



# Validation of Software Releases for CMS

Oliver Gutsche - Fermilab

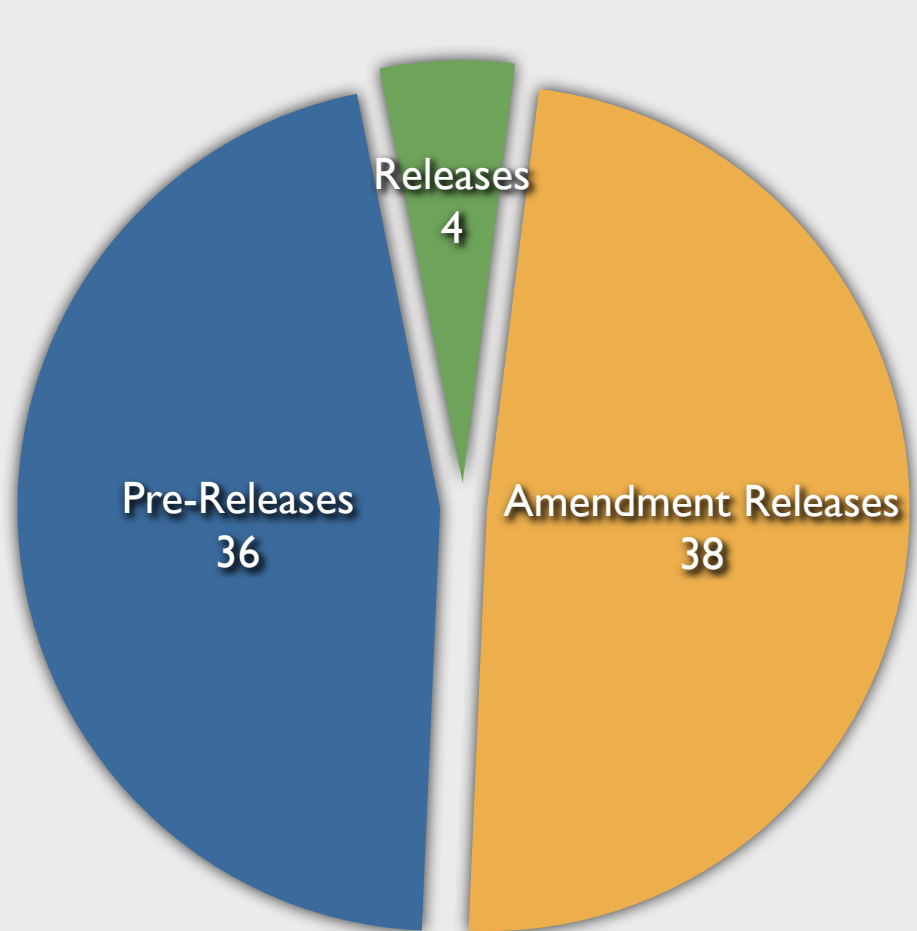
On Behalf of the CMS Offline and Computing Projects



## CMS Software

- Consists of more than **2 Million lines of code**
- Over **250 active developers** are maintaining and improving the code base
- Code is consolidated into releases**
- Feature sets are grouped into release cycles
- Rapid development per cycle: new test release approximately every week
- In 2008, CMS released 78 different releases** of the Software Stack in 4 different release cycles:

Software Releases in 2008



There are 3 types of releases:

- Pre-Release:**
  - consolidates the current state of the code
  - Purpose: test interdependencies between different software components developed in parallel
  - Not used for official production
- Release:**
  - Closes the development cycle
  - Final feature set, new features are not allowed to be added anymore
  - Used for production of MC samples and central re-processing's as well as analysis
  - Distributed and installed world-wide on all official levels of the CMS tiered computing infrastructure
- Amendment Release:**
  - Bug fix release for specific problems found in a release
  - Supersedes previous releases

