Second workshop for MARS administrators

MARS at ECMWF

Manuel Fuentes

Sebastien Villaume, Tiago Quintino, Baudouin Raoult

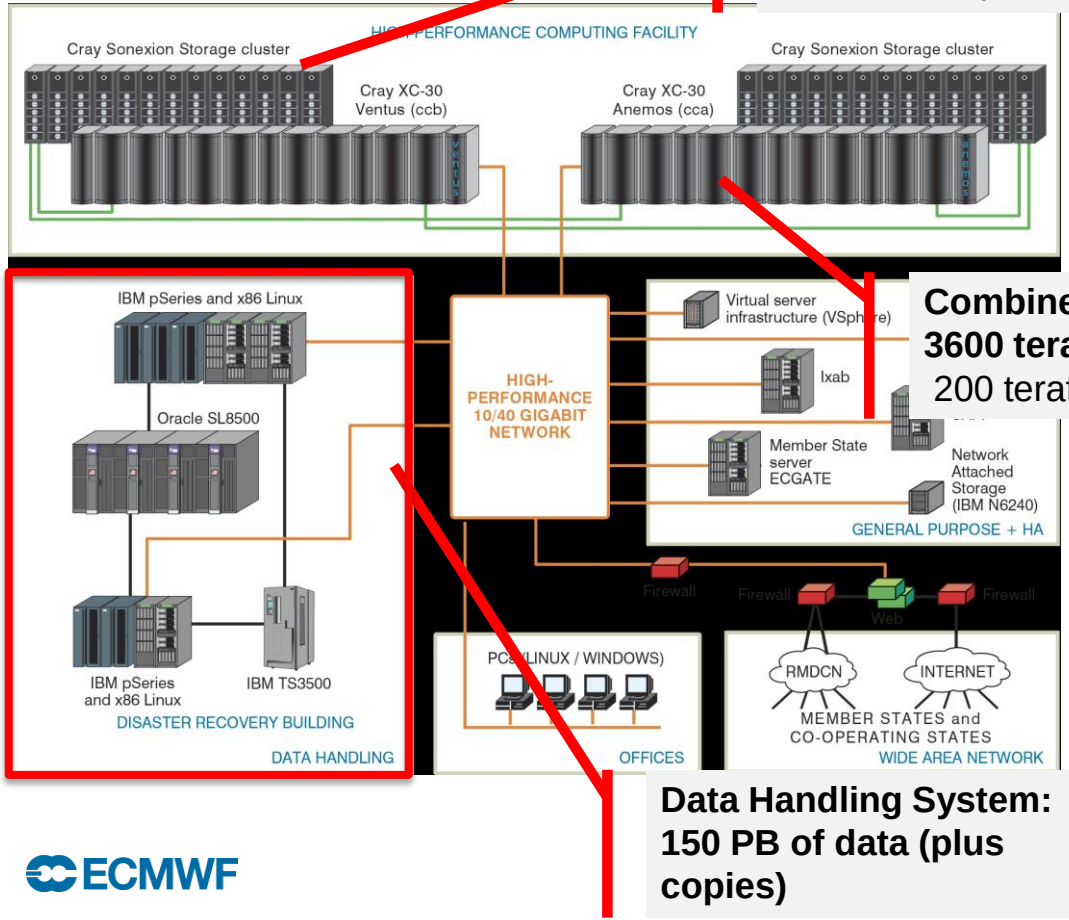
mars@ecmwf.int
mars-admin@lists.ecmwf.int

ECMWF

Outline

- Site configuration
  - Resources
  - Development workflow
- Types of data
- Logs and statistics
- Monitoring

ECMWF's Data Centre



Lustre clusters:
About 14PB (combined)

Combined Power:
3600 teraflops (peak),
200 teraflops (sustained)

Data Handling System:
150 PB of data (plus copies)

ECMWF

Current status

- Primary data:
  - 99.6 PiB for some 13.6 million files (7.6 GiB average file size)
  - 245 billion meteorological fields
- Daily archive
  - Growing about 100 TiB / day
  - Adding about 150 million fields / day
- Disk space
  - 1.4 PiB disk (prearc, cache, locked, logs, ….)
  - 2.5 TiB total disk for metadata (1.5 TiB used)
- About 1000 active users/day executing 2 million requests / day

MARS Services

- Evolved into six operational services, with different access patterns

	Holdings (PiB)	Files	Movers	Cache (TiB)	Prearc (TiB)	Daily Archive (TiB)
marsod	15	860,000	6	400	90	8-12
marsrd	67	11,000,000	5	70	220	50-70
marsode	10	175,000	2	33	80	20-25
marser	4.4	1,100,000	2	160	100	15-20
marsth	1.4	165,000	0	60	11	0.6
marsms	1.8	300,000	0	16	11	1-3
	99.6	**13,600,000**	**15**	**740**	**512**	**~ 100**

Development workflow

- Workstations (via git)
  - simple developments or configuration files
- marsdev
  - MARS related developments
- marsdhst
  - DHS/System testing: OS, HPSS, compilers, disk subsystems, ...
- marstest
  - integration: compilation and testing of versions ready for deployment
  - Reasonable size to do some realistic testing
- marsscratch
  - production test server for users, test data design

Hardware

- About 30 hosts + hot spares
  - IBM RH6.4 6 CPU (12 cores), 48GB
    - Intel(R) Xeon(R) CPU @ 2.53GHz
  - Dell RH6.7, 8 CPU (16 cores), 60GB
    - Intel(R) Xeon(R) CPU @ 2.40GHz
  - More powerful cores (marser, marsod marsrd): 12 CPU (24 cores), 96GB

- Main tape library: Oracle SL8500, 74 T10KD (8 TB), 40 T10KC (5 TB)
- Disaster Recovery System: IBM TS3500, LTO tape drives

