

Second workshop for MARS administrators

MARS at ECMWF

Manuel Fuentes

Sebastien Villaume, Tiago Quintino, Baudouin Raoult

mars@ecmwf.int

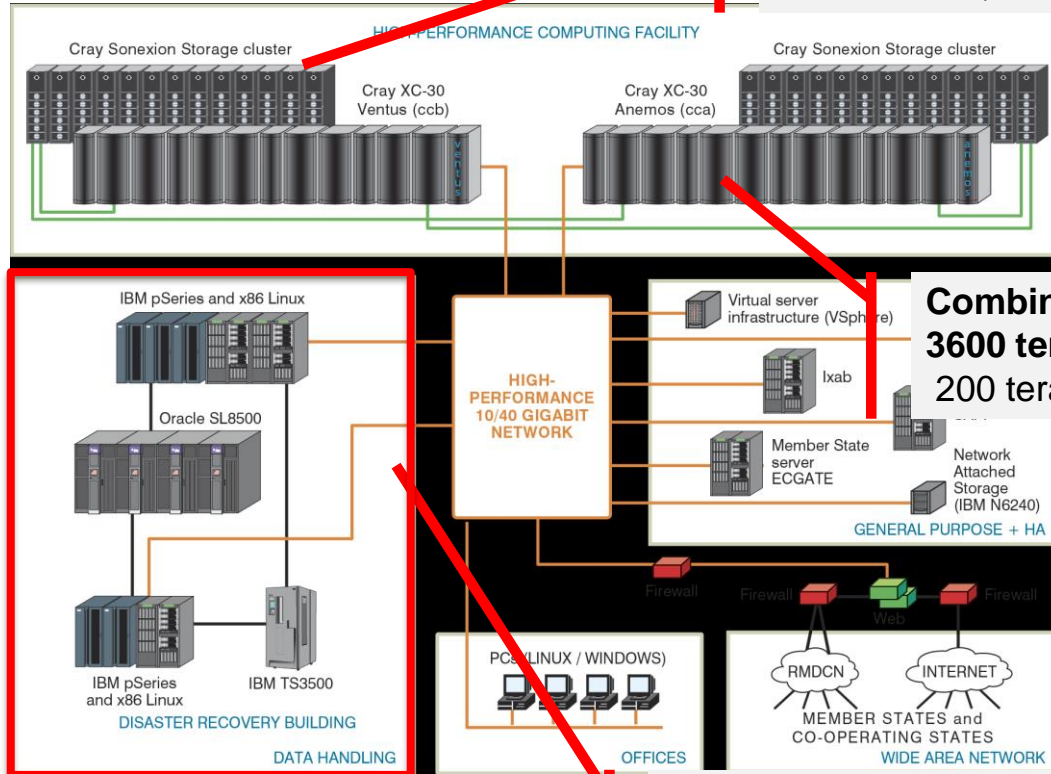
mars-admin@lists.ecmwf.int

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)

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 (marsr, 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
```

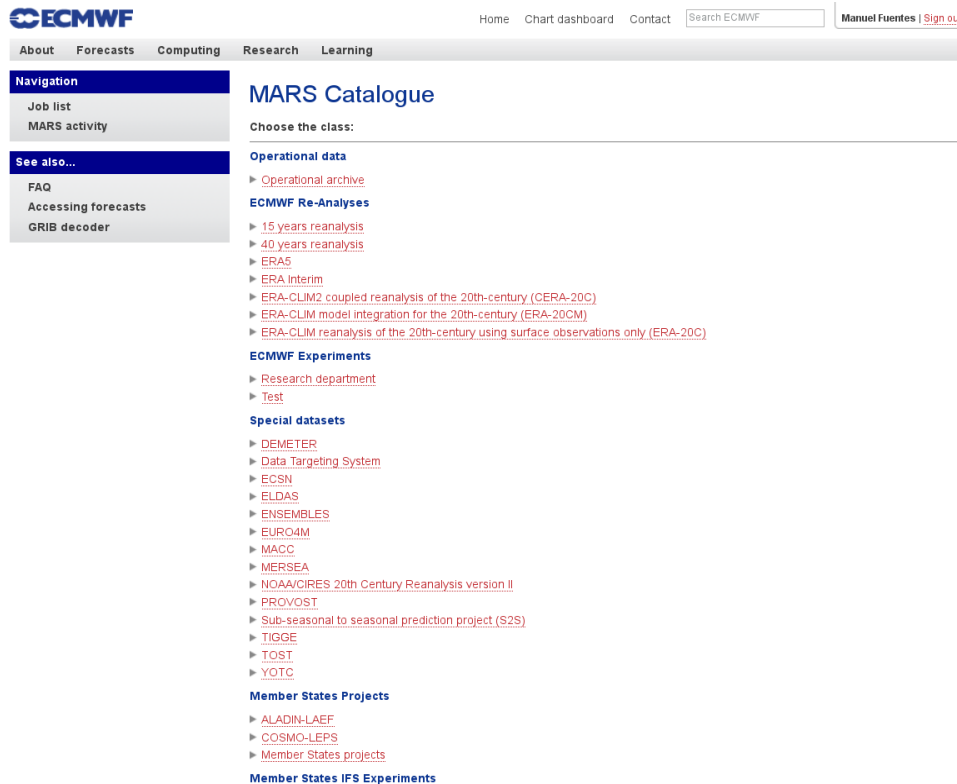
```
eckit 0.6.2 (58db2b63ed9ecc9a0fa2336ed893341d79532b6f)
```

