Deployment and Operations of the CMS Prompt Skimming System

Si Xie

Massachusetts Institute of Technology

July 22, 2010

Prompt Skimming

- CMS categorizes data recorded by the detector in primary datasets according to trigger bit selections immediately after data is recorded (prompt)
- Primary datasets splits data by physics interest and follows the distributed computing infrastructure model of CMS, while introducing minimal overlap
- CMS further skims the primary datasets promptly to facilitate common analysis selections
  - Reduces sample size and amount of data volume that needs to be analyzed
  - Reduces analysis latency
- Currently implemented for:
  - Subdetector performance analyses (ECAL, Muon systems)
  - Jet energy scale correction measurements
  - Quarkonia skim for momentum scale and tracking studies

CMS distributed Computing Infrastructure

- Multi-tiered computing infrastructure
  - Tier-0: first processing of data recorded with the CMS detector at CERN
  - Tier-1: 7 large computing centers located around the globe, each containing on average several PetaBytes of tape storage and a few thousand CPU cores for data processing

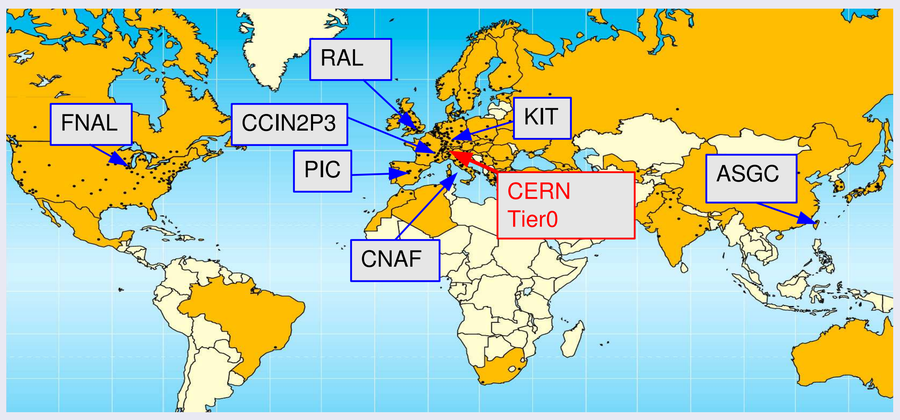

Figure: Global map showing the location of all CMS Tier1 Centers.

  - Tier-2: a large number of smaller sites located all over the world mainly dedicated to data analysis.
- Data is recorded with the detector, processed and promptly reconstructed by the Tier0, transferred and stored at the Tier-1 centers, and finally distributed to the Tier2 sites for user analysis.
- Prompt Skimming runs on data arriving from CERN at the Tier-1 sites

Workflow Design

- The skimming workflow is pre-defined by input datasets and output configurations
- The status of new prompt reconstruction data at a Tier-1 is monitored through grid interfaces to the central book-keeping database
- Arrival of new data triggers the creation and submission of new prompt skimming jobs
- Ensures prompt delivery of skimmed data as soon as new input data is available

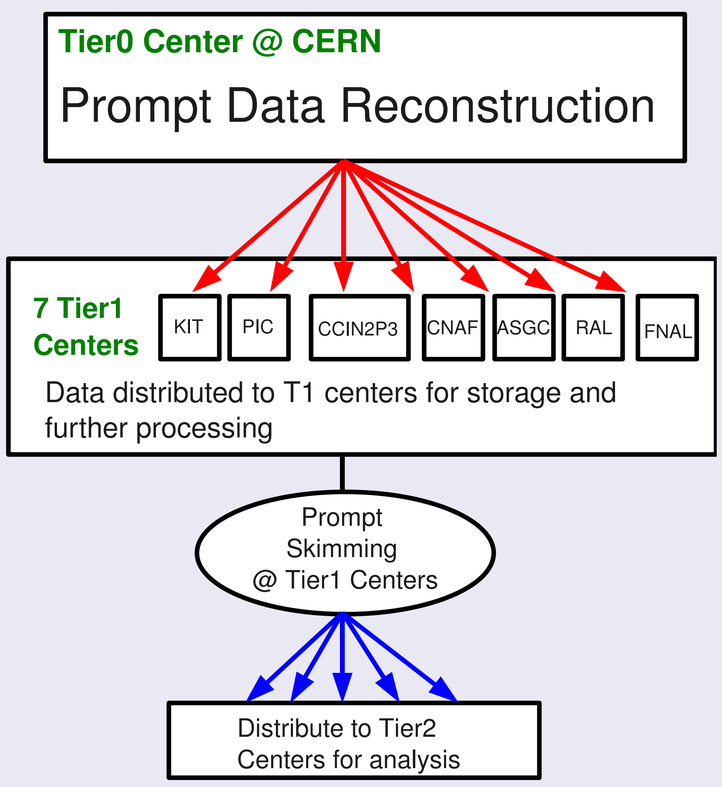

Figure: Overview of the prompt skimming workflow.

