STOCKAGE OBJET S3 DATACORE SWARM A L'IBMP

Jean-Luc EVRARD
David PFLIEGER

23 janvier 2024

1

LE STOCKAGE OBJET ET LE S3

1

QUELQUES EXPLICATIONS…

- S3 = "Simple Storage Service", invention Amazon
- Accès/gestion des fichiers par des requêtes HTTP(S)
- Les fichiers sont "à plat" : il n’existe pas de hiérarchie (même s’il existe des possibilités de la mimer)
- Les fichiers "s’autodécrivent" grâce à l’existence de métadonnées
- Le fichier et ses métadonnées constituent "l’objet"
- L’objet peut avoir sa vie propre (durée de vie, immuabilité, versioning, sceau d’intégrité…)

2

…ET QUELQUES CONFUSIONS…

- Confusion entre stockage objet et accès via le protocole HTTPS : le fait d'accéder à des fichiers via ce protocole n'implique pas un stockage réellement objet
- Le véritable stockage objet ne repose ni sur un *filesystem*, ni sur une base de données : il n'est que "l'entassement" des "objets" constitués des datas et de leurs métadonnées

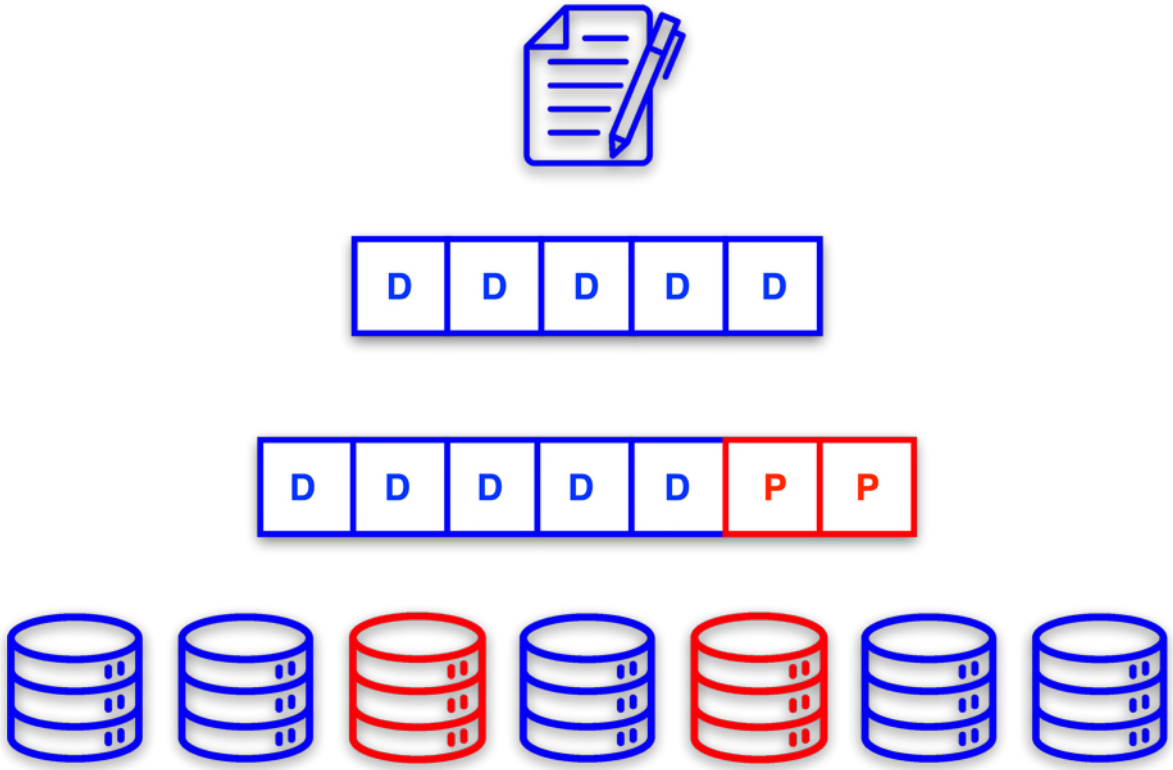
3

LES AVANTAGES DU STOCKAGE OBJET / S3

- Ils sont multiples :
  - Facilité d'accès (et de partage)
  - Facilité de fouille de données
  - Une grande résilience intrinsèque portée par une architecture physique particulière : *erasure coding* (triplicata pour les petits fichiers)
  - Une grande durabilité (fait pour des temps longs)

4

ERASURE CODING - 5:2



2

POURQUOI DU STOCKAGE OBJET DANS LA RECHERCHE ?

1 LES CONTRAINTES DE LA RECHERCHE

- La recherche est de plus en plus digitale : la production scientifique est essentiellement numérique
- Les quantités de fichiers et leurs tailles vont croissant
- Les projets scientifiques s'inscrivent dans des temps longs (5 à 10 ans)
- Le *turnover* des chercheurs est très important (CDD)
- Le partage des données est de plus en plus exigé…

2

QUELLE(S) SOLUTIONS ?

- La capacité de rétention et de sécurisation au long cours des données est primordiale tout comme la facilité d'accès
- La capacité de classification et de fouille est aussi très importante : les métadonnées ont leur rôle à jouer…
- => Le stockage objet a toutes les qualités requises pour assurer le stockage des données de la recherche et aussi le partage de ces données…

3

POURQUOI FINALEMENT DATACORE SWARM ?

