

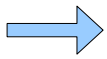
Job efficiency (monitoring)

Status report

James Casey, Daniel Filipe Rocha Da Cunha Rodrigues, Ulrich Schwickerath

- observed **bad job efficiencies** at CERN batch farms
- mismatch between efficiencies seen by us (global view) and experiments

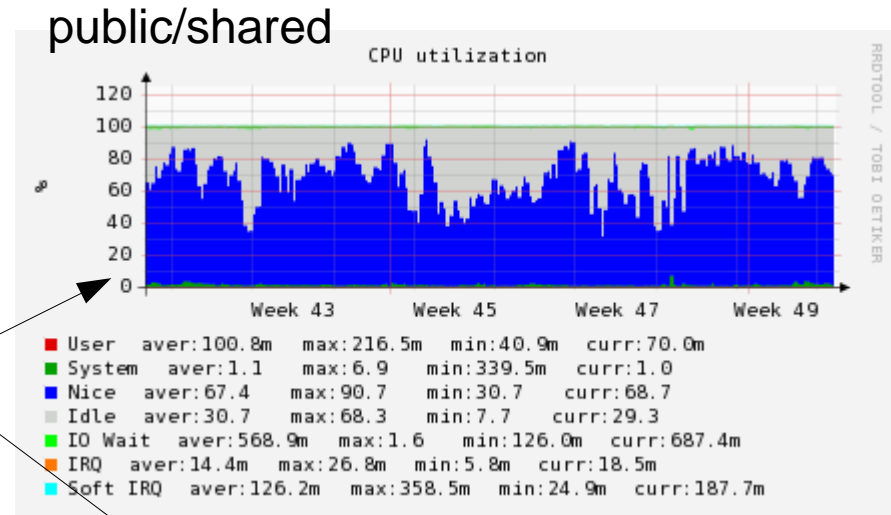
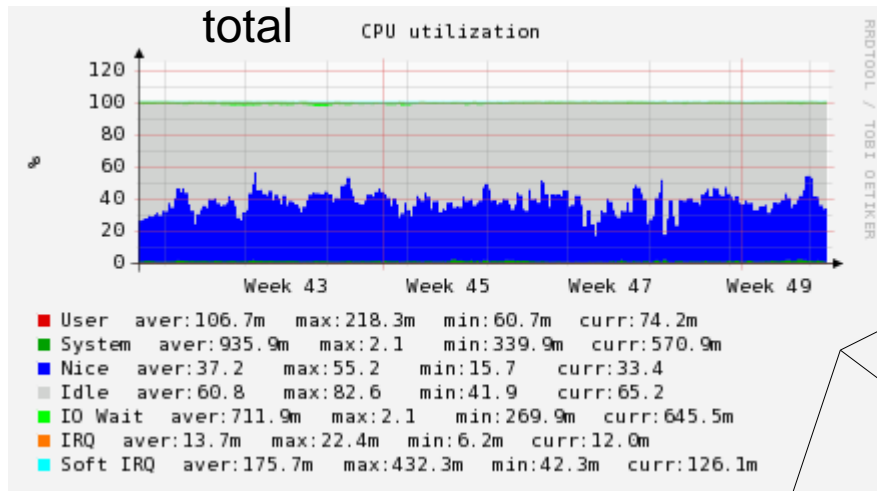
See march GDB presentation



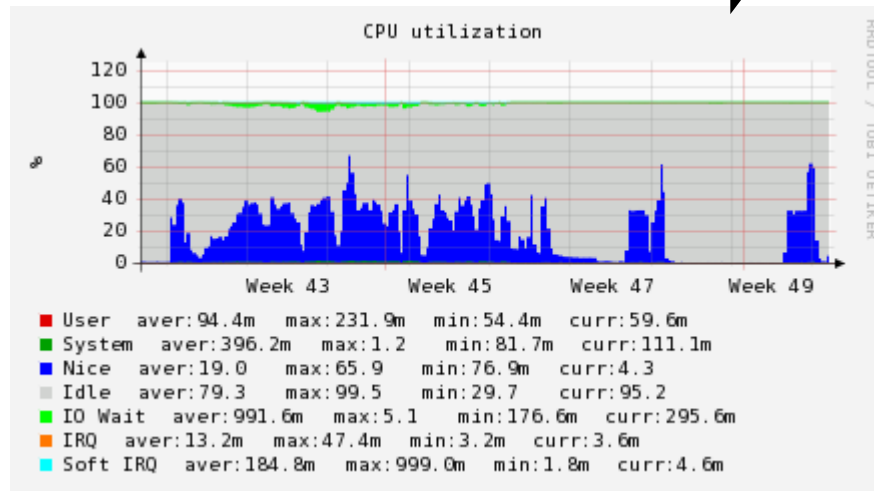
Improved monitoring of jobs based on **MSG project**

