This space left
intentionally blank

SC4 Plans

Jamie.Shiers@cern.ch

July 2005

Known Knowns

- Model for 'Production' much debated and now well understood
  - All stages from data taking leading into to end-user analysis
  - Has been exercised through experiment data challenges
  - Will also be covered during Service Phase of SC3
    - Main goal is to thoroughly stress-test the service infrastructure

- Data types, rates, flows that correspond to above all 'clear'
  - Processing, re-processing, stripping, AOD / TAG Production etc

- Roles played by different tiers, services that they offer, services that they require etc also understood
  - Services still not fully setup; in some cases software maybe…
  - Still a number of Tier2s with no clear Tier1
  - Expect to continue to make good progress on this prior to SC4

- Current plan is for 50 days of data taking in 2007 @ $x.10^{32}$ cm^{-2}s^{-1}

- **Service Challenge schedule / delivery of production system unchanged**

Known Unknowns

- End-user analysis still a mystery
  - Can easily result in significant network bandwidth / support load
  - What is the model for Analysis Facilities?
    - Dedicated PROOF farms? 100+ nodes, 50+TB disk
    - Batch mode? Single stream? Parallel?
  - We need some serious analysis facilities setup!
  - ➢ **More coordination with ROOT team and ARDA being established**

- Startup phase of LHC unknown
  - It will certainly not be like steady-state
  - Strong pressure to exploit **(needed)** distributed resources
  - There will be a strong presence at CERN, but nevertheless fundamental need to allow detector / physics groups outside have rapid / peer access to the data

- Emphasize this in paper selection for CHEP 2006

Service Challenge 4 – SC4

- SC4 starts April 2006
- SC4 ends with the deployment of the **FULL PRODUCTION SERVICE**

➢ **Deadline for component (production) delivery: end January 2006**

- Adds further complexity over SC3 – 'extra dimensions'
    - Additional components and services, e.g. COOL and other DB-related applications
    - Analysis Use Cases
    - SRM 2.1 features required by LHC experiments **← have to monitor progress!**
    - MostTier2s, all Tier1s at full service level
    - Anything that dropped off list for SC3…
    - ➢ **Services oriented at analysis & end-user**
    - What implications for the sites?

- Analysis farms:
    - Batch-like analysis at some sites (no major impact on sites)
    - Large-scale parallel interactive analysis farms and major sites
    - (100 PCs + 10TB storage) x N

- User community:
    - No longer small (<5) team of production users
    - 20-30 work groups of 15-25 people
    - Large (100s – 1000s) numbers of users worldwide

Analysis Use Cases (HEPCAL II)

- Production Analysis (PA)
  - **Goals in Context** *Create AOD/TAG data from input for physics analysis groups*
  - **Actors** *Experiment production manager*
  - **Triggers** *Need input for "individual" analysis*

- (Sub-)Group Level Analysis (GLA)
  - **Goals in Context** *Refine AOD/TAG data from a previous analysis step*
  - **Actors** *Analysis-group production manager*
  - **Triggers** *Need input for refined "individual" analysis*

- End User Analysis (EA)
  - **Goals in Context** *Find "the" physics signal*
  - **Actors** *End User*
  - **Triggers** *Publish data and get the Nobel Prize :-)*

SC4 Timeline

- Now - September: clarification of SC4 Use Cases, components, requirements, services etc.

- October 2005: SRM 2.1 testing starts; FTS/MySQL; target for post-SC3 services

- January 31st 2006: basic components delivered and in place

- February / March: integration testing

- February: SC4 planning workshop at CHEP (w/e before)

- March 31st 2006: integration testing successfully completed

- April 2006: throughput tests

- May 1st 2006: Service Phase starts (note compressed schedule!)

