



CSA07 Status and Plans

Ian Fisk
September 24, 2007



Computing Software and Analysis Challenge

Computing Software and Analysis Challenges

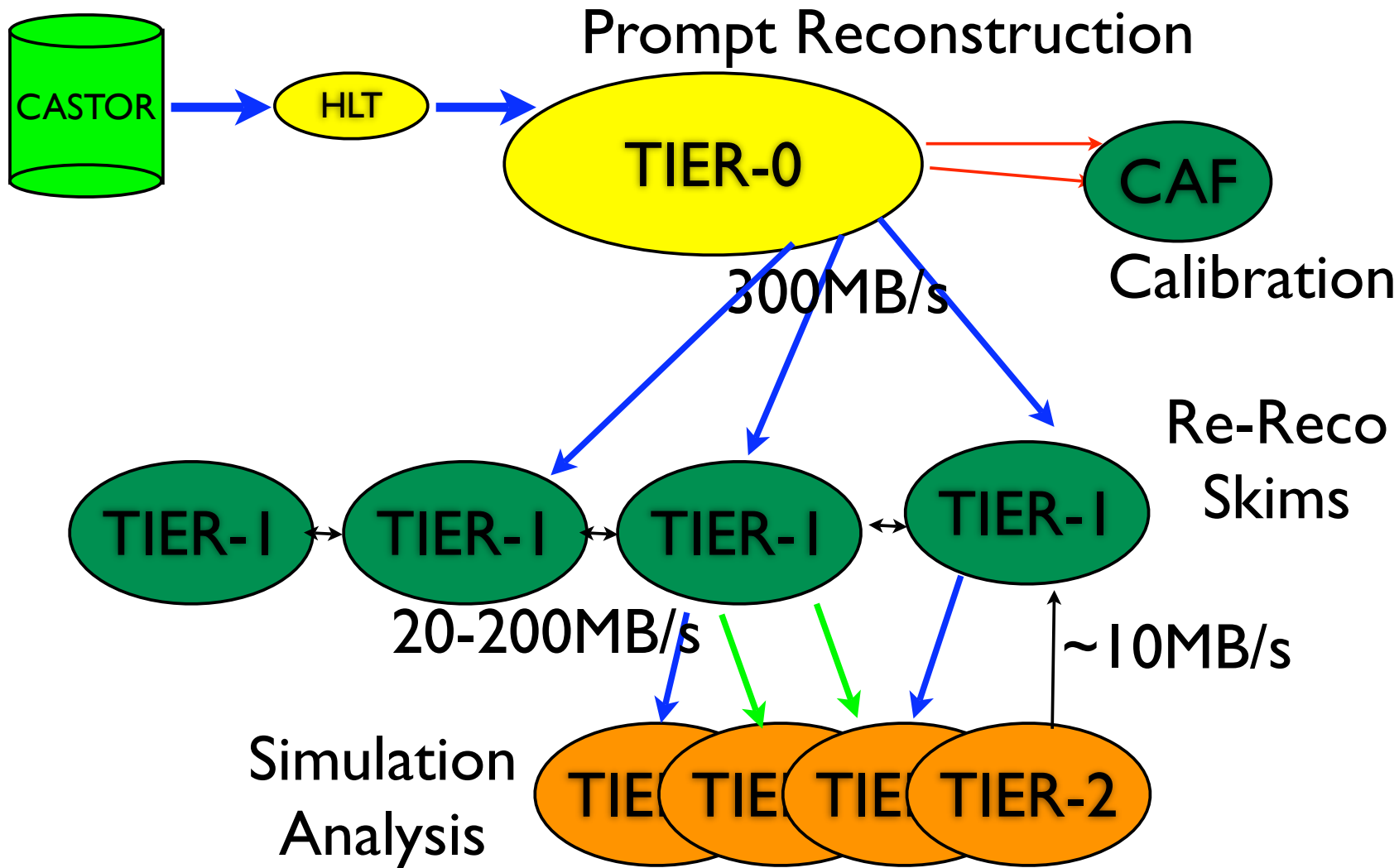
- ➔ Are opportunities to exercise aspects of the computing model and the software development program with analysis activities and users
 - Dedicated tests of components do not show interference problems
- ➔ The Challenge planned for 2007 should exceed 50% of final scale