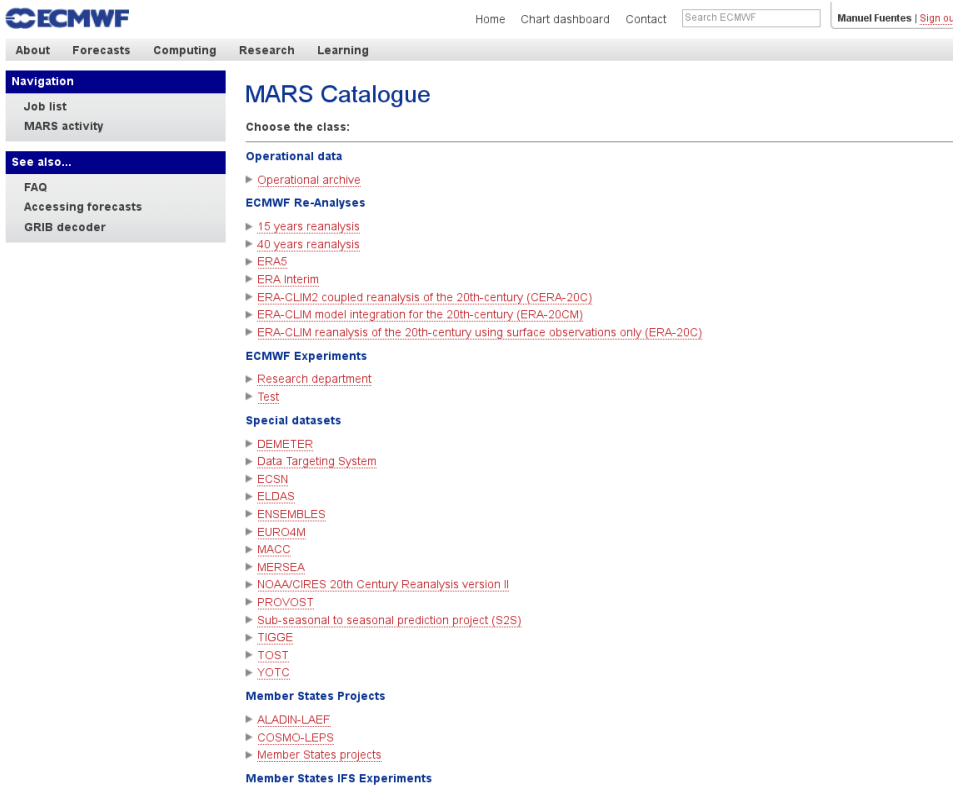
Human Resources

- Software Development
  - Baudouin Raoult (Software Architect)
  - Tiago Quintino (Data Handling Development Team Leader)
- Production Section
  - Manuel Fuentes (Products Team Leader)
  - Sebastien Villaume (MARS Analyst)
- Data Handling Team (ECFS + DHS infrastructure)
  - 5 Analysts
  - 1 Tape librarian

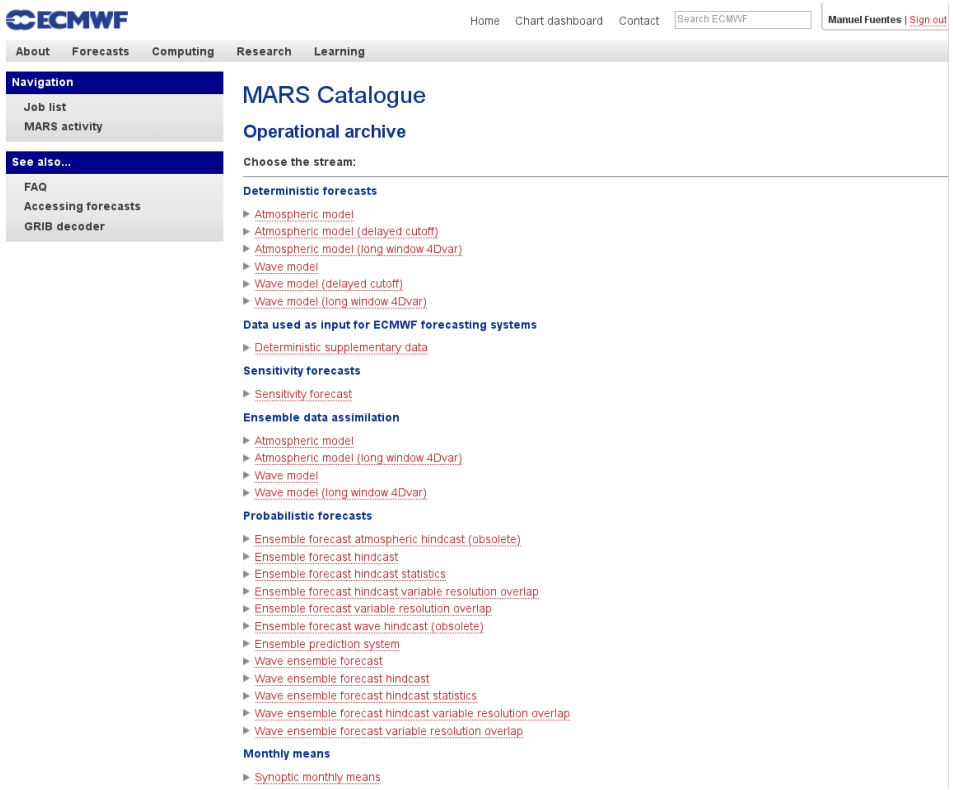
Version of packages running today

```
marsadm> version -long
MARS server 7.5.0 (3cca06ee919c0cdf1f82be58fbf03999ed47c1b4)
    build type  : Production
    timestamp   : 20160111144939
    op. system  : Linux-2.6.32-358.18.1.el6.x86_64 (linux.64)
    processor   : x86_64
    c compiler  : GNU 4.4.7
      flags     :  -pipe -O2 -g
    c++ compiler: GNU 4.4.7
      flags     :  -pipe -O2 -g
eckit 0.6.2 (58db2b63ed9ecc9a0fa2336ed893341d79532b6f)
grib_api 1.14.5 (6449e6f0cb00da95e4ff1270e217ac296fc1b590)
odb_api 0.10.2 (ab9dd19b384165df7e0aecc304d060d774d21944)
hpss 7.4.2.1 (/opt/hpss7421a_prod)
```

