DSI - CNRS

ownCoRe - Stockage Scality

Plan de la présentation

Plan de la présentation

Contexte projet

Besoin fonctionnel

- Service de partage et de synchronisation de fichiers de travail
- Offre anti-Dropbox sur un cloud souverain CNRS
- Cible de 100000 utilisateurs potentiels (agents des unités CNRS), avec un quota de 10 Go par utilisateur

Choix d'une solution applicative

- ownCloud, après diverses études

Contexte projet

Besoin technique

- Un SGBD hautement disponible
- Un système de fichiers adapté
  - avec une forte résilience, car pas de "sauvegarde",
  - unique pour toute la volumétrie,
  - capable d'être multi sites.

Plan de la présentation

Tour d'horizon

Offre riche et variée

- HDFS
- CephFS
- GlusterFS
- iRODS
- et de nombreuses solutions matérielles

Solutions préssenties

3 produits intéressants

- Dell Compellent : Le plus économique
  - Volumétrie limitée à 2Po
  - Résilience limitée
- EMC Isilon : Le plug and play
  - Matériel spécifique
  - Réseau de réplication dédié
- Scality : Le plus souple

Plan de la présentation

Eléments clés

Le RING

- Stockage logiciel d'objets
- Capacité illimitée (scale-out)
- Stockage mutualisé
- ARC

Les points forts

- Compatible tout serveur x86
- Ratio brut/utile d'environ 1.6
- Très hautement disponible
- Pas de RAID matériel

Rôles et architecture

3 types de serveurs

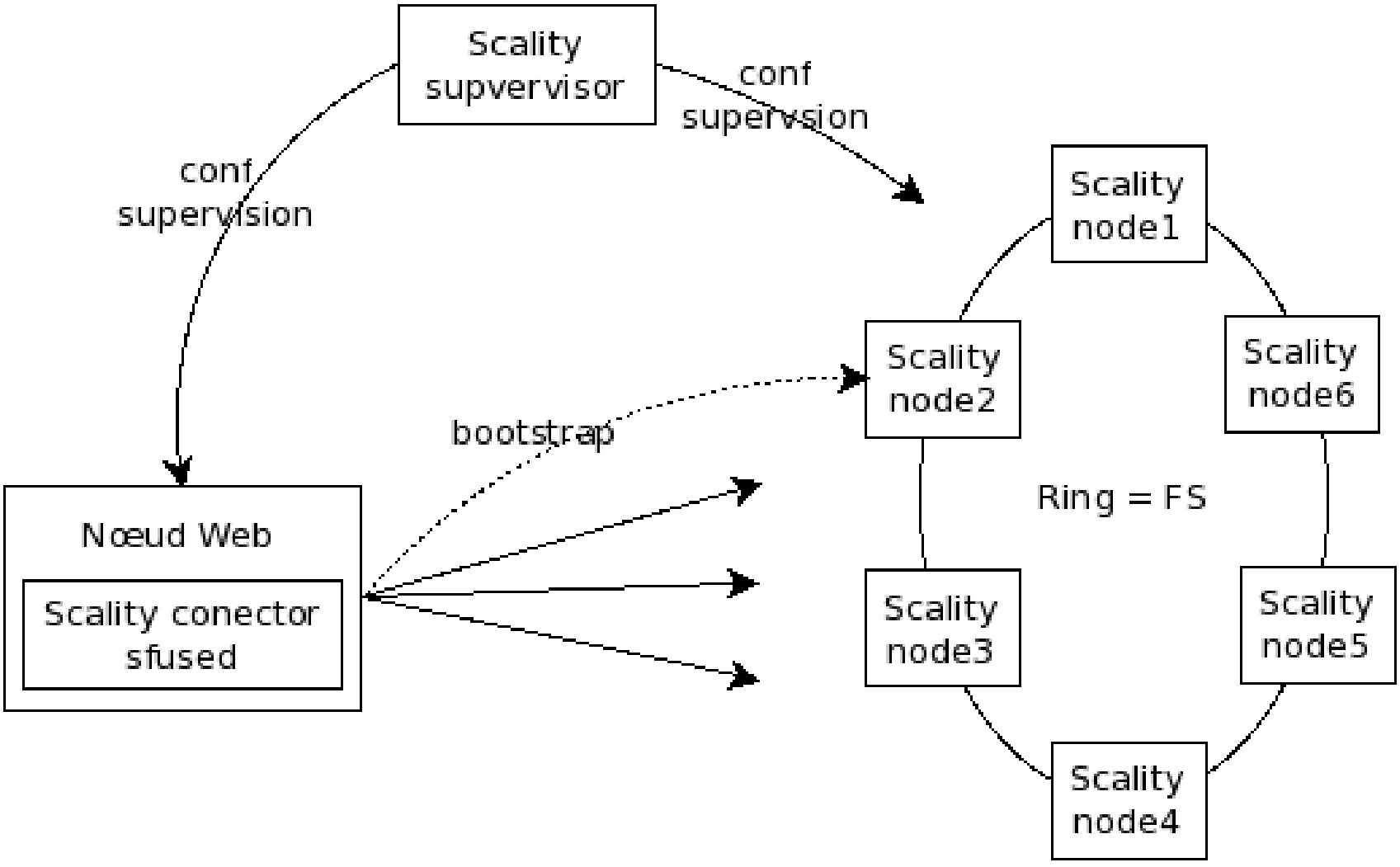
- Scality-node : Nœud de stockage
- Scality-supervisor : Manager pour la conf. et le monitoring
- Scality-connector : Accès client sur le stockage