We attempt to align the challenge activities with other activities that involve the physics user communities

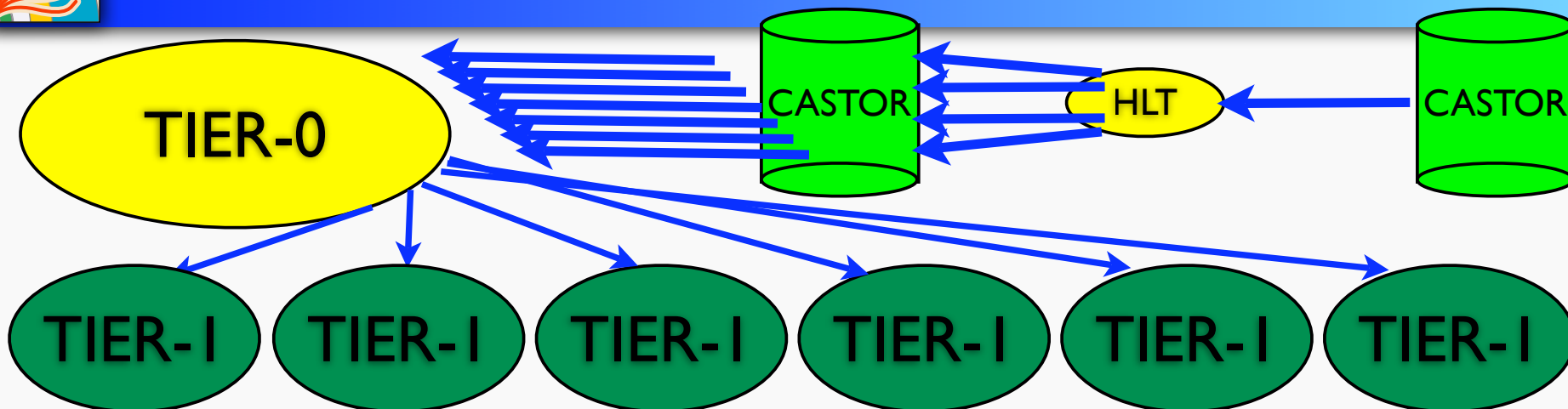
- ➔ The desire is to try to stress the system with real users
- ➔ We have a number of automated tools to generate load and stress the system, but the users make a much more realistic test
- ➔ We specifically made the computing elements of the challenge overlap with preparation of a physics analysis note



CSA07 Workflows



Tier-0 Workflow

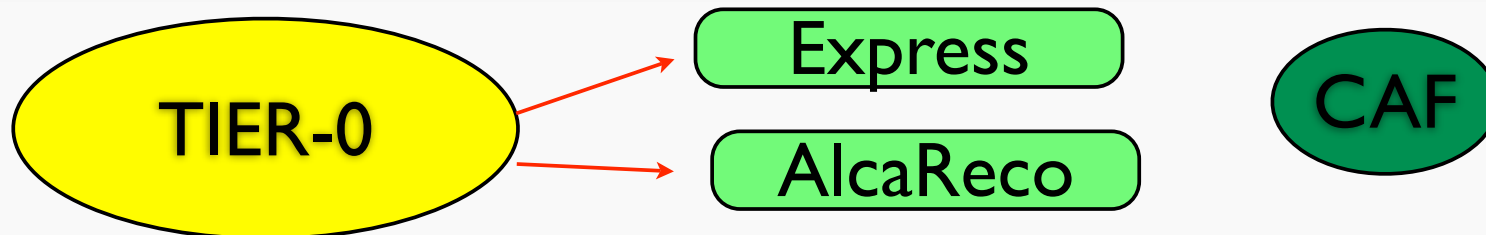


The Tier-0 workflow is similar to what was done for CSA06 with one large addition

