Patricia Méndez Lorenzo

Status of the T0 services

Presentation outlook

- **Review of CERN services**
- **Techniques for handling main types of interventions**
- **Services presented**
  - CE, WN: Ulrich Schwickerath
  - CASTOR: Miguel Coelho
  - LFC: Jean Philippe Baud, Harry Renshall
  - FTS: Gavin McCance
  - Data Bases: Maria Girone, Dirk Düllmann

Points to review

- **Hardware**
  - Power supplies in case of failures
    - Different front-ends, single point of failure, redundant configuration, UPS
  - Servers
    - Single or multiple, DNS load balanced, HA Linux, RAC
  - Network
    - Servers connected to different switches
- **Software**
  - Middleware
    - Can handle loss of one or more servers
  - Impact
    - In other services and/or users of a loss/degradation
- **Quiesce/Recovery**
  - Can the service be cleanly paused, is there built-in recovery?
- **Tests and Documentation**
  - Tried in practice, transparent interventions?
  - Documents for operations and service

Status of CEs and WNs (I)

- **Hardware**
  - WNs: “cheap hardware”
    - Each connected to different power supplies
    - Not placed in the critical area
  - CEs
    - 50% placed in the batch hardware area
      - Each CE has one power supply only
    - The rest placed in a redundant system
      - Two power supplies per CE
      - Transparent switch to the 2nd power supply
  - WNs and CEs connected to different network switches (not all of them)

Status of CEs and WNs (II)

- **Software**
  - Middleware
    - WNs: jobs affected in case of node lost
    - CEs: able to handle short interruptions
  - Impact
    - Clean impact in the individual jobs in case of hardware problems in located nodes
    - The overall service will however be ensured

Status of CEs and WNs (III)

- **Recovery**
  - Clean pause ensured
    - Standby time until the end of accepted requests
    - Drain of the queues (up to 1 week) before total stop of the system
      - New requests will not be accepted (node out of production)
- **Tests and documentation**
  - Full documentation under construction
  - Procedure tested in many cases
    - i.e migration of WNs to SLC4
    - Whole service performance ensured

Status of the FTS (I)

- **Hardware (I)**
  - Split into several components
    - Web-service
      - DNS load-balanced over 3 nodes
      - Checked by monitoring every minute and drop problematic nodes from the load balance
    - Data-transfers
      - Channel agent daemons
        - Balanced over 3 nodes with no redundancy
        - Good partition: problems in one channel do not affect the rest of channels
      - VO agent daemons
        - All available in one node with no redundancy
    - Proxy renewal
      - Load balanced in several servers
    - Monitoring
      - Placed on a single node
      - Not critical for the service operation

Status of the FTS (II)

- **Hardware (II)**
  - Service daemons randomly distributed over the hardware connected to different network switches
    - Redistribution for higher resilience to internal switch failures
  - External network: Required for transfers
    - A downtime will disable the web-service for external users and the monitoring
      - Software remains operational
      - 100% of failures on all channels
  - Internal network: individual switches that services are connected to
    - A downtime will affect the web-service in that switch
    - DNS load balance should be configured to detect the nodes from the alias
    - Transfers 100% unable

Status of the FTS (III)

- **Software**
  - FTS middleware can transparently handle loss of one or multiple servers
  - Service components well de-coupled from each other
    - Each component will keep running although other one is down

Status of the FTS (IV)

- **Recovery (I)**
  - SRM failure: Channel should be paused
  - Internal component failure: No state is lost
    - Web services
      - Poor resilience to glitches. DNS propagation is not fast enough to hide short glitches
      - Upon restart of the problematic node (including DNS propagation), the service is automatically back up
      - No loss of states
    - Data transfer Agents: no jobs or states will be lost
      - ***Channel Agents***
        - Current transfers will keep on running for short glitches
      - ***VO Agents***
        - Reliance to glitches from minutes to hours
        - Already assigned jobs will process at the normal export rate

Status of the FTS (V)

- **Recovery (II)**
  - Monitoring
    - Poor resilience to glitches
      - Glitches on the server cannot be hidden from clients
  - External and Internal network
    - Poor resilience to glitches
    - Clients will not be able to connect the service
    - Automatic recovery once the network comes back
  - Oracle DB
    - Poor resilience to glitches
    - Assuming full DB recovery, no state will be lost