Architectures possibles

- Au minimum 6 serveurs, pas de maximum
- Un ou plusieurs RING
- Géo-distribution ou réplication

On peut mixer toutes ces possibilités :-)

Schéma de principe



Maquettage avec support Scality

Réalisation d'une maquette ...

- Voir le logiciel en action
- Tester le support Scality

... la plus minimale possible

- Deux VM pour les nœuds de stockage
- Une VM pour le supervisor
- Deux VM clientes LAMP

Classe de Service

Le fichier sfused.conf

- Node de démarrage
- Nombre de copies distribuées
  - Par type de fichiers
  - Par expressions régulières
  - Par connector (donc par application)
- Niveau de mise en cache

```
{
   "general": {
      "mountpoint": "/owncore-ring",
      "file_cos": 2,
      "cat_cos": 4,
      "cat_page_cos": 2,
      "dir_cos": 4,
      "dir_page_cos": 2,
      "rootfs_cos": 4,
      [...]
   },
   "ring_driver:0" : {
      "type": "chord",
      "bstraplist": "192.168.55.21:4244,192.168.55.22:4244",
      [...]
   },
   "cache:0" : {
      "ring_driver": 0,
      "size": 2000000000,
      [...]
   },
}
```

Commandes système

Intégré au système

- Dépot RPM utilisable avec YUM
- Gestion des démons standards
- Le RPM RingSH permet d'utiliser le supervisor en CLI !
- Scripts de déploiement Salt (Puppet like)

Exemple : Initialiser un nœud de stockage

yum install scality-node -y
scality-node-config –prefix /mnt/disk –disks 2 –nodes 6 –ip <ip_scality-node>–supervisor-ip <ip_scality-supervisor>

Le Supervisor

Couple WebUI et sagentd

- Ajouter/enlever des nodes à chaud
- Etendre le RING à chaud
- Remonter les incidents matériels
- Tableau de bord
- Métrologie

SCALITY

Supervisor Administration

Logged in as root | Hardware | Logout

Local > owncore

Dashboard | Operation | Administration | Provisioning

Status

State	RUN	
Autojoin	Off	
Online	12	Nodes
	0	RS2 Connectors
Alerts	0	

Storage capacity

# Objects		
# Unique objects		1
Average size		0.00 KB
Avg size (unique)		0.00 KB
Unique	0.00 %	0 GB
Stored	0.00 %	0 GB
Used	1.84 %	0 GB
Available	98.16 %	20.00 GB
Total		21.00 GB