- ➔ We have a pre-processing step to add the HLT information
- ➔ In CSA06 we had a minbias stream, a TTbar stream, a Jets stream based on simulation information
- Now data from the Tier-0 looks more like experiment data based on the trigger selection
- There are only 6 streams plus one for all events
 - Computing model anticipated more primary datasets and no stream of all the data
 - Still work to do



Calibration and CAF Analysis

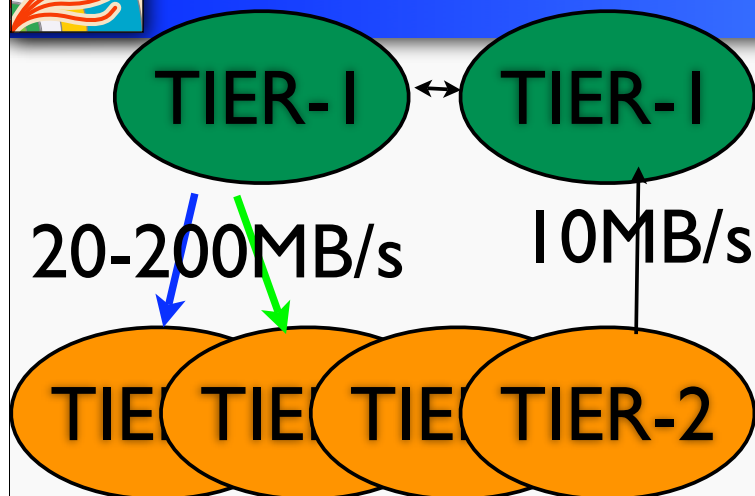


The resources identified for the CERN analysis facility in CMS this year are very modest

- ➔ ~160 CPUS
 - 75% will be used for low latency calibration and alignment work
 - 25% will be used for 3 express stream analyses
- ➔ Express stream is identified as up to 10% of the data in the CMS computing model
 - Proposal is to imbed a standard signal into the combined background
 - This is not a mock data challenge. Physics groups will know what it is.
 - MC information is available in the events



Skims and Transfers



Physics groups have defined ~50 skims to operate on the primary datasets

- ➔ In the computing model these skims are executed at the Tier-1 centers on data hosted

Since primary datasets are assigned to Tier-1 centers in the computing model there is a mesh of Tier-1 to Tier-2 transfers

- ➔ There are more than 200 Tier-1 to Tier-2 permutations

Physics groups will be looking at a mix of previously generated simulation samples

- ➔ Monitoring skimming data produced during the challenge



Transfer Commissioning

CMS has been attempting to commission the transfer system since the beginning of the year

- ➔ A series of LoadTests were developed. Good progress on infrastructure and scale testing of tools.

Earlier in the summer CMS established a task force on debugging data transfers. Establish and track technical metrics for links to be declared production ready

- ➔ The bar for production ready is very low. It is about demonstrating the functionality of all the components, and not about stressing the network
 - Essentially 3MB/s on average for 4 of 5 days.

75% of Tier-2s have at least one download link from a Tier-1. Not yet 50% have two

Commissioned connections between Tier-1 and Tier-1 centers are still missing for many permutations



