

Oracle LCG3D

Pierre-Etienne Macchi
Jean-René Rouet

Plan

▶ LCG3D

- ▶ Qu'est-ce que c'est ?
- ▶ Comment ça marche ?
- ▶ A quoi ça sert ?
- ▶ Infrastructure technique
- ▶ Prévisions

▶ Oracle au CC

- ▶ Usages
- ▶ Remarques
- ▶ Infrastructure technique
- ▶ Prévisions

LCG3D

LCG3D

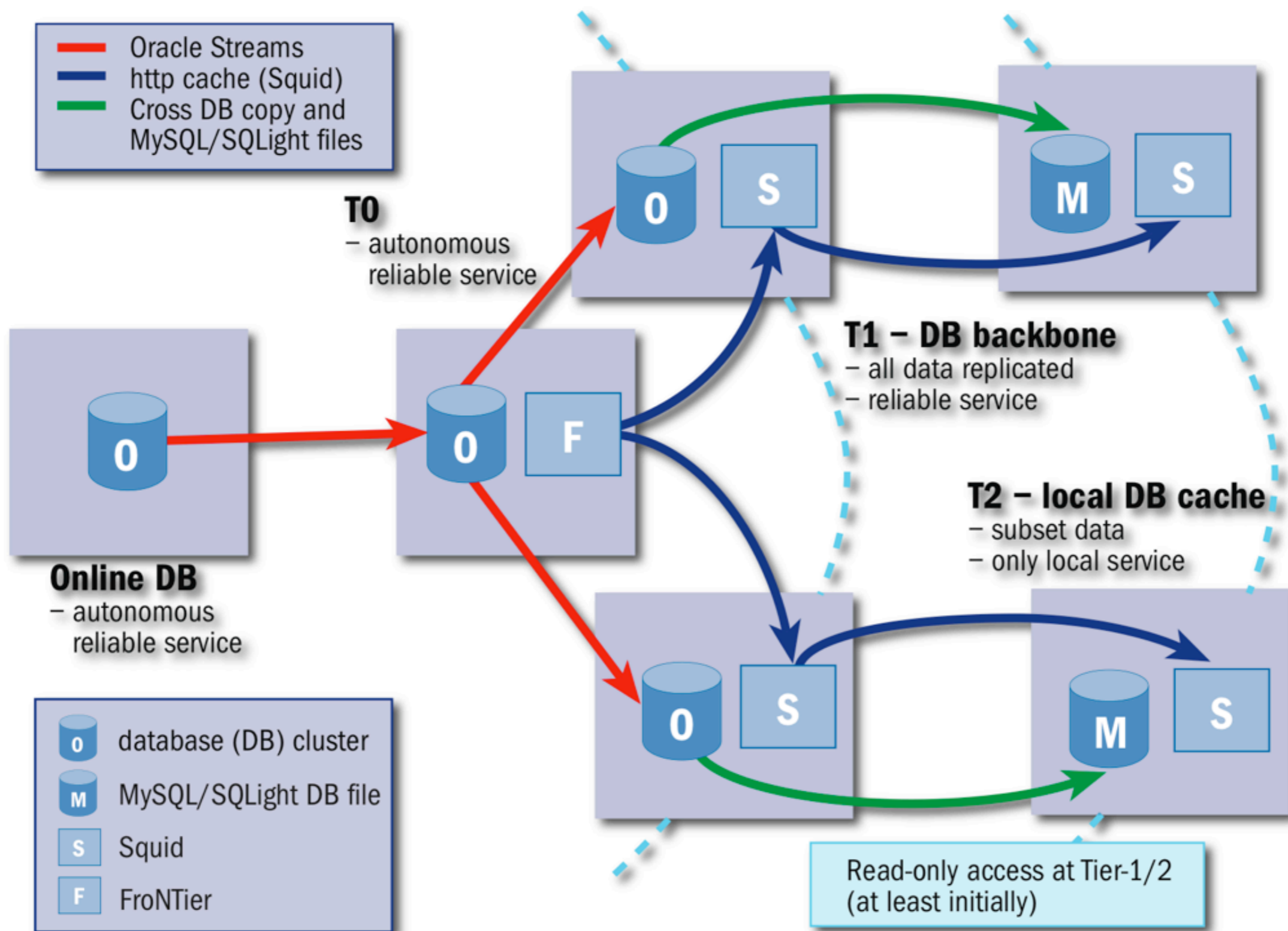
- ▶ 3D : distributed databases deployment
- ▶ Projet de réplication des base de données depuis le CERN vers les TIER I
 - ▶ basé sur les technologies Oracle
 - ▶ concerne les bases de données relatives aux données d'expérience
 - ▶ COOL* + TAGS** (ATLAS)
 - ▶ COOL* + LFC*** (LHCB)

