The Czech Distributed Tier-2 Center

Alexandr Mikula

Institute of Physics
of the
Czech Academy of Sciences

Martin Adam, Dagmar Adamová, Jiří Chudoba, Petr Horák, Jana Uhlířová, Petr Vokáč

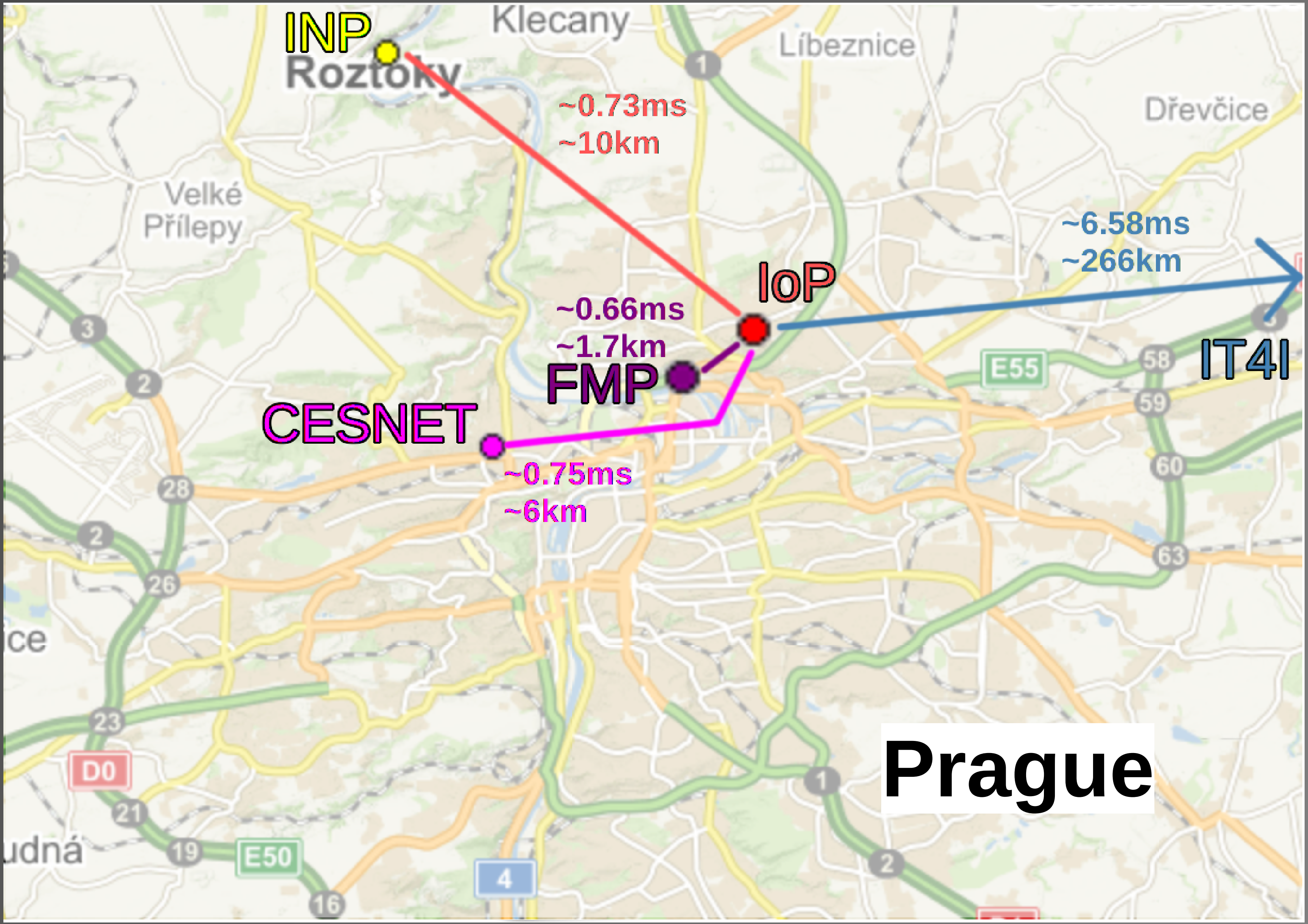
Akademie věd
České republiky

FYZIKÁLNÍ ÚSTAV
Akademie věd ČR, v.v.i.