```
grib_api 1.14.5 (6449e6f0cb00da95e4ff1270e217ac296fc1b590)
```

```
odb_api 0.10.2 (ab9dd19b384165df7e0aecc304d060d774d21944)
```

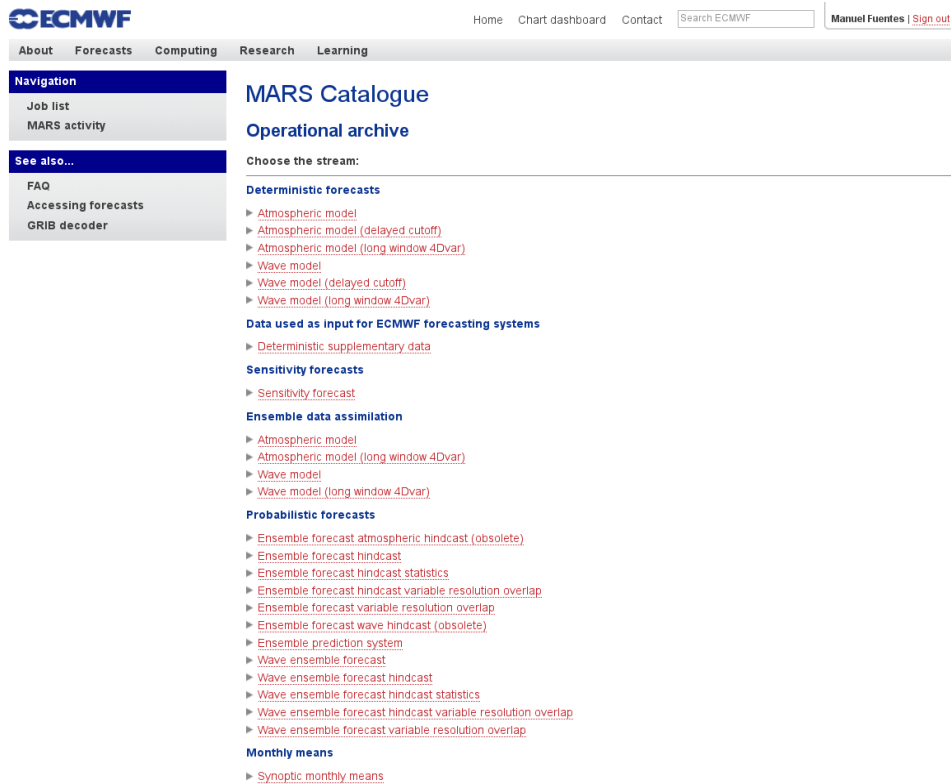
```
hpss 7.4.2.1 (/opt/hpss7421a_prod)
```

MARS Class: Projects



The screenshot shows the ECMWF MARS Catalogue website. At the top, there is a navigation bar with the ECMWF logo on the left and links for Home, Chart dashboard, Contact, a search box containing 'Search ECMWF', and a user profile for 'Manuel Fuentes' with a 'Sign out' link. Below the navigation bar is a secondary menu with 'About', 'Forecasts', 'Computing', 'Research', and 'Learning'. On the left side, there are two vertical navigation panels. The first panel, titled 'Navigation', contains links for 'Job list' and 'MARS activity'. The second panel, titled 'See also...', contains links for 'FAQ', 'Accessing forecasts', and 'GRIB decoder'. The main content area is titled 'MARS Catalogue' and features a section 'Choose the class:'. Below this, there are several categorized lists of links: 'Operational data' (with a link to 'Operational archive'), 'ECMWF Re-Analyses' (with links to '15 years reanalysis', '40 years reanalysis', 'ERA5', 'ERA Interim', 'ERA-CLIM2 coupled reanalysis of the 20th-century (CERA-20C)', 'ERA-CLIM model integration for the 20th-century (ERA-20CM)', and 'ERA-CLIM reanalysis of the 20th-century using surface observations only (ERA-20C)'), 'ECMWF Experiments' (with links to 'Research department' and 'Test'), 'Special datasets' (with links to 'DEMETER', 'Data Targeting System', 'ECSN', 'ELDAS', 'ENSEMBLES', 'EURO4M', 'MACC', 'MERSEA', 'NOAA/CIRES 20th Century Reanalysis version II', 'PROVOST', 'Sub-seasonal to seasonal prediction project (S2S)', 'TIGGE', 'TOST', and 'YOTC'), 'Member States Projects' (with links to 'ALADIN-LAEF', 'COSMO-LEPS', and 'Member States projects'), and 'Member States IFS Experiments'.

MARS stream: Forecasting Systems



The screenshot shows the ECMWF website's MARS Catalogue page. The top navigation bar includes 'Home', 'Chart dashboard', 'Contact', a search box, and a user profile for 'Manuel Fuentes | Sign out'. Below this is a secondary menu with 'About', 'Forecasts', 'Computing', 'Research', and 'Learning'. The left sidebar contains 'Navigation' (Job list, MARS activity), 'See also...' (FAQ, Accessing forecasts, GRIB decoder), and 'MARS