New Hardware Acquisition for Physics DB Services in 2011

Tier1 Service Coordination Meeting

March 17th, 2011

Luca Canali, IT-DB

Background

- Replacement of 2/3 of HW for production Physics Databases in IT
  - RAC5 and RAC6
  - H/W originally deployed in 2008
- H/W selection: desire to have a unified H/W base in IT-DB

Requirements

- Sizing for 2012-2015
  - Based on current production and growth
  - No new major service deployment requested
- Storage Capacity
  - 5-30 TB per DB, ~150TB overall estimated need
- IO Capacity
  - Random reads are the most critical
    - 10K IOPS at peak time
  - Sequential reads important too
    - ~500 MB/s reads
  - Sequential Writes less important
    - ~200 MB/s sequential write

Concentrate on NAS here

- More significant development since previous h/w acquisition
  - SAN technology is “more of the same” with newer disks; no need for 8Gb Fiber Channel.
- Proven record of stability
- Performance
  - 10 GigE connectivity
  - SSD cache to boost IOPS
- Capacity: allow large DBs with SATA disks
- Snapshots
  - ‘Filesystem Snapshots’ to be used as backup against logical corruption
- Cost /IOPS and cost/GB has gone down

NAS and PDB requirements

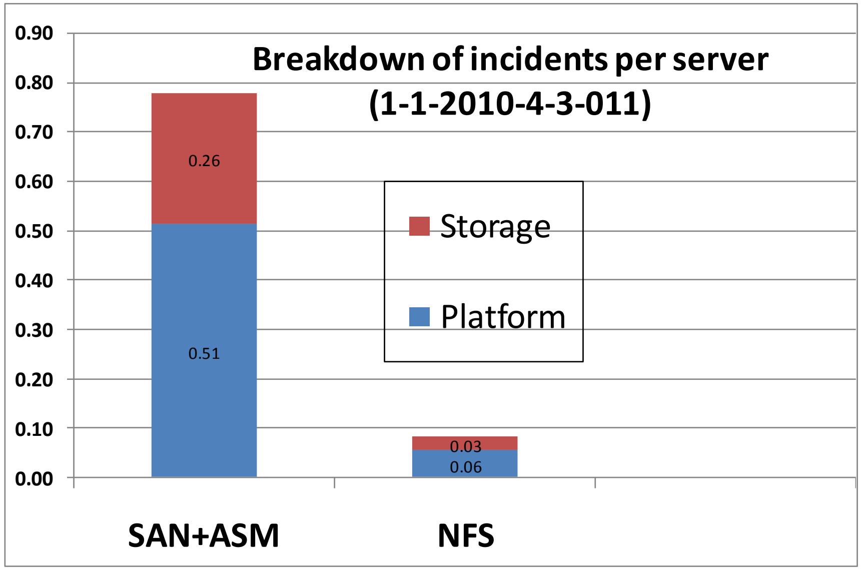
- Sizing criteria
  - Match measured production workload metrics with benchmark on HW characteristics
  - Additional evaluation based on DBAs experience
    - With storage and DB applications behaviour
    - Additional tests performed with experiments (CMS and ATLAS)
- Additional considerations
  - Response time for random IO served by NetApp SSD cache reduced 1 order magnitude
    - Achieved 40K IOPS in testing
  - Otherwise IOPSs scale with N# disks both for SAN and NAS (~100 IOPS per SATA disk)

Where do we lose

- Sequential IO with NAS
  - In particular write throughput limited by RAID DP
  - But limits are above requirements for our production DBs
- Cost: 20% more expensive
  - Total cost, including switches and servers

NAS additional advantages

- **Improved Reliability**
  - Clustered controllers
  - Mature OS and file system
- **No Fiber Channel**
  - No upgrade to 8Gbps
  - No need for FC support
- **Common H/W** across group
  - Further leverage on common procedures, e.g. Installation and monitoring



Sizing the replacement of RAC5+6

- Enough HW to run critical production that is currently on RAC5+6+7
  - Will allow to increase performance with new HW
  - Will allow move to new HW using standby DB failover
  - Will allow migration out of RAC7 in 2013
  - RAC7 will still be used for ‘low load’ DBs in his last year of life
  - Results of the running of replacement of RAC5+6 can be used to size the replacement of RAC7

Replacement of RAC5+6 with NetApp storage - proposal

- 2 setups
  - Each with 3 clusters of NetApp FAS3240
  - 6 disk shelves (of 24 disks) per cluster
  - Each controller 512GB SSD (PAM module)
  - Total: 864 disks (1TB SATA)
- In particular for 8 main prod DBs:
  - Proposed to be moved to 4x cluster
    - Isolated cluster to provide isolation
    - 144 disks per cluster -> total of 576 disks on NAS
  - currently on 1004 disks

Servers

- Evolution from current production
- In particular add more memory (48 GB)
  - Beneficial for random IOPS
- Take advantage of 10GigE
  - For storage access
  - For faster backup and restore (10Gig TSM)
- CPUs
  - Move to Westmere
  - 2xquad cores with higher frequency
- 2 setups of 24 Servers

Conclusions and Plan

- Buy NetApp based storage and new servers for Physics DB Services in 2011
  - Tier1s or online DBs will continue to have IT-DB support on SAN
  - RAC7, standby and integration on RAC9 will stay on SAN+ASM
- Next steps
  - Configure and order h/w (storage and servers)
  - Prepare installation plan with CF
  - Integrate H/W migration plan with 11.2 upgrade
    - Testing and planning activity in 2011
    - Changes to be performed during technical stop