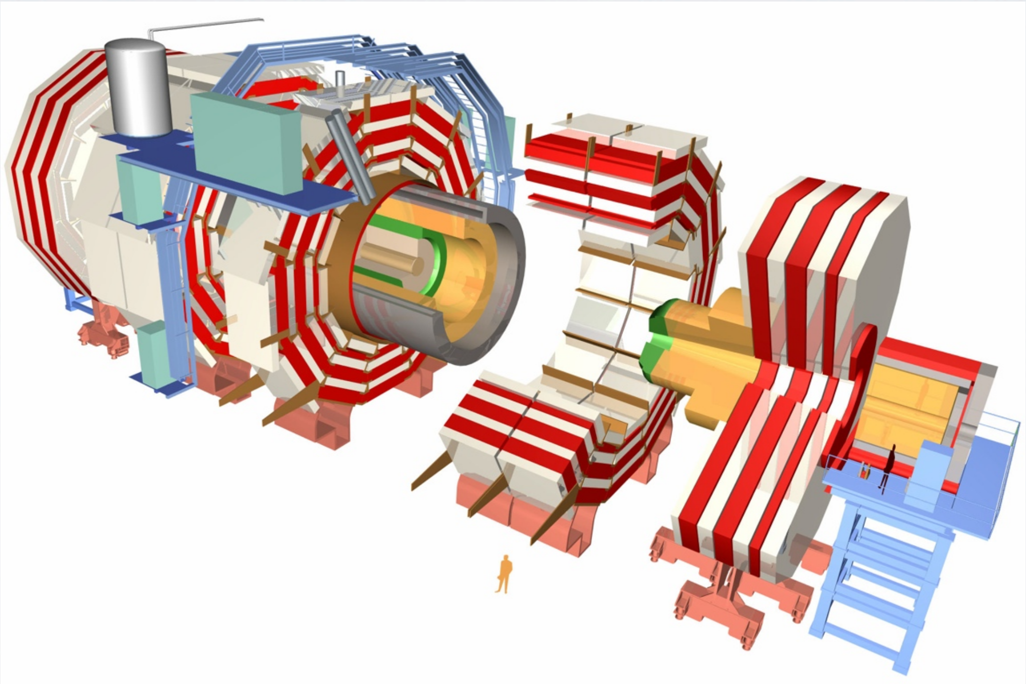
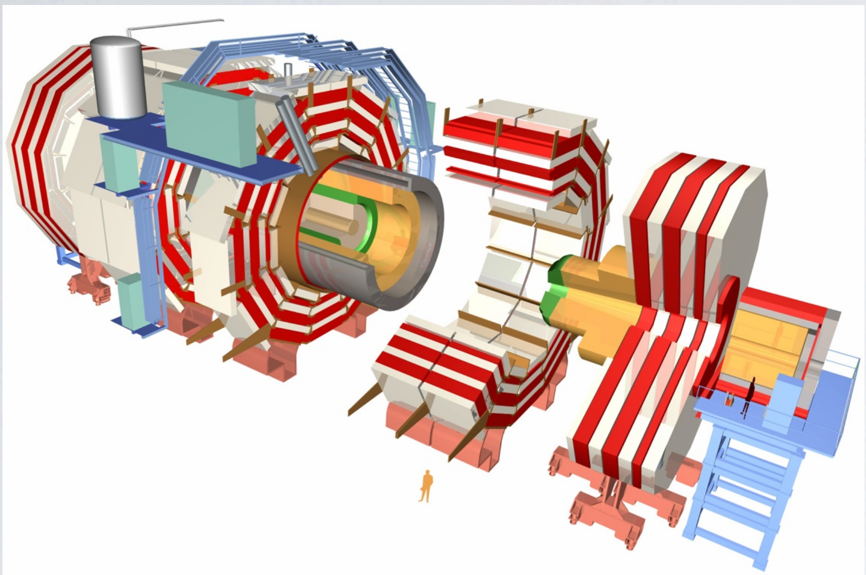
INTRODUCTION TO CMS COMPUTING



J-Term IV
8/3/09

Oliver Gutsche, Fermilab

CMS IN A NUTSHELL (CURRENT PLANNING BASIS)

