ASGC DB Services

- setup and configuration

Jason Shih
ASGC/OPS

Nov 12th, 2008
Distributed Database Operations Workshop

Academia Sinica Grid Computing

Outline

- Architecture and configurations
  - Hardware (server, storage, SAN)
  - Monitoring
  - Backup
- Applications
  - Grid services: CASTOR, LFC, FTS, SRM, 3D
- Performances & license statistics
- Remarks

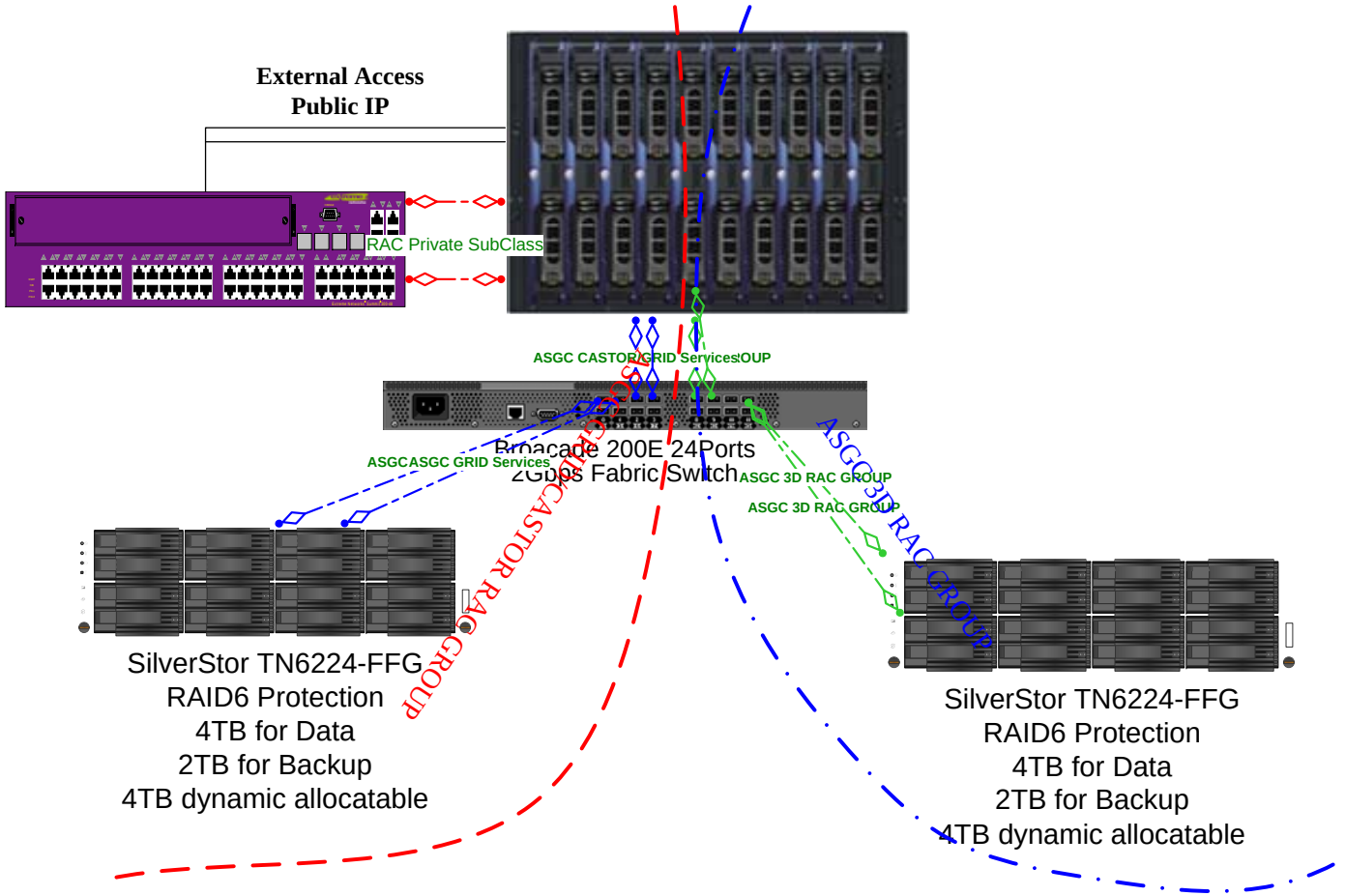
Database Services and Setup

- OS & File System
  - Oracle Unbreakable Linux 4
    - kernel: 2.6.9-42.0.0.0.1.ELsm
  - OCFS2 (*oracle home on local FS*)
- DB Engine
  - Oracle 10g RAC release 10.2.0.3.0
  - 6 nodes serving 3 databases (srmdb, gdsdb, castordb)
  - 2 nodes serving 3D (asgc3d)
- Monitoring
  - Oracle Enterprise Manager
- Backup Tools:
  - RMAM

Hardware Profile

- SAN Storage:
  - Fabric switch:
    - Brocade SAN switch E200
  - Raid Subsystem:
    - Silverstor TN-6224S-FFG RAID 6
    - 4TB for Data
    - 2TB for Backup
    - Free space that can be dynamically allocated: 4TB
- Servers
  - Quanta Blade System run EM64T
  - SMP Intel Xeon 3.0GHz
  - ECC 8GB Physical Memory
- The same profile also apply to:
  - CASTOR Services (Stager, NS, DLF, VDQM etc.)
  - Grid Services (LFC, FTS)
  - Streaming (3D)

