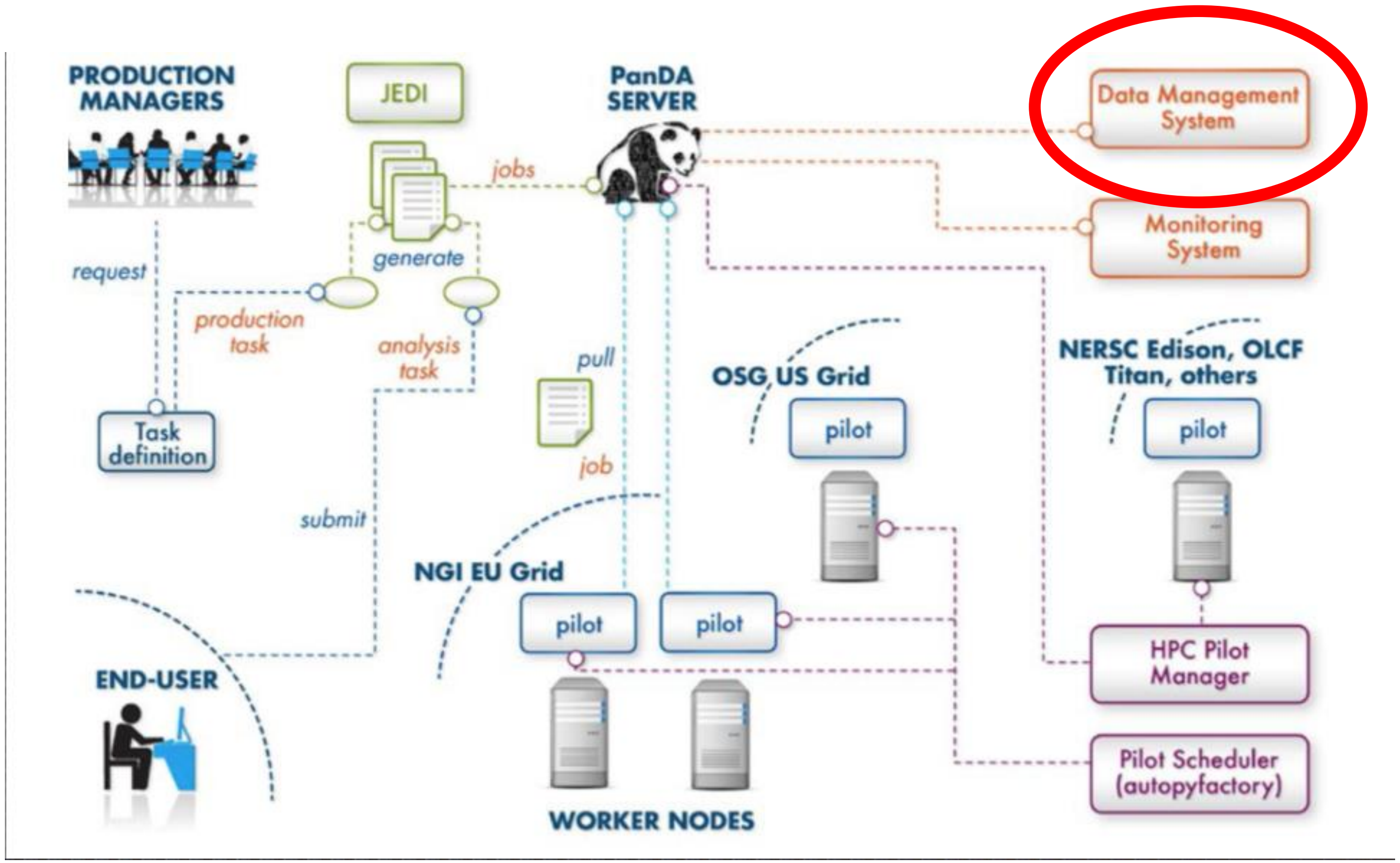
- Used by ATLAS since 2008, now adopted also by other projects
- Pilot based
- Factorized code