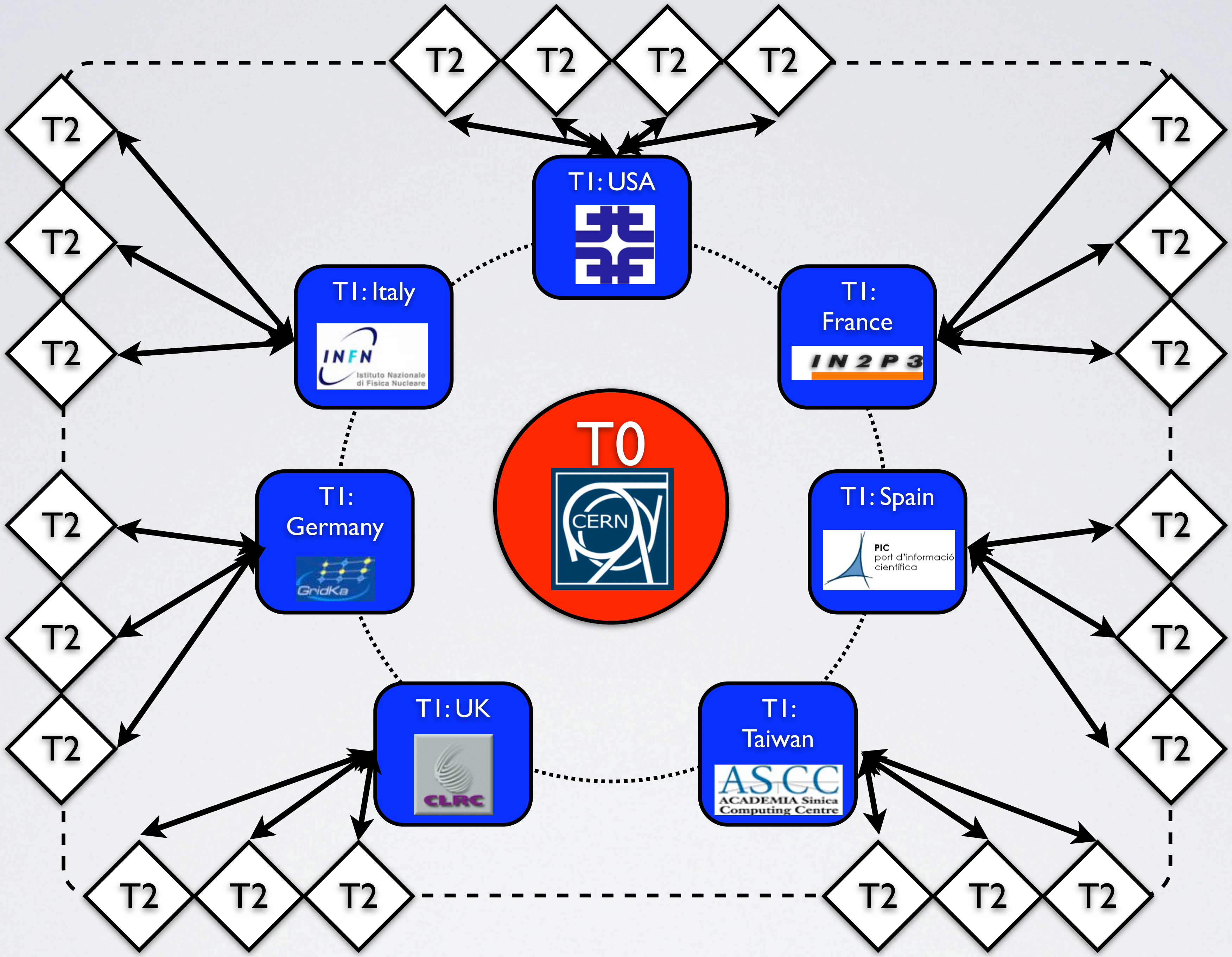


- Data recording rate: 300 Hz
- Size per event:
  - RAW: 1.5 MB (RAW+SIM 2 MB)
  - RECO: 0.5 MB
  - AOD: 0.1 MB
- Processing power per event:
  - Simulation (including reconstruction):
    - 1 event simulated and reconstructed in **100s** on 3 GHz core (1000 HS06)
  - Reconstruction:
    - 1 event reconstructed in **10s** on 3 GHz core (100 HS06)

2009/2010

- We all are eagerly awaiting the start of data taking in 2009
- Following estimates for the data taking year have been determined:
    - "2009-run": Oct'09 - Mar'10: **726M events** (already outdated, but we'll stick with it for this talk)
    - "2010-run": Apr'10 - Sep'10: **1555M events**
- Translates to:
    - **3.42 PB** RAW data,
    - **1.14 PB** RECO
        - To take re-reconstruction passes into account, this number has to be multiplied by **~3** for the data taking period 2009/2010
- This refers to collision data, MC plan for matching every recorded event with one MC event