Oracle - setup

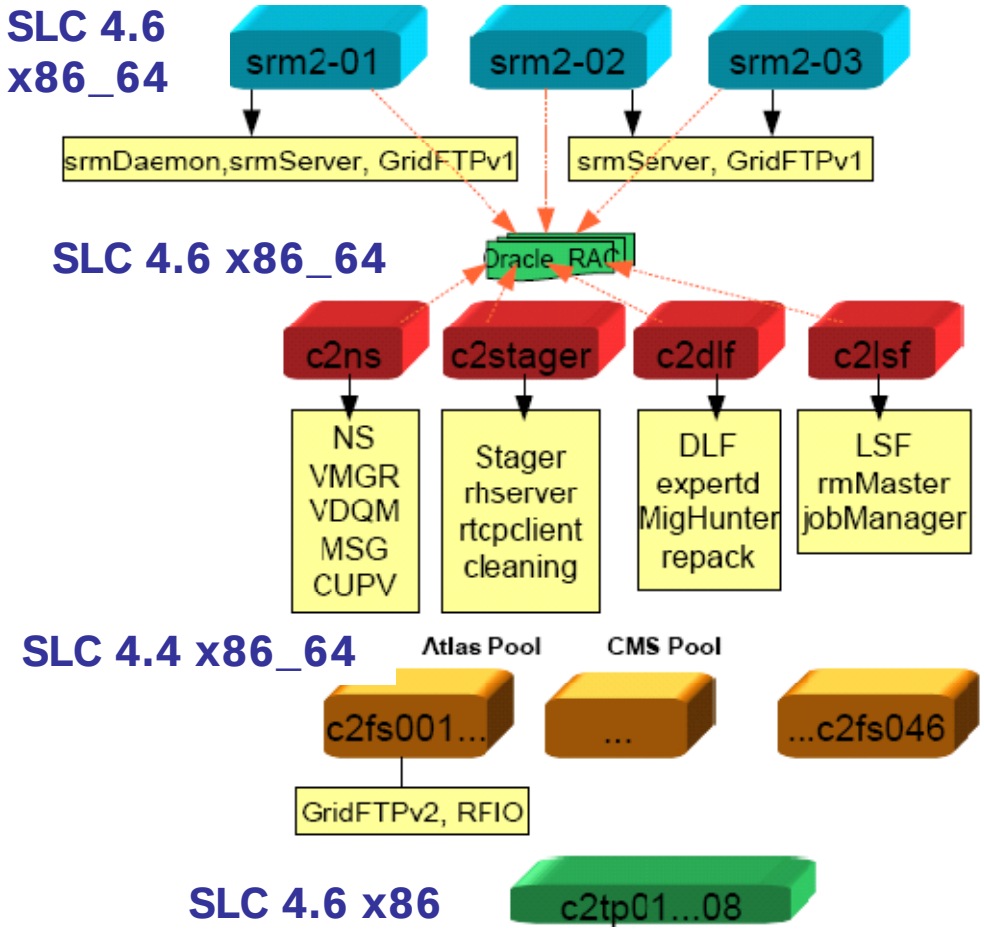


Applications

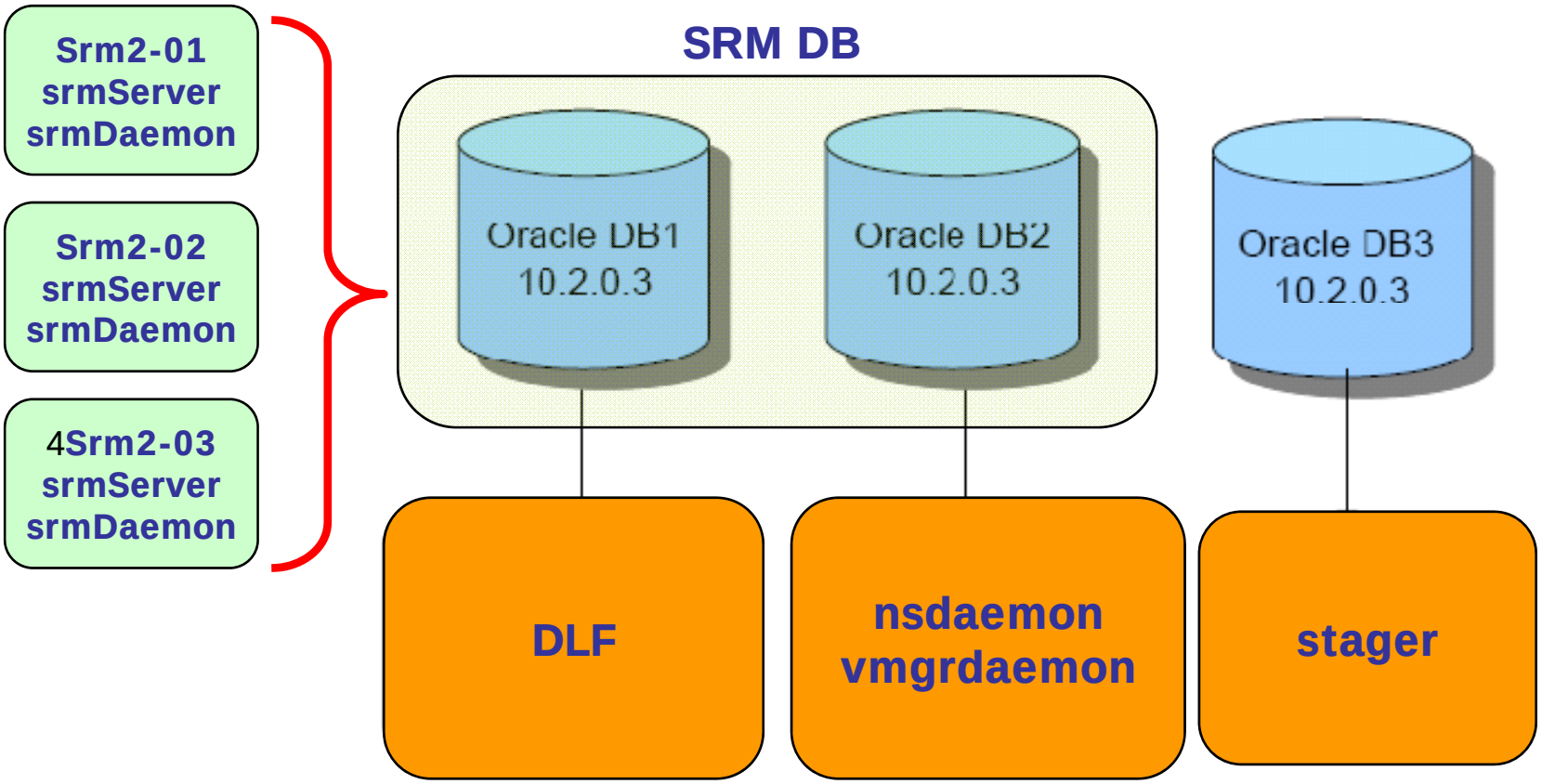
- 3D (asgc3d)
  - 2 instances
- CASTOR (castordb)
  - 3 instances serving services: DLF, NS and stager
- SRM (srmdb): *2 instances*
- LFC/FTS (gdsdb): *3 instances*