ATLAS Distributed Computing Solutions

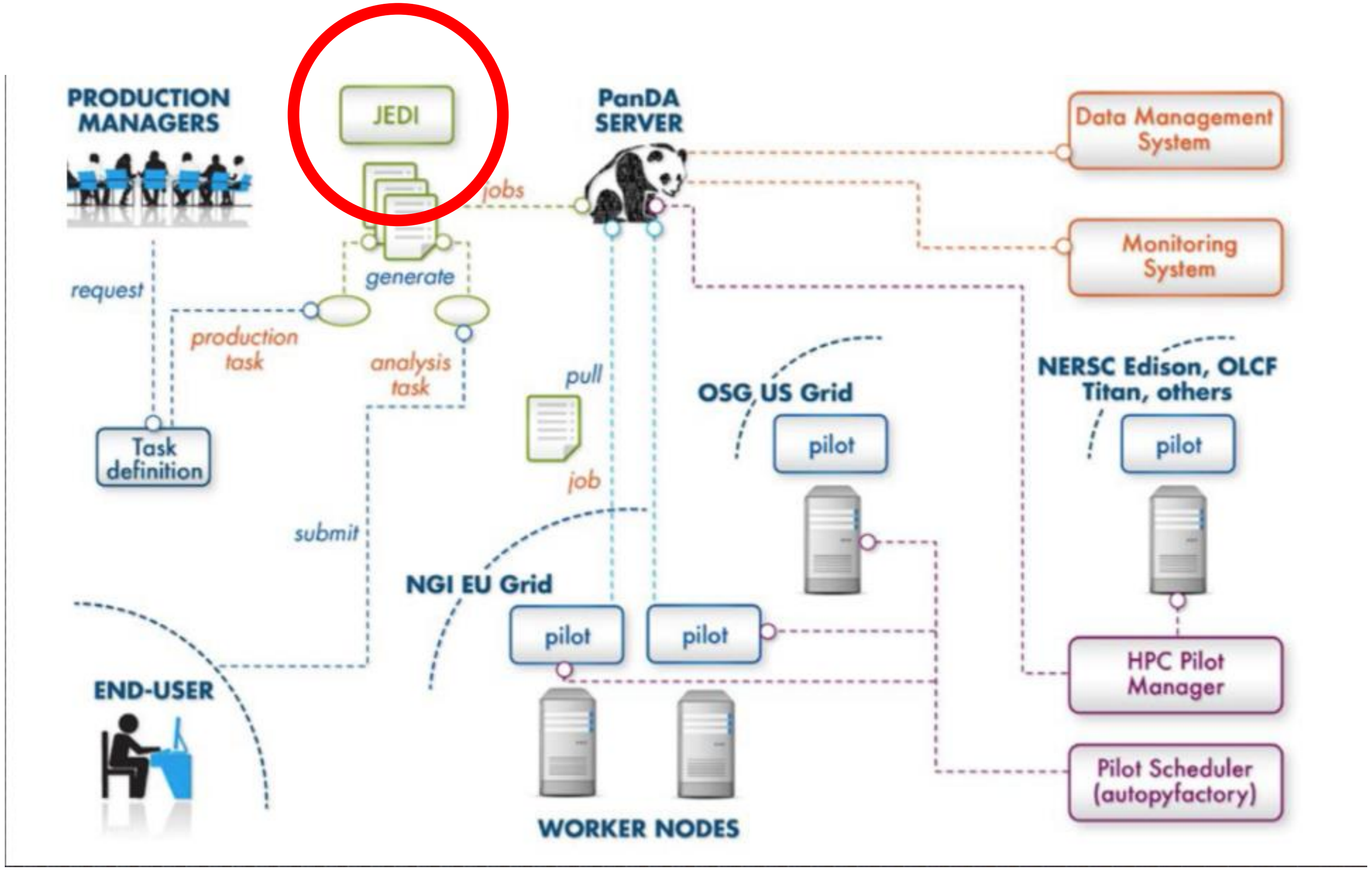
Rucio

- Distributed Data Management System
- Developed for ATLAS, but now used and evaluated by other projects
- Handles 1 Billion ATLAS files, 365 PB