Catalogue' (Operational archive). The main content area is titled 'MARS Catalogue' and 'Operational archive'. It features a 'Choose the stream:' section with several categories of links: Deterministic forecasts (Atmospheric model, Wave model), Data used as input for ECMWF forecasting systems (Deterministic supplementary data), Sensitivity forecasts (Sensitivity forecast), Ensemble data assimilation (Atmospheric model, Wave model), Probabilistic forecasts (Ensemble forecast atmospheric hindcast, Ensemble forecast hindcast, Ensemble forecast hindcast statistics, Ensemble forecast hindcast variable resolution overlap, Ensemble forecast variable resolution overlap, Ensemble forecast wave hindcast, Ensemble prediction system, Wave ensemble forecast, Wave ensemble forecast hindcast, Wave ensemble forecast hindcast statistics, Wave ensemble forecast hindcast variable resolution overlap, Wave ensemble forecast variable resolution overlap), and Monthly means (Synoptic monthly means).

ECMWF Home Chart dashboard Contact Search ECMWF Manuel Fuentes | Sign out

About Forecasts Computing Research Learning

Navigation
Job list
MARS activity

See also...
FAQ
Accessing forecasts
GRIB decoder

MARS Catalogue

Operational archive

Choose the stream:

Deterministic forecasts

- ▶ [Atmospheric model](#)
- ▶ [Atmospheric model \(delayed cutoff\)](#)
- ▶ [Atmospheric model \(long window 4Dvar\)](#)
- ▶ [Wave model](#)
- ▶ [Wave model \(delayed cutoff\)](#)
- ▶ [Wave model \(long window 4Dvar\)](#)

Data used as input for ECMWF forecasting systems

- ▶ [Deterministic supplementary data](#)

Sensitivity forecasts

- ▶ [Sensitivity forecast](#)

Ensemble data assimilation

- ▶ [Atmospheric model](#)
- ▶ [Atmospheric model \(long window 4Dvar\)](#)
- ▶ [Wave model](#)
- ▶ [Wave model \(long window 4Dvar\)](#)

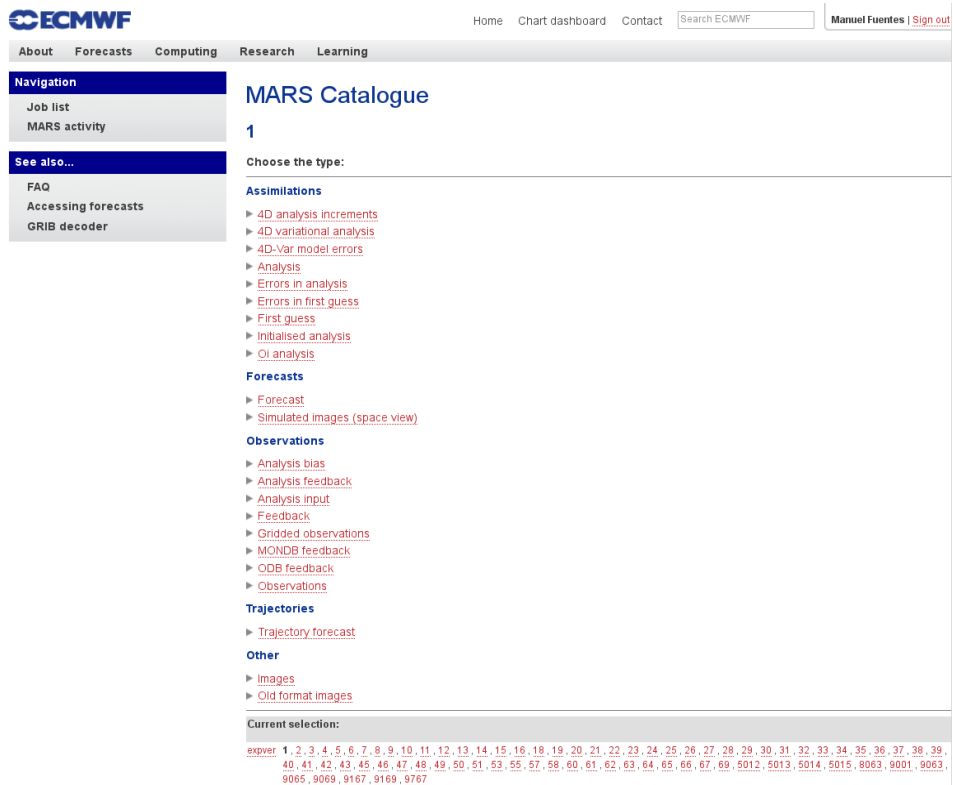
Probabilistic forecasts

- ▶ [Ensemble forecast atmospheric hindcast \(obsolete\)](#)
- ▶ [Ensemble forecast hindcast](#)
- ▶ [Ensemble forecast hindcast statistics](#)
- ▶ [Ensemble forecast hindcast variable resolution overlap](#)
- ▶ [Ensemble forecast variable resolution overlap](#)
- ▶ [Ensemble forecast wave hindcast \(obsolete\)](#)
- ▶ [Ensemble prediction system](#)
- ▶ [Wave ensemble forecast](#)
- ▶ [Wave ensemble forecast hindcast](#)
- ▶ [Wave ensemble forecast hindcast statistics](#)