Select	Name △	Status	Alerts	Compliance Score (%)	CPU Util %	Mem Util %	Total IO/sec
◉	w-rac01.grid.sinica.edu.tw	⇧	2 17	63	12.31	93.48	170.6
○	w-rac02.grid.sinica.edu.tw	⇧	1 18	63	37.16	99.42 ⚠	1549.87
○	w-rac03.grid.sinica.edu.tw	⇧	2 6	63	99.99 ✖	87.29	1598.21
○	w-rac04.grid.sinica.edu.tw	⇧	2 7	63	100 ✖	99.57 ⚠	749.88
○	w-rac05.grid.sinica.edu.tw	⇧	2 3	63	99.99 ✖	86.54	170.43
○	w-rac06.grid.sinica.edu.tw	⇧	0 11	63	4.45	92.03	147.43

Application: *CASTOR (I) - services*

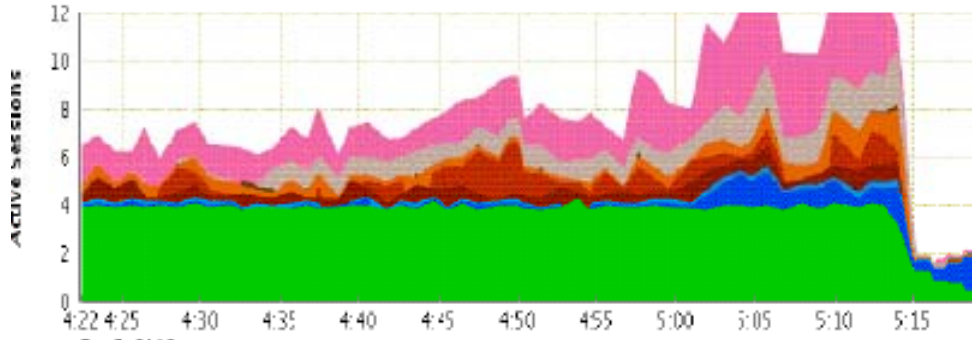


Application: *CASTOR (II) - RAC*



Application – CASTOR (III) - *performance*

- statistics collection - *twice per day*
  - Before collecting statistics:
    - cpu load > 95% & load avg. ~ 26
  - Recalculate statistics:
    - Cpu load < 10% & load avg. < 3