* : conditions database

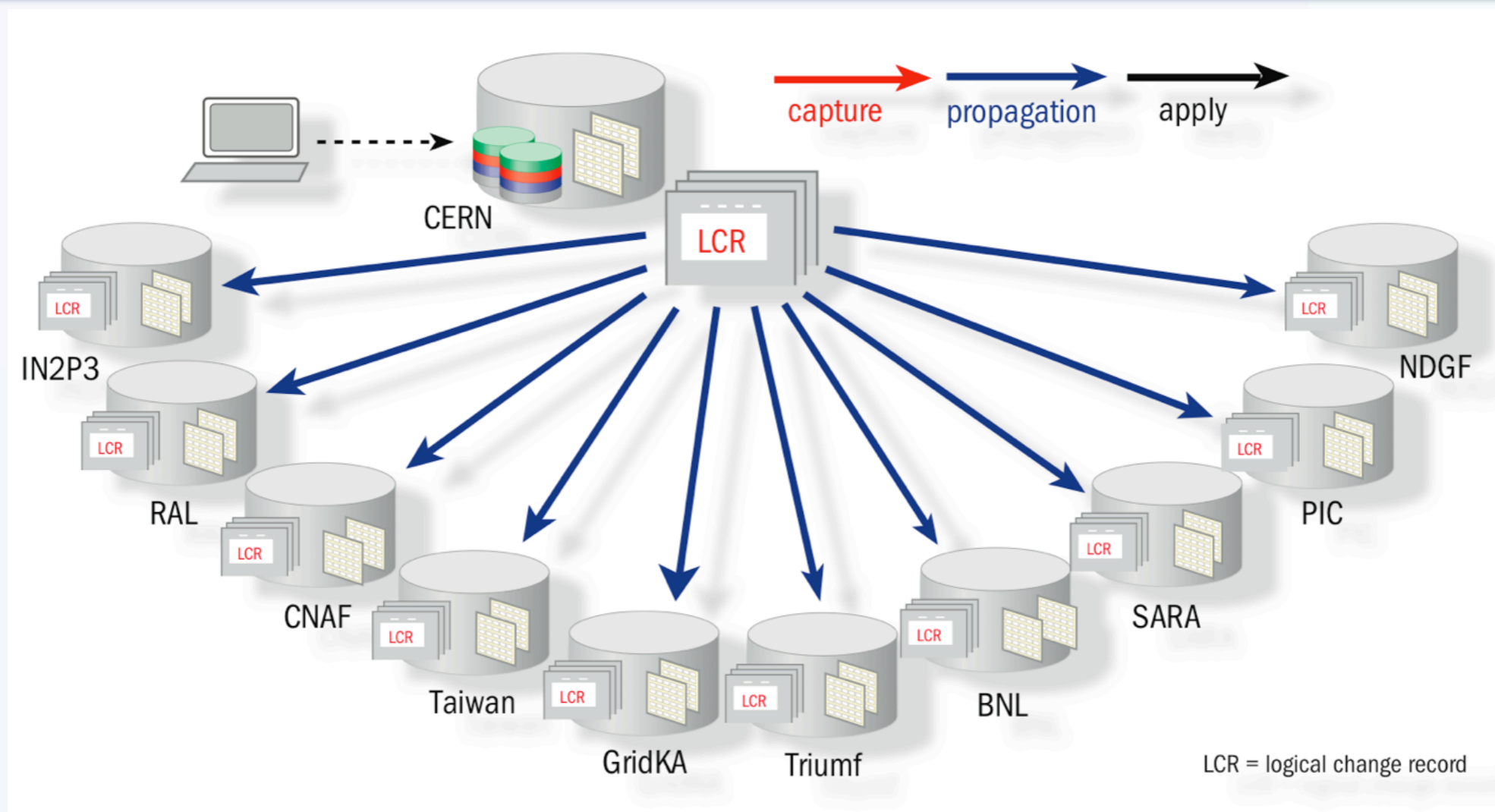
** : events metadata

*** : LHC file catalog

Architecture du service 3D



Oracle Streams



- les modifications sur la base sont capturées dans les fichiers redo log et propagées de manière asynchrone sous forme de LCRs (Logical Change Records)
- Tous les changements sont mis en queue jusqu'à leurs applications réussies sur tous les sites