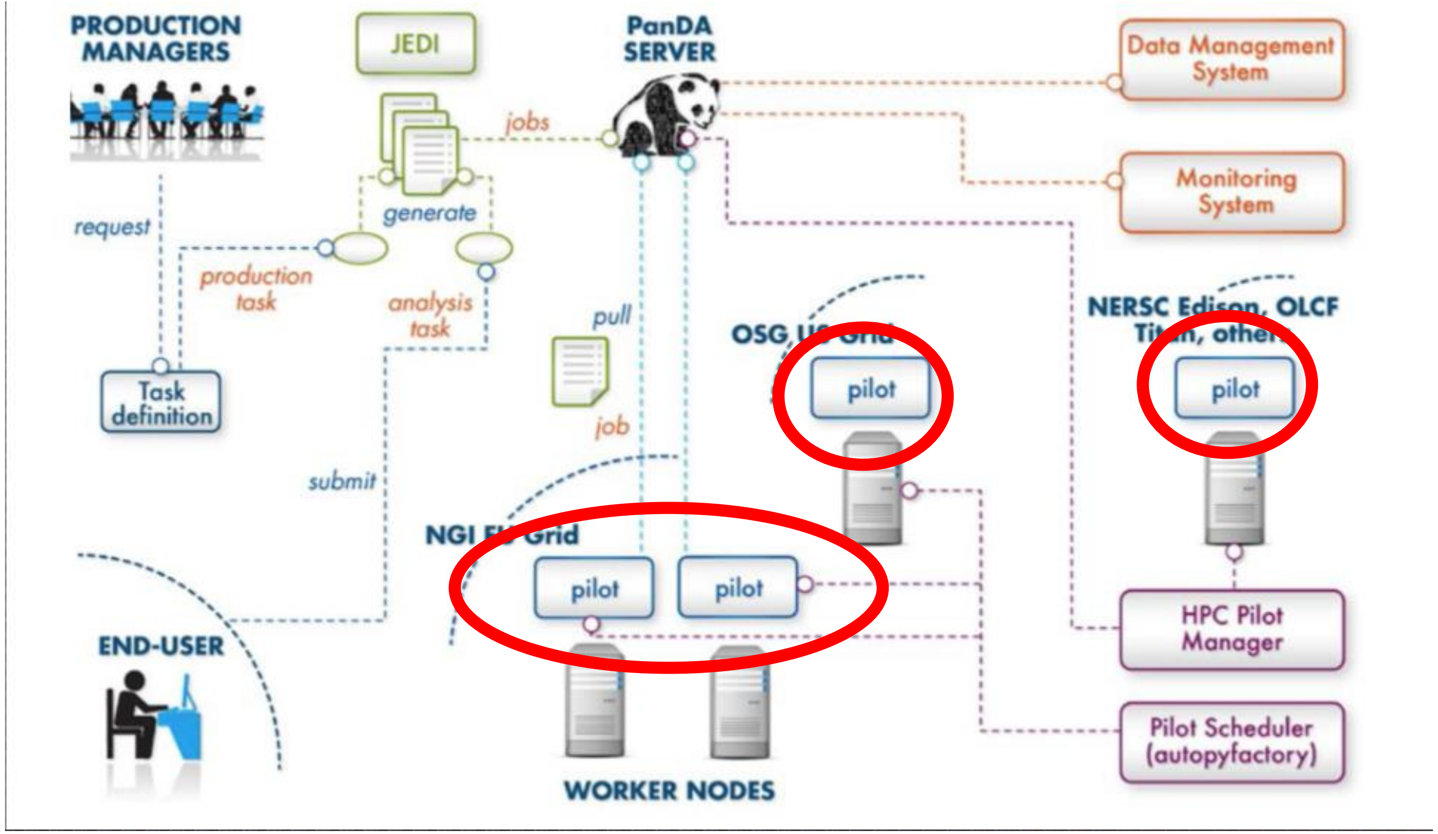
JEDI

- Job Execution and Definition Interface
- Dynamic job definition from tasks



Pilot

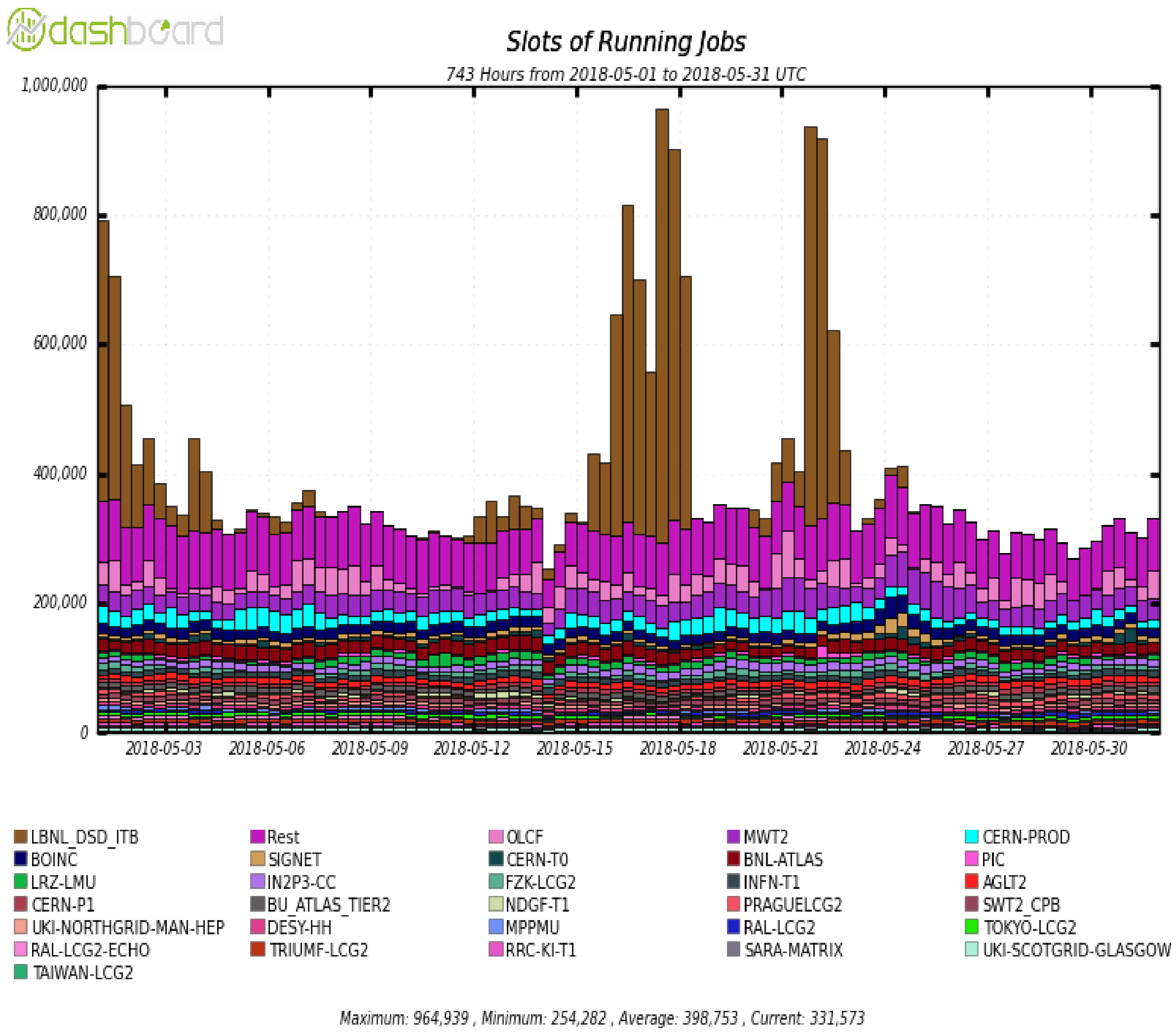
- Control and benchmark execution node
- Get jobs
- Monitor
- Stage-in, stage-out
- Cleanup