1

HASARD ET CONVICTIONS…

- Nous recherchions une solution depuis 2017 et avons exploré le marché à la recherche du Graal…
- Notre choix s'est porté à l'époque sur la solution ActiveScale développée par Western Digital…
- …Mais le rachat par Quantum de cette solution à rebattu les cartes : seule une proposition locative possible…

2

HASARD ET CONVICTIONS…

- À cette époque, DataCore rachète Caringo Swarm (anciennement dans le giron de Dell)
- Lors d'une discussion avec un représentant de DataCore, nous expliquons notre mauvaise fortune avec Quantum
- Il nous propose alors une présentation du produit
- Swarm s'avère cocher toutes les cases requises…

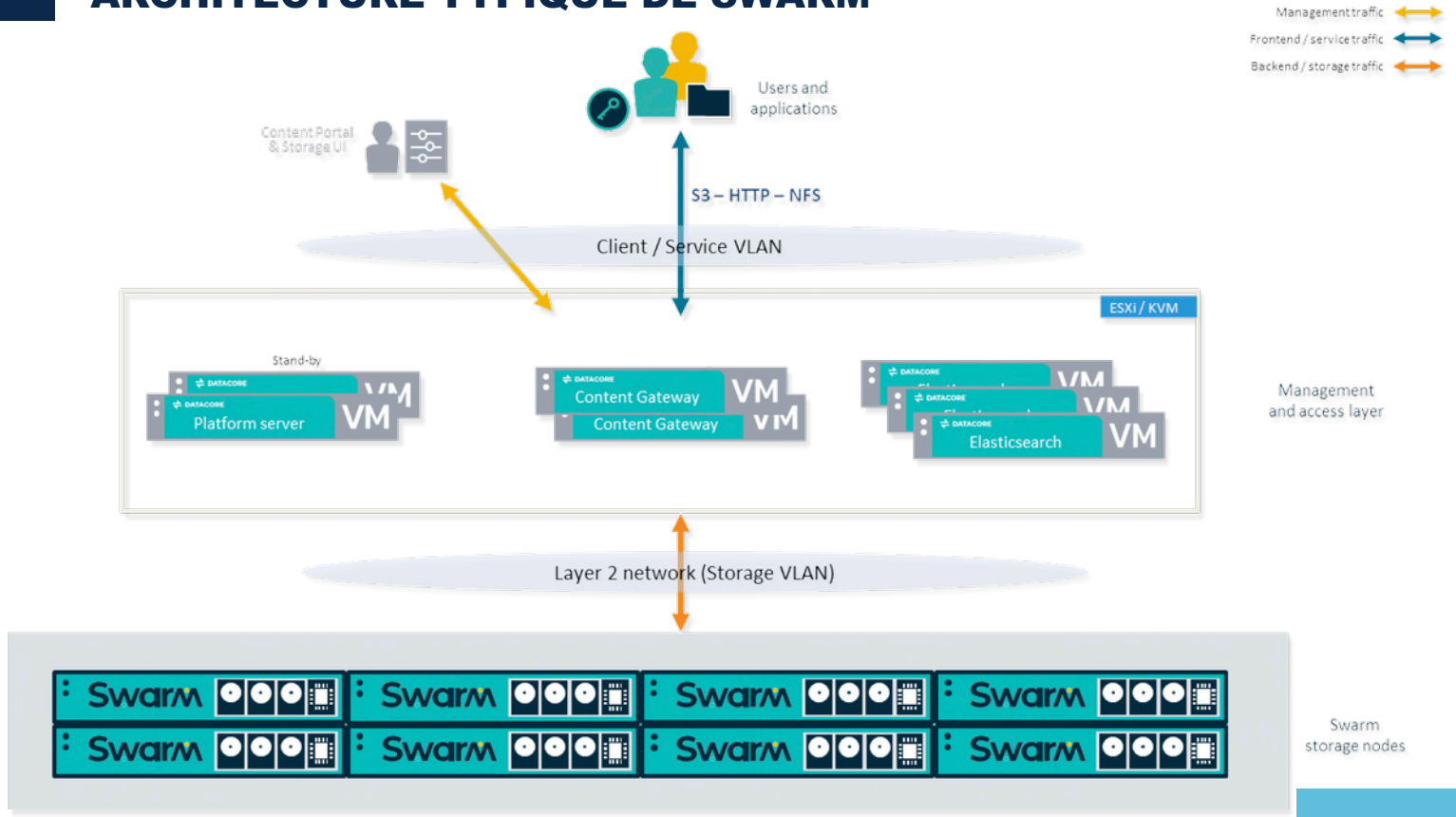
3

SWARM : L'OBJET JUSQU'AU BOUT DES DISQUES…

- Swarm est un essaim de machines avec idéalement trois "métacontrôleurs" et derrière des nœuds de stockage purs
- Pas de *filesystem* : l'OS est chargé en RAM via PXE sur les nœuds de stockage (*CastorOS*) et les disques "entassent" les données
- Pas de base de données excepté pour l'accélération des requêtes
- Les disques sont traités comme des objets à part entière : on peut les déplacer dans l'essaim de stockage…
- Protection par *erasure coding* : pas de RAID (et donc de carte…)