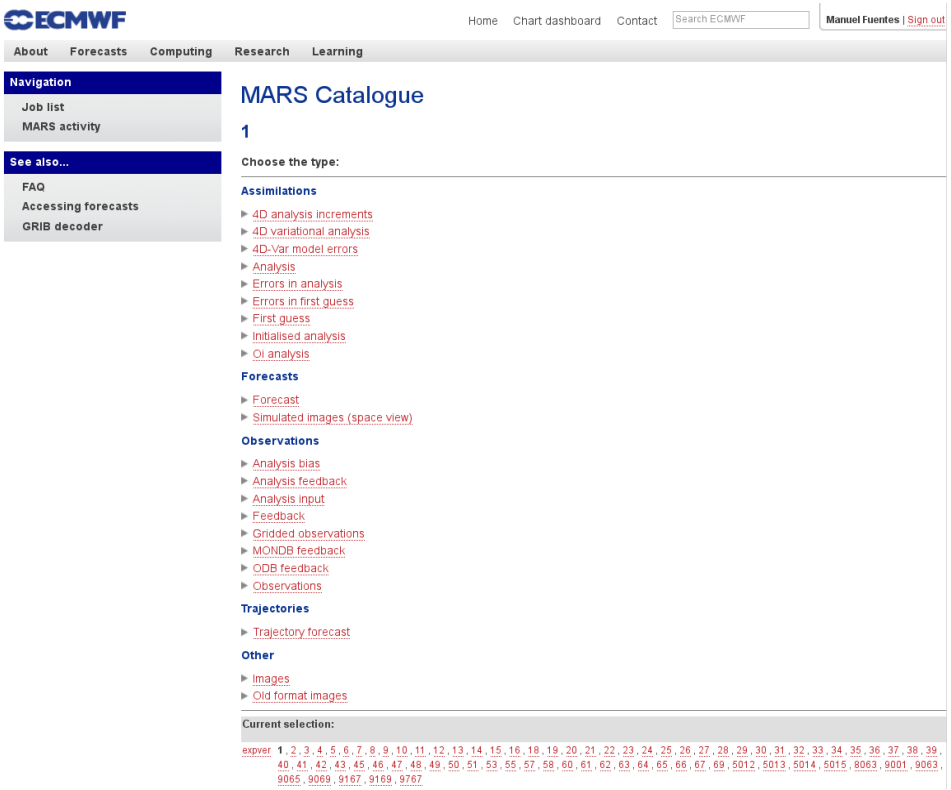
MARS Class: Projects



MARS stream: Forecasting Systems



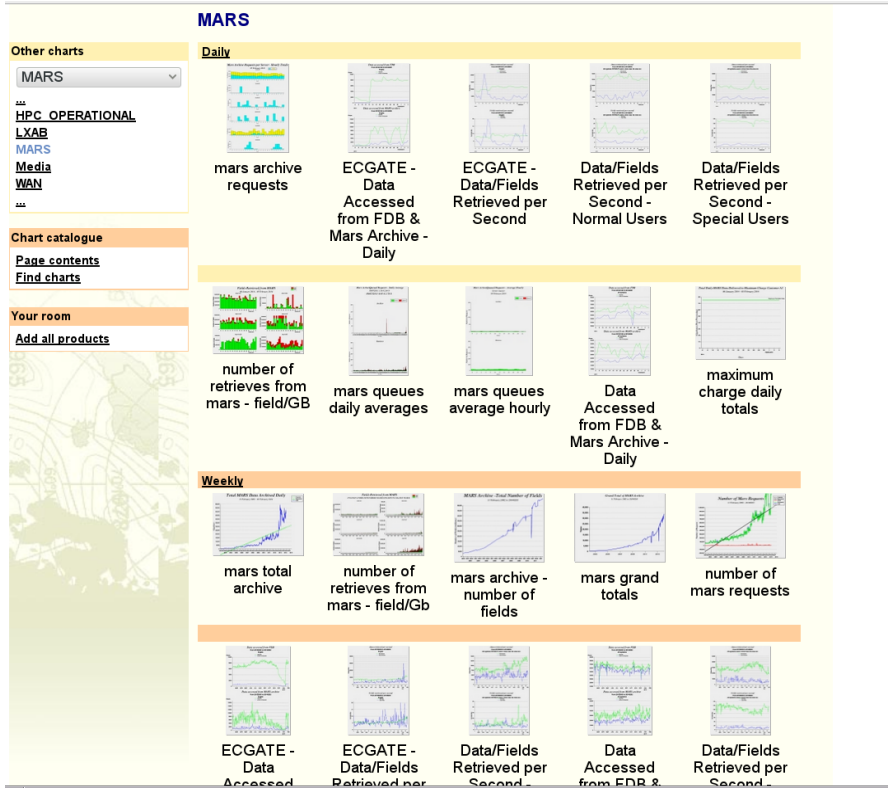
MARS type: Types of data



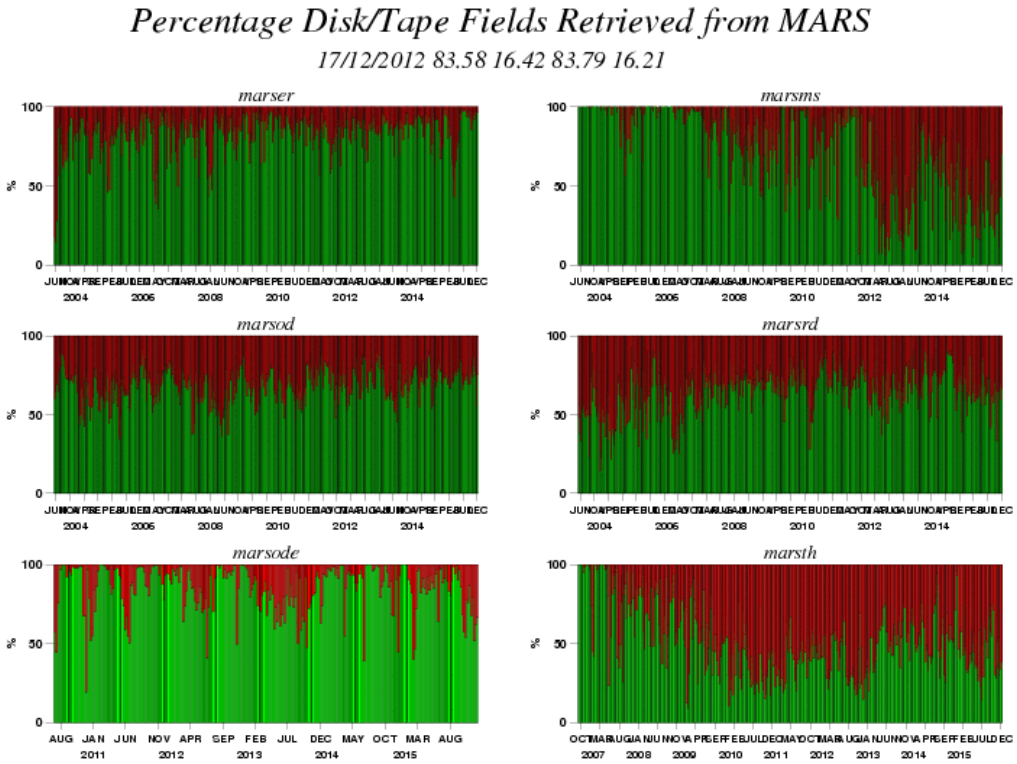
Client logfiles: $MARS_STATISTICS_FILE

$startdate="20160305";$starttime="00:06:21";$verb="retrieve";$version="20160226";$application="mars";$class="ea";$type="an";$stream="oper";$expver="2477";$retdate="19980625";$age="6463";$nbdates="1";$reqno="1";$bytes_online="24660594";$disk_files="1";$fields_online="105";$fields="105";$database="marser";$bytes="18432414";$written="6109740";$interpolated="105";$transfertime="0";$writetarget="0";$cpu="0";$elapsed="1";$status="ok";$stopdate="20160305";$stoptime="00:06:23";$user="eras";$category="all|basic|product_before_schedule";$owner="eras";$account="ecc3s|eccams|ecera|eceracli|ecrmnx";$abc="ecc3s";$environment="batch";$host="lxc14|156.136.174.166";$domain="linux.ecmwf.int";$sourcebranch="grib_odb_api";$pid="46860";$r_class="ea";$r_type="an";$r_stream="oper";$r_expver="2477";$r_repres="sh";$r_levtype="pl";$r_levelist="1|2|3|5|7|10|20|30|50|70|100|150|200|250|300|400|500|700|850|925|1000";$r_param="130|133|131|132|203.128";$r_date="19980625";$r_time="1200";$r_step="0";$r_domain="g";$r_target="./mars_request_000.grib";$r_resol="auto";$r_grid="1.5|1.5";$r_expect="any";