Supervision

► Interface web de supervision

ACTIVE STREAMS										
Stream	LCR Cap	LCR Enq	LCRs Prop	LCRs Deq	LCRs App	Latency	Capture State	Propagation State	Apply State	State
ATLDSC.CERN.CH=>ASGC3D.GRID.SINICA.EDU.TW	2313.28 /s	1650.41 /s	1805.07 /s	.15 /s	0 /s	0 sec	CAPTURING CHANGES	ENABLED	IDLE	
ATLDSC.CERN.CH=>ORCL.BNL.GOV	2313.28 /s	1650.41 /s	4252.7 /s	3161.52 /s	2130.36 /s	45 sec	CAPTURING CHANGES	ENABLED	APPLYING	
ATLDSC.CERN.CH=>CONDATLA.CR.CNAF.INFN.IT	2313.28 /s	1650.41 /s	2763.74 /s	2749.28 /s	2056.7 /s	49 sec	CAPTURING CHANGES	ENABLED	APPLYING	
ATLDSC.CERN.CH=>LCGDB1.GRIDKA.DE	2313.28 /s	1650.41 /s	1073.53 /s	1056.61 /s	350.54 /s	59 sec	CAPTURING CHANGES	ENABLED	APPLYING	
ATLDSC.CERN.CH=>ATLAS.DB1TIER1.NDGF.ORG	2313.28 /s	1650.41 /s	4227.48 /s	2199.5 /s	2879.55 /s	51 sec	CAPTURING CHANGES	ENABLED	APPLYING	
ATLDSC.CERN.CH=>DBATL.IN2P3.FR	2313.28 /s	1650.41 /s	4252.75 /s	2621.26 /s	1011.26 /s	42 sec	CAPTURING CHANGES	ENABLED	APPLYING	
ATLDSC.CERN.CH=>ATLASPIC.PIC.ES	2313.28 /s	1650.41 /s	1982.31 /s	1561.65 /s	1062.02 /s	47 sec	CAPTURING CHANGES	ENABLED	IDLE	
ATLDSC.CERN.CH=>OGMA.GRIDPP.RL.AC.UK	2313.28 /s	1650.41 /s	1036.08 /s	933.63 /s	685.68 /s	16 min 46 sec	CAPTURING CHANGES	ENABLED	APPLYING	
ATLDSC.CERN.CH=>SARADB.GRID.SARA.NL	2313.28 /s	1650.41 /s	1712.64 /s	1275.68 /s	623.31 /s	48 sec	CAPTURING CHANGES	ENABLED	APPLYING	
ATLDSC.CERN.CH=>TRAC.TRIUMF.CA	2313.28 /s	1650.41 /s	2521.51 /s	1840.88 /s	969.19 /s	45 sec	CAPTURING CHANGES	ENABLED	APPLYING	
ATONR.CERN.CH=>ATLR.CERN.CH	137.74 /s	0 /s	0 /s	0 /s	0 /s	0 sec	CAPTURING CHANGES	ENABLED	IDLE	
ATONR.CERN.CH=>ATLR.CERN.CH	1038.51 /s	928.2 /s	912.22 /s	1071.37 /s	976.69 /s	3 sec	CAPTURING CHANGES	ENABLED	APPLYING	

- ▶ Les travaux de calcul accèdent aux bases de données répliquées dans leur tier I
- ▶ le service doit être permanent dans chaque tier I
- ▶ Le CERN n'absorbe que la charge des travaux s'exécutant au CERN