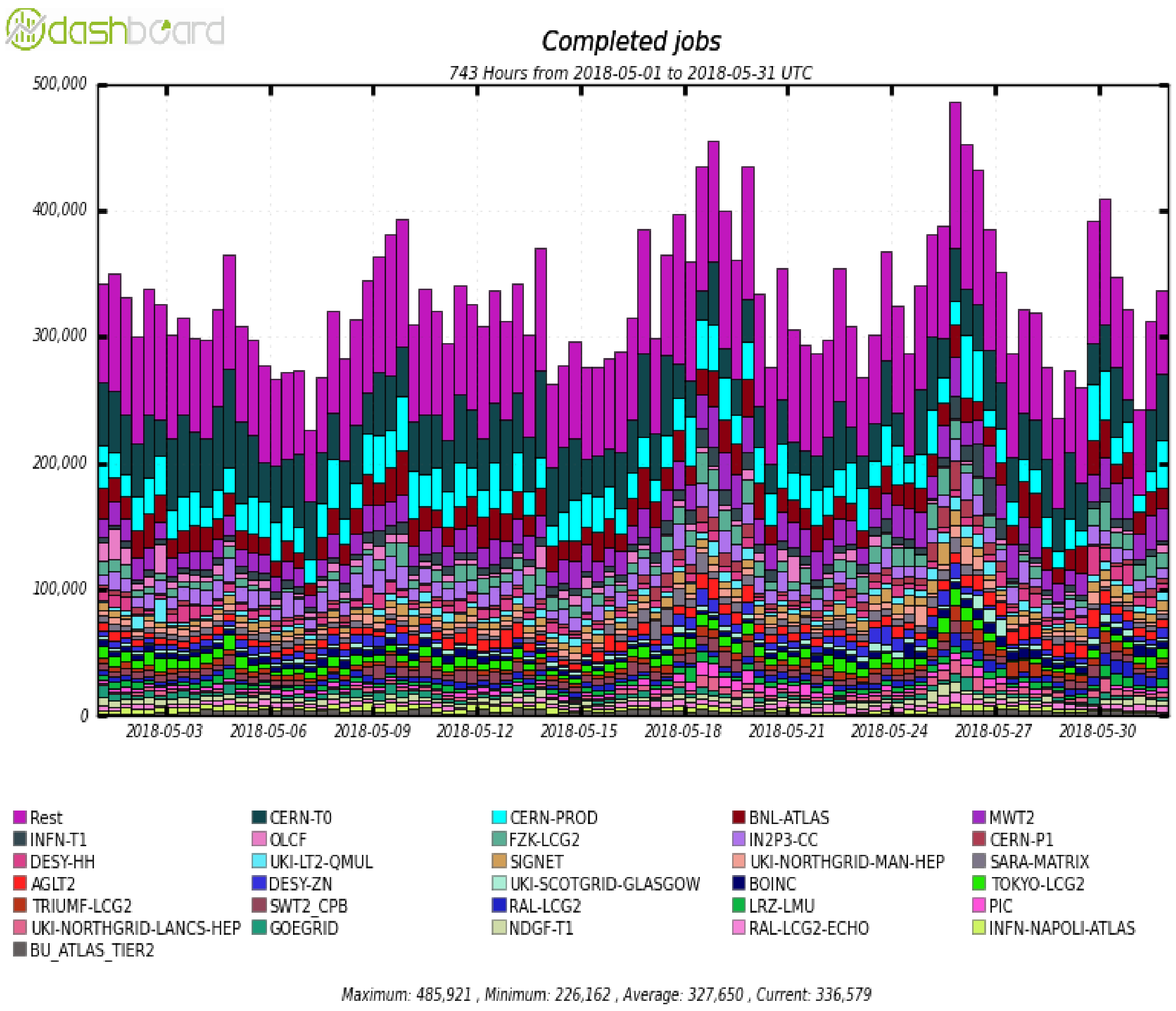
Differences between HPC and HTC

- Architecture
- Authentication
- Network access
- Access model

HPCs



dashboard
Slots of Running Jobs
743 Hours from 2018-05-01 to 2018-05-31 UTC
1,000,000
800,000
600,000
400,000
200,000
0
2018-05-03
2018-05-06
2018-05-09
2018-05-12
2018-05-15
2018-05-18
2018-05-21
2018-05-24
2018-05-27
2018-05-30
LBNL_DSD_ITB
Rest
OLCF
MWT2
CERN-PROD
BOINC
SIGNET
CERN-T0
BNL-ATLAS
PIC
LRZ-LMU
IN2P3-CC
FZK-LCG2
INFN-T1
AGLT2
CERN-P1
BU_ATLAS_TIER2
NDGF-T1
PRAGUELCG2
SWT2_CPB
UKI-NORTHGRID-MAN-HEP
DESY-HH
MPPMU
RAL-LCG2
TOKYO-LCG2
RAL-LCG2-ECHO
TRIUMF-LCG2
RRC-KI-T1
SARA-MATRIX
UKI-SCOTGRID-GLASGOW
TAIWAN-LCG2
Maximum: 964,939 , Minimum: 254,282 , Average: 398,753 , Current: 331,573