- allows monitoring of job efficiencies at various execution steps
- very flexible
- back end easy to use
- in place at CERN since ~6 months
- first tests done by Dirac 3 developers
- prototype presented already in August
- usable since then already

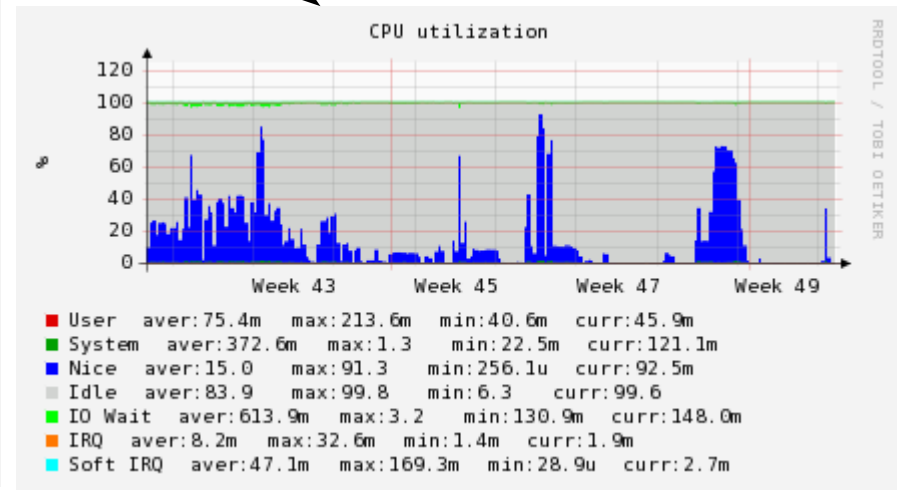
Current efficiencies snapshot and issues last two months



dedicated CMS (cmsproc)



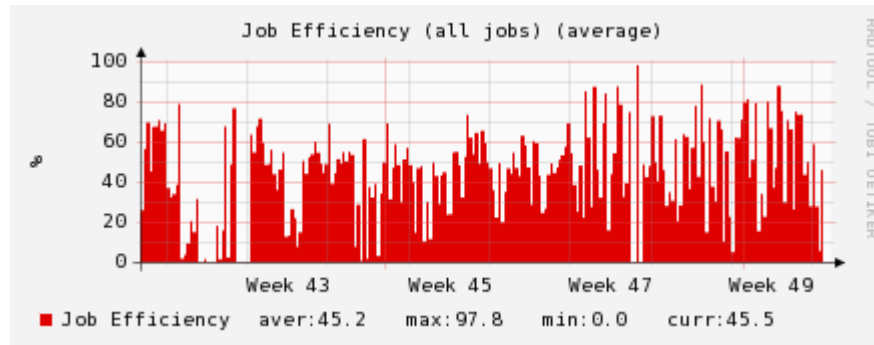
dedicated ATLAS (itdc)



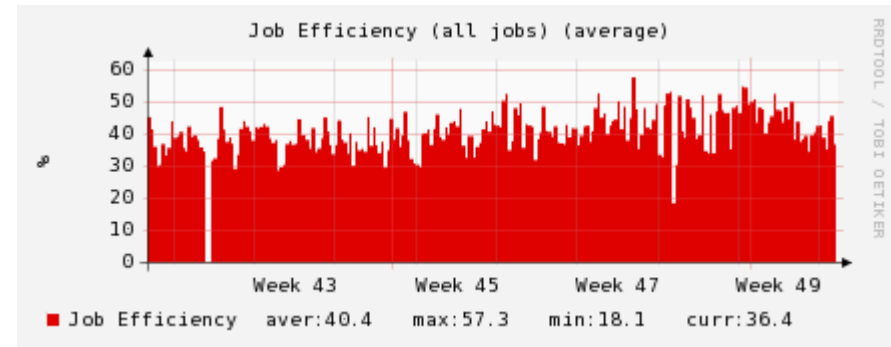
-> under discussion with the individual experiments