- ▶ [Wave ensemble forecast hindcast variable resolution overlap](#)
- ▶ [Wave ensemble forecast variable resolution overlap](#)

Monthly means

- ▶ [Synoptic monthly means](#)

MARS type: Types of data



The screenshot shows the ECMWF MARS Catalogue page. At the top, there is a navigation bar with the ECMWF logo and links for Home, Chart dashboard, and Contact. A search box labeled 'Search ECMWF' and a user profile for 'Manuel Fuentes | Sign out' are also present. Below the navigation bar, there are several menu items: 'About', 'Forecasts', 'Computing', 'Research', and 'Learning'. The 'Navigation' menu is expanded, showing 'Job list' and 'MARS activity'. The 'See also...' menu is also expanded, showing 'FAQ', 'Accessing forecasts', and 'GRIB decoder'. The main content area is titled 'MARS Catalogue' and displays a list of data types under the heading 'Choose the type:'. The list is organized into several categories: 'Assimilations', 'Forecasts', 'Observations', 'Trajectories', and 'Other'. Each category contains a list of specific data types with a right-pointing arrow icon. At the bottom of the page, there is a 'Current selection:' section with a list of 'exper' numbers.

ECMWF Home Chart dashboard Contact Search ECMWF Manuel Fuentes | Sign out

About Forecasts Computing Research Learning

Navigation
Job list
MARS activity

See also...
FAQ
Accessing forecasts
GRIB decoder

MARS Catalogue

1

Choose the type:

Assimilations

- ▶ [4D analysis increments](#)
- ▶ [4D variational analysis](#)
- ▶ [4D-Var model errors](#)
- ▶ [Analysis](#)
- ▶ [Errors in analysis](#)
- ▶ [Errors in first guess](#)
- ▶ [First guess](#)
- ▶ [Initialised analysis](#)
- ▶ [O_i analysis](#)

Forecasts

- ▶ [Forecast](#)
- ▶ [Simulated images \(space view\)](#)

Observations

- ▶ [Analysis bias](#)
- ▶ [Analysis feedback](#)
- ▶ [Analysis input](#)
- ▶ [Feedback](#)
- ▶ [Gridded observations](#)
- ▶ [MONDB feedback](#)
- ▶ [ODB feedback](#)
- ▶ [Observations](#)

Trajectories

- ▶ [Trajectory forecast](#)

Other

- ▶ [Images](#)
- ▶ [Old format images](#)

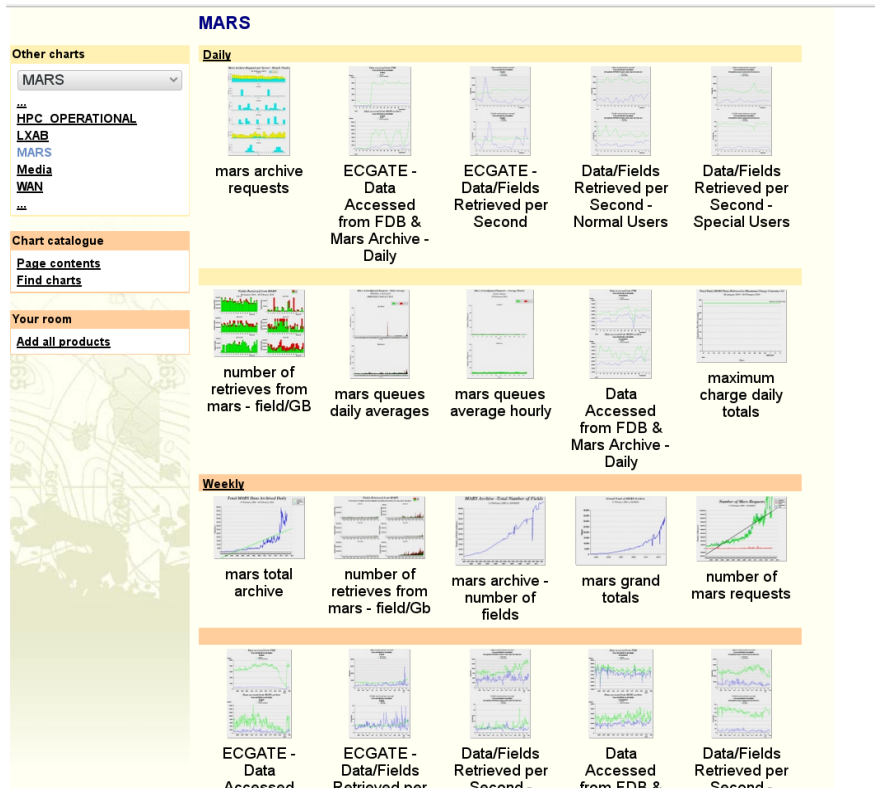
Current selection:

exper 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 53, 55, 57, 58, 60, 61, 62, 63, 64, 65, 66, 67, 69, 5012, 5013, 5014, 5015, 8063, 9001, 9063, 9065, 9069, 9167, 9169, 9767

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";
```