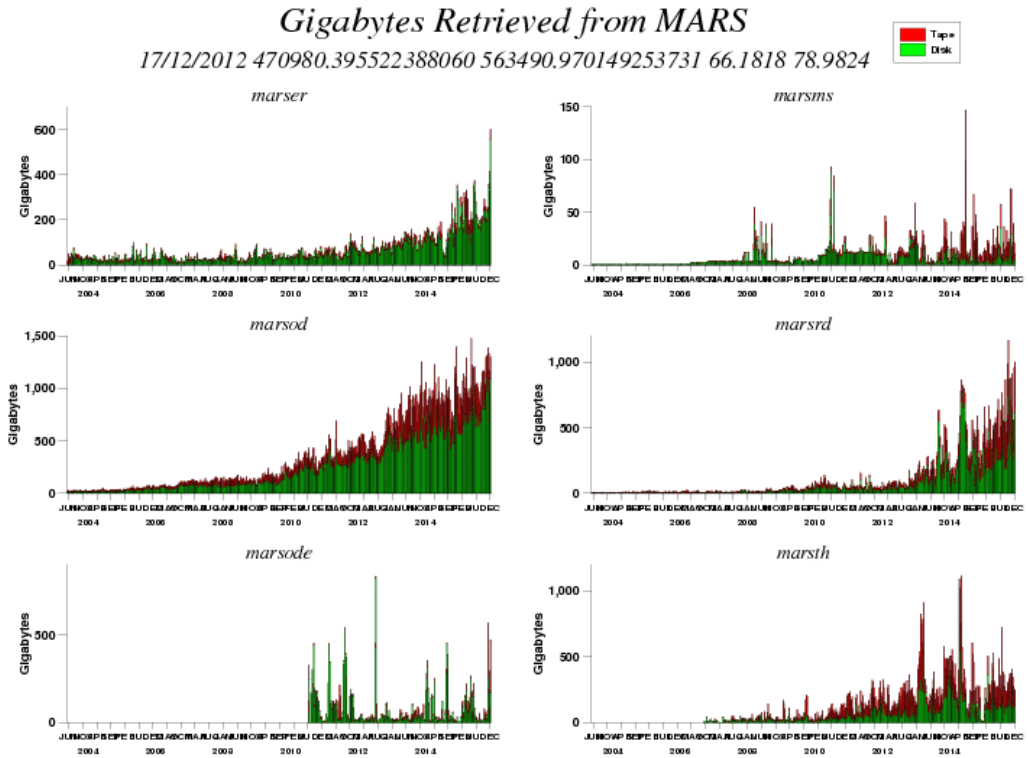
Client side statistics



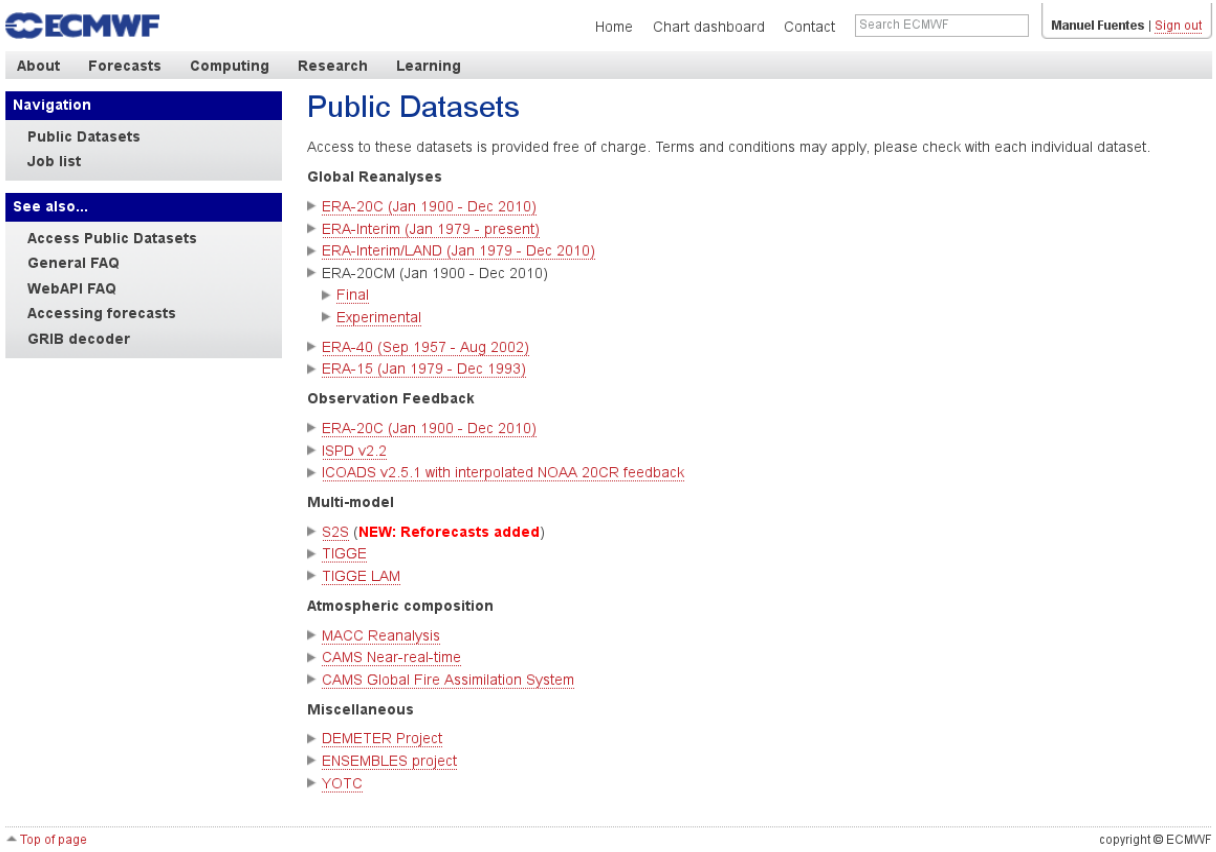
Client side statistics



Client side statistics



Public Data Server



Data Server statistics

Pages /... / Monthly stats

Edit Favourite Watch

S2S monthly since May 2015

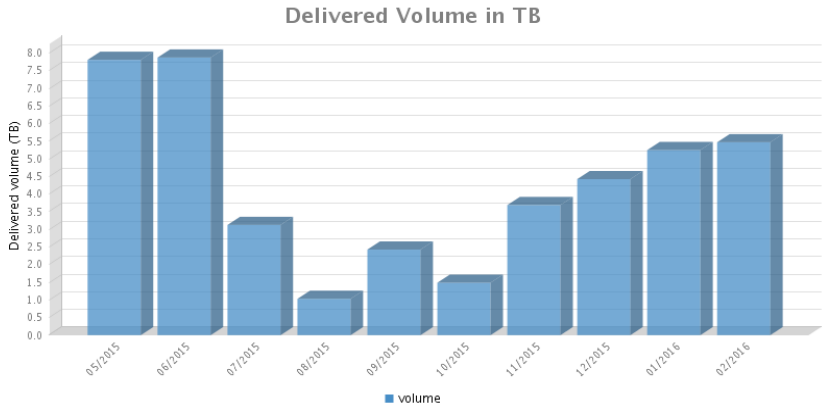
Created by Matthew Manoussakis, last modified on Feb 25, 2016

In brief

Results in brief

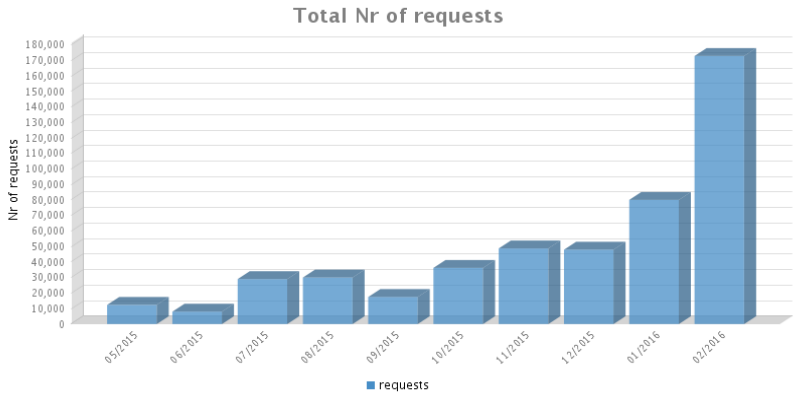
Total fields	1406482726
Total requests	482623
Delivered Volume (TB)	42.5800235271453857

Delivered volumes

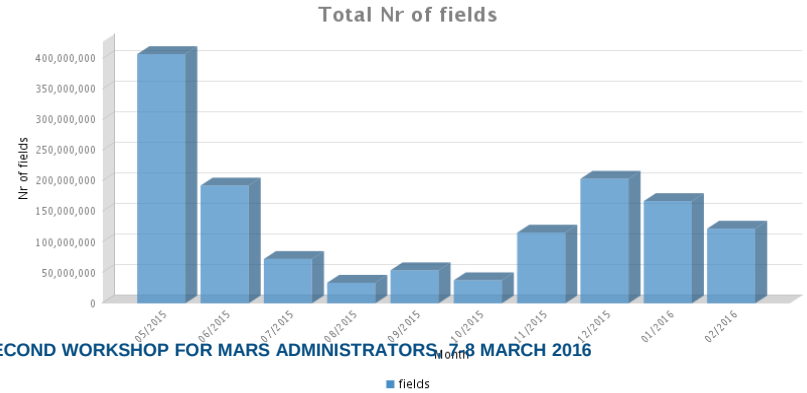


Data Server statistics

Nr of requests



Nr of fields



Captors

```
marsadm> captors
$VAR1 = {
  'tapes.hpss.read.media_requests.205986012' => 1,
  'tapes.hpss.read.volumes.queued.UD054200' => 1,
  'tapes.hpss.read.volumes.queued.UD025500' => 27,
  'tapes.hpss.read.volumes.queued.UD052700' => 1,
  'mars.retrieve.host.atls17' => 3,
  'tapes.hpss.read.volumes.picked.UD066800' => 3,
  'mars.retrieve.user.max.larson@me.com' => 1,
  'mars.cost.fields' => '2768',
  'tapes.hpss.read.media_requests.205985283' => 1,
  'mars.retrieve.user.rajeshpv@msn.com' => 2,
  'tapes.hpss.read.request_id.205983305.media.UD025500' => 27,
  'mars.retrieve.host.ecgb08' => 1,
  'watermarks.locked.max' => '98',
  'mars.archive.host.ccbppn042' => 1,
…
```

Ganglia

Created by Manuel Fuentes, last modified on Dec 15, 2015

Entry point for ganglia http://dhs-ganglia.ecmwf.int/ganglia/

Entry point for MARS displays in ganglia:

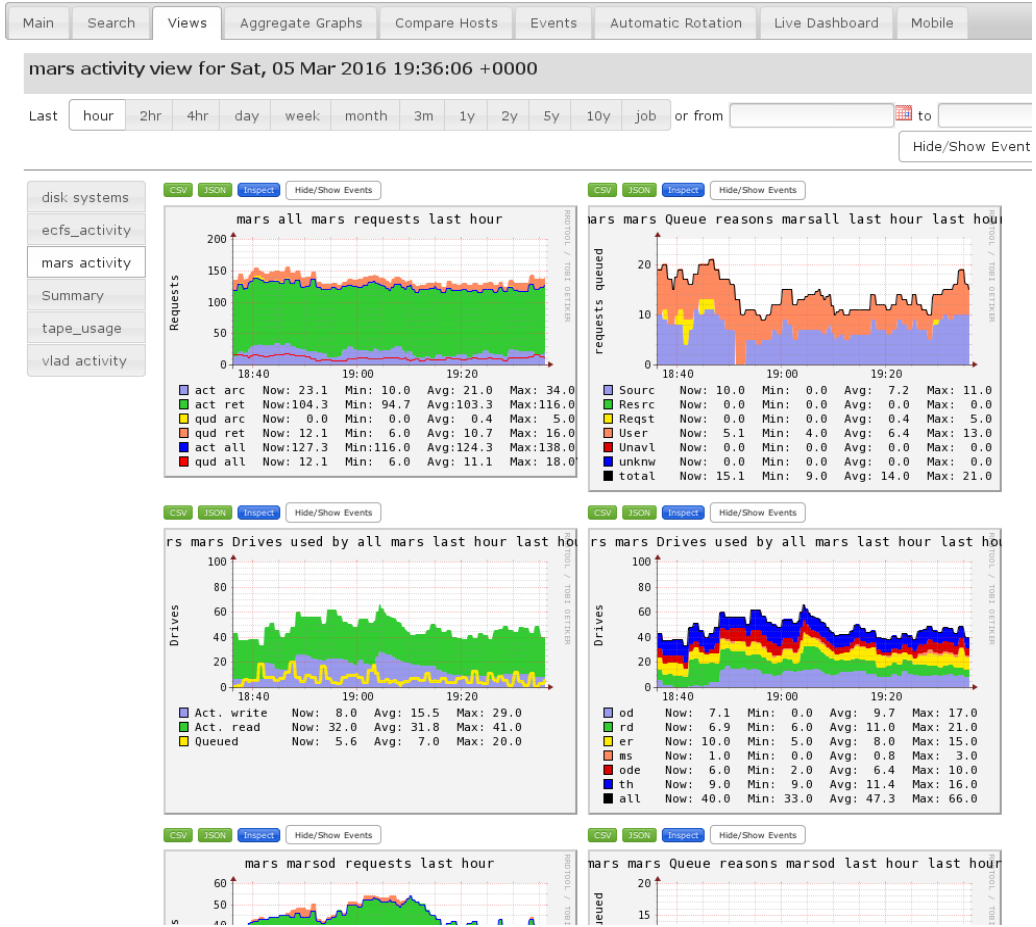
- MARS systems
- Disk systems
- Tape usage
- DHS systems under RH6.4, contains MARS monitoring at the top, HOST monitoring at the bottom. The table below shows a direct link to the -core machine. For the -mvrnn machines, first find the machine name for the mover (using the Mapping for mars-services and hostnames), then edit the URL.

All services	drives	queues	reasons for queueing	mismatch	mismatch	CPU/IO usage on -core
marsod	drives	queues	reasons	cache	prearc	dhs0025
marsrd	drives	queues	reasons	cache	prearc	dhs1101
marsode	drives	queues	reasons	cache	prearc	dhs1106
marser	drives	queues	reasons	cache	prearc	dhs1108
marsth	drives	queues	reasons	cache	prearc	dhs1208
marsms	drives	queues	reasons	cache	prearc	dhs1128

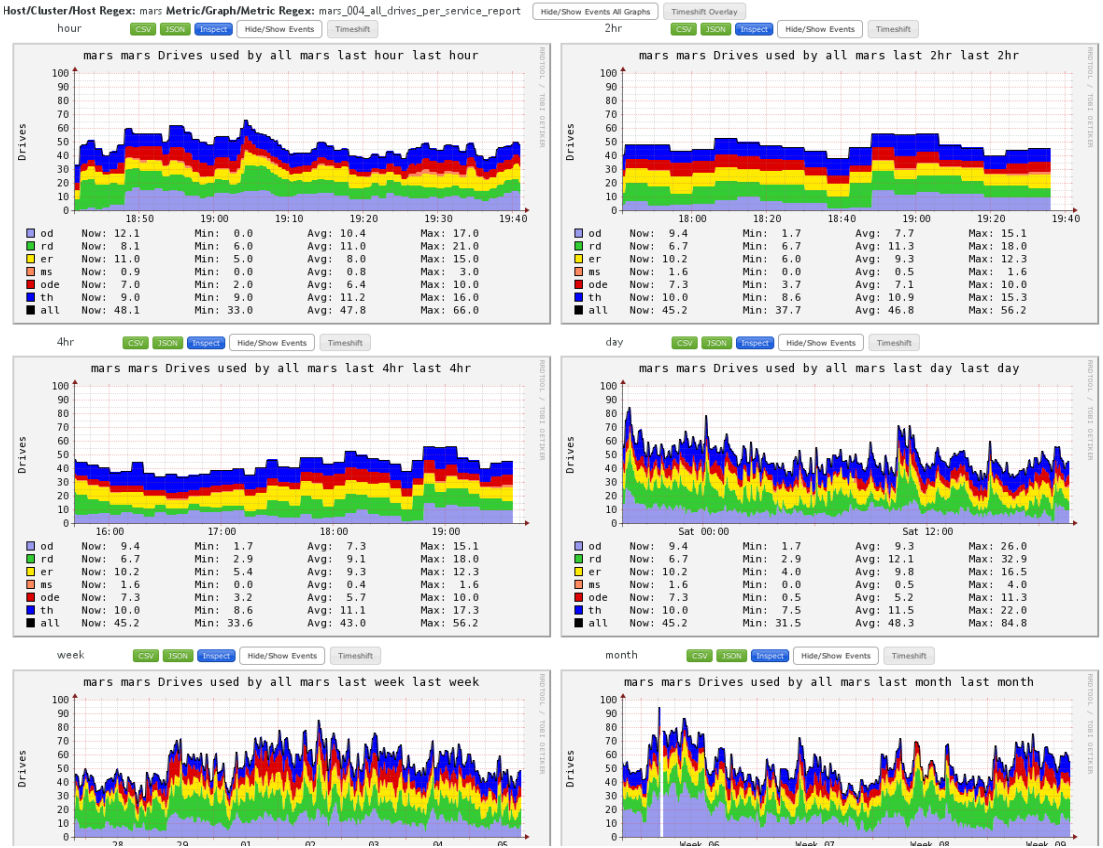
Explanation of the reasons for queueing:

- **source**, indicates the source of the request, eg, we have limits on requests coming from Data Servers
- **resource**, indicates there is a resource limitation, usually tape drives. We have limits in the number of concurrent requests accessing more than x number of tapes.
- **requests**, eg, total number of requests of a given type. There is a certain number of slots for archive requests, another for retrieve requests, etc...
- **user**, total number of requests for a given user
- **unavailable**, indicates that a given system is unavailable. This could be HPSS not available or a MARS mover serving some data from disk not available
- ***unknown***, none of the above

Ganglia


Main
Search
Views
Aggregate Graphs
Compare Hosts
Events
Automatic Rotation
Live Dashboard
Mobile
mars activity view for Sat, 05 Mar 2016 19:36:06 +0000
Last
hour
2hr
4hr
day
week
month
3m
1y
2y
5y
10y
job
or from
to
Hide/Show Event
disk systems
ecfs_activity
mars activity
Summary
tape_usage
vlad activity
CSV
JSON
Inspect
Hide/Show Events
mars all mars requests last hour
Requests
act arc Now: 23.1 Min: 10.0 Avg: 21.0 Max: 34.0
act ret Now:104.3 Min: 94.7 Avg:103.3 Max:116.0
qud arc Now: 0.0 Min: 0.0 Avg: 0.4 Max: 5.0
qud ret Now: 12.1 Min: 6.0 Avg: 10.7 Max: 16.0
act all Now:127.3 Min:116.0 Avg:124.3 Max:138.0
qud all Now: 12.1 Min: 6.0 Avg: 11.1 Max: 18.0
ars mars Queue reasons marsall last hour last hou
requests queued
Sourc Now: 10.0 Min: 0.0 Avg: 7.2 Max: 11.0
Resrc Now: 0.0 Min: 0.0 Avg: 0.0 Max: 0.0
Reqst Now: 0.0 Min: 0.0 Avg: 0.4 Max: 5.0
User Now: 5.1 Min: 4.0 Avg: 6.4 Max: 13.0
Unavl Now: 0.0 Min: 0.0 Avg: 0.0 Max: 0.0
unknw Now: 0.0 Min: 0.0 Avg: 0.0 Max: 0.0
total Now: 15.1 Min: 9.0 Avg: 14.0 Max: 21.0
rs mars Drives used by all mars last hour last ho
Drives
Act. write Now: 8.0 Avg: 15.5 Max: 29.0
Act. read Now: 32.0 Avg: 31.8 Max: 41.0
Queued Now: 5.6 Avg: 7.0 Max: 20.0
rs mars Drives used by all mars last hour last ho
od Now: 7.1 Min: 0.0 Avg: 9.7 Max: 17.0
rd Now: 6.9 Min: 6.0 Avg: 11.0 Max: 21.0
er Now: 10.0 Min: 5.0 Avg: 8.0 Max: 15.0
ms Now: 1.0 Min: 0.0 Avg: 0.8 Max: 3.0
ode Now: 6.0 Min: 2.0 Avg: 6.4 Max: 10.0
th Now: 9.0 Min: 9.0 Avg: 11.4 Max: 16.0
all Now: 40.0 Min: 33.0 Avg: 47.3 Max: 66.0
mars marsod requests last hour
mars mars Queue reasons marsod last hour last hour

Ganglia: long term statistics



Opsview: Operator monitoring

OPSVIEW

mar

dashboard monitoring modules settings

Service Status > Opsview » Systems Software Section » DHS » marser-core

Host	Service	Status	Last Check	#	Status Information
marser-core	BS - Disks	OK	2016-03-05 20:10:00	1/1	OK
	BS - Drives	OK	2016-03-05 20:10:00	1/1	OK
	BS - HPSS	OK	2016-03-05 20:10:00	1/1	OK
	BS - Locks	OK	2016-03-05 20:10:00	1/1	OK
	BS - Mismatch	OK	2016-03-05 20:10:00	1/1	OK
	BS - Shutdown	OK	2016-03-05 20:10:00	1/1	OK
	Connectivity - LAN	OK	2016-03-05 20:11:45	1/3	OK - 136.156.164.31: rta 0.296ms, lost 0%
Totals	7	7 OK			