Analysis Tools

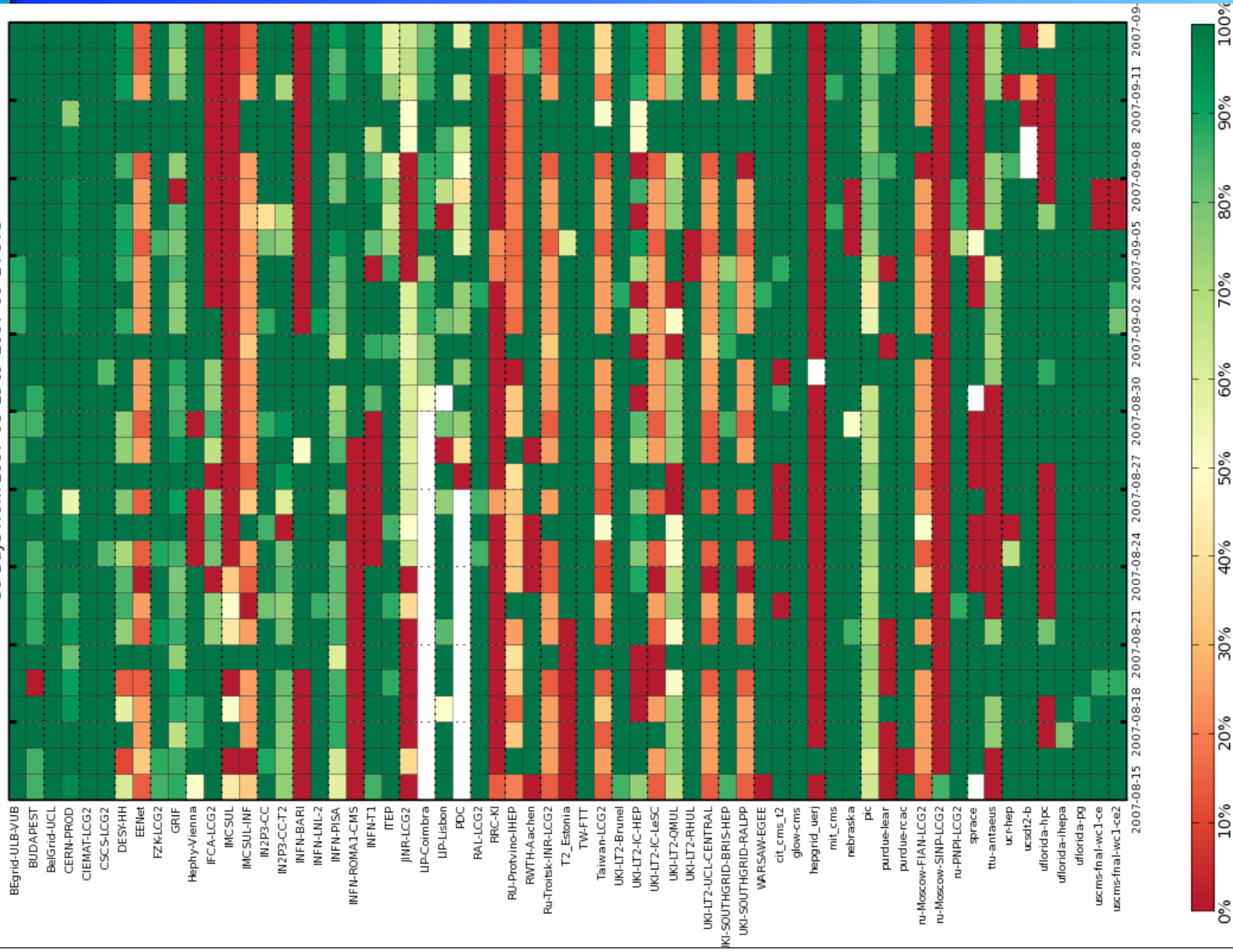
Data is accessible remotely using grid tools and CMS has a reasonably mature product called CRAB

- ➔ The CMS Remote Analysis Builder
 - Submits to both the EGEE and OSG sites
- ➔ Given a CMS application environment and a dataset name, CRAB will configure package the differences between the user code and the distributed application, interact with the data management system to determine the logical files required and the location of the data, and submit and monitor the jobs
- ➔ More than 300 users have submitted a CRAB job in CMS
 - User feedback is generally positive