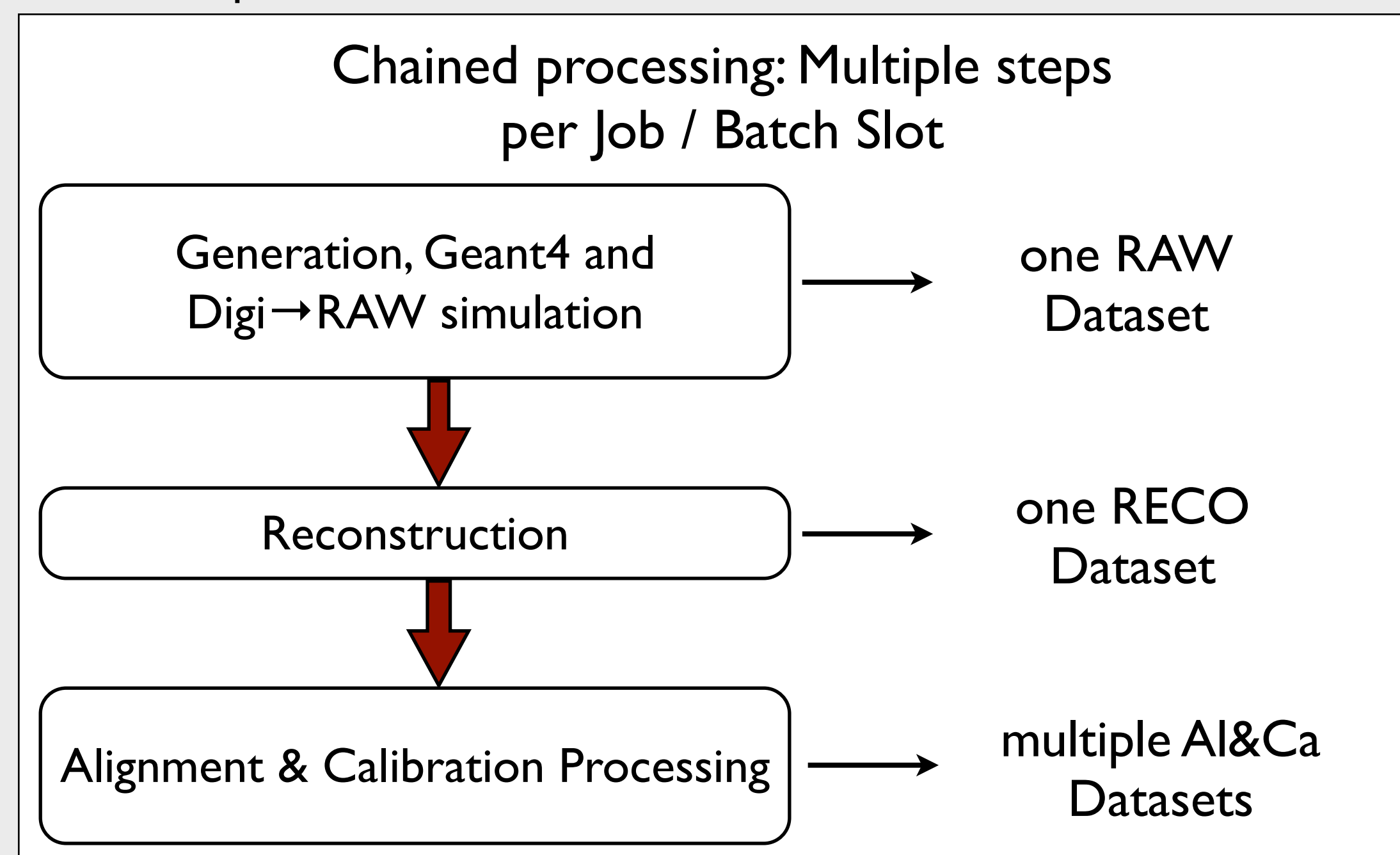
## CMS Release Validation

- Purpose:**
  - Guarantee** that all components of a release **work together without interference or failures** during execution
  - Check** that release **can be used by the global production infrastructure** of CMS
  - Validate the correctness of produced physics output**
  - Validate the performance** of a release in terms of
    - Algorithmic performance
    - Stability at larger scales (e.g. Number of events)
    - Memory and time consumption
- Organization:**
  - Produce Monte Carlo (MC) simulation samples and reconstruction passes of detector data** once per release
    - Provide samples to software experts **time near** for validation and bug fixes
    - Use **dedicated resources** where software can be installed instantaneous after release announcement:
      - CERN:** 500 priority batch slots co-shared with T0
      - Fermilab:** opportunistic cycles of 5000 batch slots of T1 center in parallel to central production
  - Group samples in **Validation Sets**:
    - "Standard Set": produced within 24 hours** at CERN to enable rapid feedback before the next release (~1 week)
    - "High Statistics Set": produced within 1 week** at Fermilab
  - Balance statistics and number of samples to stay within resource limits (detector data samples have been added only recently):

	Full Simulation			Fast Simulation		
Generation	Particle Gun	Physics Process	# Events / Sample	Particle Gun	Physics Process	# Events / Sample
Standard Set	8	24	9k	6	1	27k
High Statistics Set	12	19	25k	0	8	100k