Version: 4.6.3

This software is supported and certified by Opsview Limited

OPSVIEW

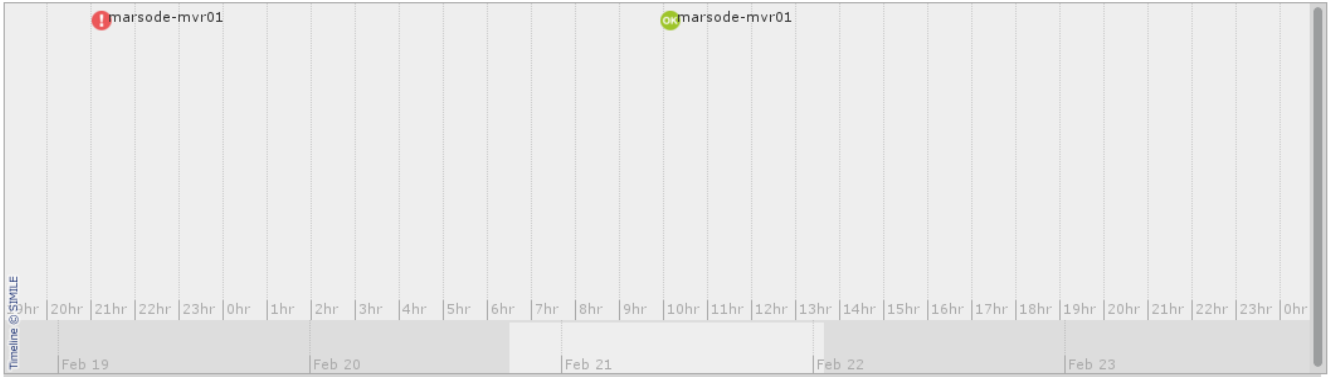
mar | search

dashboard | monitoring | modules | settings | help

Events

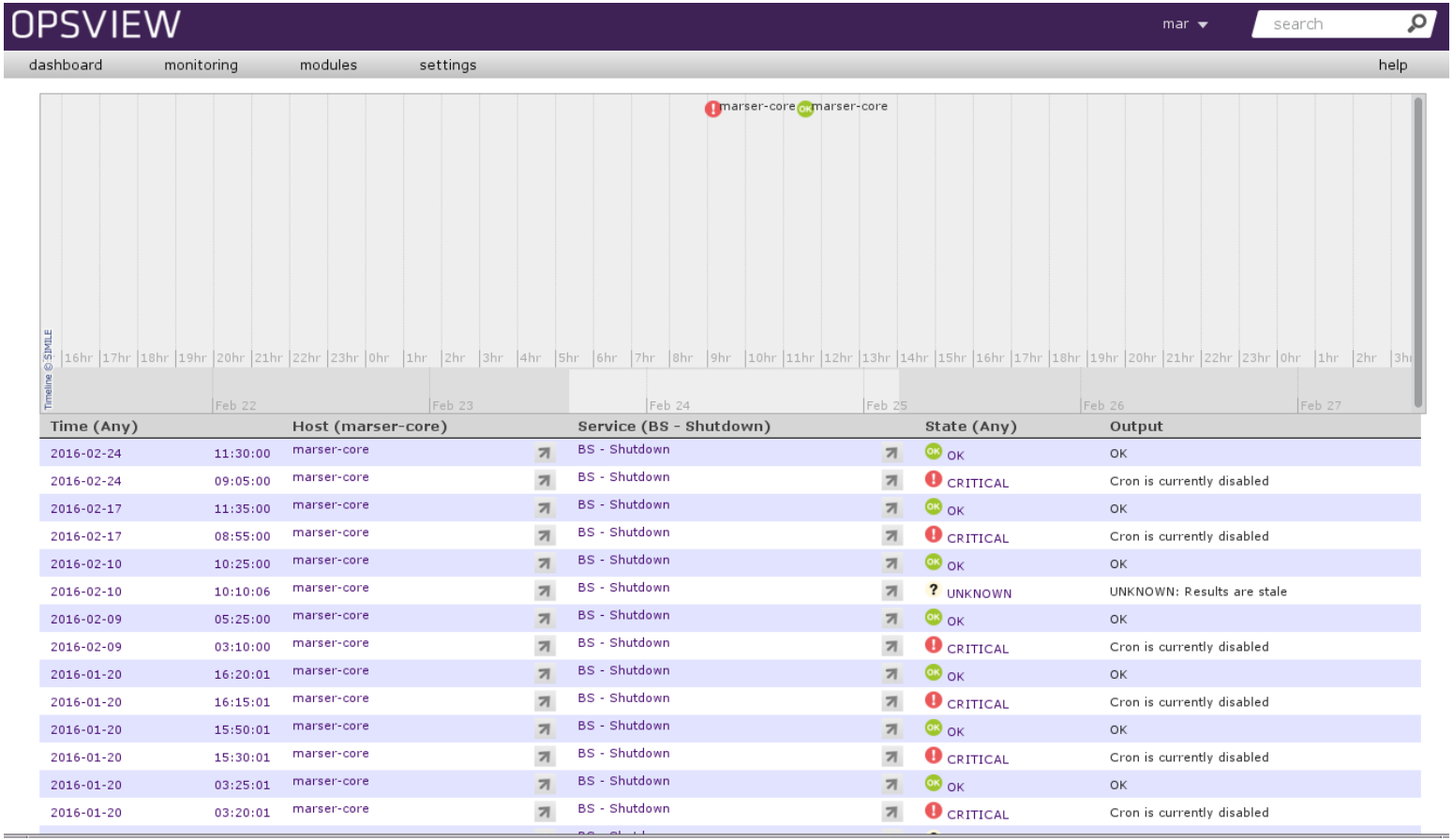
FILTERS: Host: marsode-mvr01, Service: BS - Disks, State Type: HARD

Export Data: Select



Time (Any)		Host (marsode-mvr01)	Service (BS - Disks)	State (Any)	Output
2016-02-21	10:05:01	marsode-mvr01	BS - Disks	OK	OK
2016-02-20	21:10:02	marsode-mvr01	BS - Disks	CRITICAL	Filesystem /data/mars_p_d12_1_16 is full
2016-02-10	10:25:01	marsode-mvr01	BS - Disks	OK	OK
2016-02-10	10:10:06	marsode-mvr01	BS - Disks	UNKNOWN	UNKNOWN: Results are stale
2015-12-09	12:10:00	marsode-mvr01	BS - Disks	OK	OK
2015-12-09	06:15:26	marsode-mvr01	BS - Disks	UNKNOWN	UNKNOWN: Results are stale
2015-11-25	09:40:03	marsode-mvr01	BS - Disks	OK	OK
2015-11-25	06:20:21	marsode-mvr01	BS - Disks	UNKNOWN	UNKNOWN: Results are stale
2015-11-20	15:20:01	marsode-mvr01	BS - Disks	OK	OK
2015-11-20	14:42:53	marsode-mvr01	BS - Disks	UNKNOWN	UNKNOWN: Results are stale
Time (Any)		Host (marsode-mvr01)	Service (BS - Disks)	State (Any)	Output

< 1 >



Time (Any)		Host (marser-core)	Service (BS - Shutdown)	State (Any)	Output
2016-02-24	11:30:00	marser-core	BS - Shutdown	OK	OK
2016-02-24	09:05:00	marser-core	BS - Shutdown	CRITICAL	Cron is currently disabled
2016-02-17	11:35:00	marser-core	BS - Shutdown	OK	OK
2016-02-17	08:55:00	marser-core	BS - Shutdown	CRITICAL	Cron is currently disabled
2016-02-10	10:25:00	marser-core	BS - Shutdown	OK	OK
2016-02-10	10:10:06	marser-core	BS - Shutdown	UNKNOWN	UNKNOWN: Results are stale
2016-02-09	05:25:00	marser-core	BS - Shutdown	OK	OK
2016-02-09	03:10:00	marser-core	BS - Shutdown	CRITICAL	Cron is currently disabled
2016-01-20	16:20:01	marser-core	BS - Shutdown	OK	OK
2016-01-20	16:15:01	marser-core	BS - Shutdown	CRITICAL	Cron is currently disabled
2016-01-20	15:50:01	marser-core	BS - Shutdown	OK	OK
2016-01-20	15:30:01	marser-core	BS - Shutdown	CRITICAL	Cron is currently disabled
2016-01-20	03:25:01	marser-core	BS - Shutdown	OK	OK
2016-01-20	03:20:01	marser-core	BS - Shutdown	CRITICAL	Cron is currently disabled

Thank you for your attention

Any questions ?