- September 1st 2006: Initial LHC Service in stable operation

- Summer 2007: first LHC event data

SC4 Milestones (TDR)

Date	Description
31 Jan 06	**All required software for baseline services deployed and operational at all Tier-1s and at least 20 Tier-2 sites.**
30 Apr 06	**Service Challenge 4 Set-up: Set-up complete and basic service demonstrated. Performance and throughput tests complete: Performance goal for each Tier-1 is the nominal data rate that the centre must sustain during LHC operation (see** Table 7.2 below) CERN-disk → network → Tier-1-tape. Throughput test goal is to maintain for three weeks an average throughput of 1.6 GB/s from disk at CERN to tape at the Tier-1 sites. All Tier-1 sites must participate. The service must be able to support the full computing model of each experiment, including simulation and end-user batch analysis at Tier-2 centres.
31 May 06	**Service Challenge 4: Start of stable service phase, including all Tier-1s and 40 Tier-2 centres.**
30 Sept 06	**1.6 GB/s data recording demonstration at CERN: Data generator→ disk → tape sustaining 1.6 GB/s for one week** using the CASTOR mass storage system.
30 Sept 06	**Initial LHC Service in operation:** Capable of handling the full target data rate between CERN and Tier-1s (see Table 7.2). The service will be used for extended testing of the computing systems of the four experiments, for simulation and for processing of cosmic-ray data. During the following six months each site will build up to the full throughput needed for LHC operation, which is twice the nominal data rate.
1 Apr 07	**LHC Service Commissioned: A series of performance, throughput and reliability tests completed to show readiness to operate continuously at the target data rate and at twice this data rate for sustained periods.**

SC4 Use Cases (?)

Not covered so far in Service Challenges:

- T0 recording to tape (and then out)

- Reprocessing at T1s

- Calibrations & distribution of calibration data

- HEPCAL II Use Cases

- Individual (mini-) productions (if / as allowed)

Additional services to be included:

- Full VOMS integration

- COOL, other AA services, experiment-specific services (e.g. ATLAS HVS)

- PROOF? xrootd? (analysis services in general…)

- Testing of next generation IBM and STK tape drives

September SC3.5 workshop
(Most likely a 'virtual' one…)

- SC3 experience
  - Sites
  - Experiments
  - Outlook for remainder of service phase
  - Requirements gathering from site + experiment view points + report (by two rapporteurs from above sessions)

- SC4 preparation
  - (recent) experiment goals / plans in terms of HEPCAL use cases
  - proof / xrootd / roles / plans
  - LCG SRM status
  - targets for SC4
  - T1 plans for incorporating T2s
  - T2 plans

Remaining Challenges

- Bring core services up to robust 24 x 7 standard required

- Bring remaining Tier2 centres into the process

- Identify the additional Use Cases and functionality for SC4

- Build a cohesive service out of distributed community

- Clarity; simplicity; ease-of-use; functionality

Work Items (= Milestones?)

- **SRM developers to produce plan for SRM 2.1 features**
  - Delivery of first versions Oct 2005; production Jan 2006
- **T1s to provide 10Gbit/s network links to CERN**
  - On-going… but (v.) late in some countries
- **T1s/T2s to setup File Catalog services**
  - Primarily LFC (ALICE, ATLAS)
- **Service model for FTS / LFC etc on MySQL**
  - T1(/2?) to prototype
- **Resolution of VO boxes**
- **Agreement on SC4 Use Cases**
- **…**

Actions

- **September GDB**
    - First list of Use Cases by experiment to be tested in SC4
    - List of additional services required by site (over SC3)
    - LCG SRM 2.1 implementation plans by concerned SRM

- **September ROOT workshop**
    - Discussion of analysis wrt ROOT / PROOF for SC4

- **October GDB**
    - Revised list of Use Cases for SC4
    - Site responses / deployment plans for additional services
    - Clear (I mean clear) response from Service Providers
    - LCG SRM 2.1 implementation status (test releases at this time)
    - T2 plan by T1
    - Identification of 'T1-less' T2s - resolution

Job Done…