CMS Frontier Experience

Distributed Database Workshop
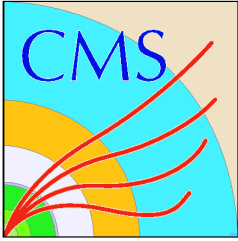
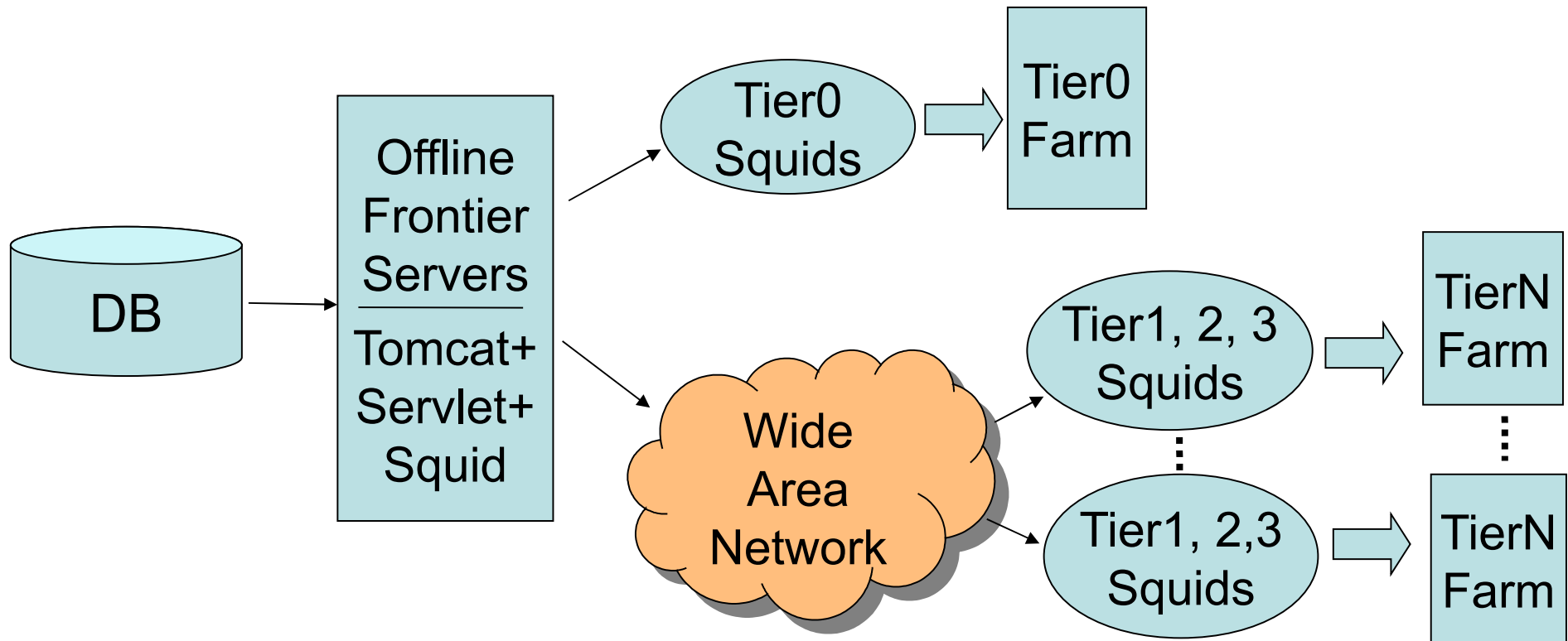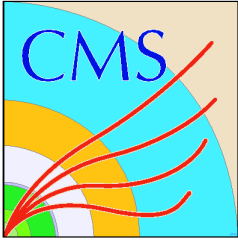Dave Dykstra, Fermilab
dwd@fnal.gov
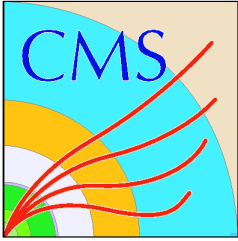
Overview

- CMS uses Frontier to read **all** conditions, both Offline and Online, for all workflows
- Offline has been in production for nearly 4 years
  - Many improvements along the way
- The only Offline outages have been brief DB, power, or network outages at CERN
  - Occasionally individual Frontier servers have wedged but software improvements keep reducing those cases