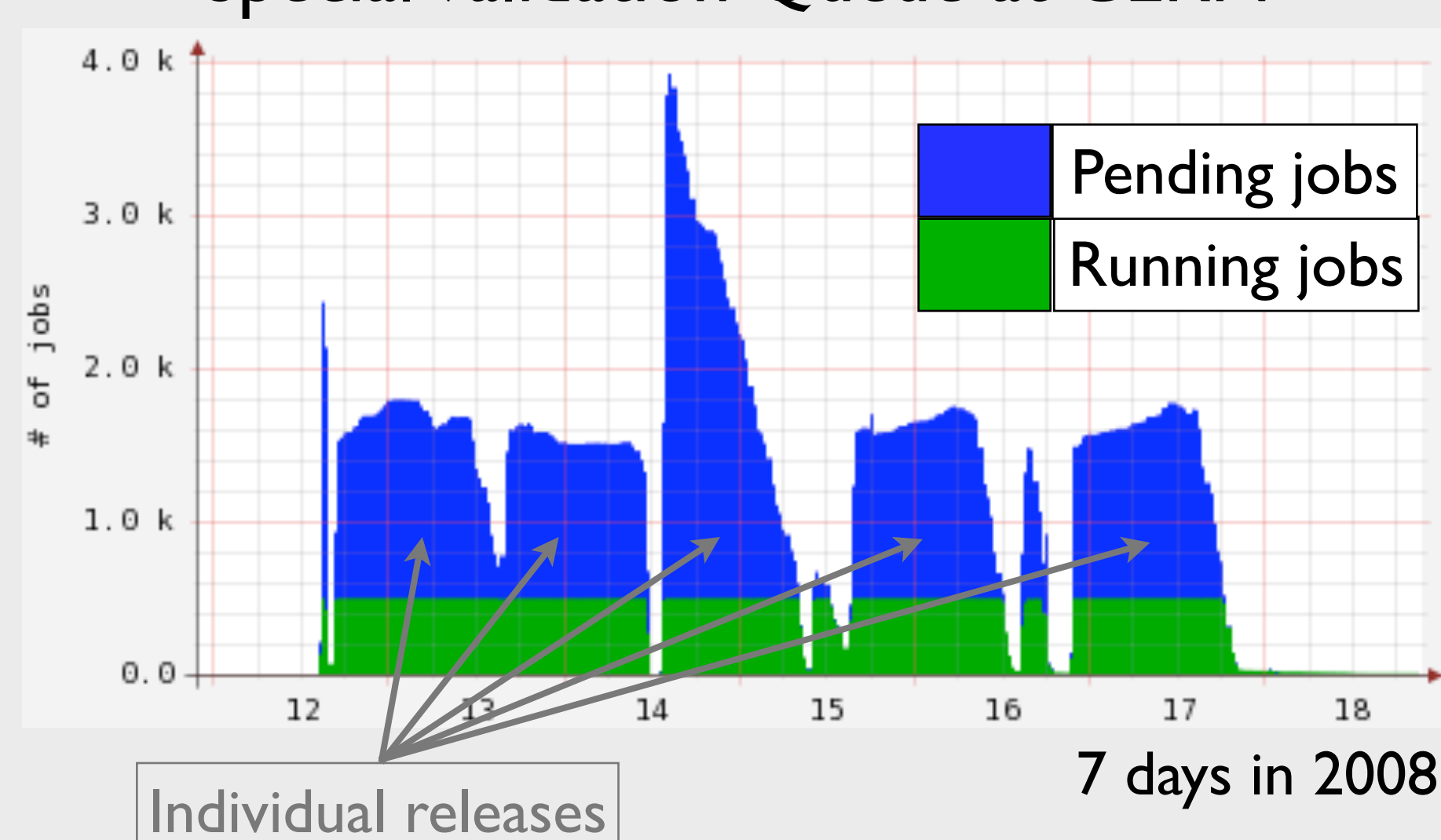
## Validation Set Production

- Production Infrastructure had to be optimized** to produce all requested samples within resource constraints at CERN:
- Introduced "Chained Processing":**
  - Use one batch slot to process different steps of sample production consecutively
  - Local output file used as input of next step and also copied to tape

