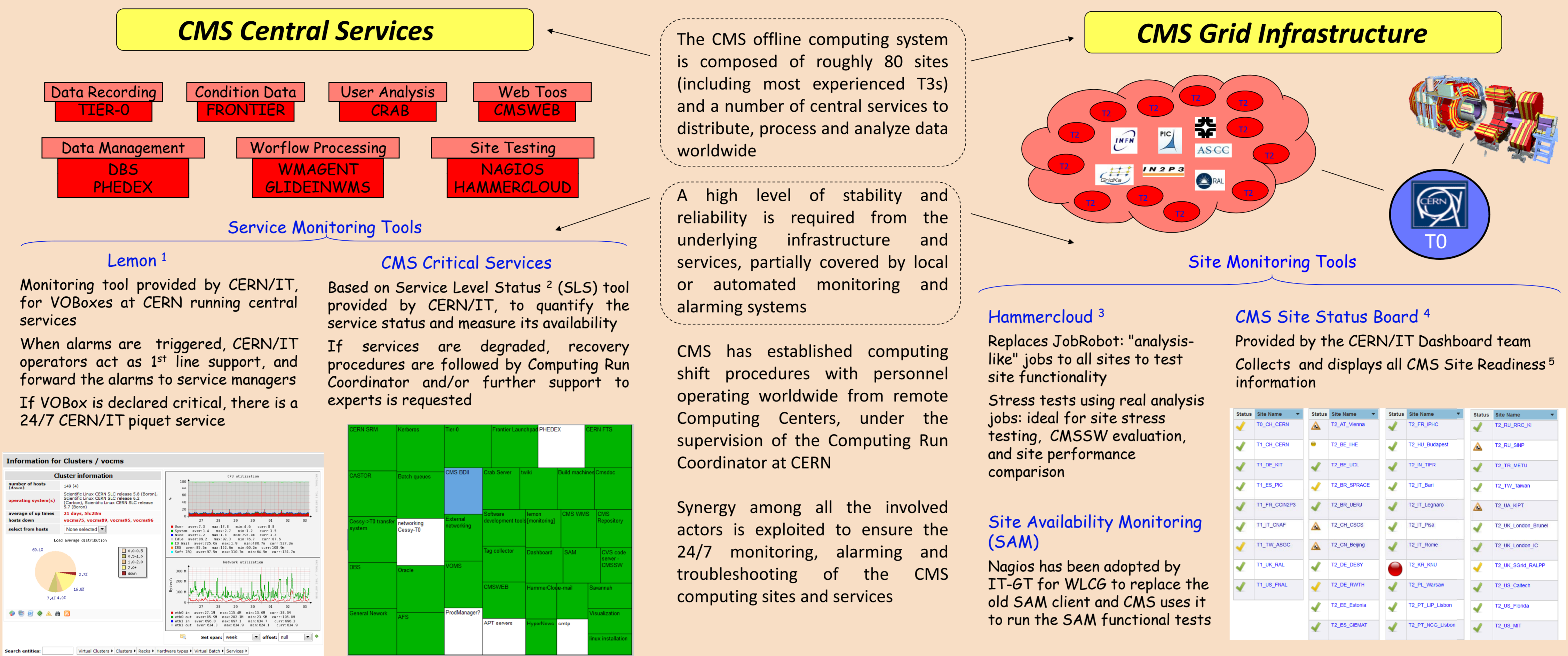# Monitoring techniques and alarm procedures for CMS services and sites in WLCG

Poster Presenter:
Jorge.Molina-Perez@cern.ch

Tommaso Boccali (INFN), Daniele Bonacorsi (Univ. Bologna), Ignas Butenas (Vilnius), E. Fajardo (Uniandes), Josep Flix (PIC/CIEMAT), Ran Du (Beijing), Diego Gomes (UERJ), Oliver Gutsche (Fermilab), Rapolas Kaselis (Vilnius), Markus Klute (MIT), Peter Kreuzer (RWTH/CERN), Nicolo Magini (CERN), **Jorge Molina-Perez (CERN/UCSD)**, Andrea Sciaba (CERN), Weizhen Wang (Beijing)

CMS Central Services

Data Recording	Condition Data	User Analysis	Web Toos
TIER-0	FRONTIER	CRAB	CMSWEB

Data Management	Worflow Processing	Site Testing
DBS / PHEDEX	W/MAGENT GLIDEINWMS	NAGIOS HAMMERCLOUD

The CMS offline computing system is composed of roughly 80 sites (including most experienced T3s) and a number of central services to distribute, process and analyze data worldwide

A high level of stability and reliability is required from the underlying infrastructure and services, partially covered by local or automated monitoring and alarming systems

CMS Grid Infrastructure

T0

Service Monitoring Tools

Lemon [1]

Monitoring tool provided by CERN/IT, for VOBoxes at CERN running central services

When alarms are triggered, CERN/IT operators act as 1st line support, and forward the alarms to service managers

If VOBox is declared critical, there is a 24/7 CERN/IT piquet service

CMS Critical Services

Based on Service Level Status [2] (SLS) tool provided by CERN/IT, to quantify the service status and measure its availability

If services are degraded, recovery procedures are followed by Computing Run Coordinator and/or further support to experts is requested