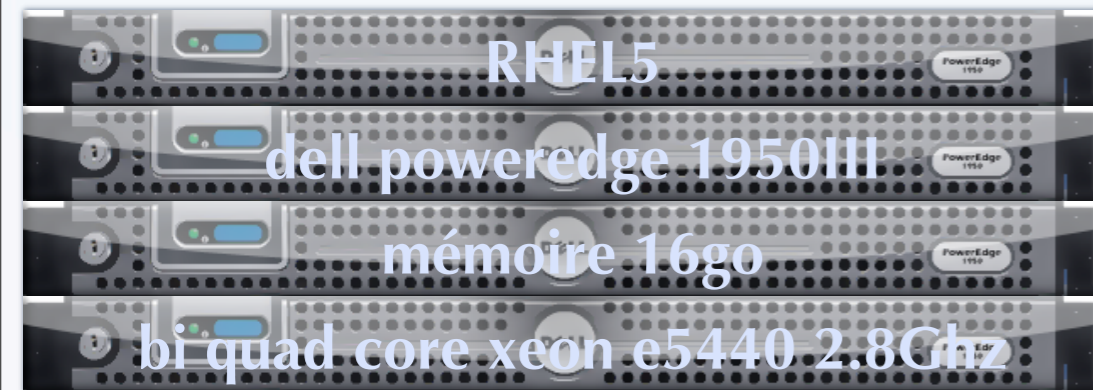
LCG 3D au CC



double attachement FC



double attachement FC



ATLAS



LHCB

Prévisions LCG3D

- ▶ ATLAS (actuellement 225Go)
 - ▶ année nominale 6,9To (TAGS + COOL)
- ▶ LHCB (actuellement 13Go)
 - ▶ année nominale 700Go (COOL + LFC)
- ▶ 2014 : 45To
- ▶ Achat d'un SAN fin 2008 de plus grande capacité
- ▶ Questions ?

Oracle

Usage

- ▶ cluster mutualisé
 - ▶ bases de données mutualisées
 - ▶ développement (35Go)
 - ▶ production (630Go)
 - ▶ bases de données dédiées
 - ▶ services du centre (2Go)
 - ▶ cmstracker (44Go)
 - ▶ edms (12Go)
 - ▶ services de grille (98Go)

Usages (2)

▶ Cluster dédiés

▶ OPERA

▶ Atlas

▶ LHCB

▶ (bientôt SRB)

Usage (3)

▶ Quelques gros utilisateurs en volume

- ▶ Antares (180Go)
- ▶ Griflemont (80Go)
- ▶ Dcache stats (45Go)
- ▶ Bioemergence (30Go)
- ▶ SRB (25Go)

Usage (4)

▶ Quelques gros utilisateurs en cpu

▶ antares

▶ srb

Remarques

▶ Performances

- ▶ gros volume + sql non optimal = très mauvaises performances

▶ rangement des données

- ▶ nécessité pour les gros volumes
- ▶ il faut connaître le volume prévu
- ▶ les réorganisations sont possibles mais très longues

Remarques (2)

