



CMS Distributed Production Prototype

Michael Thomas
March 30, 2006



Distributed Production



(Centralized) Production Manager maintains all work to be done

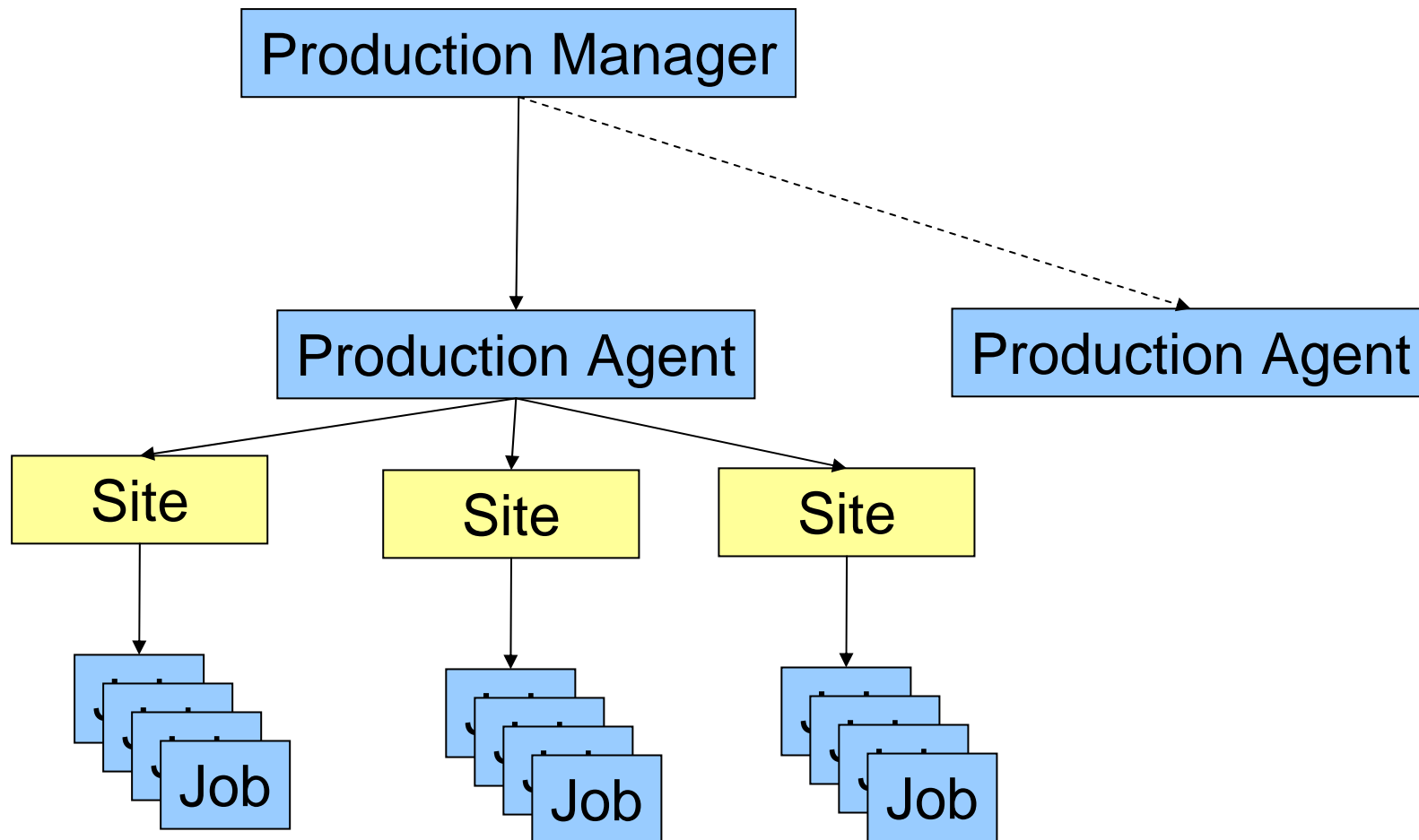
Remote Production Agents pull work from the Production Manager, run and track jobs

Jobs are monitored and resubmitted in case of failure

System can be used for some types of Analysis (non-interactive)