CMS has established computing shift procedures with personnel operating worldwide from remote Computing Centers, under the supervision of the Computing Run Coordinator at CERN

Synergy among all the involved actors is exploited to ensure the 24/7 monitoring, alarming and troubleshooting of the CMS computing sites and services

Site Monitoring Tools

Hammercloud [3]

Replaces JobRobot: "analysis-like" jobs to all sites to test site functionality

Stress tests using real analysis jobs: ideal for site stress testing, CMSSW evaluation, and site performance comparison

Site Availability Monitoring (SAM)

Nagios has been adopted by IT-GT for WLCG to replace the old SAM client and CMS uses it to run the SAM functional tests

CMS Site Status Board [4]

Provided by the CERN/IT Dashboard team

Collects and displays all CMS Site Readiness [5] information

Roles for 24/7 Service Operations

CMS Computing Operator at CERN (pool of ~15 people)
- Routine service operations and monitoring
- Responding as On-Call Expert during day operations
- VOCMS cluster administration and management

Computing Run Coordinator – CRC (from a pool of ~15 people)
- 1 CRC/week
- Applies Critical Service Recovery Procedures
- Triggers phone calls to On-Call Experts

Computing Shift Person – CSP (from a pool of ~140 people)
- 3 CSP/day == 24/7 coverage
- Follows general monitoring instructions

CMS local Site admins (typically 1-5 people per site)

CMS local Site contact (more critical for T1s)
- Conducts overall site trouble shooting for CMS specific issues
- Addresses CMS specific operations to central site admins

Communication Tools

Service Now
- Single Service Desk at CERN with standard processes for all service providers
- Available for incident management and request fulfillment

CMS Computing LogBook
- Electronic LogBook provided by CERN/IT
- Subdivided in several categories according to main CMS Computing workflows

Savannah
- Problem tracking tool provided by WLCG
- Subdivided in CMS "trackers" and "squads", the latter are grouping experts to be targeted for solving site or services issues

GGUS
- Problem tracking tool provided by WLCG, mapped to all sites and Regional Operations Centers (ROCs)
- Featuring various ticket attributes : "User" / "Team" / "Alarm"; the latter are only for Tier-0/Tier-1 and can trigger 24/7 phone calls to local site experts

CMS VOBox Central Infrastructure

VOCMS cluster
- Hosts CMS Central Services (e.g. improvement of the submission infrastructure and Operations effort make possible to send up to 50k parallel production running jobs [6] into the Grid)
- ~ 145 voboxes (~ 8.3 kHS06) at CERN, progressive virtualization of nodes (41 %)
- Lemon monitoring, quattor administration (~ 60 profiles) allows automated server re-installation from scratch → production machine ready in a couple of hours

Critical Services Documentation
- Project developed in collaboration with CMS service managers
- Service criticality defined according to the expected response time
- Definition of recovery procedures to be applied by CERN/IT operators and/or the CRC when service experts are not reachable
- After the recovery procedure is validated for a service, a new SLS entry is included in the CMS Critical Services gridmap

Criticality Level	Meaning	Response Time (h)
10	CMS stops operating	0.5
10	CMS stops transferring from Cessy	0.5
8	T0 Production stops	0.5
7	T1/T2 Production/analysis stops	2

Worldwide Computing Operations Centers

~140 Computing Shift Persons located at ~30 institutions covering the Asian, European, and American zones

CMS Remote Operations Centre at Fermilab

CMS Experiment Control Room

CMS Centre at CERN: monitoring, computing operations,

Dedicated 24/7 computing shift personnel is contributing to detect and react timely on any unexpected error and hence ensure that CMS computing operations are carried out in an efficient and sustained manner

Service and Site troubleshooting procedure

Service/Sites Monitoring
Mapped to SLS   Mapped to SAM

CSP checks every 2 hours Status Green ?
No / Yes

E-LogBook & Ticketing tool

Expert Computing Operations
CMS Core Computing experts / CMS Site Contacts:
- Apply routine service / infrastructure operations and monitoring
- Respond as On-Call Experts to Alarms

CERNIT

Expert answer within 1 hour ?
No / Yes

Service/Site Alarm Procedure
Computing Run Coordinator (CRC) reachable 24/7 for :
- Critical Service recovery procedure
- Priority (GGUS-Team) ticket to site

Core System Alarming (CERN)
Integrated in Lemon

Problem solved ?
No / Yes

(*) highly critical alarms to Tier-0/1s are sent via GGUS-Alarm tickets and can trigger phone calls. Similarly for CMS Service On-Call Operators.
(***) CRC, Service Expert or Site Contacts actions are systematically reported back to the E-LogBook for transparency purposes.

High Reliability of CMS Sites

- Since LHC collisions started, jobs efficiency has increased progressively, in particular during the last year

- The average of CMS Site Readiness has been around 80% during the last 2 years

Efficiency Good jobs
104 Weeks from Week 11 of 2010 to Week 11 of 2012

Site Readiness Status for CMS Tier-2 sites

- Service availability level has been satisfactory since the beginning of LHC operation

- Critical Service operations are now well integrated into the 24/7 computing shift monitoring/alarming/recovering procedures

CHEP'2012 conference, New York, 21-25 May 2012