CC IoP CAS

- Distributed site located in CZ
- Several institutions involved
  - Institute of Physics CAS (IoP)
  - Institute of Nuclear Physics CAS (INP)
  - Faculty of Mathematics and Physics CUNI (FMP)
  - CESNET
  - IT4Innovations

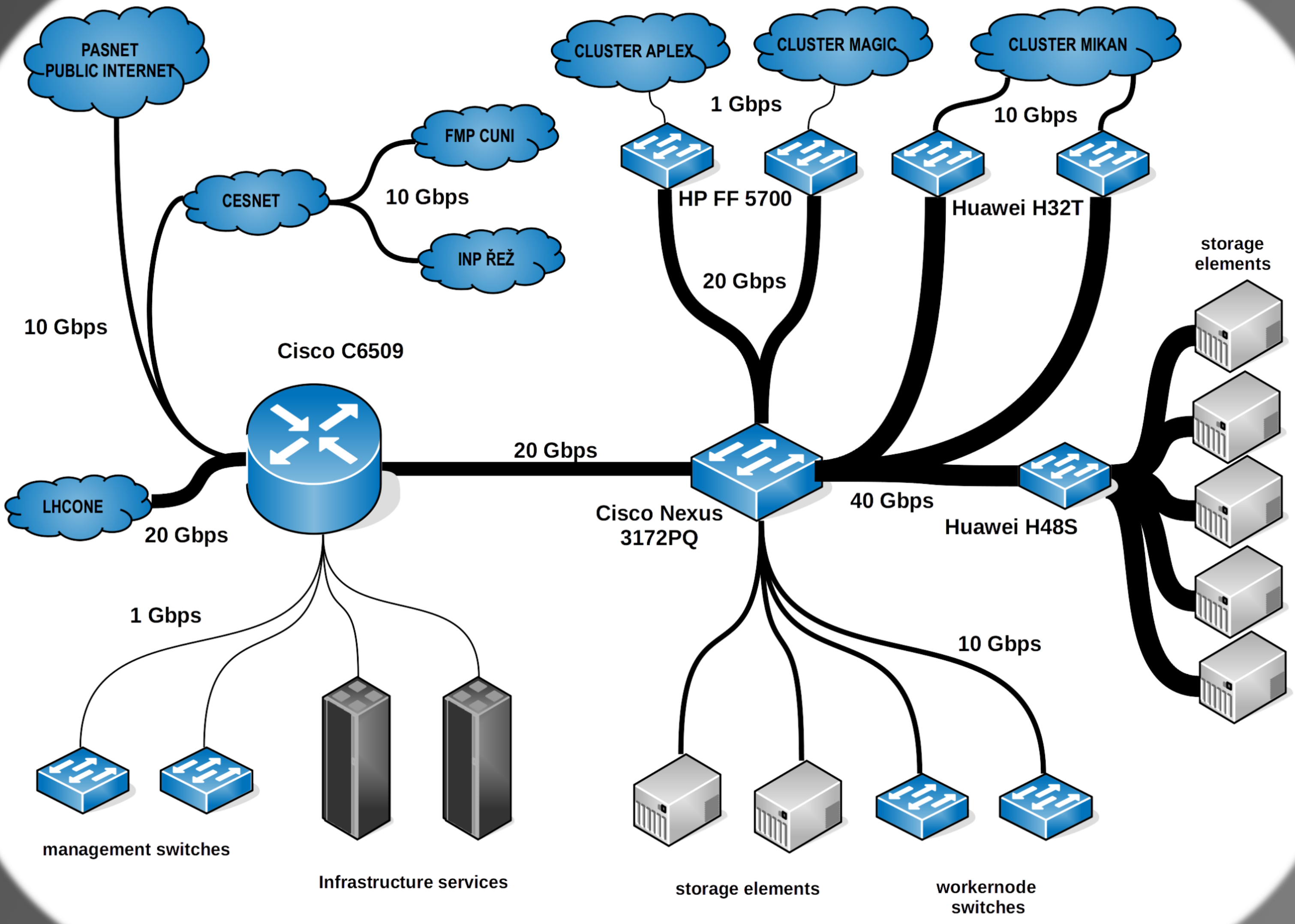
CC IoP CAS


INP
Roztoky
Klecany
Líbeznice
Dřevčice
~0.73ms
~10km
Velké
Přílepy
~6.58ms
~266km
IoP
~0.66ms
~1.7km
FMP
IT4I
E55
CESNET
~0.75ms
~6km
ice
D0
udná
E50
4
Prague