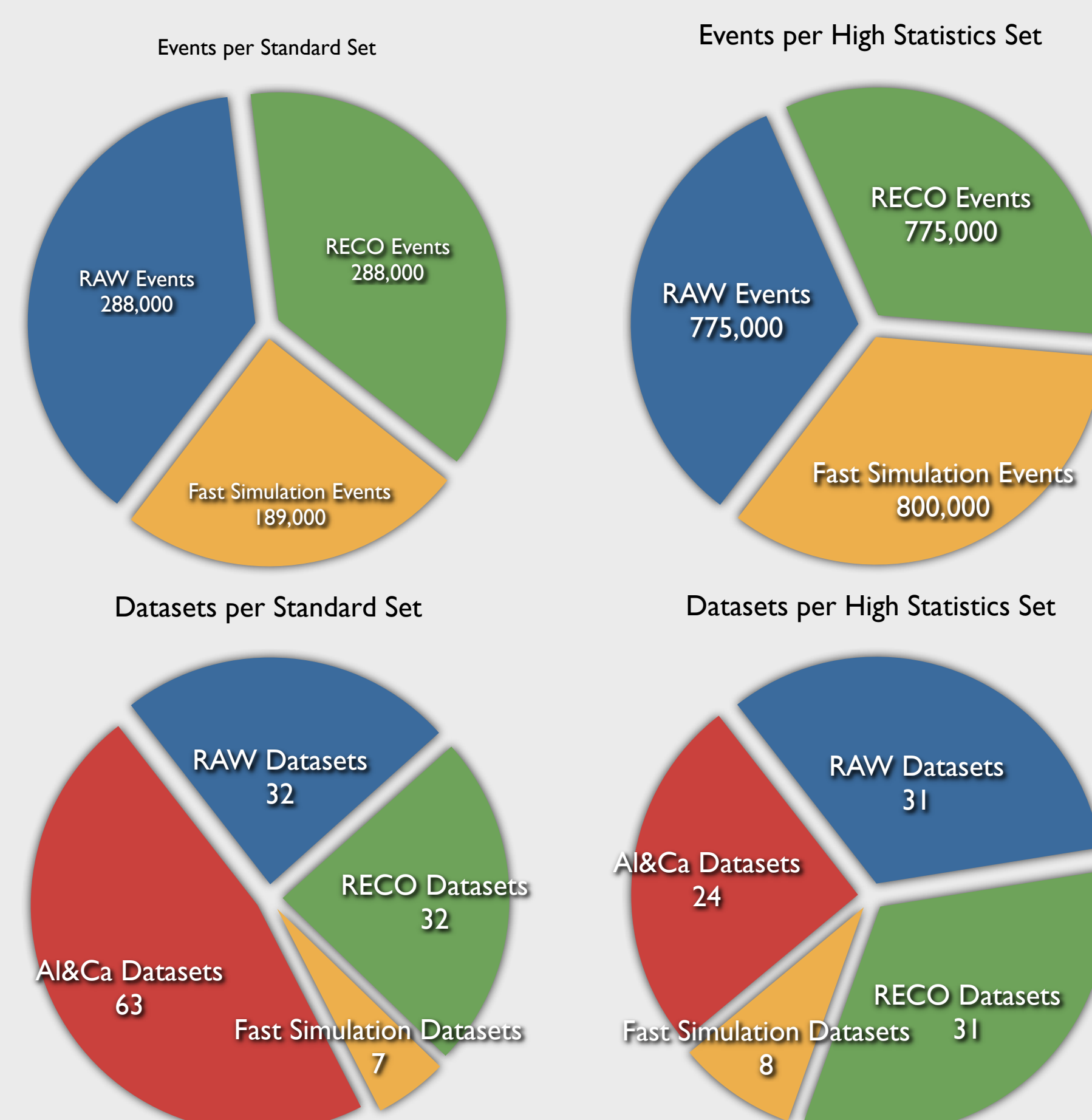


- With Chained Processing**
  - Usage of CERN batch slot allocation optimal
  - Production of Standard Set takes 24 hours