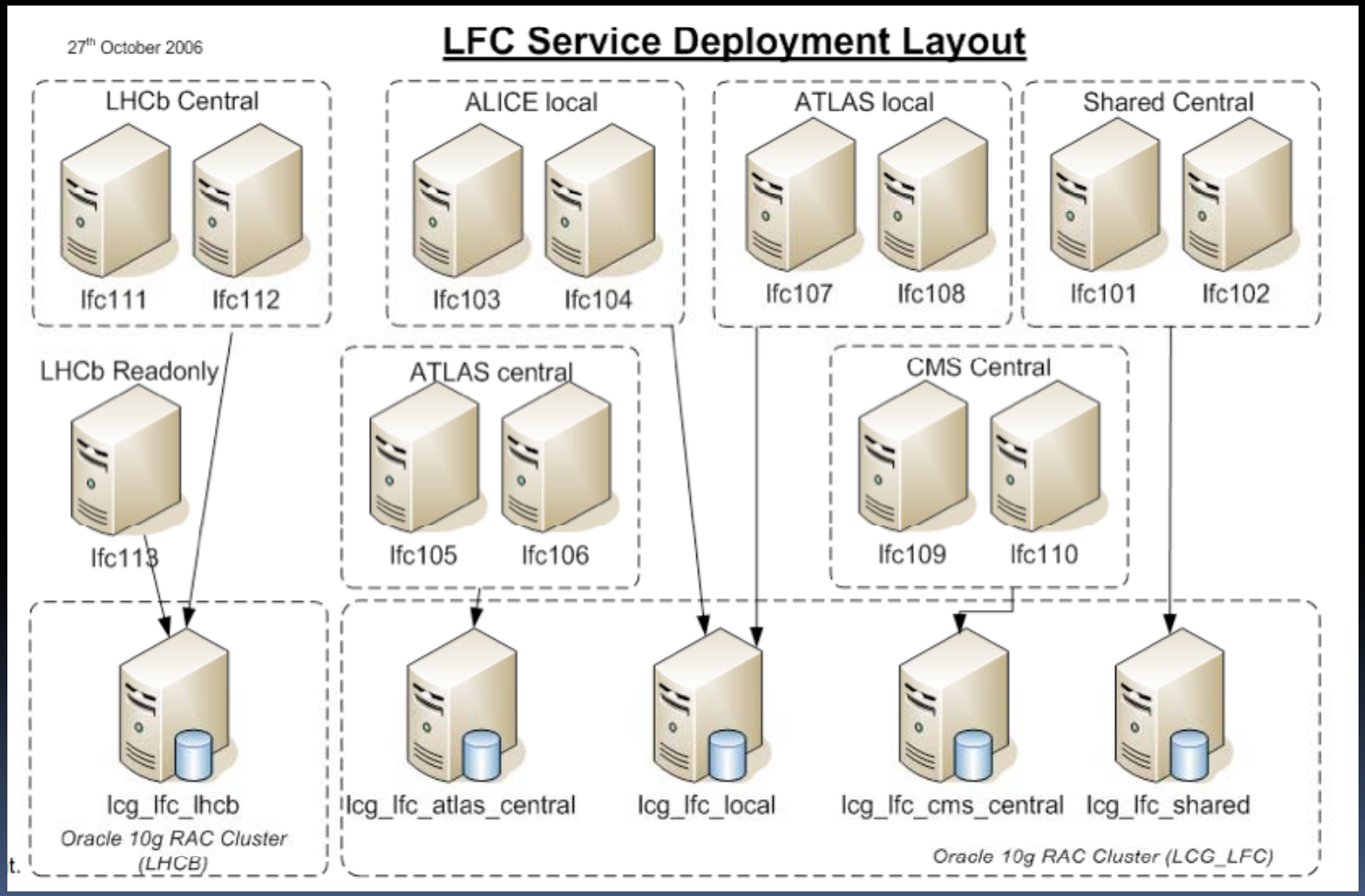
Status of the FTS (VI):

- **Tests and Documentation**
  - Full performance tested in case of patch interventions
  - Zero user-visible downtime
    - Automatic interventions: web-service, Oracle DB, Internal and external networks
    - Manual interventions: fts-transfers, monitoring
  - Totally documented

Status of the LFC (I)

- **Dependencies**
  - LCG Oracle DB
    - Same 10g Oracle RAC for ALICE, ATLAS and CMS
    - Separated for LHCb
- **Hardware**
  - LFC servers placed in the UPS region
    - 10 min of recovery
    - Movement to diesel region before end of 2007
  - Servers DNS load-balanced
  - Different network switches placed in different racks

LFC Layout at CERN



Status of the LFC (II)

- **Software**
  - Middleware
    - Individual mode
      - Recovery mechanism among servers
    - Session (SM) and transaction (TM) modes
      - Connection lost in server fails
      - DB commit of individual commands in SM
      - DB commit at the end of the session in TM
  - Impact
    - Experiments Data Managements affected
    - Jobs submission affected
      - RB connects LFC for matchmaking purposes

Status of the LFC (III)

- **Recovery**
  - Updates shutdown
    - Accepted requests processed until the end
    - DB schema upgrade forces the full service will be down (up to 1h)
    - Middleware upgrade totally transparent for the users
      - While server upgrade, the 2nd one ensures the service
      - Upgrade chain: Certification testbed->PPS->Production
- **Tests and Documentation**
  - Procedures tested and documented