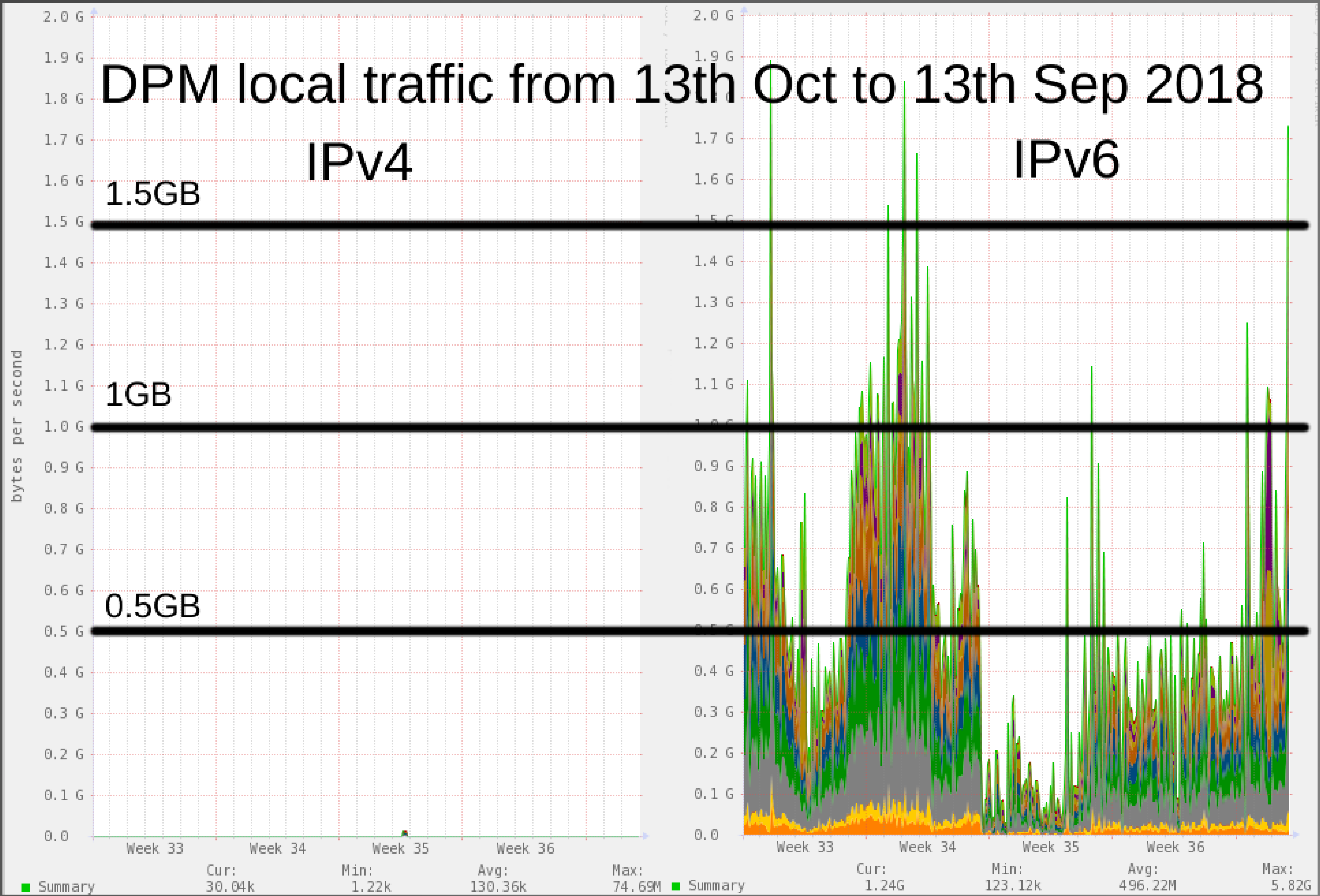
Network Connectivity

- Most of services are dual stack IPv4/IPv6
  - Internal communication mostly IPv6
  - WNs behind IPv4 NAT
  - External IPv6 communication is growing