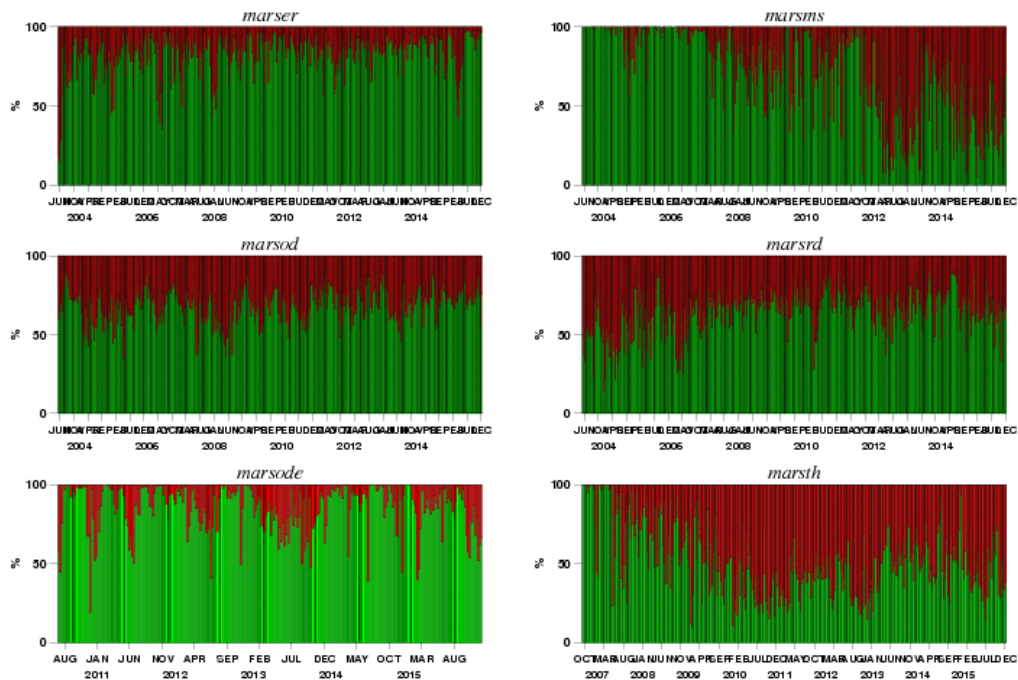
Client side statistics



Client side statistics

Percentage Disk/Tape Fields Retrieved from MARS

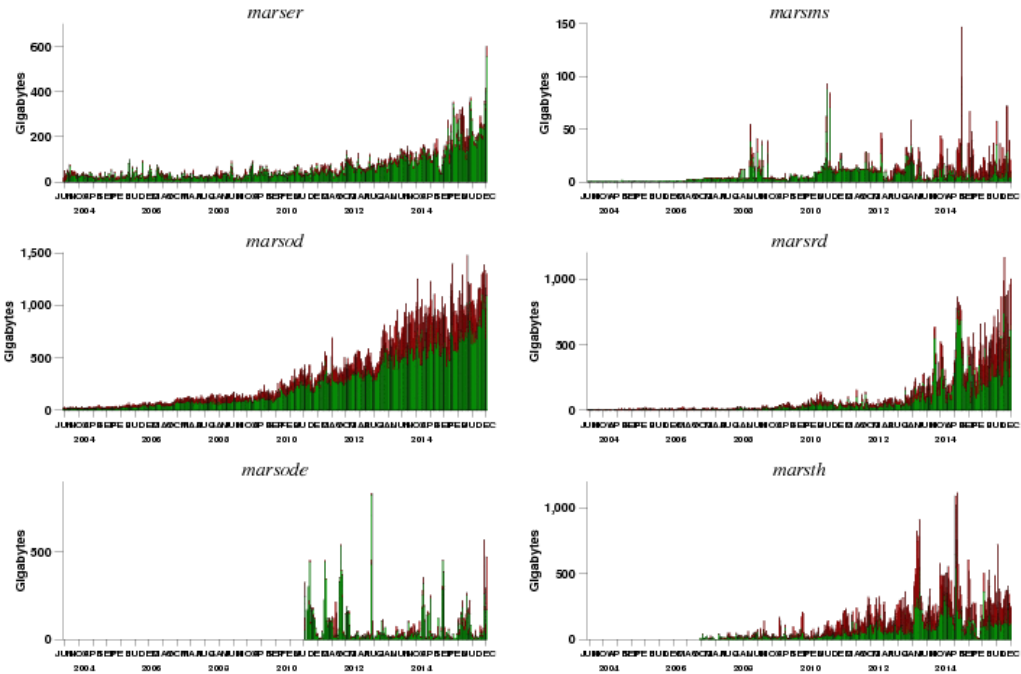
17/12/2012 83.58 16.42 83.79 16.21



Client side statistics

Gigabytes Retrieved from MARS

17/12/2012 470980.395522388060 563490.970149253731 66.1818 78.9824



Navigation

[Public Datasets](#)[Job list](#)

See also...