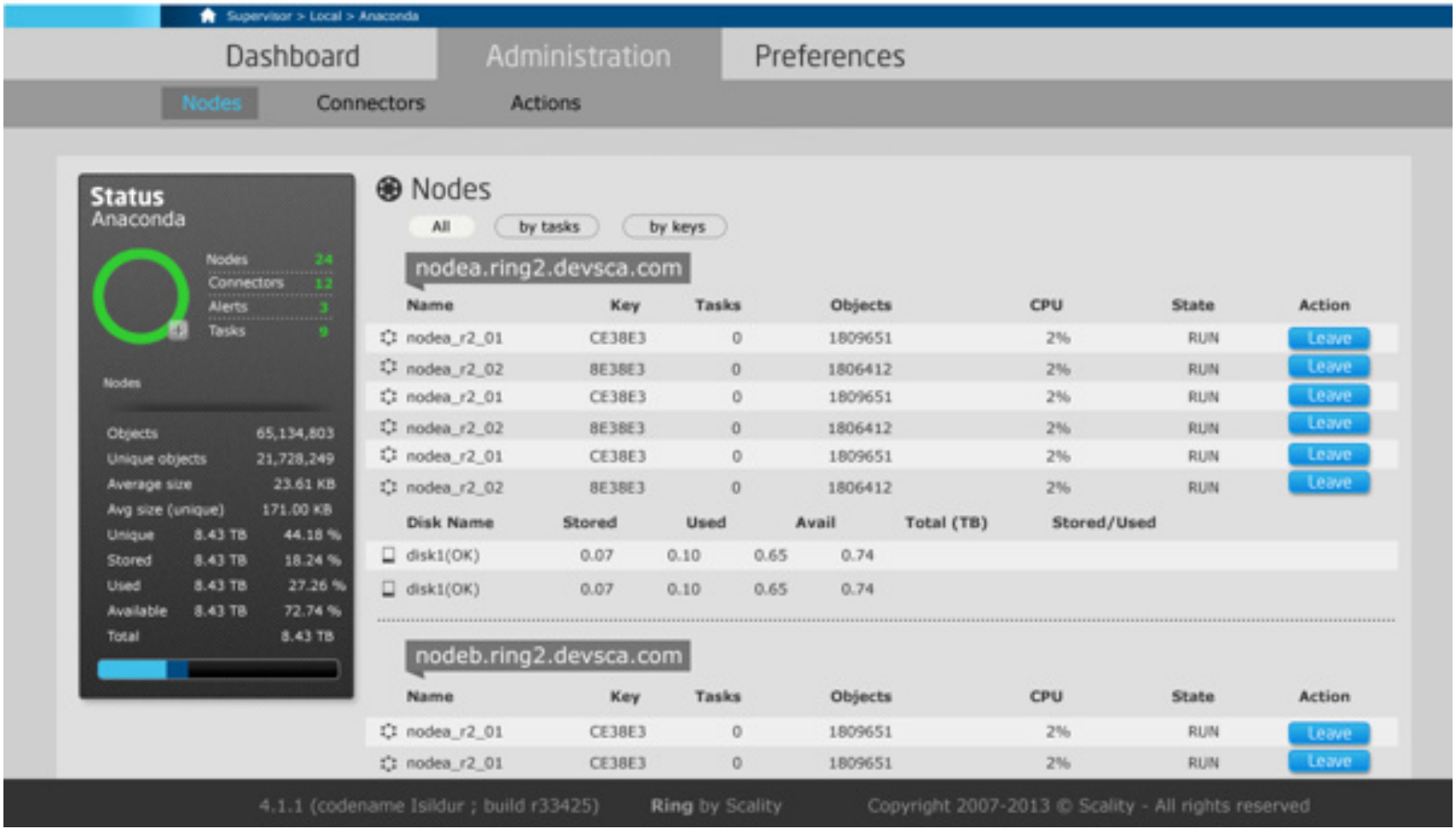
Welcome to the provisioning wizard, step one: choose the servers you want to configure.

	Server	Zone	State	# nodes
☑	192.168.55.21		RUN	6
☑	192.168.55.22		RUN	6

Warning! We recommend having at least 6 servers.