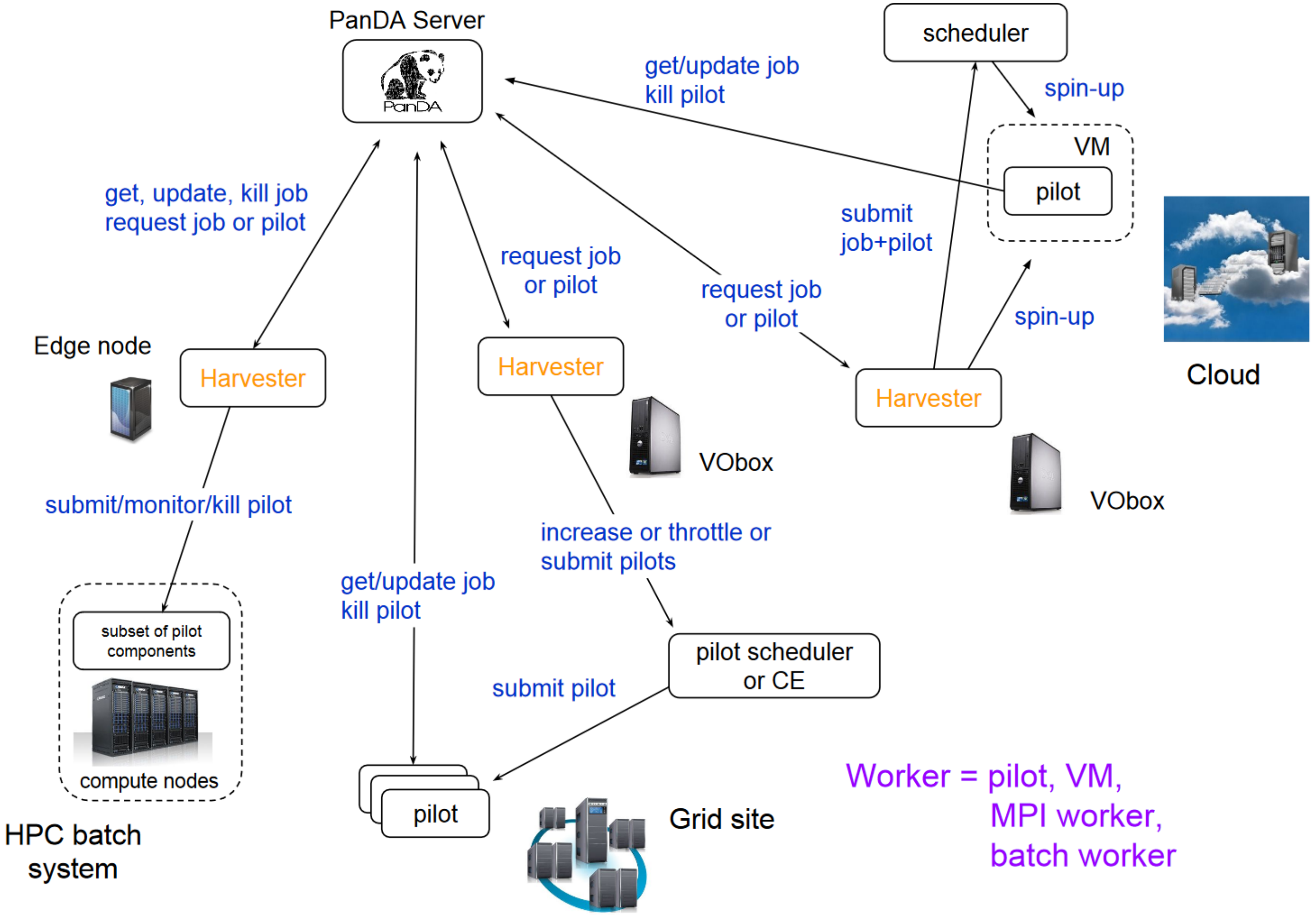


dashboard
Completed jobs
743 Hours from 2018-05-01 to 2018-05-31 UTC
500,000
400,000
300,000
200,000
100,000
0
2018-05-03
2018-05-06
2018-05-09
2018-05-12
2018-05-15
2018-05-18
2018-05-21
2018-05-24
2018-05-27
2018-05-30
Rest
CERN-T0
CERN-PROD
BNL-ATLAS
MWT2
INFN-T1
OLCF
FZK-LCG2
IN2P3-CC
CERN-P1
DESY-HH
UKI-LT2-QMUL
SIGNET
UKI-NORTHGRID-MAN-HEP
SARA-MATRIX
AGLT2
DESY-ZN
UKI-SCOTGRID-GLASGOW
BOINC
TOKYO-LCG2
TRIUMF-LCG2
SWT2_CPB
RAL-LCG2
LRZ-LMU
PIC
UKI-NORTHGRID-LANCS-HEP
GOEGRID
NDGF-T1
RAL-LCG2-ECHO
INFN-NAPOLI-ATLAS
BU_ATLAS_TIER2
Maximum: 485,921 , Minimum: 226,162 , Average: 327,650 , Current: 336,579

Schematic View

Harvester

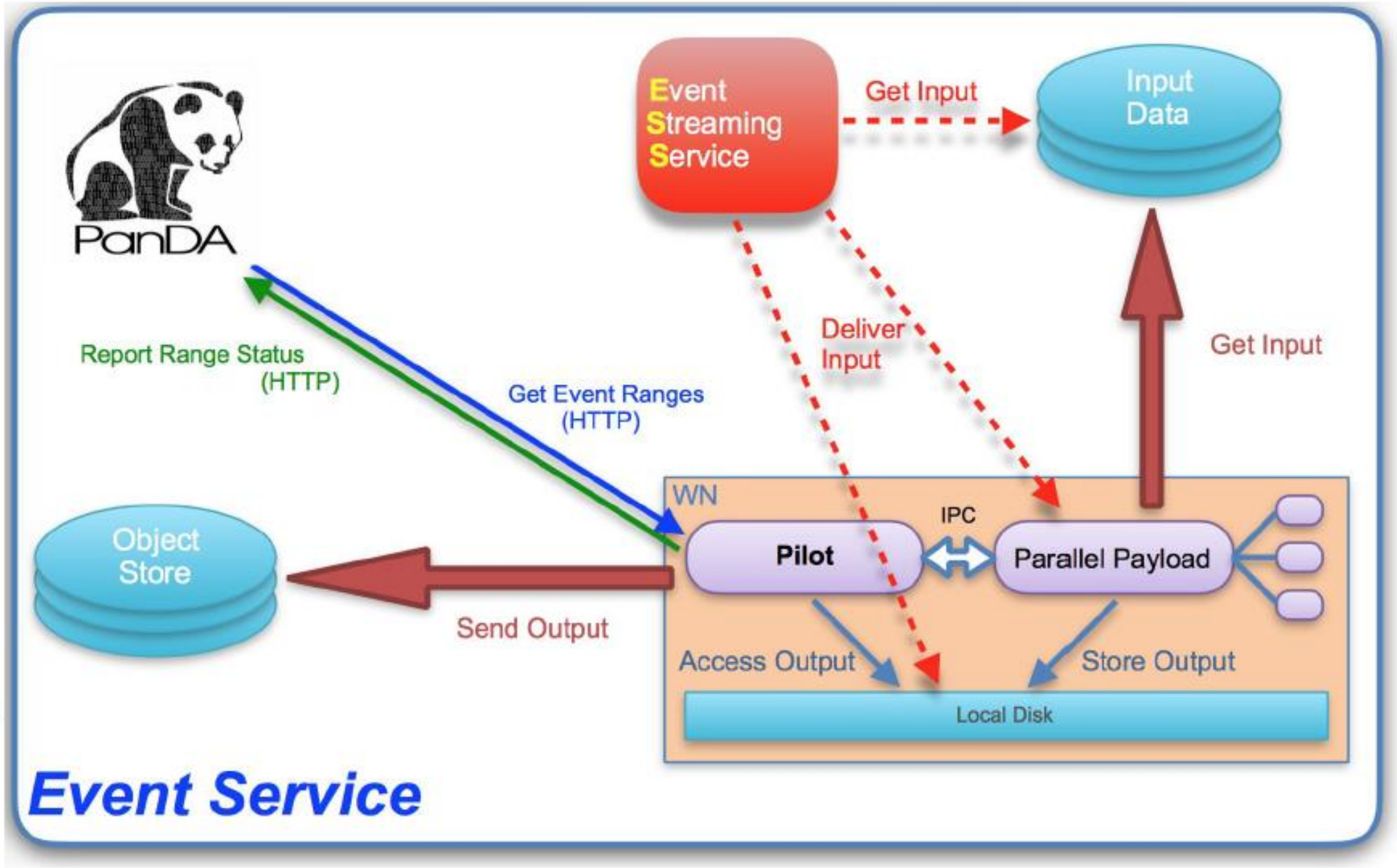
- New component between Pilot and PanDA
- Running on site
- Stateless service plus DB



chudoba@fzu.cz

Event Service

- reduces lost CPU time in case of premature job termination
- important for backfilling HPC resources