4

ARCHITECTURE TYPIQUE DE SWARM



4

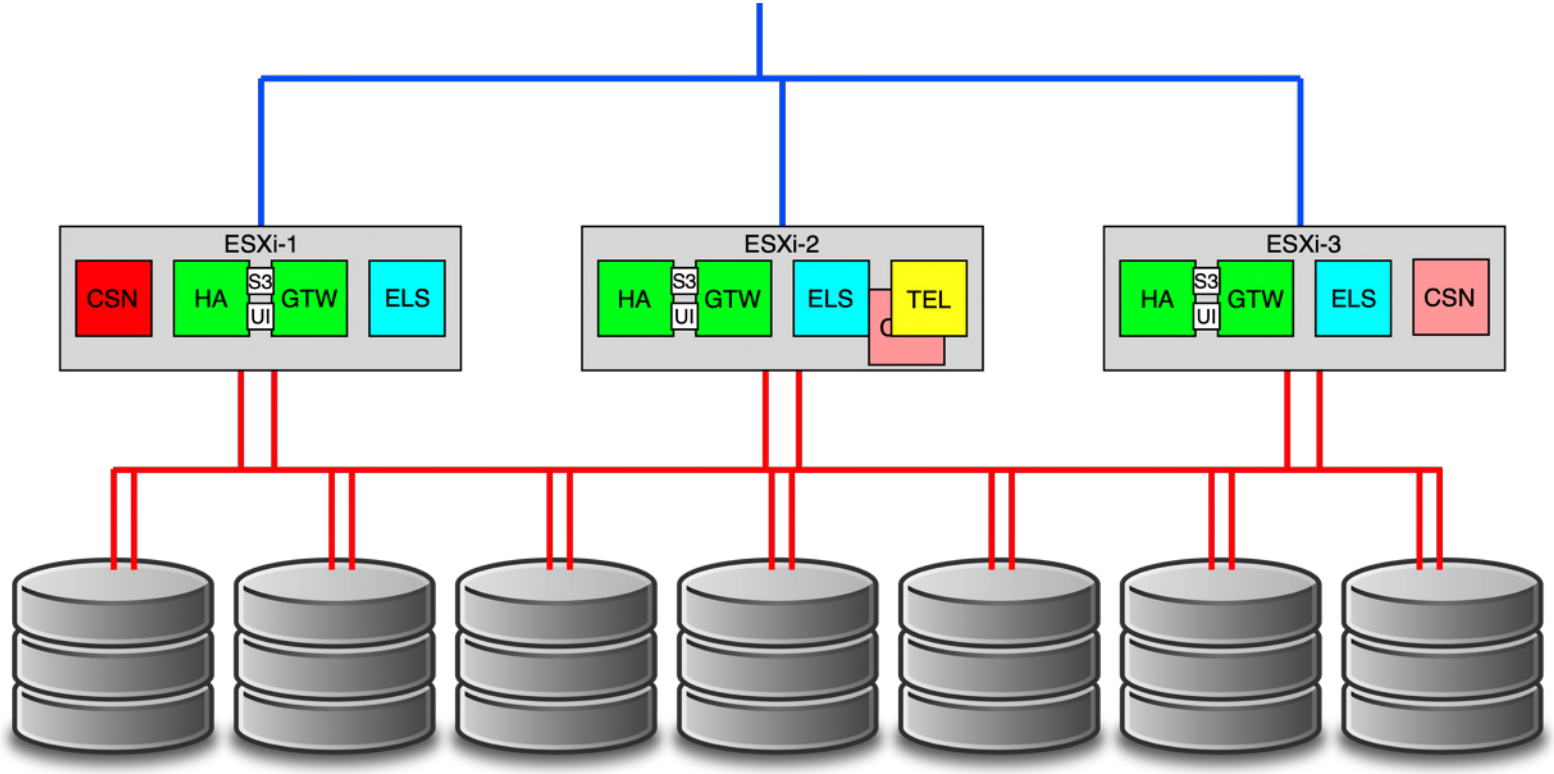
L'INSTALLATION À L'IBMP

1

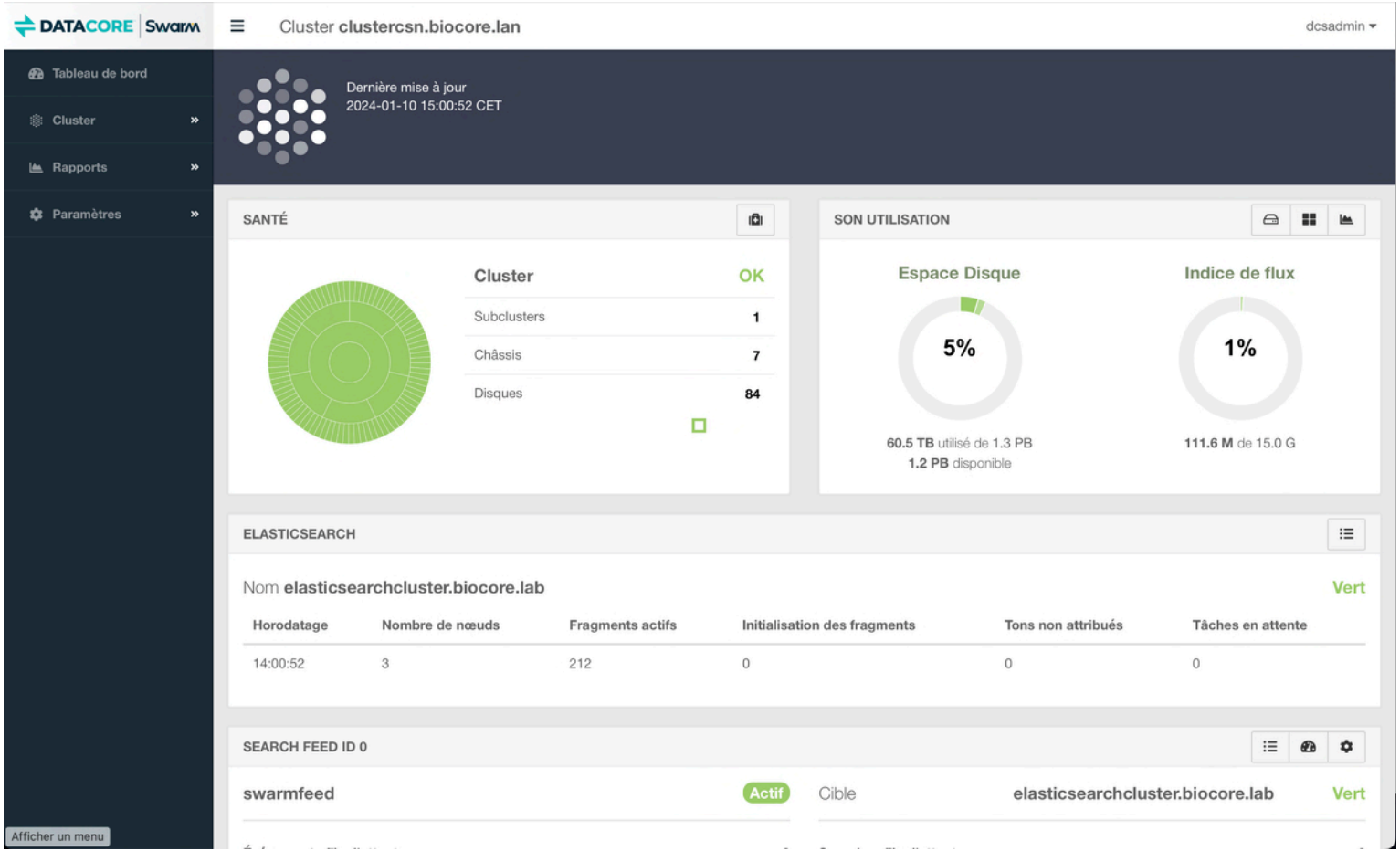
QUE FAIRE AVEC 145 K€ ?

- 850To utiles licenciés "à vie" + 3 ans de "maintenance"
- 10 serveurs DELL garantis 7 ans :
  - 3 R6515 (services) AMD 24c 3,2GHz/256Go/7,6To + SSD
  - 7 R7515 (stockage) AMD 24c 2,65GHz/128Go/192To
- 1 switch 10/100 Gbps FS S5860 (48/8)
- 3 licences VMware ESXi