[Access Public Datasets](#)[General FAQ](#)[WebAPI FAQ](#)[Accessing forecasts](#)[GRIB decoder](#)

Public Datasets

Access to these datasets is provided free of charge. Terms and conditions may apply, please check with each individual dataset.

Global Reanalyses

- ▶ [ERA-20C \(Jan 1900 - Dec 2010\)](#)
- ▶ [ERA-Interim \(Jan 1979 - present\)](#)
- ▶ [ERA-Interim/LAND \(Jan 1979 - Dec 2010\)](#)
- ▶ [ERA-20CM \(Jan 1900 - Dec 2010\)](#)
 - ▶ [Final](#)
 - ▶ [Experimental](#)
- ▶ [ERA-40 \(Sep 1957 - Aug 2002\)](#)
- ▶ [ERA-15 \(Jan 1979 - Dec 1993\)](#)

Observation Feedback

- ▶ [ERA-20C \(Jan 1900 - Dec 2010\)](#)
- ▶ [ISPD v2.2](#)
- ▶ [ICOADS v2.5.1 with interpolated NOAA 20CR feedback](#)

Multi-model

- ▶ [S2S \(NEW: Reforecasts added\)](#)
- ▶ [TIGGE](#)
- ▶ [TIGGE LAM](#)

Atmospheric composition

- ▶ [MACC Reanalysis](#)
- ▶ [CAM5 Near-real-time](#)
- ▶ [CAM5 Global Fire Assimilation System](#)

Miscellaneous

- ▶ [DEMETER Project](#)
- ▶ [ENSEMBLES project](#)
- ▶ [YOTC](#)

Data Server statistics

Pages /... / Monthly stats

[Edit](#) [Favourite](#) [Watch](#) ...

S2S monthly since May 2015

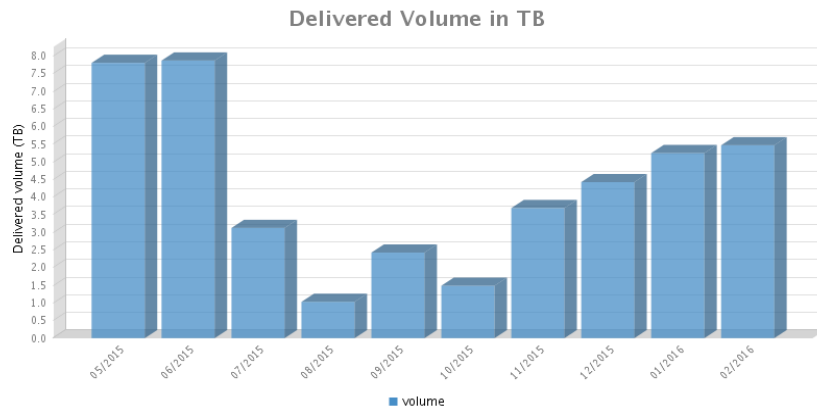
Created by Matthew Manoussakis, last modified on Feb 25, 2016

In brief

Results in brief

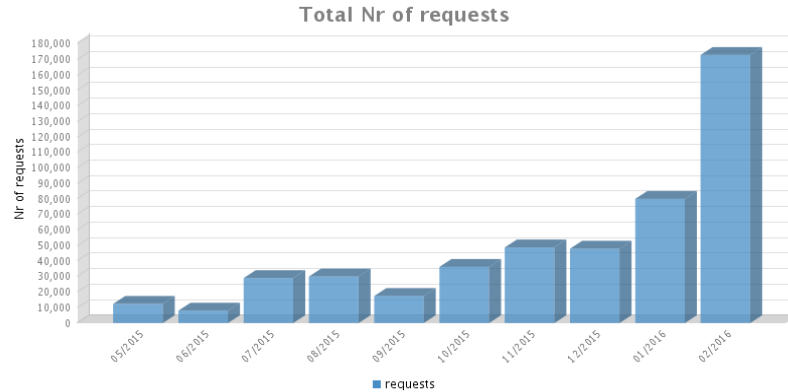
Total fields	1406482726
Total requests	482623
Delivered Volume (TB)	42.5800235271453857