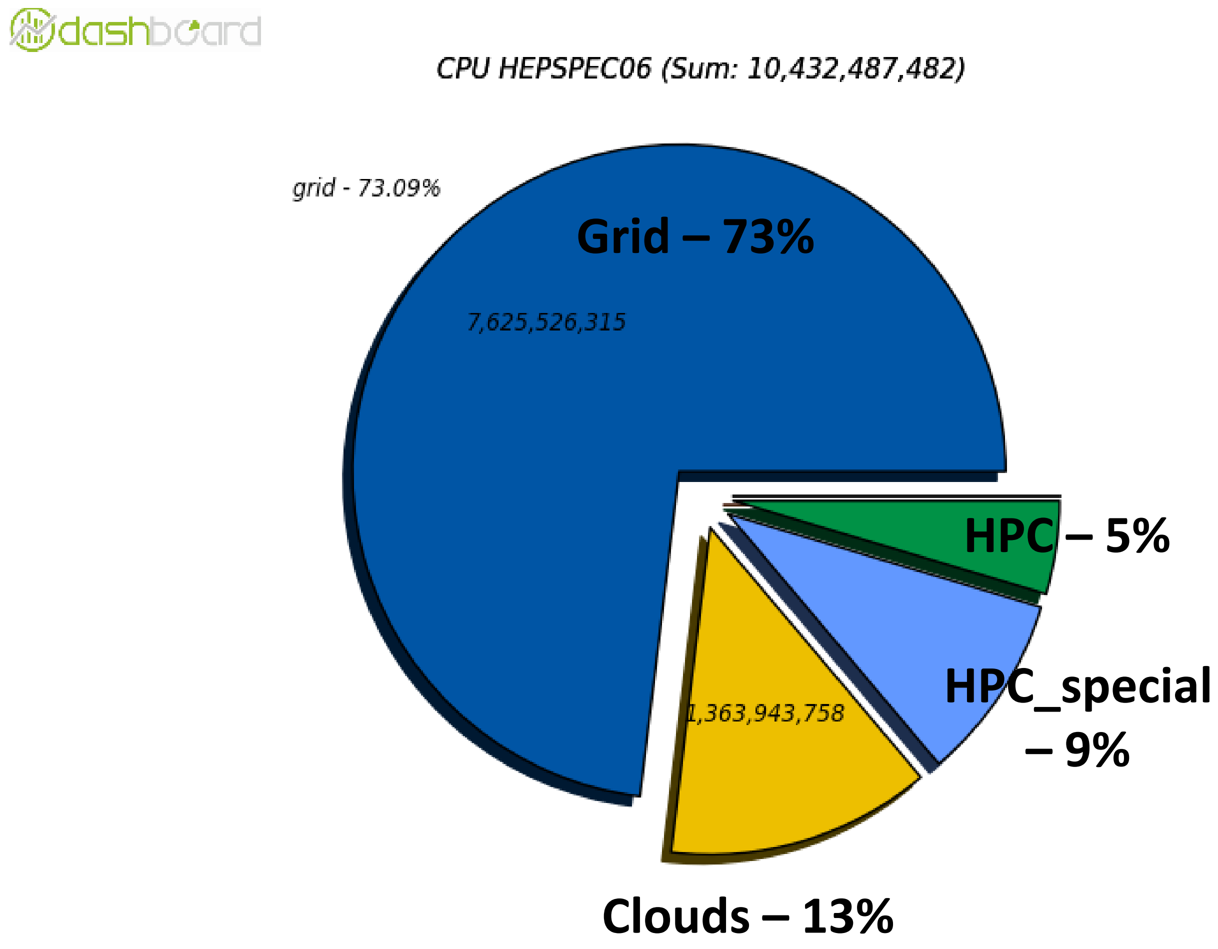
chudoba@fzu.cz

Clouds

- Successful business model
- Huge resources provided by commercial clouds
  - All LHC computing need covered by less than 1%
  - But significantly more expensive (apart from spot market)
  - Option for peak demands
- Private clouds enable better sharing with other groups

Clouds for ATLAS

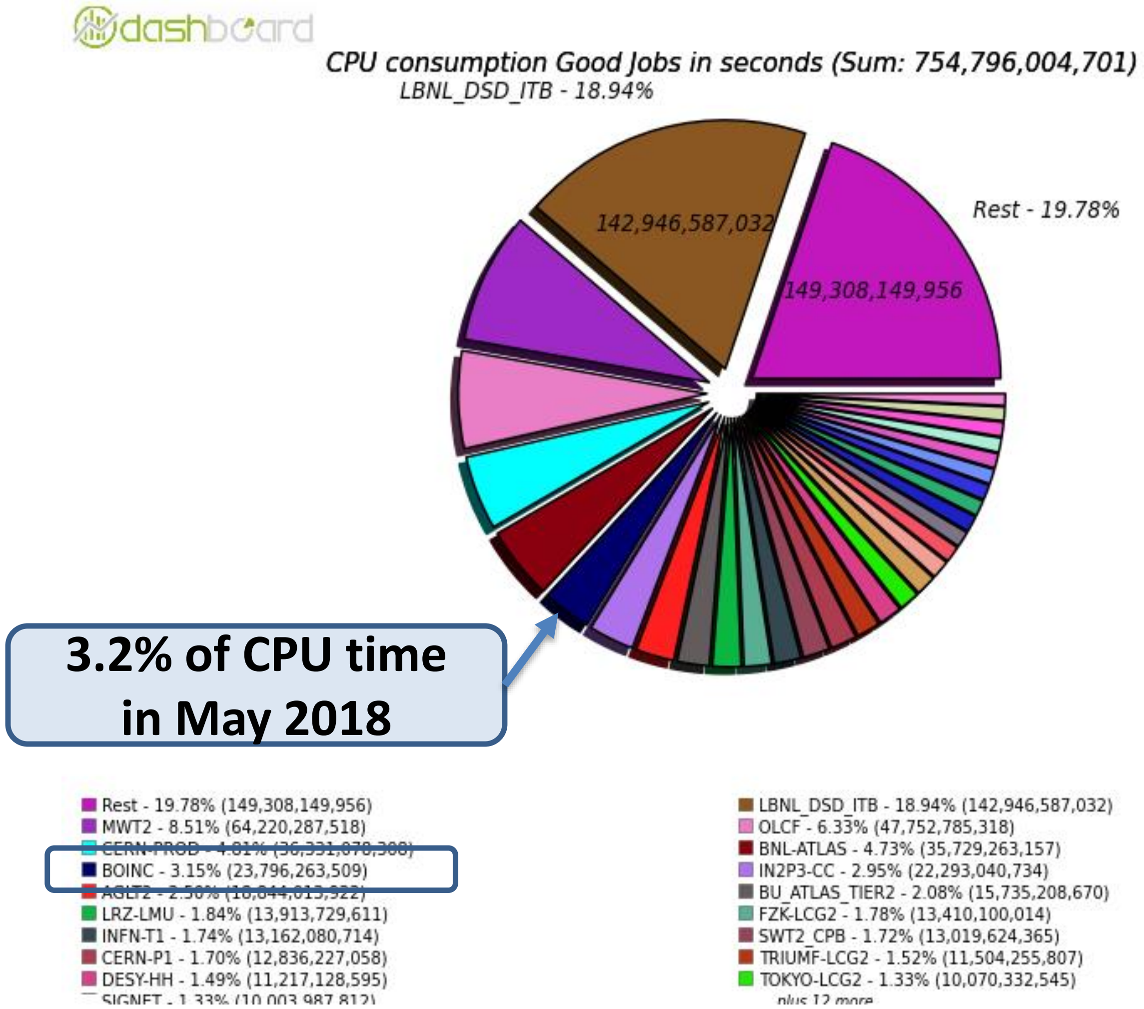
- Fully integrated in the ATLAS infrastructure
  - "Standard" CE exposed by clouds
  - Details hidden by local setup
    - HTCondor, ARC
- Cloud Data storage used mostly for log files
  - Not (yet) suitable for big volume data storage