Production System





Production System



Production/Workflow Manager

Workflow Spec

Production Agent

Job Spec

Job

Application

*

*

*



Production System



Workflow Spec

Request/Assignment level template

Job Spec

Has Unique ID

Job

Corresponds to CTDR Task

Application

Used by the (central) Production Manager process



Production System II



Workflow Spec

Processing Job Definition

Job Spec

Has a unique ID within a workflow/request

Job

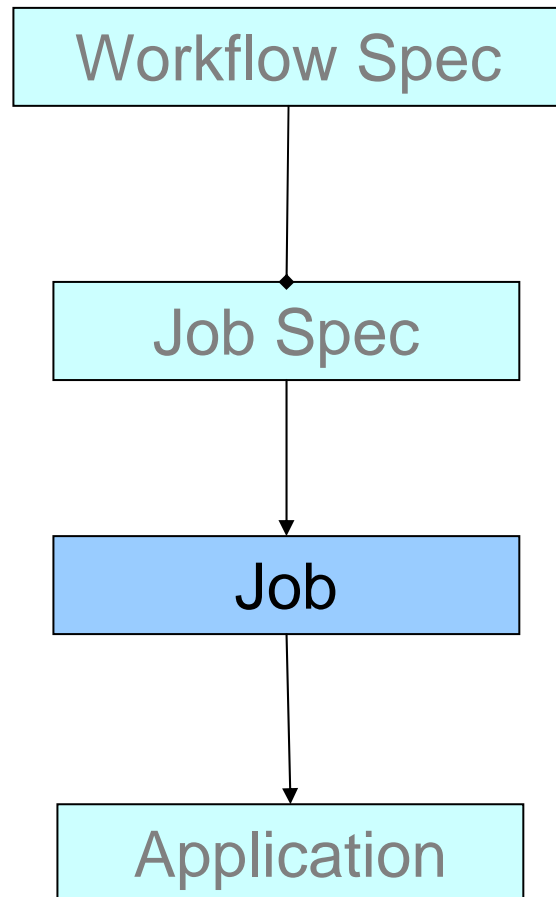
Used to create physical jobs

Application

Dealt with by the Production Agent



Production System III



Physical job created from Job Spec

Combination of JobSpecID and batch/grid job id are unique

Corresponds to a CTDR job

Multiple jobs created from the same JobSpec will be the same job in terms of physics (retries, etc.)



Production System III



Workflow Spec

**Application like nodes within a job,
managed by SHREEK**

Job Spec

**CMSSW, StageOut, Rescue types of
tasks**

Job

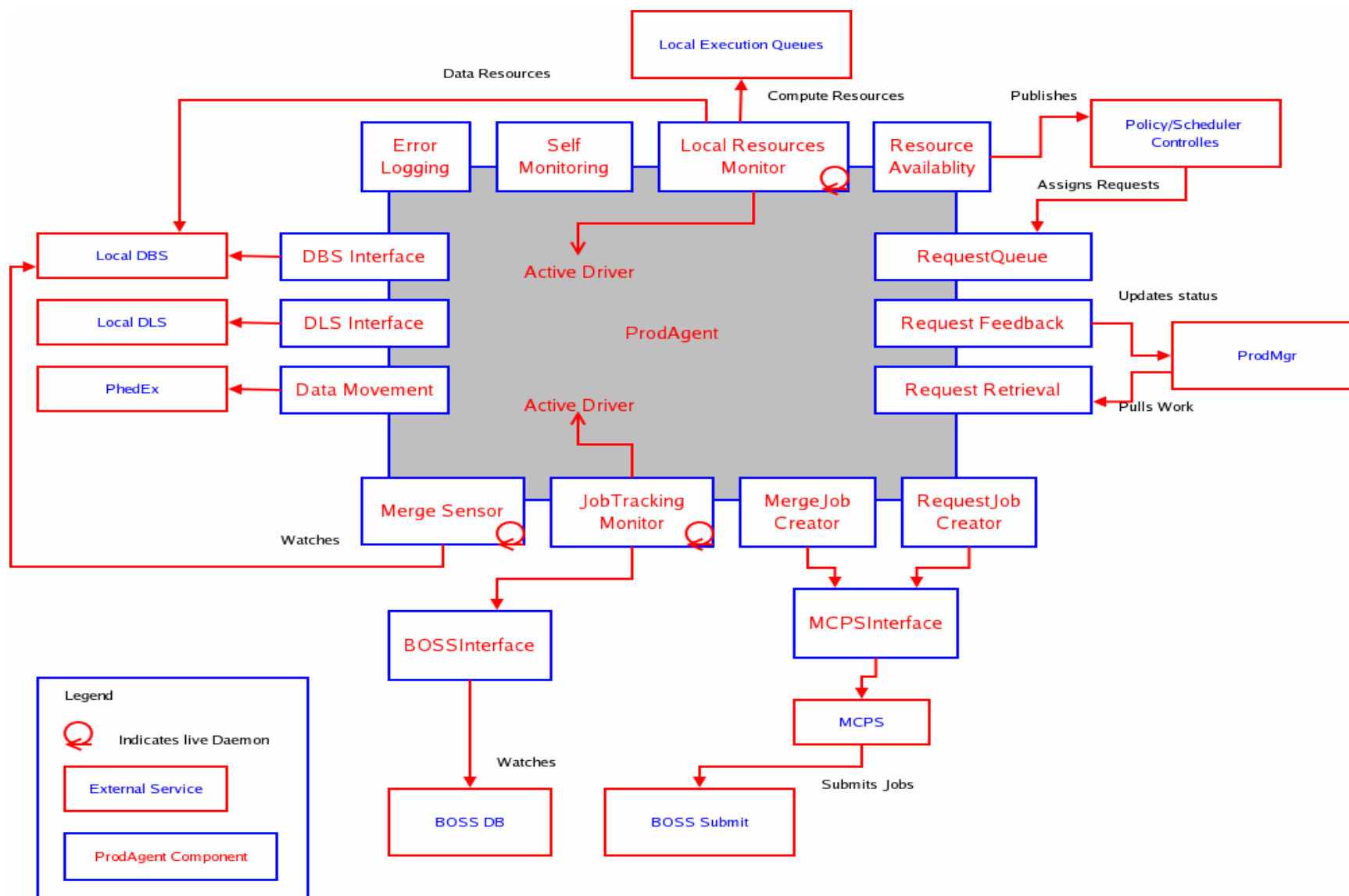
**May be multiple CMSSW apps run in a
single job**