Special Validation Queue at CERN

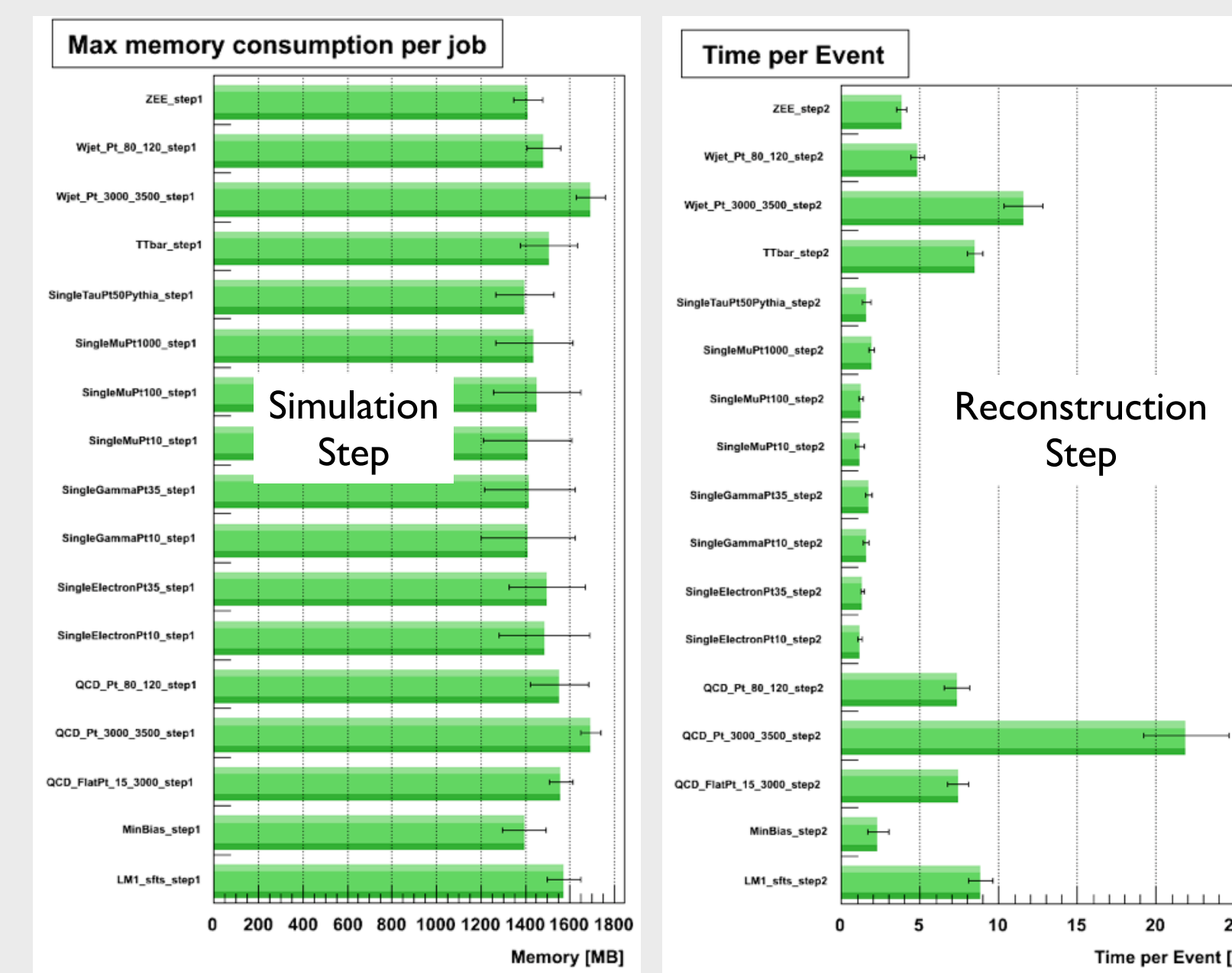


## Validation Sets



## Performance Overview

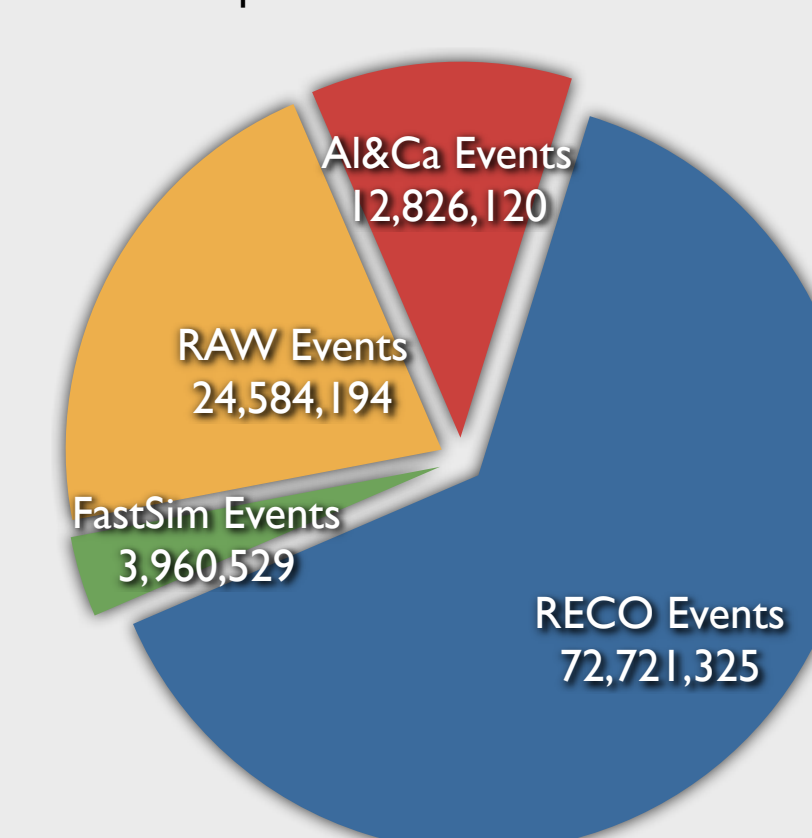
- Extract for every release: **Performance Overview**
  - Extract average memory and time consumption
  - Compare different samples in one release and between releases. Example overview for one release:



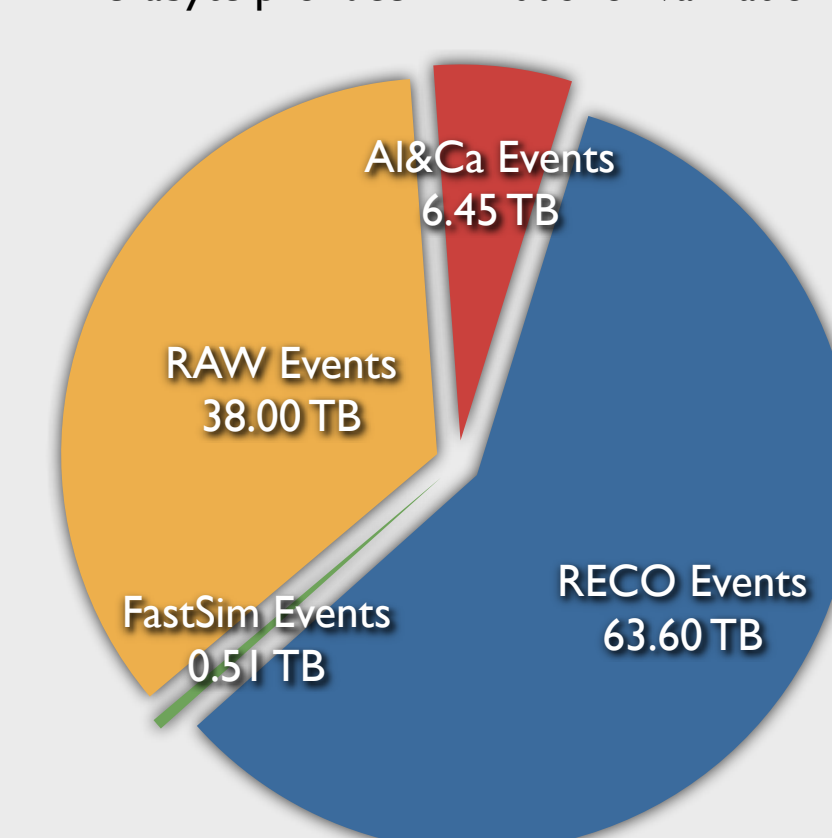
## 2008 Release Validation

- In 2008, CMS produced for Validation:**
  - More than **5100 Datasets** consisting of:
    - More than **114 Million events**
    - Almost **110 TB of MC and data samples**

Events produced in 2008 for Validation



Terabyte produced in 2008 for Validation



- Release Validation is an integral part of the CMS software development process**
  - Developers rely on timely provision of reference samples to validate their software components:
    - 2/3 of the samples** have been validated **within 2 days**
    - 1/3 within 6-10 days**
  - Demand exceeds production capabilities** many times
- Release Validation process contributes significantly to ensure stable production and analysis**
  - Releases usable by the distributed production infrastructure** of CMS
  - Stable over larger scales and event numbers**
- In 2008, Release Validation contributed significantly to the timely and successful delivery of software releases** for data taking, central production activities and analysis

### Related CHEP talks and posters

**Talk:** 260 - Software integration and development tools in the CMS experiment  
Software Components, Tools and Databases - Thursday 26 March 2009 16:50  
Presenter: LANGE, David (LLNL)

**Poster:** 82 - CMS production and processing system - Design and experiences  
Poster session - Tuesday 24 March 2009 08:00  
Presenter: Mr. VAN LINGEN, Frank (California Institute of Technology)

**Poster:** 156 - Experience Building and Operating the CMS Tier-I Computing Centers  
Poster session - Monday 23 March 2009 08:00  
Presenter: GRANDI, Claudio (INFN Bologna)