Network connectivity



IPv4 and IPv6 internal traffic on



DPM local traffic from 13th Oct to 13th Sep 2018

IPv4

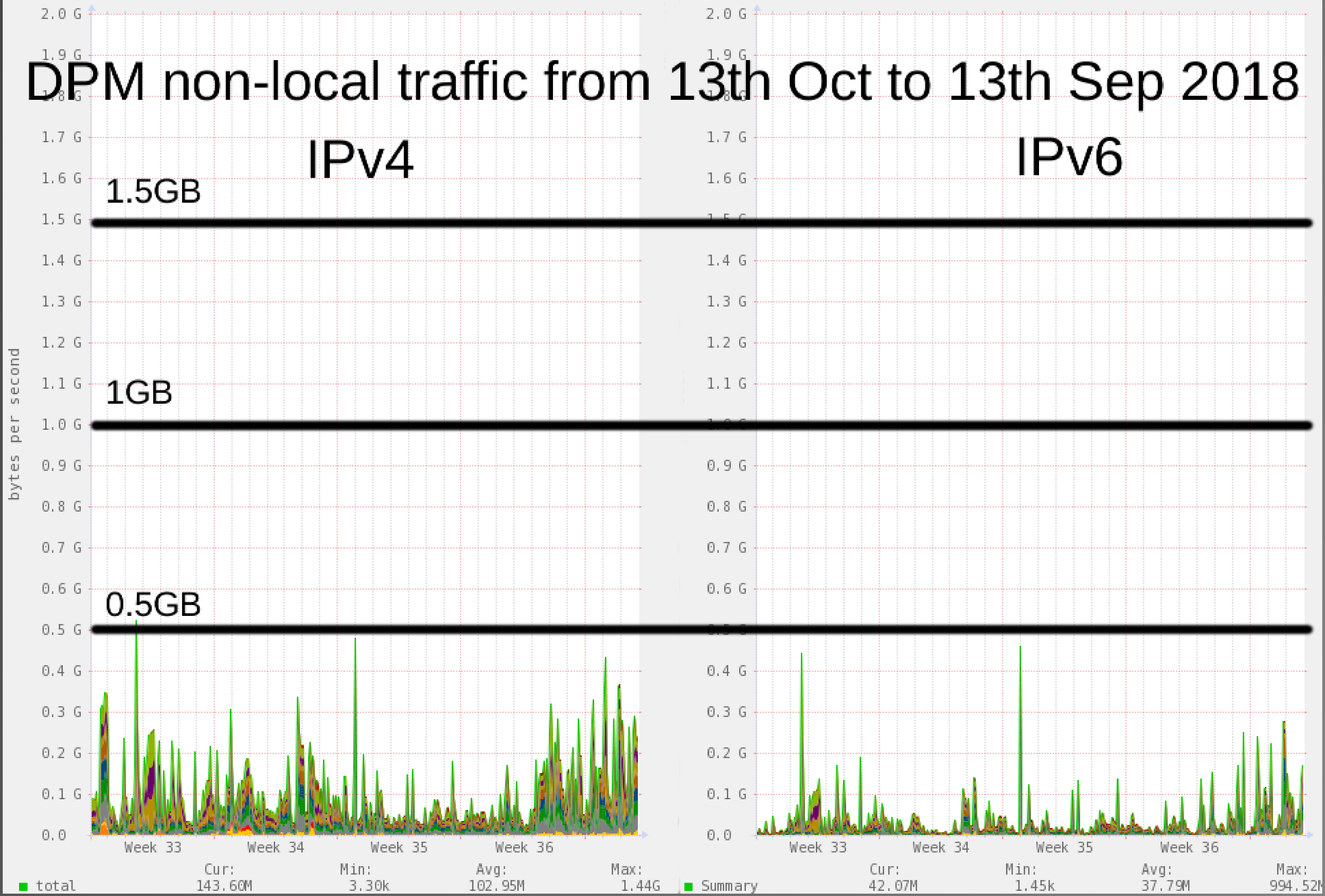
IPv6

1.5GB

1GB

0.5GB

IPv4 and IPv6 external traffic on



DPM non-local traffic from 13th Oct to 13th Sep 2018