CMS Offline Frontier architecture

DB → Offline Frontier Servers / Tomcat+ Servlet+ Squid

Tier0 Squids → Tier0 Farm
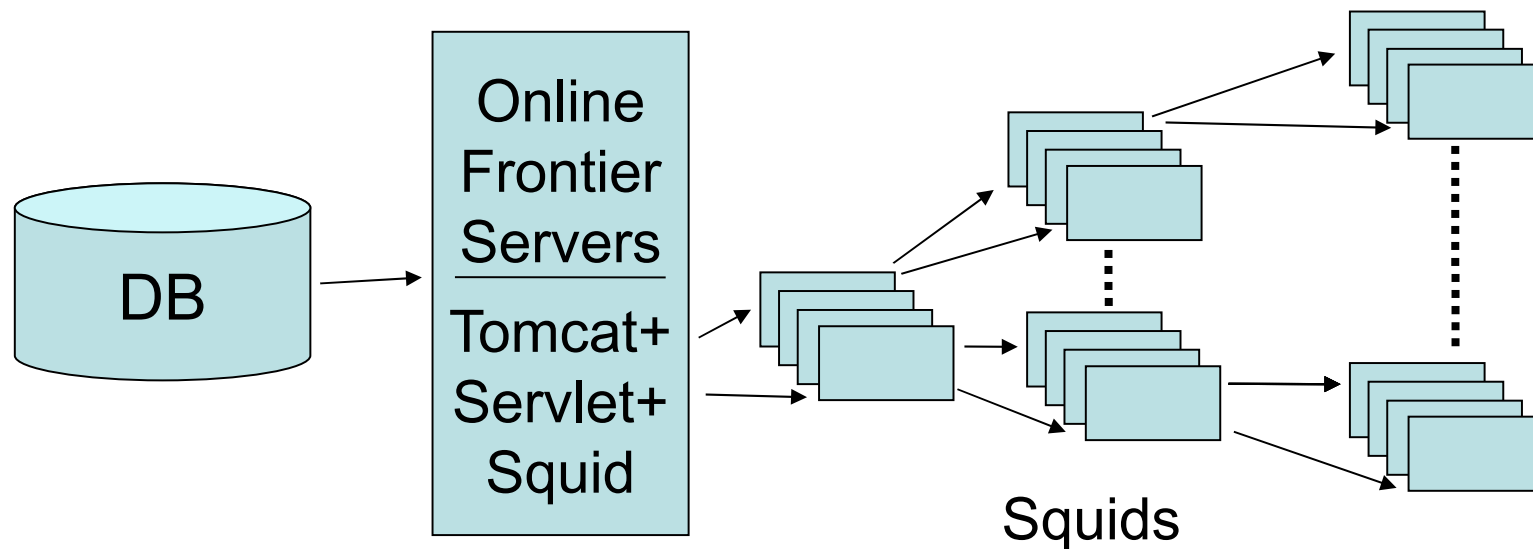
Wide Area Network → Tier1, 2, 3 Squids → TierN Farm

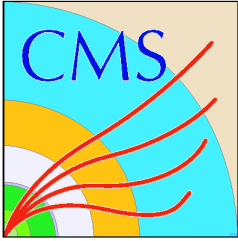Tier1, 2,3 Squids → TierN Farm

- One set of 3 Frontier Servers for the world
- Cache expirations vary from 5 minutes to a year

CMS Online Frontier architecture

- Hierarchy of squids on every worker node
- Blasts data to all 1400 worker nodes in parallel
- Caches only for 30 seconds
- 2 servers for reliability, shared with s/w repository

Statistics

- CMS jobs do roughly 2K requests at startup
- Over the last week:
  - Offline launchpads averaged 6.5K req/min, 0.7MB/s
  - Squids globally averaged 500K req/min, 500MB/s
    - 250 average job starts per minute, 120MB per job
    - 360K jobs per day
  - Almost all Offline Frontier usage less than ¼ CPU hour on one database CPU core for a whole week
    - One special usage, reading Luminosity data, doesn't have many repeat readers and used 13 DB CPU hours
 - Online jobs read conditions in about 30 seconds
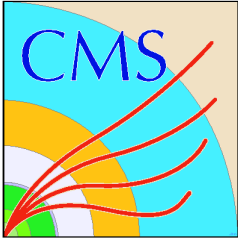
Monitoring

- **For central Frontier servers**
  - Service Level Status (SLS)
  - Awstats collects stats regarding hits on servers
    - New automatic analyzer (more on next slide)
- **For worldwide squids**
  - SAM tests try all squids
- **For both**
  - MRTG collects usage stats every 5 minutes
  - Automatic email when MRTG sees squid down
- **Part-time operations people watch for problems**
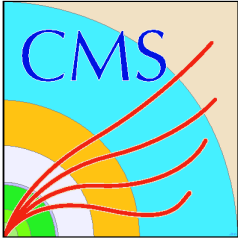
Recent updates

- New automatic analyzer of awstats
  - Sends email to administrator of a squid that has had greater than 5000 direct hits from IP addresses at its site onto the central servers for the last 2 hours
  - Sends email no more than once every 24 hours
  - Uses internet service to discover "ISP" for addresses
- Fix for Frontier servers getting stuck for 2 hours after a DB node crash
  - Set kernel parameter tcp_keepalive_time = 120
- One server moved to VM on critical power
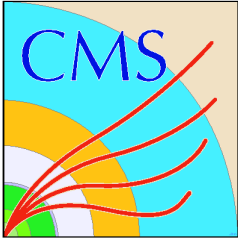
Summary

- **CMS use of Frontier has been very solid**
  - Frontier is in a very low maintenance mode
  - Performance and robustness is excellent
- **CMS avoids a lot of complexity that ATLAS has:**
  - Streaming conditions to many Tier 1s
  - Databases and Frontier servers administered by many different people
  - Separate infrastructure for reprocessing conditions
  - Separate infrastructure for large conditions files
  - Separate infrastructure for Online conditions