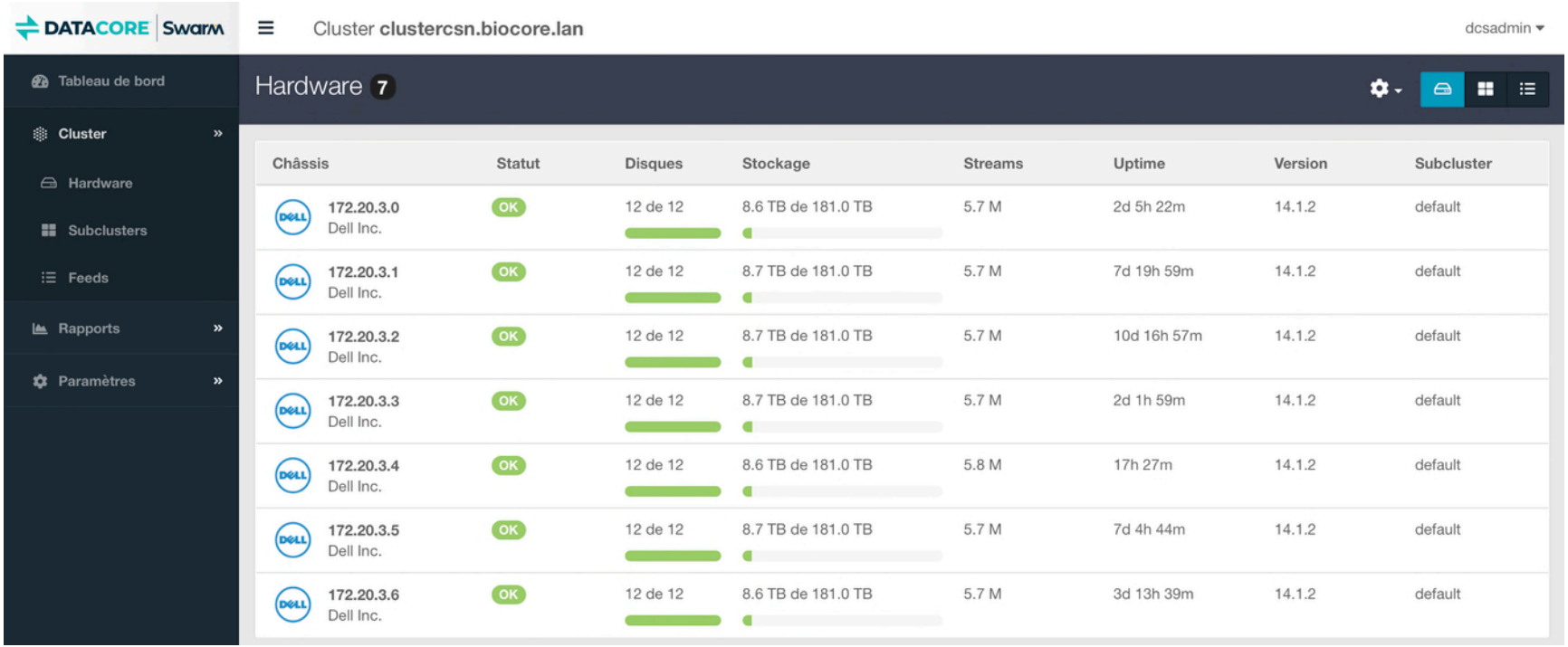
2 SCHÉMA D'INSTALLATION À L'IBMP



3

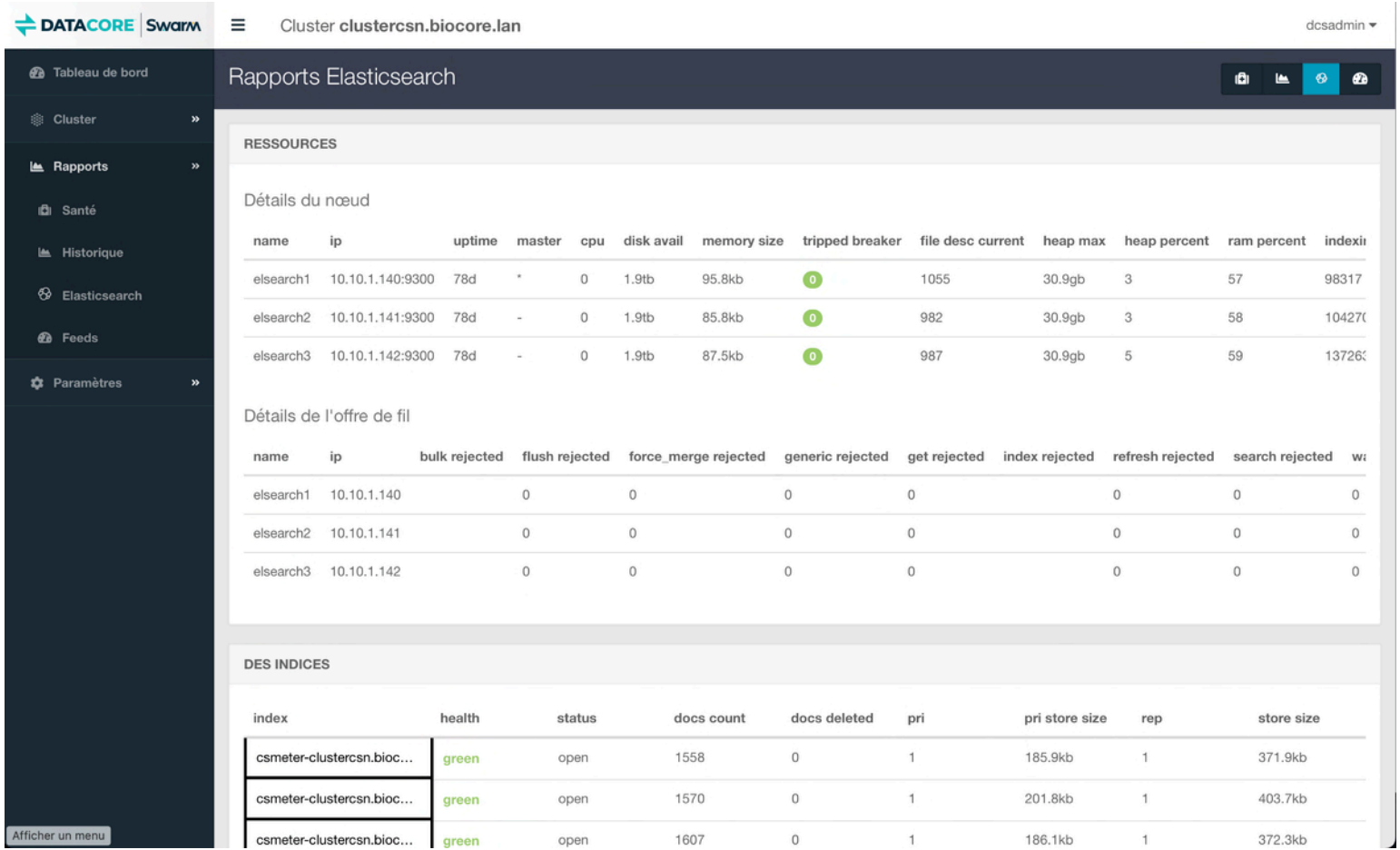

DATACORE Swarm
Cluster clustercsn.biocore.lan
dcsadmin
Tableau de bord
Cluster
Rapports
Paramètres
Dernière mise à jour
2024-01-10 15:00:52 CET
SANTÉ
Cluster
OK
Subclusters
1
Châssis
7
Disques
84
SON UTILISATION
Espace Disque
5%
60.5 TB utilisé de 1.3 PB
1.2 PB disponible
Indice de flux
1%
111.6 M de 15.0 G
ELASTICSEARCH
Nom elasticsearchcluster.biocore.lab
Vert
Horodatage
Nombre de nœuds
Fragments actifs
Initialisation des fragments
Tons non attribués
Tâches en attente
14:00:52
3
212
0
0
0
SEARCH FEED ID 0
swarmfeed
Actif
Cible
elasticsearchcluster.biocore.lab
Vert
Afficher un menu