IPv4

IPv6

1.5GB

1GB

0.5GB

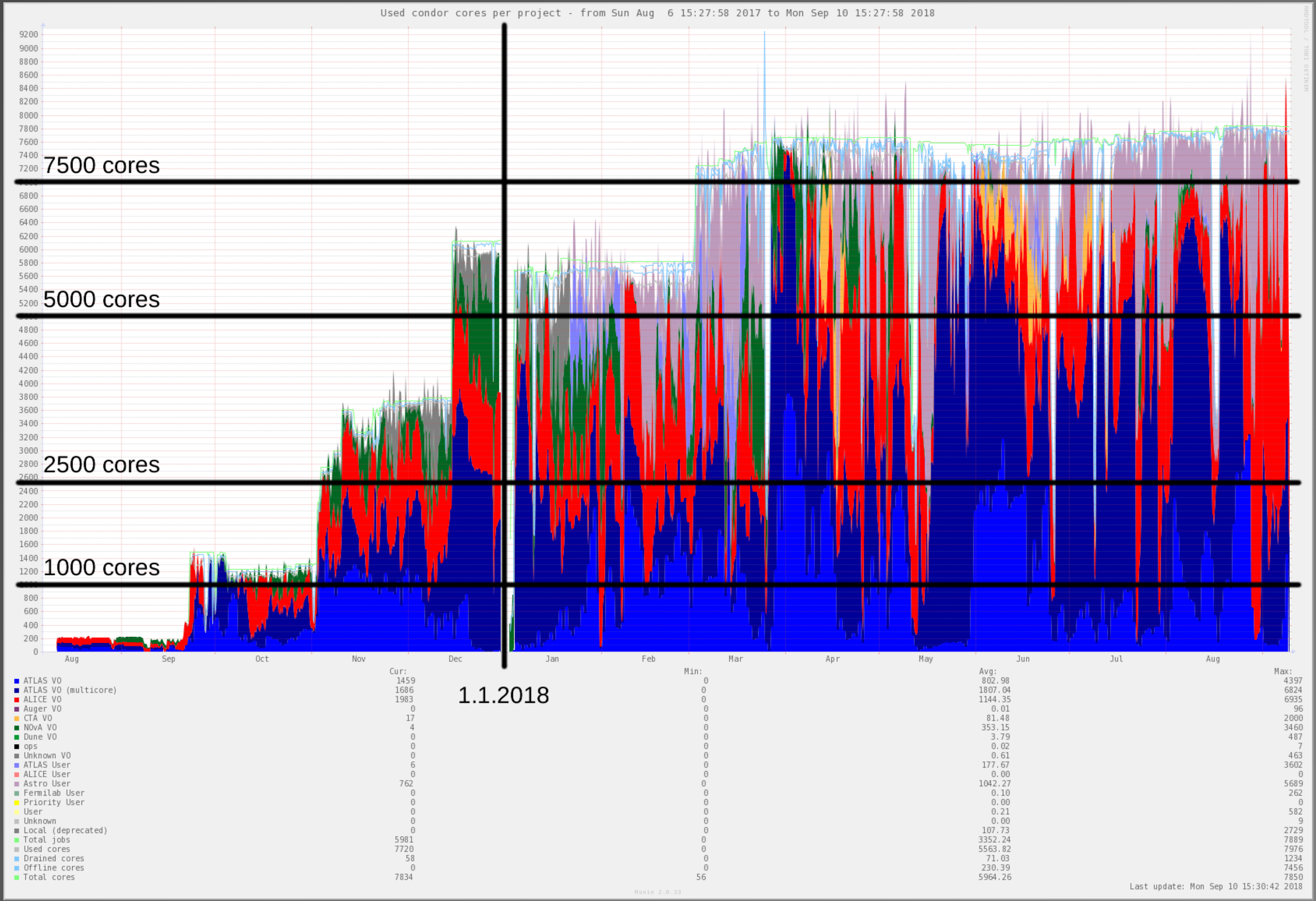
Supported Projects

- Tier 2 site
  - ATLAS and ALICE
- Other projects
  - Pierre Auger observatory
  - Cherenkov telescope array
  - NOvA
  - DUNE
- Local users