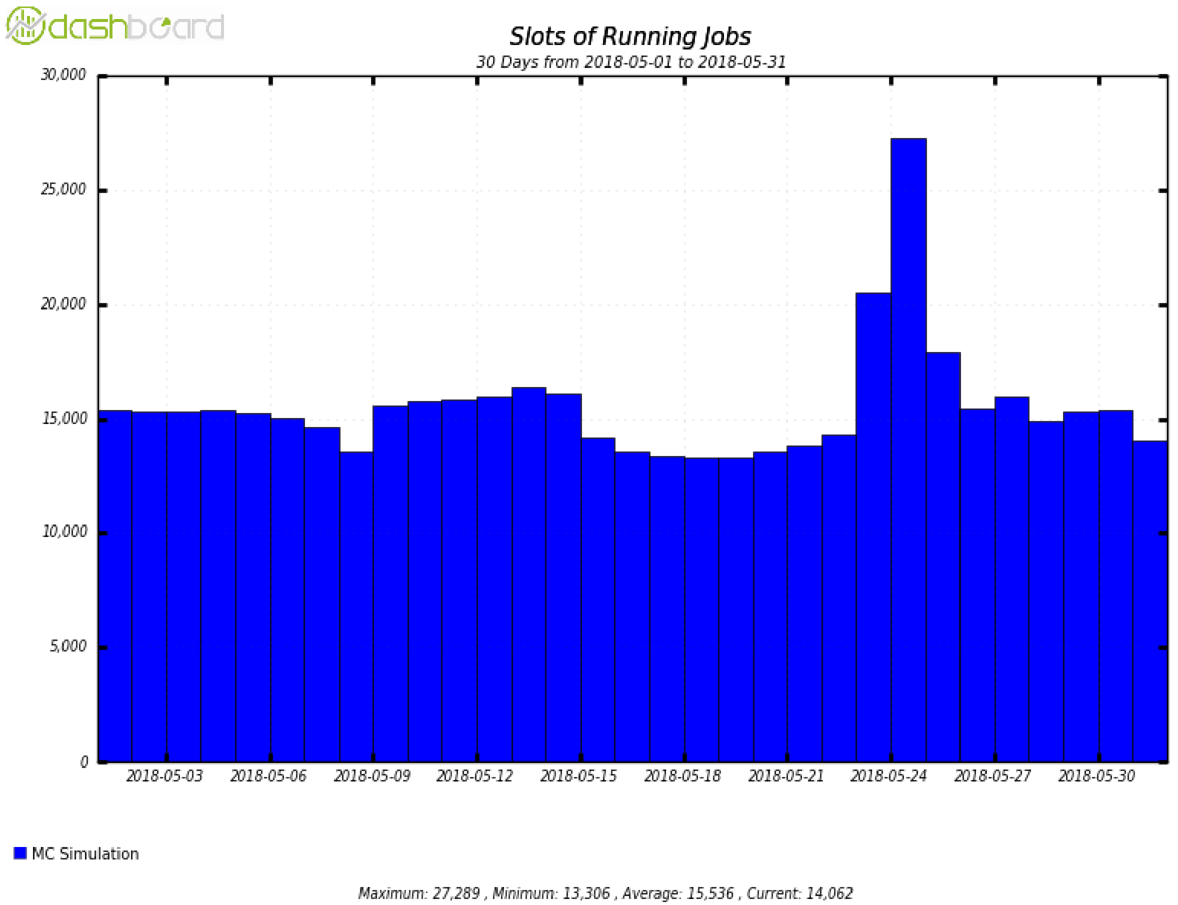
ATLAS CPU consumption per resource, January - May 2018