CMS COMPUTING MODEL: TIERS



- Distribute computing resources and interconnect computing centers
  - Leverage national and local resources (hardware and manpower)
  - Smaller sites are easier to administrate and operate
  - Spend tax payers money in own country rather than concentrate in one place

CMS COMPUTING MODEL

- **Tier 0 (T0)** at CERN **(20% of all CMS computing resources)**
  - Record and prompt reconstruct collision data
  - Calculate condition and alignment constants, provide prompt physics feedback on special resources available at CERN: **CAF**
    - Access to the CAF is controlled and has to be specially granted
  - Store data on tape (archival copy, no general access for anyone)
  - **Central processing only, *no user access***

CMS COMPUTING MODEL

- **Tier 1 (T1)**: regional centers in 7 countries **(40% of all CMS computing resources)**
  - ASGC (Taiwan), CNAF (Italy), FZK (Germany), FNAL (USA), IN2P3 (France), PIC (Spain), RAL (Great Britain)
  - Store recorded data and produced MC on tape
    - Central operations and specialized workflows are allowed access
    - Every T1 site gets a fraction of the data and MC according to its respective size
    - Every T1 site holds a full set of **A**nalysis **O**bject **D**ata (AOD, subset of RECO output which should be sufficient for 90% of all analyses)
  - **Central processing only, *no user access***
    - Centrally skim data to reduce data size and make data more easily handleable
      - A skim contains only events fulfilling a defined skim selection, skims can be based on immutable quantities like trigger bits and trigger objects or RECO objects like reconstructed electrons, muons, jets, etc.
    - Rereconstruct data with newer software and conditions/alignment constants
  - FNAL is CMS' biggest T1 site
    - Provides also special resources similar to the CERN CAF reachable only through the CMSLPC interactive login nodes: **LPCCAF**

CMS COMPUTING MODEL

- **Tier 2 (T2)**: local computing centers at Universities and Laboratories **(40% of all CMS computing resources)**
  - Generate and simulate MC events (**50%** of the available resources)
  - **Physics group activities like group skims and user analysis** (**50%** of the available resources)
    - Data is transferred from the T1 custodial storage to the T2 sites per request by physics groups or users
    - Users can access the data and analyze the events
- **Tier 3 (T3):** small computing resources of Universities
  - No expectations by CMS:
    - Sites are not required to produce MC or host data for analysis access
    - Variety of setup possibilities, access to data via transfer to the T3 sites is possible
    - Supported in the US by the US CMS software and computing project