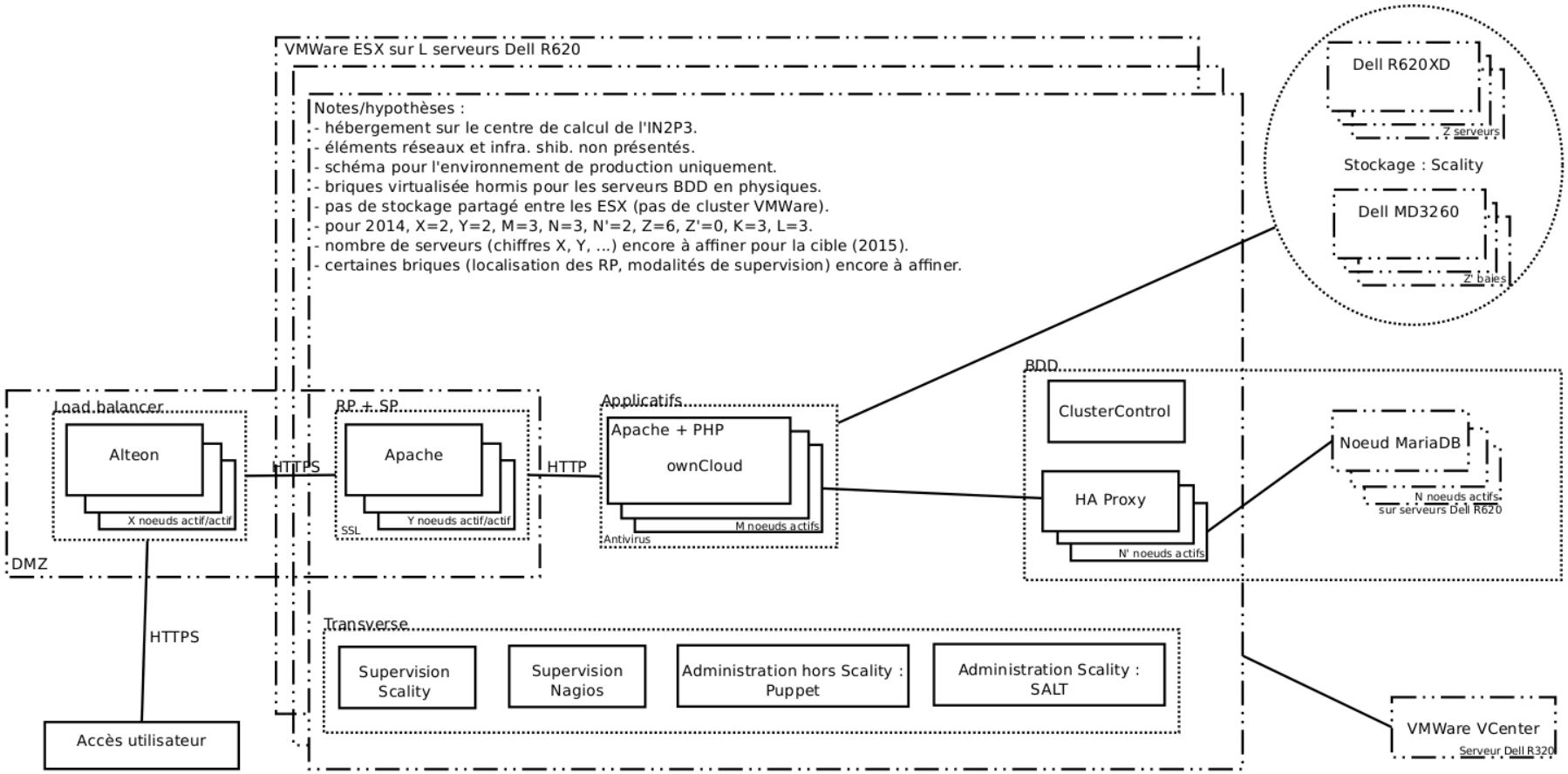
Configure

[Manual provisioning]