BOINC

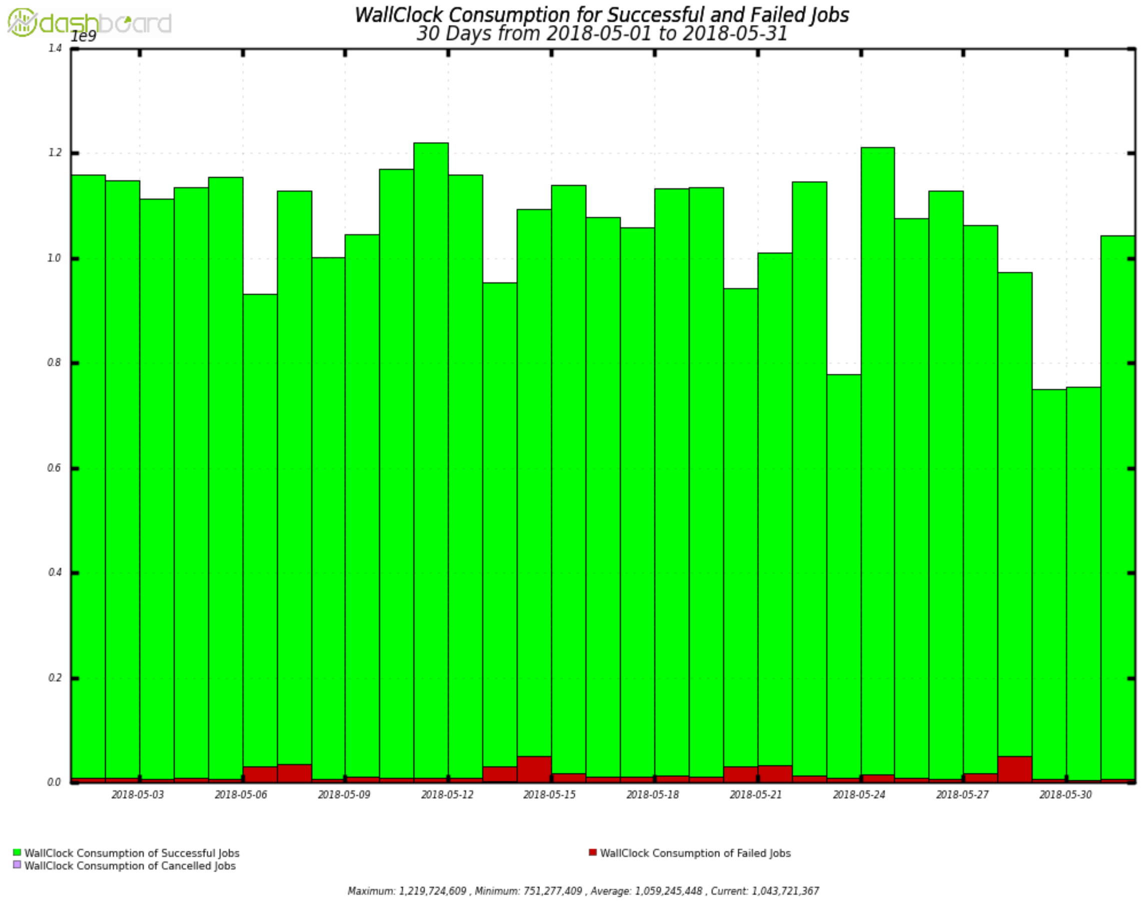
- Open-source software for volunteer computing
- Uses the idle time of (personal) computers
- VirtualBox used for various platforms (Windows, Mac, Linux)
- Easy to use for laymen
- Also great for outreach



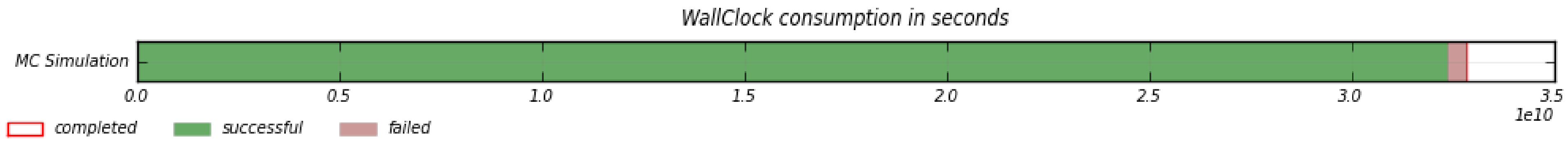
ATLAS@home



dashboard
Slots of Running Jobs
30 Days from 2018-05-01 to 2018-05-31
30,000
25,000
20,000
15,000
10,000
5,000
0
2018-05-03 2018-05-06 2018-05-09 2018-05-12 2018-05-15 2018-05-18 2018-05-21 2018-05-24 2018-05-27 2018-05-30
MC Simulation
Maximum: 27,289 , Minimum: 13,306 , Average: 15,536 , Current: 14,062



dashboard
WallClock Consumption for Successful and Failed Jobs
30 Days from 2018-05-01 to 2018-05-31
1e9
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
2018-05-03 2018-05-06 2018-05-09 2018-05-12 2018-05-15 2018-05-18 2018-05-21 2018-05-24 2018-05-27 2018-05-30
WallClock Consumption of Successful Jobs
WallClock Consumption of Cancelled Jobs
WallClock Consumption of Failed Jobs
Maximum: 1,219,724,609 , Minimum: 751,277,409 , Average: 1,059,245,448 , Current: 1,043,721,367



WallClock consumption in seconds
MC Simulation
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
1e10
completed
successful
failed

ATLAS@Home

- New use cases on clusters
  - clusters not supporting ATLAS VO
  - clusters supporting ATLAS VO
- Example of Beijing Tier2 site
  - Grid jobs walltime utilization 88%, cputime 66%
  - additional 23% of cpu utilization

Rank	Name	Recent average credit	Total credit	Country
1	Agile Boincers	2,567,907	1,172,053,424	Switzerland
2	TRIUMF-LCG2	631,874	54,506,414	Canada
3	BEIJING-LCG2	479,779	130,474,893	China
4	CharityEngine1	303,429	28,874,365	International
5	WLCG Performance-Test Cluster	251,080	83,857,386	Switzerland
6	LRZ-LMU	246,891	10,176,386	Germany

Conclusions

- ATLAS Offline Computing successfully uses various resources with still increasing automation of workflows
- But it remains a huge distributed effort in terms of manpower for development and operations
- We greatly appreciate also non-pledged resources

Thank you for your attention!

chudoba@fzu.cz