Computing power

- HTCondor Batch and ARC CE
- (HTCondor CE for OSG stack)
  - Own accounting data processing for APEL
- 285 Worker nodes
- 7800 cores
- 77k HEPSPEC06
- 47 Worker nodes off site at MFF
- ARC CE for the Salomon IT4I cluster
  - Opportunistic resource for ATLAS

Computing power



7500 cores

5000 cores

2500 cores

1000 cores

1.1.2018

	Cur:	Min:	Avg:	Max:
■ ATLAS VO	1459	0	802.98	4397
■ ATLAS VO (multicore)	1686	0	1807.04	6824
■ ALICE VO	1983	0	1144.35	6935
■ Auger VO	0	0	0.01	96
■ CTA VO	17	0	81.48	2000
■ NOvA VO	4	0	353.15	3460
■ Dune VO	0	0	3.79	487
■ ops	0	0	0.02	7
■ Unknown VO	0	0	0.61	463
■ ATLAS User	6	0	177.67	3602
■ ALICE User	0	0	0.00	0
■ Astro User	762	0	1042.27	5689
■ Fermilab User	0	0	0.10	262
■ Priority User	0	0	0.00	0
■ User	0	0	0.21	582
■ Unknown	0	0	0.00	9
■ Local (deprecated)	0	0	107.73	2729
■ Total jobs	5981	0	3352.24	7889
■ Used cores	7720	0	5563.82	7976
■ Drained cores	58	0	71.03	1234
■ Offline cores	0	0	230.39	7456
■ Total cores	7834	56	5964.26	7850