Ring
Administration Interface
Log off
Welcome, Joe Madureira
Supervisor > Local > Anaconda
Dashboard
Administration
Preferences
Nodes
Connectors
Actions
Status
Anaconda
Nodes 24
Connectors 12
Alerts 3
Tasks 9
Nodes
Objects 65,134,803
Unique objects 21,728,249
Average size 23.61 KB
Avg size (unique) 171.00 KB
Unique 8.43 TB 44.18 %
Stored 8.43 TB 18.24 %
Used 8.43 TB 27.26 %
Available 8.43 TB 72.74 %
Total 8.43 TB
Nodes
All
by tasks
by keys
nodea.ring2.devsca.com
Name Key Tasks Objects CPU State Action
nodea_r2_01 CE38E3 0 1809651 2% RUN Leave
nodea_r2_02 8E38E3 0 1806412 2% RUN Leave
nodea_r2_01 CE38E3 0 1809651 2% RUN Leave
nodea_r2_02 8E38E3 0 1806412 2% RUN Leave
nodea_r2_01 CE38E3 0 1809651 2% RUN Leave
nodea_r2_02 8E38E3 0 1806412 2% RUN Leave
Disk Name Stored Used Avail Total (TB) Stored/Used
disk1(OK) 0.07 0.10 0.65 0.74
disk1(OK) 0.07 0.10 0.65 0.74
nodeb.ring2.devsca.com
Name Key Tasks Objects CPU State Action
nodea_r2_01 CE38E3 0 1809651 2% RUN Leave
nodea_r2_01 CE38E3 0 1809651 2% RUN Leave
4.1.1 (codename Isildur ; build r33425) Ring by Scality
Copyright 2007-2013 © Scality - All rights reserved


Hardware
Jack KIRBY
Log off
> Supervisor > Local > Anaconda
Dashboard
Administration
Preferences
General
Tasks
Tiering
Connectors
RS2 templates
Status
Anaconda
Nodes 24
Connectors 12
Alerts 3
Tasks 9
Objects 65,134,803
Unique objects 21,728,249
Average size 23.61 KB
Avg size (unique) 171.00 KB
Unique 8.43 TB 44.18 %
Stored 8.43 TB 18.24 %
Used 8.43 TB 27.26 %
Available 8.43 TB 72.74 %
Total 8.43 TB
Ring Viewer
Anaconda
Node Status All Running Other On error
Server Status All Running Other On error
Size view
Node 5.23.24.09.93/234.02.21
Key : 71996744B2B777D0F1F57B41BFBD F090F733A8A3
Reachable Yes
State On error
Activity Stand-by
Storage 67% used

Architecture prévue sur ownCoRe



Questions et réponses

Logo prévisionnel de l'Offre De Service

Merçi de votre attention !

Liens utiles

Présentation du projet ownCoRe

- Support : `http://xstra.u-strasbg.fr/lib/exe/fetch.php?media=doc:josy-cloud:josy-cloud-2014-projet-owncore.pdf`
- Vidéo : `http://webcast.in2p3.fr/videos-josy2014_etat_d_avancement_du_projet_cloud_du_cnrs_owncore`

Template CNRS pour Latex/Beamer

`https://aresu.dsi.cnrs.fr/spip.php?article178`

Compléments sur l'architecture ownCoRe

Installation prévue pour les pilotes ownCoRe

- 2 RING Scality, 1 de data, 1 de metadata (celui de metadata est sur des disques SSD)
- Pas de stockage des disques de VM sur Scality (pas adapté)
- ARC(9,3) au lieu de ARC(4,2) pour optimiser les coûts
- 12 nœuds (au sens Scality) par serveur physique
- Connecteur Scality en mode fichier (SFUSED) avec 8Go de RAM sur les VM clientes
- 2 interfaces réseaux 10GBs sur chaque serveur de stockage, configurées en actif/passif, sur un réseau dédié

Compléments sur l'architecture ownCoRe

Evolution de la configuration pour le futur d'ownCoRe

- PRA entre 2 sites serveurs en mode réplication asynchrone entre RING Scality (faute de pouvoir faire à court terme une géo-localisation du RING)
- Passage du mode fichier (SFUSED) au mode objet depuis les VM clientes ownCloud, dès que ce dernier supportera cela en natif
- Chiffrage du système de fichiers via Scality (dans la roadmap produit) envisageable