4

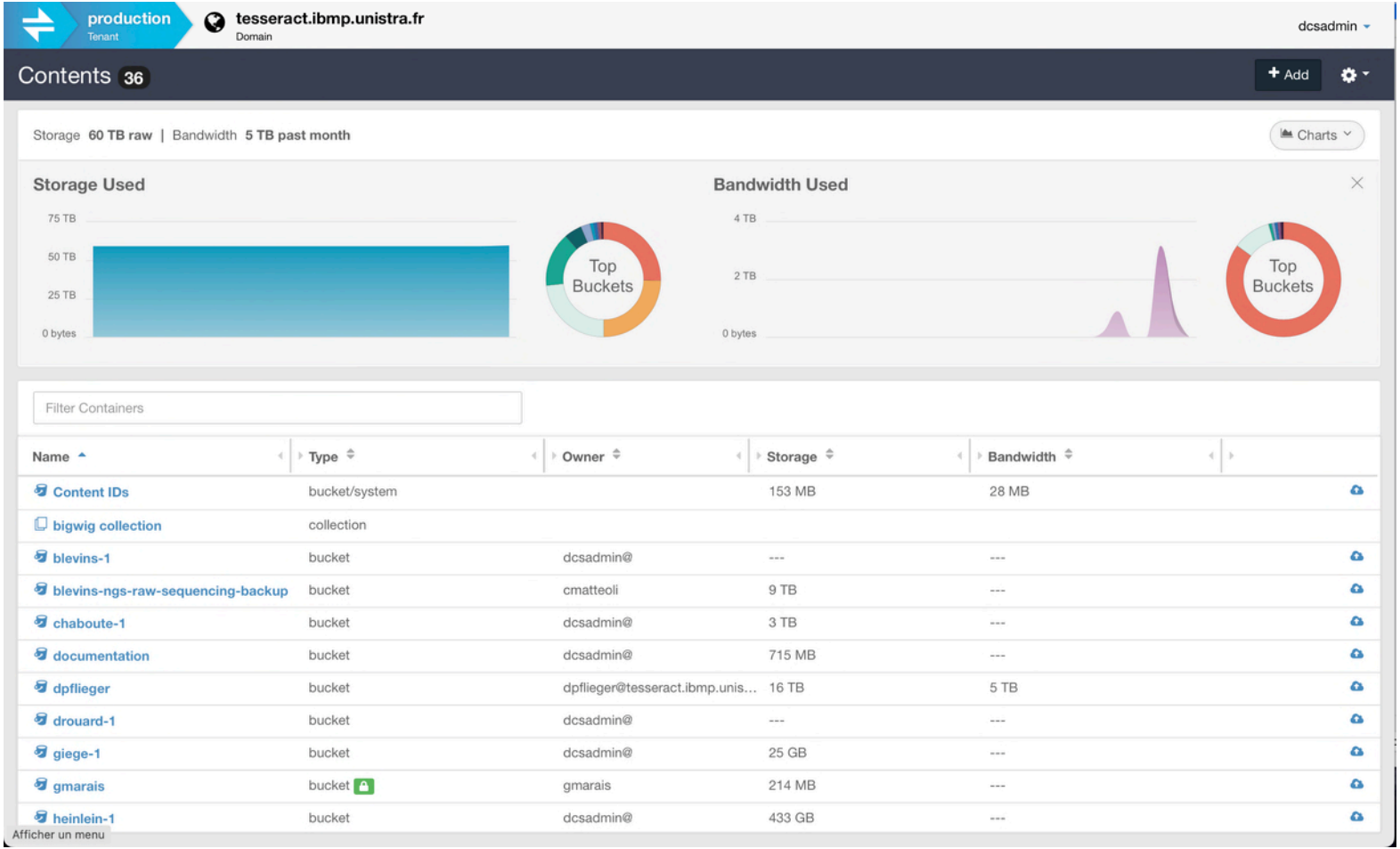


Châssis	Statut	Disques	Stockage	Streams	Uptime	Version	Subcluster
172.20.3.0 Dell Inc.	OK	12 de 12	8.6 TB de 181.0 TB	5.7 M	2d 5h 22m	14.1.2	default
172.20.3.1 Dell Inc.	OK	12 de 12	8.7 TB de 181.0 TB	5.7 M	7d 19h 59m	14.1.2	default
172.20.3.2 Dell Inc.	OK	12 de 12	8.7 TB de 181.0 TB	5.7 M	10d 16h 57m	14.1.2	default
172.20.3.3 Dell Inc.	OK	12 de 12	8.7 TB de 181.0 TB	5.7 M	2d 1h 59m	14.1.2	default
172.20.3.4 Dell Inc.	OK	12 de 12	8.6 TB de 181.0 TB	5.8 M	17h 27m	14.1.2	default
172.20.3.5 Dell Inc.	OK	12 de 12	8.7 TB de 181.0 TB	5.7 M	7d 4h 44m	14.1.2	default
172.20.3.6 Dell Inc.	OK	12 de 12	8.6 TB de 181.0 TB	5.7 M	3d 13h 39m	14.1.2	default