Last update: Mon Sep 10 15:30:42 2018

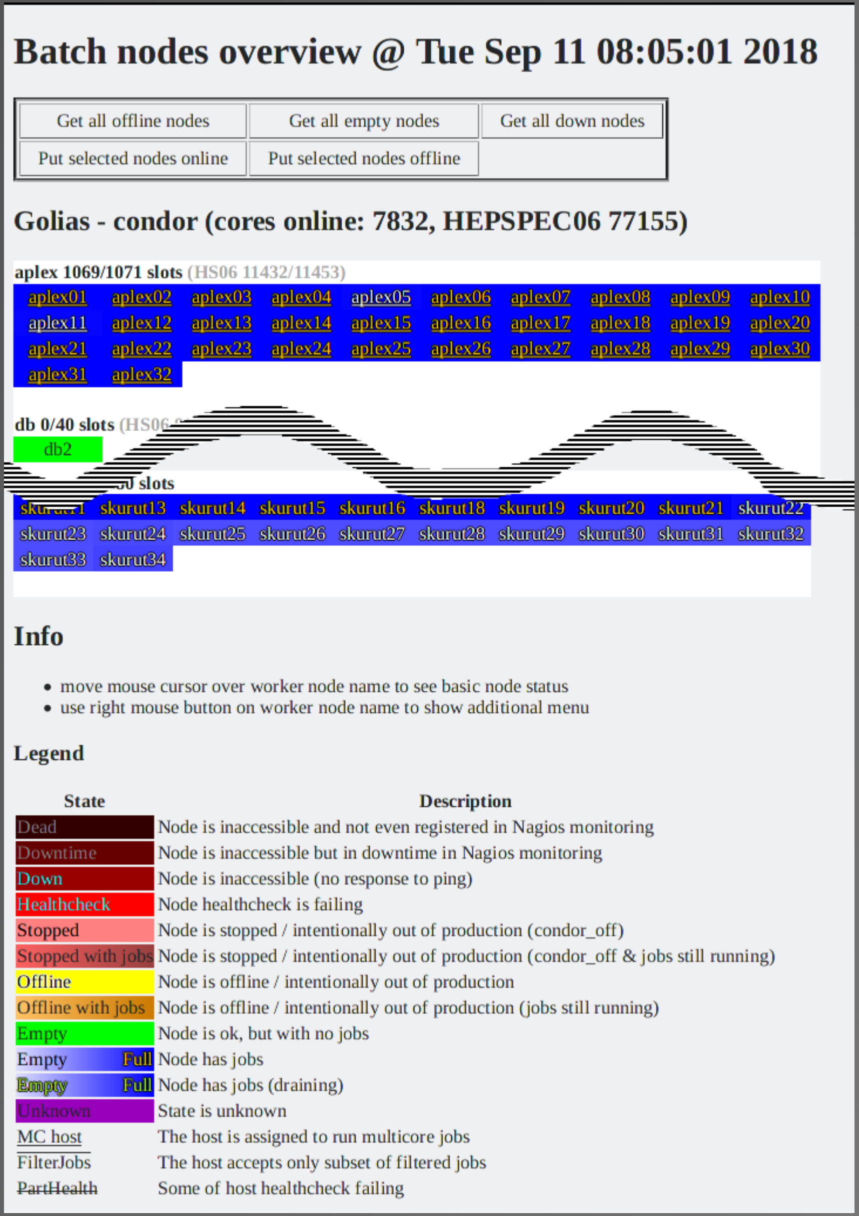
Other resources

- NFS servers
  - 5 servers
  - 212 TB available space
- 40TB Hitachi SAN array
- KVM Hypervisors
  - 3 AMD based severs with 128GB RAM
  - 2 Intel based servers with 256GB RAM
  - FC connection to SAN array
  - 42 virtual machines

Monitoring

- Nagios with Check-MK
- Munin
- Observium
- Ganglia
- Puppet Explorer
- On site developed scripts

Monitoring



Batch nodes overview @ Tue Sep 11 08:05:01 2018

Golias - condor (cores online: 7832, HEPSPEC06 77155)

aplex 1069/1071 slots (HS06 11432/11453)

aplex01	aplex02	aplex03	aplex04	aplex05	aplex06	aplex07	aplex08	aplex09	aplex10
aplex11	aplex12	aplex13	aplex14	aplex15	aplex16	aplex17	aplex18	aplex19	aplex20
aplex21	aplex22	aplex23	aplex24	aplex25	aplex26	aplex27	aplex28	aplex29	aplex30
aplex31	aplex32								

db 0/40 slots (HS06

db2

slots

skur	skurut13	skurut14	skurut15	skurut16	skurut18	skurut19	skurut20	skurut21	skurut22
skurut23	skurut24	skurut25	skurut26	skurut27	skurut28	skurut29	skurut30	skurut31	skurut32
skurut33	skurut34								

Info

- move mouse cursor over worker node name to see basic node status
- use right mouse button on worker node name to show additional menu

Legend

State	Description
Dead	Node is inaccessible and not even registered in Nagios monitoring
Downtime	Node is inaccessible but in downtime in Nagios monitoring
Down	Node is inaccessible (no response to ping)
Healthcheck	Node healthcheck is failing
Stopped	Node is stopped / intentionally out of production (condor_off)
Stopped with jobs	Node is stopped / intentionally out of production (condor_off & jobs still running)
Offline	Node is offline / intentionally out of production
Offline with jobs	Node is offline / intentionally out of production (jobs still running)
Empty	Node is ok, but with no jobs
Empty Full	Node has jobs
Empty Full	Node has jobs (draining)
Unknown	State is unknown
MC host	The host is assigned to run multicore jobs
FilterJobs	The host accepts only subset of filtered jobs
~~PartHealth~~	Some of host healthcheck failing

Software management

- Puppet configuration management
  - Git history control
- Spacewalk package management
- Migration to the Foreman

Storage

- 12 DPM Servers
  - 3.8 PB available storage
  - ATLAS, Auger, CTA, ...
- XrootD
  - ALICE experiment
  - 1.6 PB available storage
  - 3 servers on site
    - 1.3 PB
  - 4 servers off site at INP in Řež
    - 340.6 TB