Delivered volumes



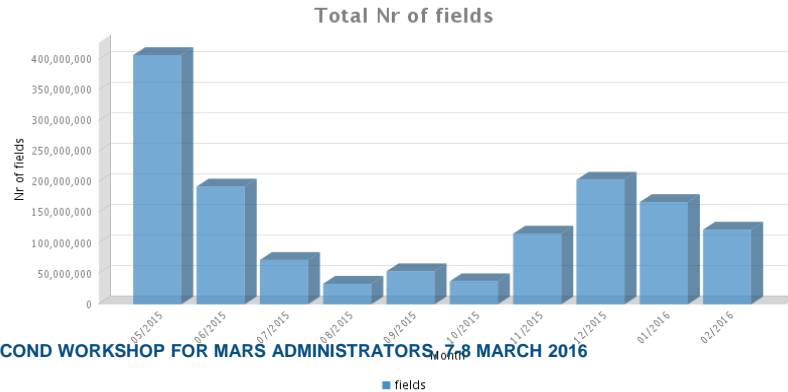
Data Server statistics

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,
```

```
...
```



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 -mvrrn 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
marsrer	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

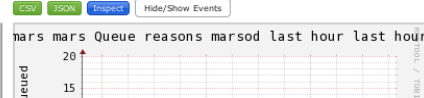
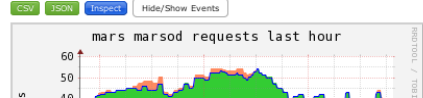
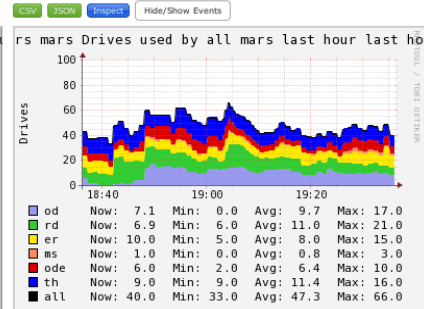
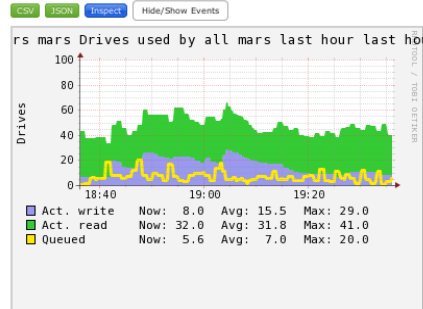
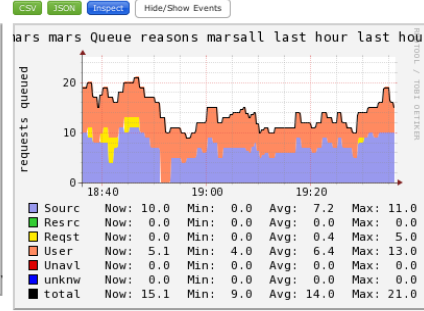
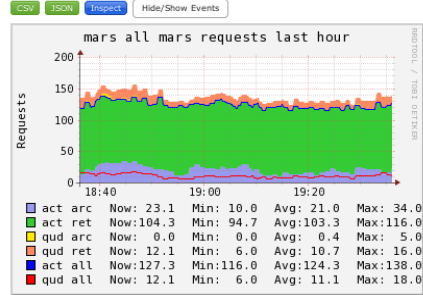
[Main](#)
[Search](#)
[Views](#)
[Aggregate Graphs](#)
[Compare Hosts](#)
[Events](#)
[Automatic Rotation](#)
[Live Dashboard](#)
[Mobile](#)

mars activity view for Sat, 05 Mar 2016 19:36:06 +0000

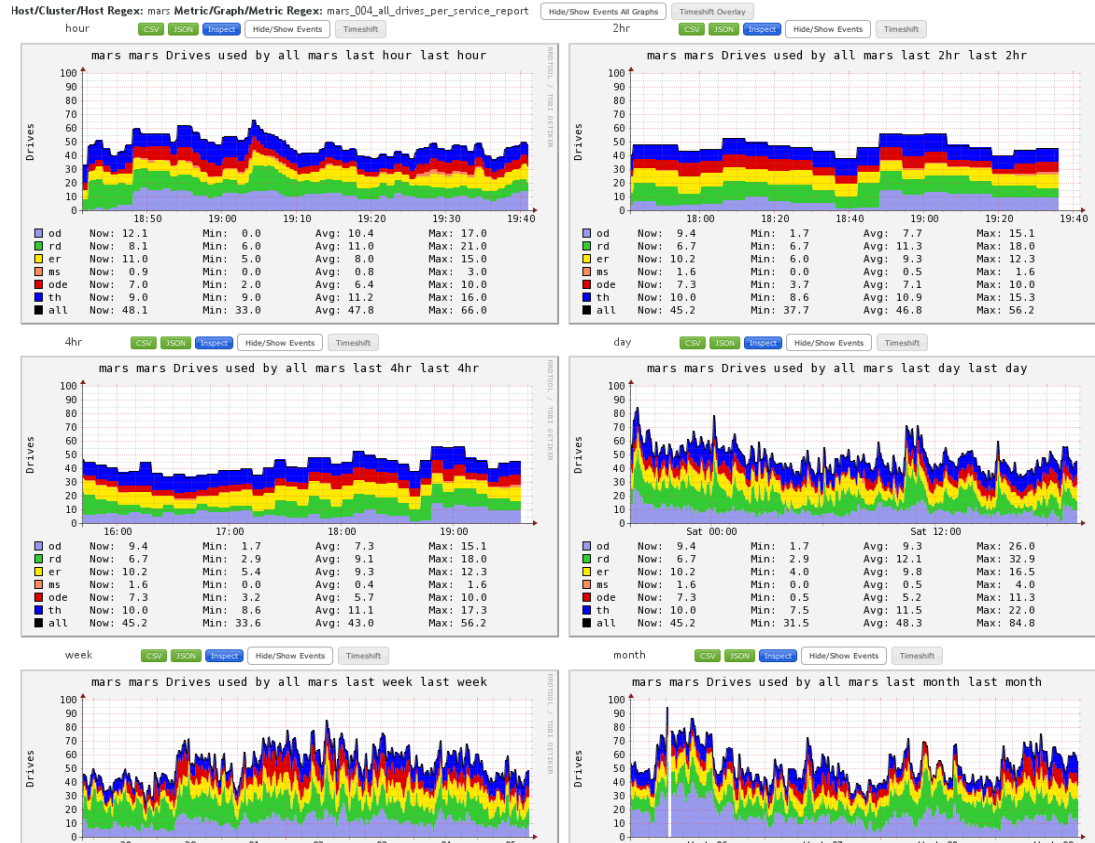
Last 2hr 4hr day week month 3m 1y 2y 5y 10y job or from to

[Hide/Show Event](#)

- disk systems
- ecfs_activity
- mars activity**
- Summary
- tape_usage
- vlad activity



Ganglia: long term statistics