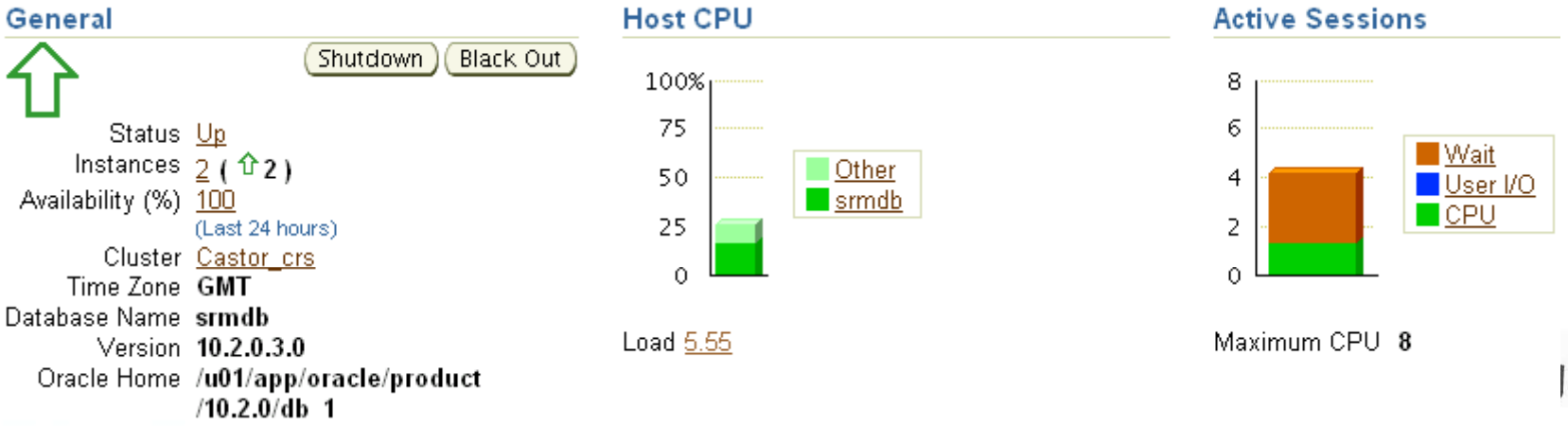


Backup Policy

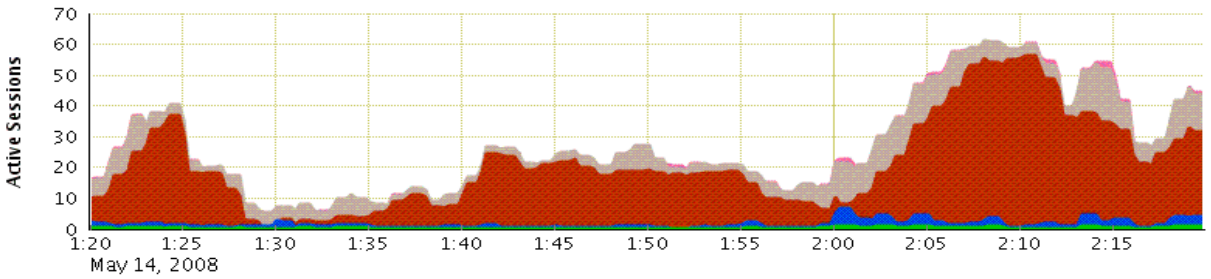
- Incremental backup
  - incremental level=0 (Mon 0:00)
  - Differential incremental level=1 (every week day)
- Daily backup via cron job
  - Customized script
    - alternative: RMAN GUI
- E-mail notification
  - To all DB OPS list
  - status report inc:
    - backup status
    - restore verification testing (every Sat.)
    - delete obsolete backups (every Sat.)
- Retention policy
  - keep 1 full backups each week for 3 weeks

Monitoring

- Nagios probes:
  - Dummy login check for all RAC nodes
  - Oracle deadlocks (per 20min)
    - Alarm trigger if session lock > 10min.
  - Generic NRPE host plugins (CPU load, cache, swap)
- Grid Control
  - Castordb, srmdb, gdsdb



Operation events:



- SRM DB deadlock
  - ERROR: CGSI-gSOAP: Error reading token data
    - From Exp scope: *Too many threads busy with Castor at the moment*
  - Workaround:
    - S2 DB patch provide by SRM dev.
  - Prevention:
    - Plug-in & SMS alarm
    - Increasing SOAP backlog in S2 config.
    - Increasing sessions numbers
- disk copy stuck in “WAITDISK2DISK” state
  - Force flushing pending request more than 1k sec help resuming all pending staging request.
  - Impact also found for CMS transfers during CCRC
    - Data transfers will stage from production disk pool to wanout pool
    - Manual fix able to resume the data transfers