Operations

- The status of individual prompt skimming jobs are tracked via a combination of components from CMS' old workfload management system (ProdAgent, message queue based) and new system (WMAgent, state machine)
- Job submission is performed using the glideIn WMS Pilot Grid Submission infrastructure
- The workflow management is executed remotely at FNAL
- A minimal access to the Tier0 database at CERN is needed for synchronization

Operational Performance

- Latency goal for prompt skimming is 24 hours after prompt reconstruction workflow is completed
- Average latency performance in 2010 data taking

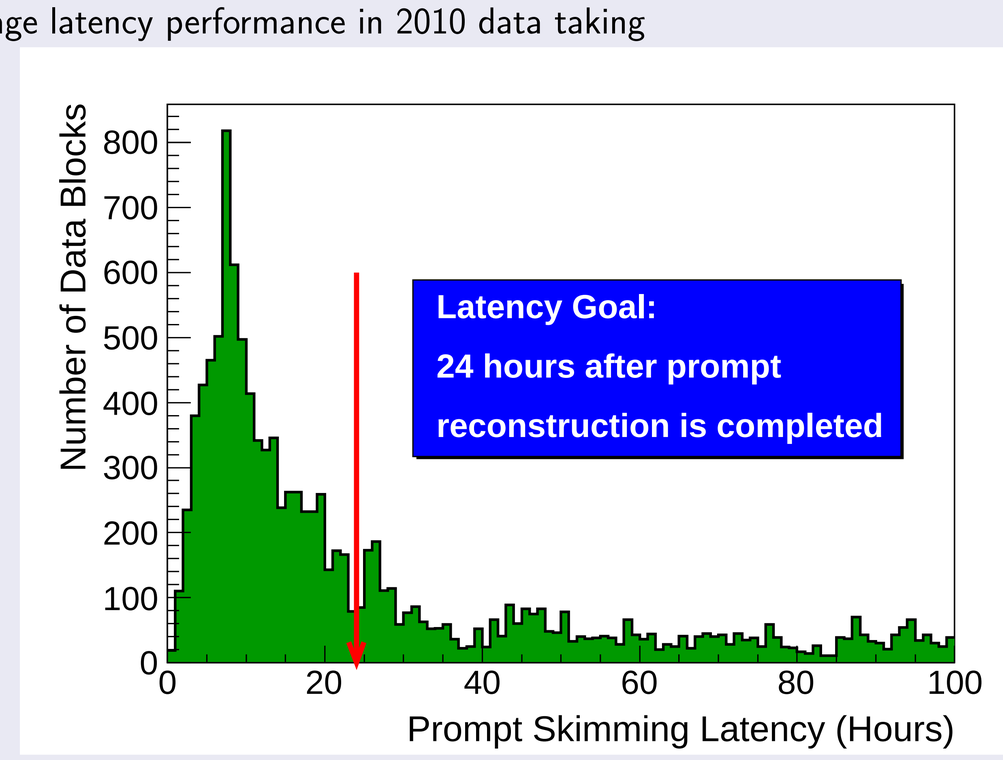

Figure: Histogram of the prompt skimming latency per file block. The long tails are mainly attributed to issues in data transfers and tape migration.

- Volume of data produced in 2010 data taking

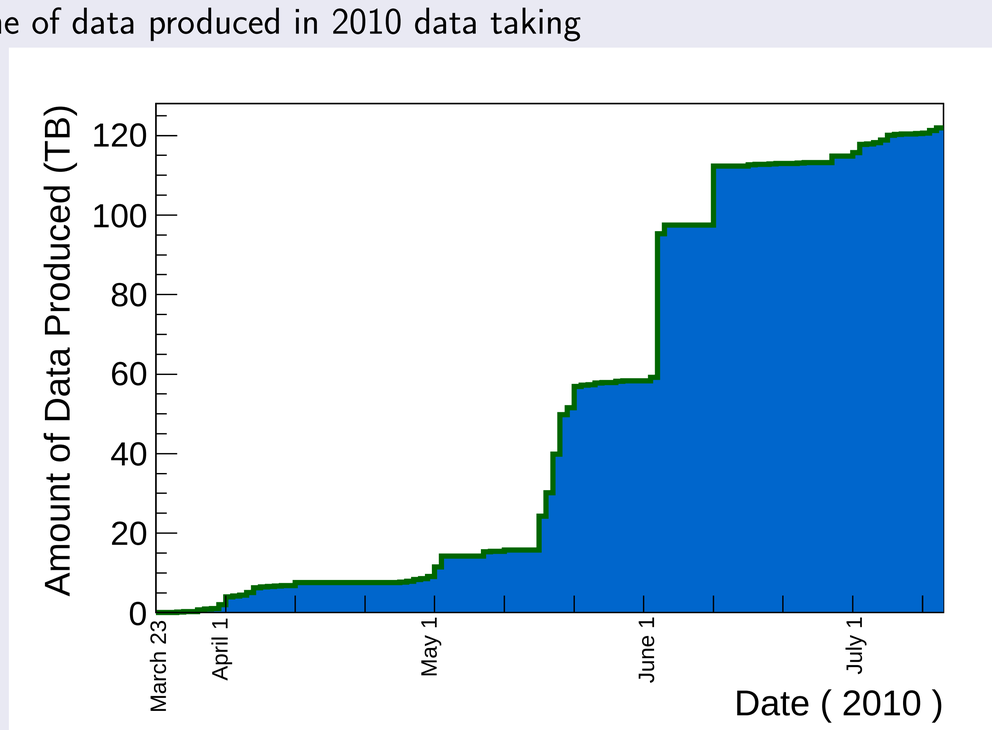

Figure: Volume of promptly skimmed data produced as a function of time.

Operational Issues and Lessons

- Workflow output too large
  - Skim efficiency was too high for early data workflows
  - Caused an overload of the IO and tape writing capacity at the Tier-1 sites
  - Partially resolved by a redefinition of primary datasets and a redefinition of skim workflows
- Workflow creation trigger
  - Prompt skim jobs only created when files are on tape at the Tier1 and at the Tier0.
  - An unnecessarily large latency waiting for tape migration
  - A possible need to track information at the Tier0 in a more fine grained way.

Future Design and Operation Plans

- Plan to move to a fully operational state machine of file and job tracking - "WMAgent"
- A much more reliable and user friendly workflow management system
- The prompt skimming workflows will be one of the first implementations of the WMAgent in production