Opsview: Operator monitoring

OPSVIEW mar ▾

dashboard monitoring modules settings help

Keywords > MARS

Failures: 0


Host States

cca-odb	ccb-odb	marser-core	marser-mvr01	marser-mvr02	marsms-core	marsod-core	marsod-mvr01
marsod-mvr02	marsod-mvr03	marsod-mvr04	marsod-mvr05	marsod-mvr06	marsode-core	marsode-mvr01	marsode-mvr02
marsrd-core	marsrd-mvr01	marsrd-mvr02	marsrd-mvr03	marsrd-mvr04	marsrd-mvr05	marssc-core	marsth-core

Version: 4.6.3

This software is supported and certified by **Opsview Limited**
© 2015 **Opsview Limited**. All Rights Reserved



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

Version: 4.6.3

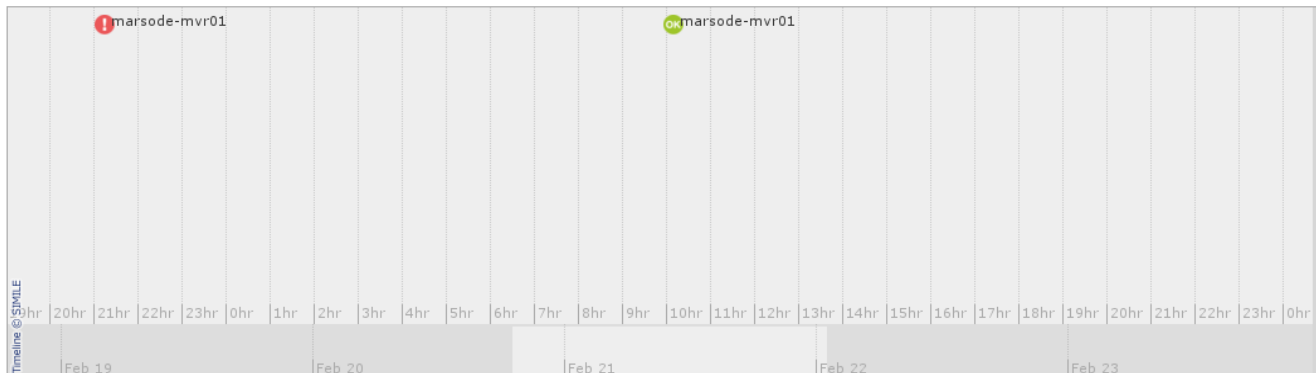
This software is supported and certified by **Opsview Limited**
 © 2015 **Opsview Limited**. All Rights Reserved



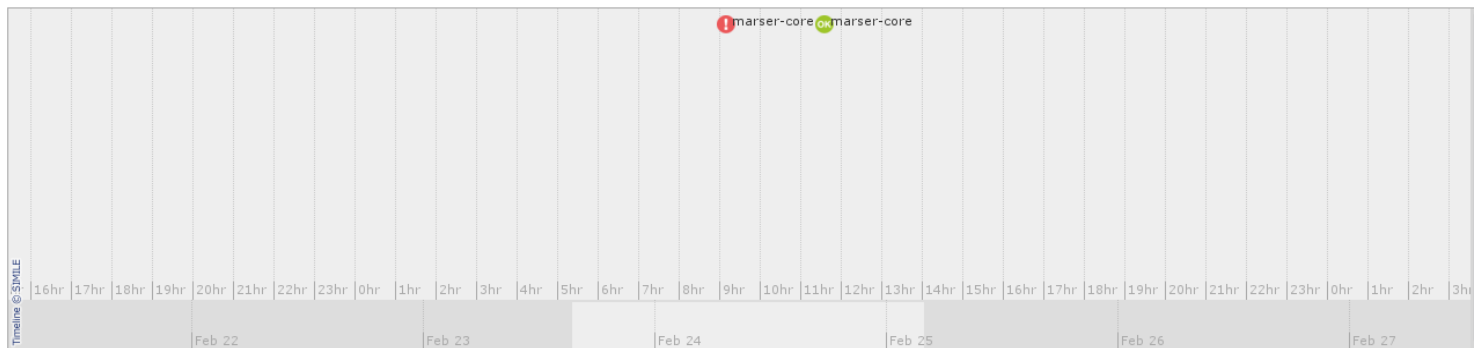
Events



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

 Export Data:


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 (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 ?