Status of the DB (I)

- **Hardware**
  - DB services placed in diesel region
  - Three service layers
    - Development, pre-production/validation, production
  - Multiple servers doted of redundant power supplies and Oracle load balanced
  - Similar hardware and configuration for validation and production layers
    - 2-node clusters for validation of experiment applications
    - 6-8 node clusters for production redundant configured
  - Two networks available
    - Internal network redundant
    - Public network very stable with max time recovery of 2h

Status of the DB (II)

- **Software**
  - Middleware
    - Not DB, but service based
  - Impact
    - Impact in VOMS, FTS, LFC, Phedex, etc
    - FTS example
      - ***Web-service***, data-transfer and monitoring synchronize their states with Oracle DB cluster
      - Web-service will die
        - Message: “Can`t connect to DB”
      - ***Data-transfer*** stopping on all channels
      - ***Monitoring*** will suffer degradation

Status of the DB (III):

- Recovery
  - Nodes
    - Deployment of new Oracle versions once-twice per year
      - Performed following Oracle recommendations
      - The new version is installed on the validation RAC where it can be tested for a month
  - Data backup
    - Well established infrastructure for tape and disk recovery following the provided strategy of Oracle RMAN 10g
    - Special Oracle features are used to reduce latency and weight of the DB backups
- Tests and Documentation
  - Fully documented

Status of CASTOR (I)

- **Hardware (I)**
  - Complex system broken in three separate areas
    - Central services
    - Disk cache
    - Tape backend
  - Use of DLF (daemon + Oracle RAC DB) for logging messages
  - The system foreseen different front-ends
    - Not all of them load balanced
      - Request handler not yet load-balanced
  - There is a single point of failure

Status of CASTOR (II)

- **Hardware (II)**
  - The RAC DBs are on critical power
    - Used on name servers and stager DBs
    - All components have redundant power supplies
  - Multiple servers, DNS load balanced for name server
    - Planned to extend it to disk cache components
  - Most of the instances head nodes shared a given network switch
    - 3 switches in total for instances headnodes
    - Disk and tape servers spread across multiple switches

Status of CASTOR (III)

Daemon/activity	Description	Critical	Single point of failure
Name server	Oracle RAC and load balanced daemons	YES	NO
Cupv	Access control	YES	
Message deamon			
Vdqm			
Vmgr			
Tapes servers	Worker nodes for tape access	NO	NO
Tape driver	Interface between tapes servers and servers	NO	NO
Tape robots	Tape access for storage	NO	NO
LSF	Per instance scheduler	YES*	YES*
Rtcpclientd	Per instance tape interface	NO	NO
MigHunter	Per instance hunter of to be migrated	NO	NO
Stager	Per instance stager	YES*	YES* **
Request handler	Per instance request handert	YES*	YES* **
rmmaster	Per instance LSF job submitter	YES*	YES* YES**
rmMasterDeamon	Per instance monitoring aggregation	YES*	YES*
diskservers	Disk cache	NO	NO

* Not globally critical but critical for a given instance

** Work undergoing for running multiple deamons

Status of CASTOR (IV)

- **Software**
  - Middleware
    - The lose of disk and tape servers are handled by the software
    - DNS load balanced servers are mandatory
  - Impact
    - Affecting data access/recording
- **Recovery**
  - Procedures available for a clean startup and stop of services
- **Tests and documentation**
  - Most of the software upgrade is not transparent

Summary of the services (I)

	Hardware	Software	Recovery	Tests/Docum
CE/WN	Different power supplies CE fully redundant end 2007	Able to handle short glitches Overall service ensured	Clean pause ensured	Ongoing
CASTOR	Different front-ends Multiple servers DNS load balanced Single points of failures	Able to handle lose of disk and tape servers	Procedures available for a clean start and stop Software upgrade not transparent	YES

Summary of the services (II)

	Hardware	Software	Recovery	Tests/Docum
DB	Diesel region Oracle load balanced Multiple networks	Not DB based	Deployment infrastructure following Oracle setup Tape disk backups fully defined	YES
LFC	DNS load-balanced in UPS region Moving to diesel end 2007	Middleware able to handle load balance in individual mode	Transparent upgrades for software Stop of service in schema upgrades	YES
FTS	Components DNS load-balanced but monitoring Redistribution of internal switches needed	Able to handle loss of servers	Components poor resilience to glitches No state or jobs lost	YES

Summary of the talk

- We have tried to give a general view of the T0 status before the data taking
  - It is a preliminary check - more work needs to be done
  - Provide an homogeneous picture for all services
- **We have to continue including the status of other services**
  - VOMS, myproxy, WMS, etc
- **Check of T1 services foreseen**
- **Workshop foreseen Nov 2007**
- **Should be driven be experiment priorities, i.e. CMS critical services list**