Individual job efficiencies by VO

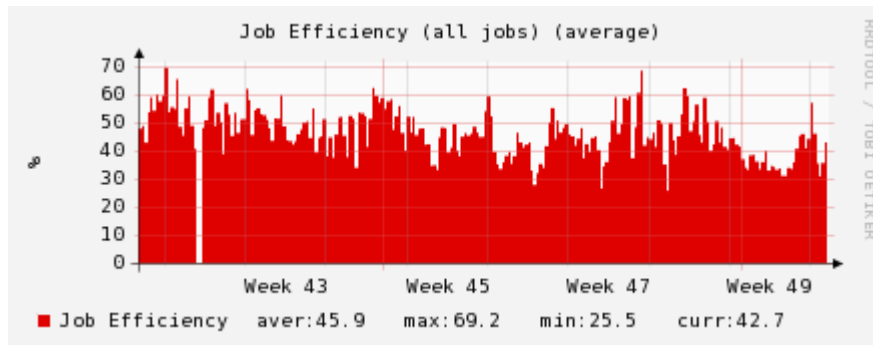
Alice



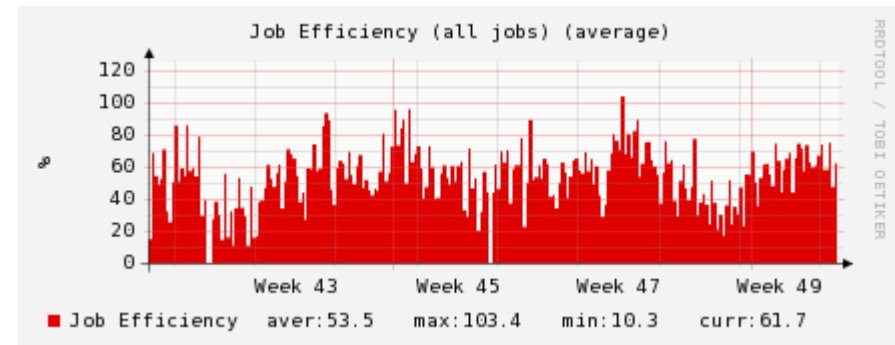
Atlas



CMS



LHCb



both grid and local jobs
last two months shown

What is currently recorded by IT/FIO (CERN) ?

- Job start record (just before user job is started)
- Job end record (just after the user job terminated)
- LSF accounting information (what LSF sees, bulk upload once per day)

New backend script : **msg-tag**

- added on request of LHCb
- encapsulates automatically retrievable data

```
> msg-tag --help
```

```
Usage: /usr/bin/msg-tag --context=<context> --cputime=<cputime in seconds>
```

where:

context is a tag (as a string), indicating the context the job is in right now
cputime : a number indicating the CPU time used so far, including all sub processes (real time)

optional arguments:

```
--walltime=<walltime> : a number indicating the total time the job has been running so far (real time)  
--vojobid=<Vo job ID> : VO internal identifier  
--vosite=<Vo site name> : site name in VO language  
--memoryusage=<Memory usage in MB>: max memory usage so far  
--config=</path/to/config>: configuration file location  
--debug : print diagnostics, don't upload the records  
--state : current job state  
--verbose : be verbose
```

<https://twiki.cern.ch/twiki/bin/view/FIOgroup/FsLSFJobEfficiencyMon>

- first early adaptor (thanks a lot!): Dirac 3
- unfortunately, nobody else yet

Accessing the data

```
> sqlplus LCG_SAM_MESSAGING_R@int11r
```

```
SQL*Plus: Release 10.2.0.3.0 - Production on Mon Dec 8 17:21:48 2008
```

```
Copyright (c) 1982, 2006, Oracle. All Rights Reserved.
```

```
Enter password:*****
```

```
SQL> describe LCG_SAM_MESSAGING.usagejobdetails;
```

Name	Null?	Type
-----	-----	-----
OLD_LOCALJOBID	NOT NULL	NUMBER <- this is obsolete now
OWNERDN	NOT NULL	VARCHAR2(1000)
WNHOSTNAME		VARCHAR2(1000)
STATE		VARCHAR2(20)
CONTEXT		VARCHAR2(20)
CPUUSAGE		NUMBER
MEMORYUSAGE		NUMBER
CPUFACTOR		NUMBER
EXITCODE		NUMBER
WALLTIME		NUMBER
SUBMITTIME		VARCHAR2(100)
FINISHTIME		VARCHAR2(100)
INSERTTIME		DATE
VOJOBID		VARCHAR2(20)
VOSITE		VARCHAR2(1000)
SITENAME		VARCHAR2(1000)
RECORDTIMESTAMP		VARCHAR2(100)
GRIDJOBID		NUMBER
BATCHUSER		VARCHAR2(100)
LOCALJOBID	NOT NULL	VARCHAR2(38)

state: jobstart,jobend etc

context: jobwrapper (used by fio), ...

www.cern.ch/it-fio

- **dealing with Job arrays: (on solution)**
 - ◆ same job ID but different index
 - ◆ required a schema change on the master to distinguish them
 - ◆ impact: dirac3 records were initially rejected due to this
- **upload of historical LSF data (on solution)**
 - ◆ scalability issues (limit ~100k records at a time)
 - ◆ not critical because not really needed (nightly uploads)
- **Access to the data too complicated (missing)**
 - ◆ currently no default display available
 - ◆ first step: direct database access (read only)

Job efficiencies:

- **abominable** with a slight slope downwards
- Box usage is reasonable on shared resources but
- CERN has big issues with idle T0 dedicated resource

Job instrumentation with MSG:

- in place
- can be used by the experiments
- only one early adopter (LHCb Dirac3)
- several issues identified, most of them solved or under solution
- it is designed to be portable. Usable also at other sites
- **please USE it**