▶ sql optimisé

- ▶ c'est le meilleur gain qu'on puisse attendre

- ▶ attention au code sql généré automatiquement par une couche logicielle

▶ batch contre interactif

- ▶ des temps d'attente acceptables en batch

- ▶ ne le sont plus en interactif

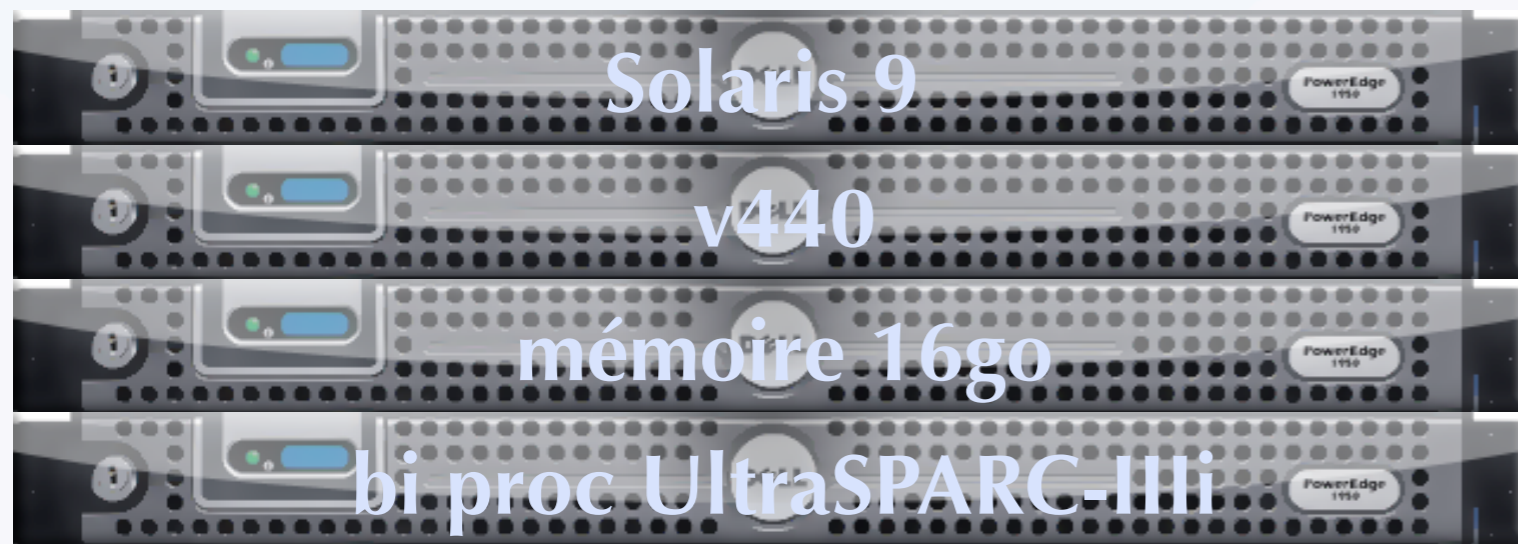
Infrastructure



double attachement FC

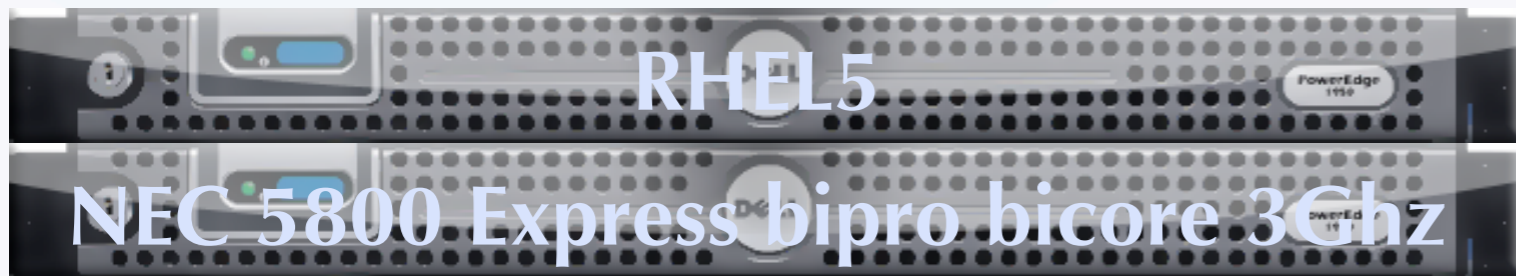
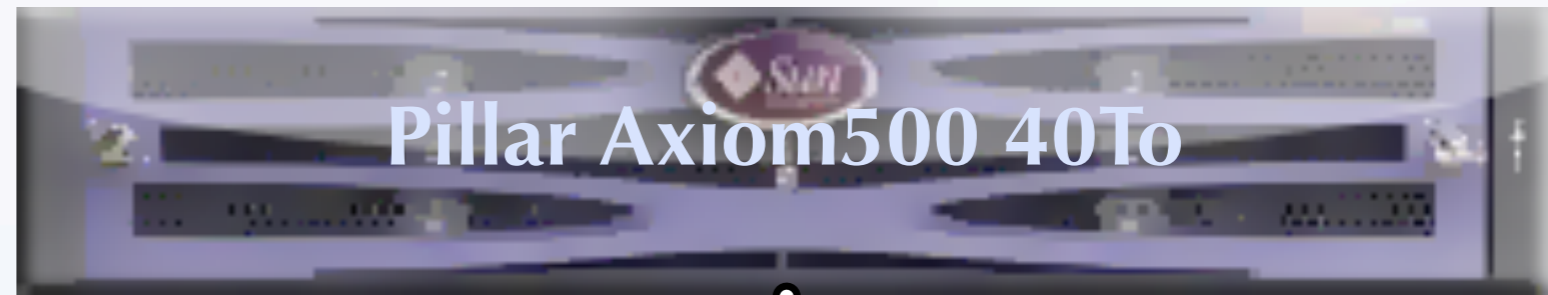


double attachement FC



Cluster mutualisé

Infrastructure (2)



Opera

Prévisions

- ▶ Monitoring plus fin des usages
 - ▶ par utilisateur
 - ▶ par profil
- ▶ Création d'un cluster pour SRB (Irods)
- ▶ Augmentation du volume de disque
 - ▶ passage à 2To.
- ▶ Questions ?

Conclusion

- ▶ ~2 ETP au CC pour toutes les bases de données
- ▶ Une partie des graphiques sont extraits de la présentation de Dirk Duellman (CERN) : [Production Experience with Distributed Deployment of Databases for the LHC Computing Grid \(http://indico.cern.ch/contributionDisplay.py?contribId=171&sessionId=24&confId=3580\)](http://indico.cern.ch/contributionDisplay.py?contribId=171&sessionId=24&confId=3580)