Application

**Each app has a unique node name
within the workflow**



Production Agent





Monitoring



Prod. Manager receives notifications from Prod. Agent

- ★ **Allocated**
- ★ **Final states: success, failure**

Prod. Agent monitors the high level status of jobs via Message Service and reports to CMS Dashboard

- ★ **Created**
- ★ **Running**
- ★ **Finished**
- ★ **Fail (requeue)**

Job Tracker monitors the job status via BOSS

- ★ **Queued**
- ★ **% completed**
- ★ **Output text**

Job publishes data

- ★ **BOSS wrapper captures stdout**
- ★ **Job may use ApMon directly to publish to MonALISA**



Data Analysis



CRAB – CMS Remote Analysis Builder

Provides CMS users with

- ★ Framework to run analysis on datasets hosted by T1/T2 centers
 - ★ No detailed knowledge about Grid necessary
- ★ Uses Grid infrastructure
 - ★ Authentication by Grid certificates
 - ★ Job interaction using Grid middleware



CRAB



CRAB splits User interaction into steps

- ★ Job Creation
- ★ Job Submission
- ★ Job Status check
- ★ Job output retrieval

CRAB takes care of User code

- ★ Pack up user executables and binaries
- ★ Ship user code to worker nodes for execution
- ★ Preparation of software environment on worker node and execution



Clarens UI



Clarens Java Webstart Client

File Certificate Connection Manager View Manager

Main Ahsan

clarens

MCPS.util
MCPS.workflow
MCPS_MON
MCPS_MON.generated
https://hepgrid2.dynalias.net
https://hepgrid2.dynalias.net
http://hepgrid2.dynalias.net:8
http://hepgrid2.dynalias.net:8
Submit
agent
astro
http://tg-tst-h.caltech.teragrid
http://tg-tst-h.caltech.teragrid
http://tg-tst-h.caltech.teragrid
https://tg-tst-h.caltech.teragri
https://tg-tst-h.caltech.teragri
https://tg-tst-h.caltech.teragri
astro2
boss
boss_back
http://DHCP-112-197.caltech

File Servers

http://fntst-1.fnal.gov:8080
web/
.clarens_access
.clarens_accessc
index.html
http://t2sam01.sdsc.edu:8
web/
test/
system/
submissions/
stylesheets/
registry/
proxy/
javascript/
images/
group/
file/
empty.html
empty.html.2005-0
intro.html
intro.html.2005-03
intro.html~

Name : /web/intro.html Type : 1
Size : 1374 Modified : Tue Jan 13 13:44:30 PST

Clarens Web Interface

How to use this page

Click on one of the service links in the navigation panel on the left. If the panel is empty, your browser may not be supported.

To return to this page, click on the words Service Index in the navigation panel.

File browsing

- Any directory can be browsed by appending `auto_index.html` to the URL
- If an `index.html` file exists in a directory, that will be loaded when a directory URL is specified
- If no `index.html` file exists, a directory listing will be displayed instead
- Try this out now.

Output >>
Sat Sep 10 16:59:30 PDT 2005 : Connected to http://discover.gridservice.info/xmlrpc
Sat Sep 10 16:59:30 PDT 2005 : Connection Active : http://discover.gridservice.info/xmlrpc
Sat Sep 10 16:59:50 PDT 2005 : Connection Active : http://discover.gridservice.info/xmlrpc
Sat Sep 10 17:00:10 PDT 2005 : Connection Active : http://discover.gridservice.info/xmlrpc
Sat Sep 10 17:00:30 PDT 2005 : Connection Active : http://discover.gridservice.info/xmlrpc

Messages Command Window



More Information



CMS Production System

- ★ <http://lynx.fnal.gov/runjob/CMSProductionSystem>

CRAB

- ★ <http://cmsdoc.cern.ch/cms/ccs/wm/www/Crab/>