CMS SAM Tests

SAM Site Quality
30 Days from 2007-08-15 to 2007-09-14 UTC

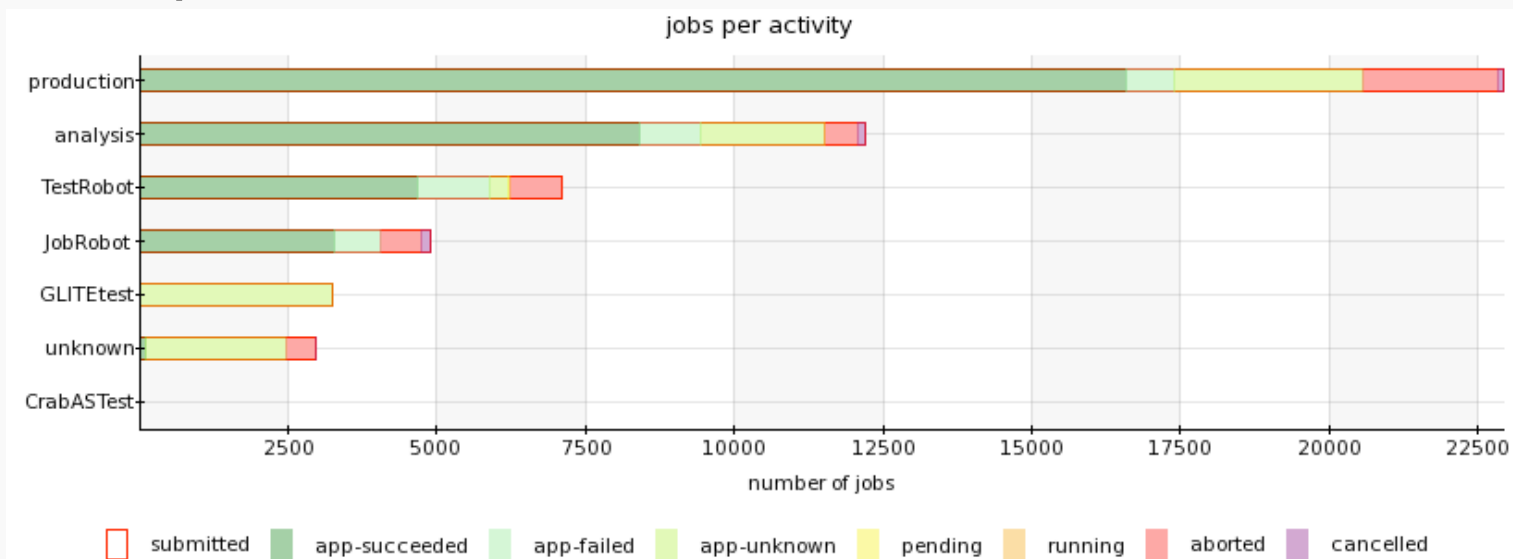




Analysis Activity

The computing model estimates 100k-200k submissions per day in the full scale system

- ➔ Roughly 25% are organized production processing and 75% analysis processing
- ➔ CSA07 goal is 100k jobs per day
 - In CSA06 we hit 50k jobs per day, but with extensive use of the job robot
- ➔ Activity the week before CMS Week





Outlook

Sample preparation for the challenge has been going on for months

- ➔ Splitting of data into primary datasets should begin today
- ➔ Reconstruction of samples will hopefully begin tomorrow

From a resource standpoint we are in reasonable shape

- ➔ Still very large variations in the sizes of the sites

CSA07 is an opportunity to push analysis into the Tier-2 centers

- ➔ Tier-2 centers are the primary resource for analysis in the computing model

Site commissioning has gone well

- ➔ A few persistent problems to solve

Transfer commissioning is still a work in progress

- ➔ The situation is improving, but there is a lot of work to do