DPM
Disk Pool Manager
lcgdm.web.cern.ch/dpm

XRootD

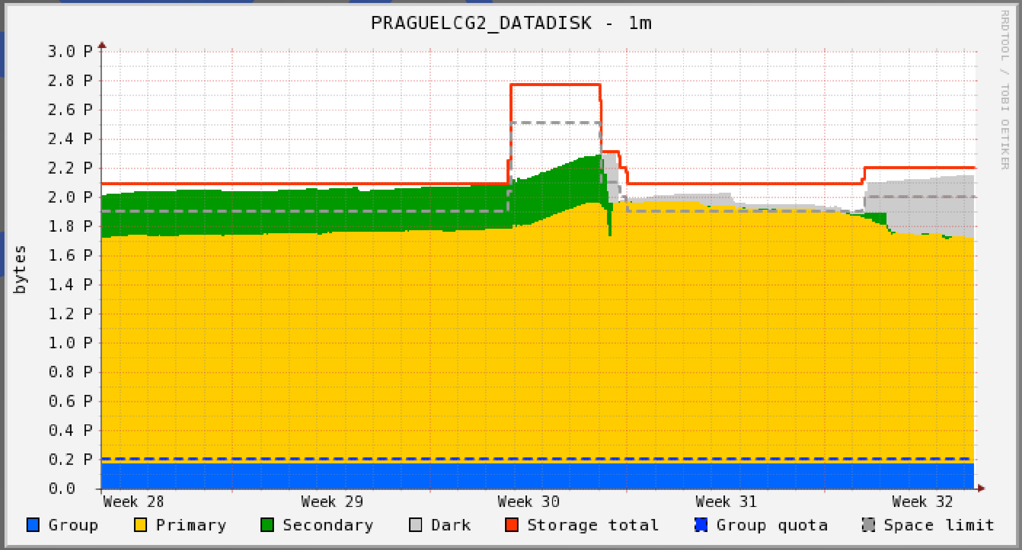
Software testing

- ZFS as an alternative for HW raid
  - Mixed results based on HW (type and age)
  - Still in testing
  - Some features are not marked stable
  - CDDL license

Software testing

- DPM
  - Used version 1.10.3
  - DOME flavour core
  - Several bugs encountered and reported

DPM
Disk Pool Manager
lcgdm.web.cern.ch/dpm



DPM short intro

- Storage management
- Easy to setup and maintain
- Multiple VO support
- X509 authentication
- Multi protocol support
- Great user/admin support
- http://lcgdm.web.cern.ch/dpm

DPM DOME migration experience

- Easy to migrate
- Downtime needed
- In production for ATLAS on several sites besides our
  - Gaining some experience with it before recommending migration
  - Several bugs found, reported and fixed 1.10.4
    - Now in EPEL testing
    - Great work of Petr Vokáč
  - No major breakdown
- Own fast scripts for detecting dark and lost data
- Developers plan to drop SRM support
  - Active in WLCG DOMA group to bring TPC for xrootd and http
  - Built-in new non-SRM storage reporting

DPM performance comparison

command	protocol	legacy (s)	DOME (s)
ls	SRM	0.48	0.46
	XROOT	0.28	0.22
	DAVS	0.24	0.9
xattr	SRM	0.59	2.85
	XROOT	0.91	0.33
	DAVS	N/A	N/A
copy 1M file => DPM	SRM	27.10	17.11
	XROOT	17.10	3.69
	DAVS	17.22	3.88
copy 1M DPM => file	SRM	23.51	18.36
	XROOT	4.65	5.01
	DAVS	5.43	6.06
rm	SRM	9.68	9.32
	XROOT	7.32	2.70
	DAVS	8.20	3.26

DPM DOME bugs in 1.10.3

- Host not known error on el7
- SRM/DOME reported space discrepancy
  - Own script for fixing the difference
- Many dome_dirlist requests exhaust the MySQL handles
- dpm-xrootd does not create dir structure in the disk server
- HTTP authentication on dual stack
- The periodic cache cleanup may cause crashes

Thank you for your attention