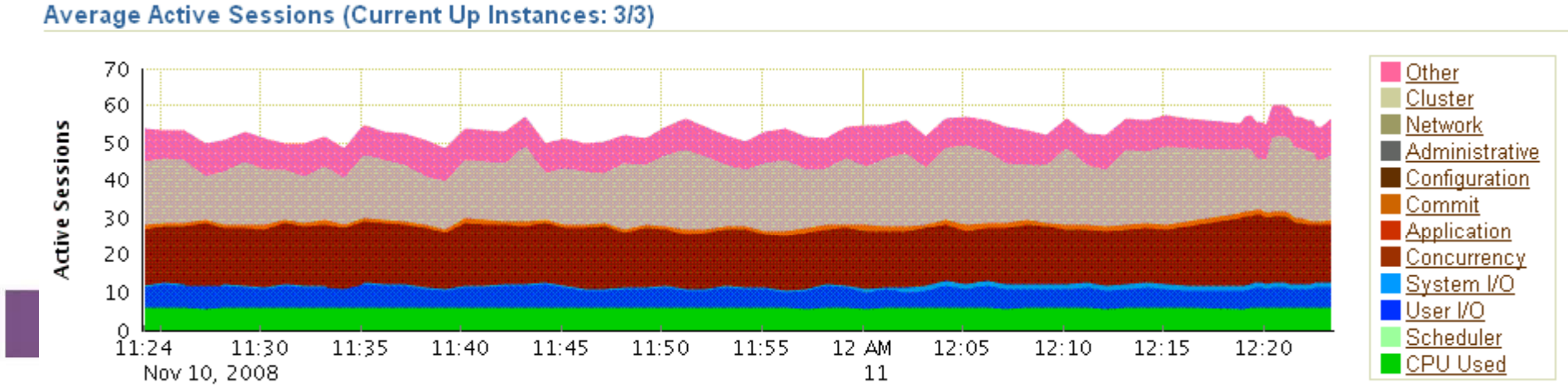
Complete Actions – Q1-2

- LFC
  - Before:
    - LFC query stuck when > 1K files in the directory
      - Known LFC issue
      - Restarting MySQL helps solves the problem
  - Migration from MySQL to Oracle – earlier of Feb
- CASTOR
  - Add one RAC nodes – mid of Mar
  - RAC hardware migration (backend storage)
- FTS
  - Migrate from single DB to RAC – Apr
  - In parallel: *Add 2 WS frontend and FTS upgrade to 2.0*

Services - Nodes/Util/Load

Services	nodes	Avg. CPU Util (%)/load	Lost 1 node CPU Util (%)
asgc3d	2	10%	20~25%
castordb	3	25%/3.0	40~45%
srmdb	2	25%/3.8	-
gdsdb	3	50%/4.5	70~100%

SRMDB



Oracle Licenses stats

Service	Nodes	CPU(s)	2yr(CPU)	1yr(CPU)
asgc3d	2	4	2	0
castordb	3	6	2	2
gdsdb	3	6	6	2
backup	1	1	-	-
OMS	1	0	-	-
Total	10	17	11(10+1)	5(4+1)

- 2 yr estimate: we assume that we can accept a load of 7

- Approx. double DB load if disk capacity (utilization) increase from 1.2PB to 2.4PB in 2 years

- expect FTS requests in the next two years to increase 3 times - increase transfer requests from T2, increase transfer rates

Future remarks

- SPOF:
  - Chassis management blade (SOL + RPM)
  - Dual controller of raid subsystem
  - Fabric:
    - Dual port FC cards + two SAN switches
- Database management
  - Disaster recovery
  - Limited trouble shooting experiences
  - Need production DB administration (hire DBA)