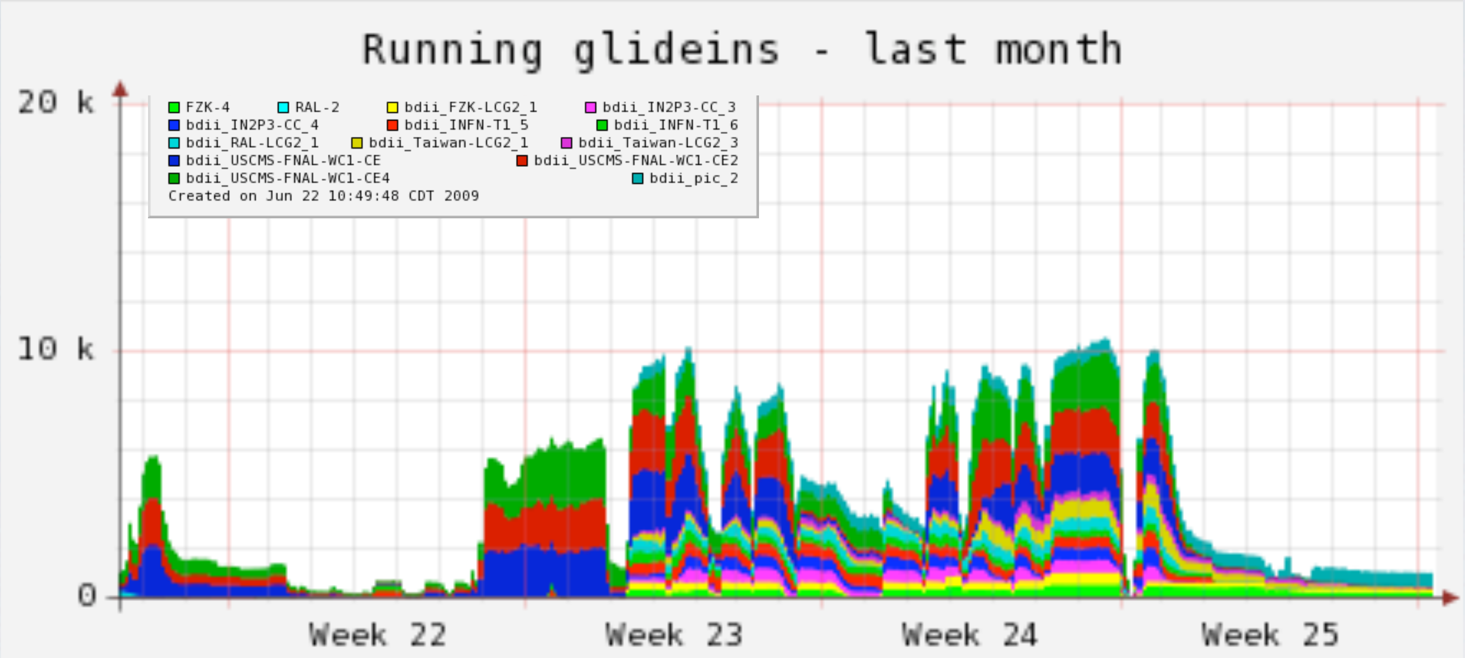
CMS DISTRIBUTED COMPUTING MODEL

- "Data driven" computing model
  - Data and MC samples are distributed centrally, jobs (processing, analysis) "go" to the data
- **Requires very fast network connections between the different centers:**
  - **T0→T1**: handled via the LHC-OPN (**O**ptical **P**rivate **N**etwork) consisting of dedicated network links (at least 10 Gbit/s)
    - Used to distribute the recorded data for storage on tape at T1 sites
    - Rule: always have two copies of each data event on two independent safe storage medium
      - This requires in time transfers to the T1 sites because the disk buffer space at the T0 is not large enough to sustain data taking for longer times
  - **T1→T1**: also handled via the OPN
    - Redistribute parts of the data produced during rereconstruction (AOD)
  - **T1→T2**: handled via national high speed network links
    - Transfer datasets for analysis to T2 sites
  - **T2→T1**: handled via national high speed network links
    - Transfer produced MC to T1 for storage on tape

T0 WORKFLOWS

- The online system of the detector records events and stores them in binary files (streamer files)
- There are 2 processing paths in the T0:
  - **Bulk**: latency of several days
    - Repacking of the binary files into ROOT files and splitting of the events into Primary Datasets according to trigger selections (RAW data tier)
    - Reconstruction of the RAW data for the first time (Prompt Reconstruction) (RECO data tier) including AOD extraction
    - Special alignment and calibration datasets are produced and copied directly to the CERN Analysis Facility (**CAF**)
    - All RAW, RECO and AOD data is stored on tape at CERN and transferred to the T1 sites for storage on tape
  - **Express**: latency of 1 hour
    - All the steps above combined into a single process run on 10% of all events selected online from all the recorded data
    - Output is copied to the CAF for express alignment and calibration workflows and prompt feedback by physics analysis

T1 WORKFLOWS



Currently up to 10,000 jobs in parallel on the T1 sites

- All T1 centers store a fraction of the recorded data on tape called the custodial data fraction
  - MC produced at the T2 sites is archived on tape at the T1 sites as well
- The T1 processing resources are used for
  - Skimming of data to reduce the data size by physics pre-selections
    - More easily handleable samples for the physicists
  - Rereconstruction of data with new software versions and updated alignment and calibration constants
- The T1 sites serve data to the T2 sites for analysis requested by users, physics groups or central operations

T2 WORFKLOWS

- 35-40 T2 sites serve 2000 physicists and provide access to data and MC samples for analysis
  - Each CMS physicist can access data at all the T2 sites (CRAB)
  - There is still some regional association between physicist and T2 site to support local resources for storage of user files (/store/user)
  - There is also association between physics groups (top, higgs, …) and T2 sites to organize data and MC sample placement
  - A typical T2 sites has ~800 batch slots and 300 TB of disk space
    - T2 sites don't have tape
  - Half of the resources of all T2 sites are reserved for MC production which is handled centrally

DATA OPERATIONS

- The Data Operations project handles all central processing on the different tier levels
  - Project is lead by Markus Klute (MIT) and Oliver Gutsche (FNAL)
- Data Operations has 5 main tasks:

1. Host laboratory processing (T0)
   - David Mason (FNAL) and Josh Bendavid (MIT)
2. Distributed re-processing & skimming (T1)
   - Kristian Hahn (MIT) and Guillelmo Gomez Ceballos (MIT)
3. Distributed MC production (T2)
   - Maarten Thomas (Aachen) and Ajit Mohapatra (Wisconsin)
4. Data transfers and integrity
   - Paul Rossman (FNAL) and Si Xie (MIT)
5. Release validation & data certification
   - Oliver Gutsche (FNAL) and NN

DATA OPERATIONS

- Each Data Operations team member gains intimate knowledge about the inner workings of CMS computing:
  - Thorough training to gain detailed knowledge about the computing infrastructure and software tools
  - Expert knowledge about dataset bookkeeping and location information
  - Detailed overview how and where data is processed and becomes available
  - Detailed experience with distributed data processing via the GRID
  - Directly applicable to physics analysis

DATA OPERATIONS TEAM AND SERVICE CREDIT

- Two L2 coordinators for the Data Operations project
  - Markus Klute (MIT) and Oliver Gutsche (FNAL)
- Two L3 coordinator positions per Data Operation task:
  - Coordinate and organize the specific task
  - Train and supervise operators
  - Expected to spend 50% of their time working for the data operations project: **earn 50% service credit**
- Operators:
  - Operate the computing tools to process workflows on the CMS tier structure
  - Expected to spend 25% of their time working for the data operations project: **earn 25% service credit**
- Communication is very important and is carried out by a variety of means:
  - One central meeting per week plus personal interaction, E-Log, mails and chat (very important)
- After an initial training at CERN or FNAL (2-3 weeks), L3 and operator tasks **can be performed remotely also from home institutes**

SEARCH FOR NEW MEMBERS

- The Data Operations project is looking for new members as operators and L3 coordinators
  - Currently, we are looking for 2 new coordinators for the release validation and data certification task and a new coordinator for the MC production task
  - L3 coordinator positions require larger experience with the CMS computing systems and intention for a longer term involvement (> 2 year)
- The benefits of earning necessary service credit in the Data Operations project are numerous:
  - Detailed knowledge about the computing tools applicable for physics analysis
  - Possibility to fulfill responsibilities remotely also from home institutes
- If you are interested and want to join the Data Operations project, please contact:
  - Markus Klute: klute@mit.edu
  - Oliver Gutsche: gutsche@fnal.gov

CENTRAL COMPUTING SHIFTS

- CMS Computing is - and has to be treated as - a CMS sub detector. Core computing people engineered and built it.
- Especially during the startup of the LHC, the computing infrastructure has to be constantly monitored to identify problems and trigger actions/calls.
- A general CMS Computing shift was created monitoring all the aspects and component of the overall computing machinery.
  - The shifts are meant to cover the monitoring of the computing systems 24/7 in 3 8 hour shifts
  - The shifter is called CSP (**C**omputing **S**hift **P**erson)

CSP DUTIES

- The CSP follows periodically (every 1-2 hours) instructions documented on a TWiki during his 8 hour shift:
  - https://twiki.cern.ch/twiki/bin/view/CMS/ComputingShifts
  - Reports observations and problems to the Elog
  - Triggers Savannah tickets to processed by experts and site admins
- Currently, CSP shifts can be taken in all CMS centers (CERN, FNAL ROC, DESY, …) which fulfill the CMS Center requirements and have a permanent video link for communication with the detector control room and other centers
  - http://cmsdoc.cern.ch/cmscc/index.jsp

CSP SHIFT TEAM AND SERVICE CREDIT

- CSP shifters earn CMS service credit by taking CSP shifts corresponding to a 25% service credit
- Especially in the US, the CSP shifter team is very small and new members are more than welcome
- If you are interested in taking CSP shifts for CMS, please contact
  - Peter Kreuzer: peter.kreuzer@cern.ch
  - Oliver Gutsche: gutsche@fnal.gov

SUMMARY

- CMS computing needs are significant, CMS expects to record in 2009/2010 alone more than 6 PetaByte of collision data
- All data and MC has to be processed and accessed via the distributed tiered computing infrastructure and the GRID
- Various opportunities are open for CMS collaborators to contribute to the central Data Operations team and the general CMS computing shifts
  - Working for Data Operations is an excellent preparation for distributed physics analysis and provides valuable experience and knowledge about the details of CMS data and MC
  - All contributions earn CMS service credit
  - If you are interested to join, please contact us.

THERE IS MORE

- Dave Mason, Monday, 3rd August 2009, 5 PM:
  - Data Operations - Where is the Data? How do I Access it? Is it OK? DBS
- Eric Vaandering, Tuesday, 4th August 2009, 12.10 PM:
  - The CMS Computing Model and User Tools