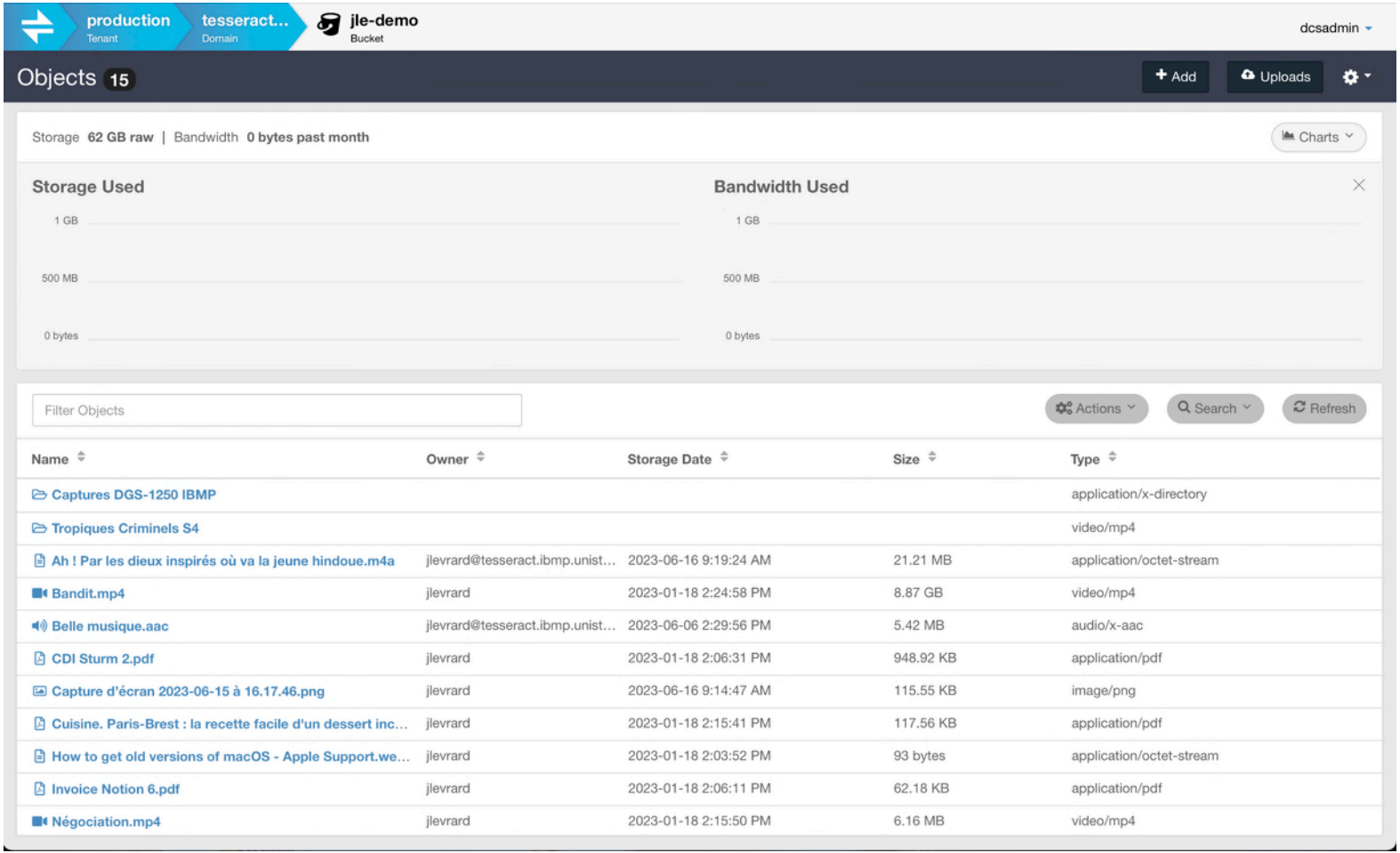
5


DATACORE Swarm
Cluster clustercsn.biocore.lan
dcsadmin
Tableau de bord
Cluster
Rapports
Santé
Historique
Elasticsearch
Feeds
Paramètres
Afficher un menu
Rapports Elasticsearch
RESSOURCES
Détails du nœud
name ip uptime master cpu disk avail memory size tripped breaker file desc current heap max heap percent ram percent indexi
elsearch1 10.10.1.140:9300 78d * 0 1.9tb 95.8kb 0 1055 30.9gb 3 57 98317
elsearch2 10.10.1.141:9300 78d - 0 1.9tb 85.8kb 0 982 30.9gb 3 58 10427(
elsearch3 10.10.1.142:9300 78d - 0 1.9tb 87.5kb 0 987 30.9gb 5 59 13726
Détails de l'offre de fil
name ip bulk rejected flush rejected force_merge rejected generic rejected get rejected index rejected refresh rejected search rejected w
elsearch1 10.10.1.140 0 0 0 0 0 0 0
elsearch2 10.10.1.141 0 0 0 0 0 0 0
elsearch3 10.10.1.142 0 0 0 0 0 0 0
DES INDICES
index health status docs count docs deleted pri pri store size rep store size
csmeter-clustercsn.bioc... green open 1558 0 1 185.9kb 1 371.9kb
csmeter-clustercsn.bioc... green open 1570 0 1 201.8kb 1 403.7kb
csmeter-clustercsn.bioc... green open 1607 0 1 186.1kb 1 372.3kb

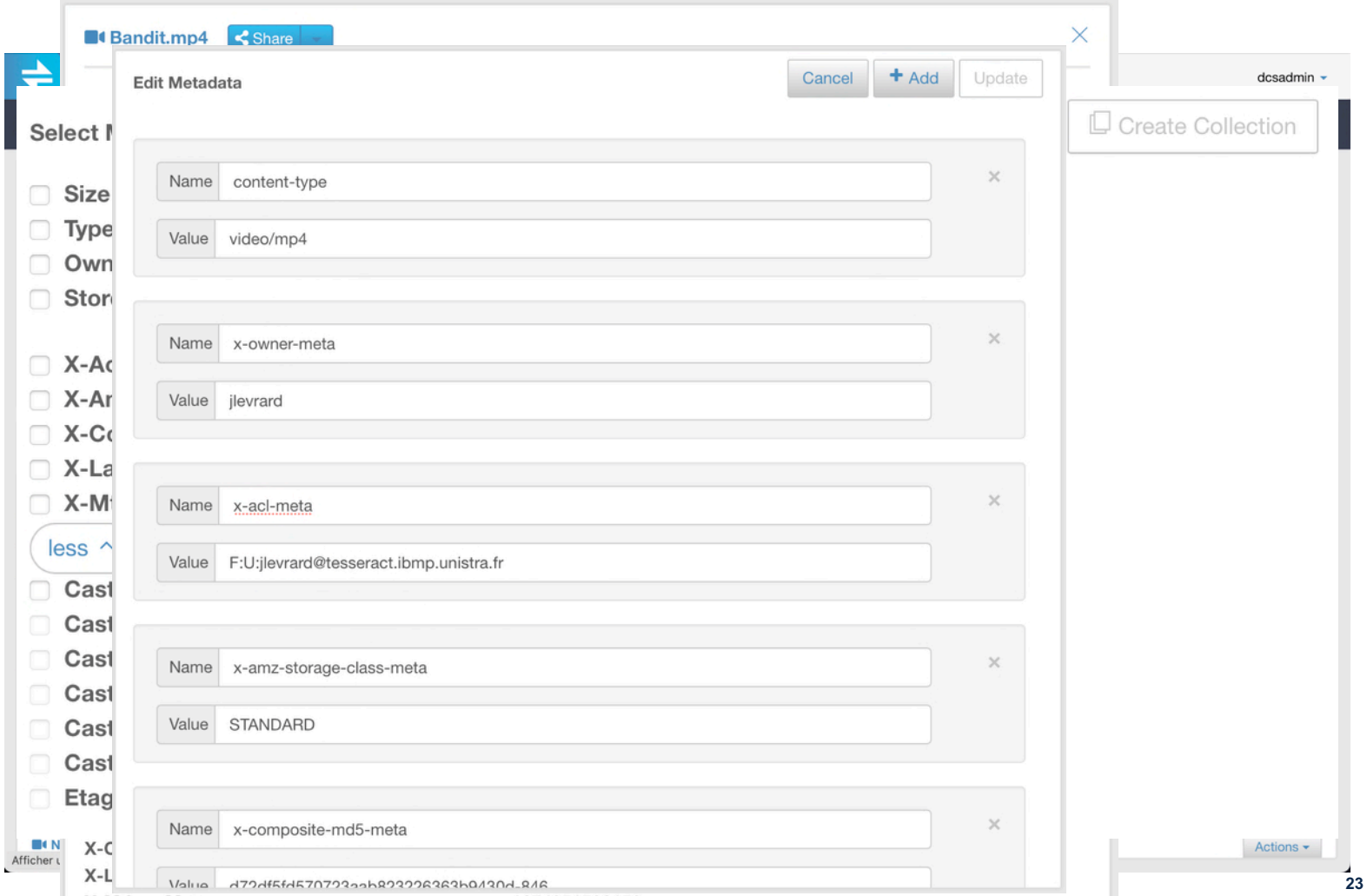
6


production
Tenant
tesseract.ibmp.unistra.fr
Domain
dcsadmin
Contents 36
+ Add
Storage 60 TB raw | Bandwidth 5 TB past month
Charts
Storage Used
75 TB
50 TB
25 TB
0 bytes
Top Buckets
Bandwidth Used
4 TB
2 TB
0 bytes
Top Buckets
Filter Containers
Name
Type
Owner
Storage
Bandwidth
Content IDs bucket/system 153 MB 28 MB
bigwig collection collection
blevins-1 bucket dcsadmin@ --- ---
blevins-ngs-raw-sequencing-backup bucket cmatteoli 9 TB ---
chaboute-1 bucket dcsadmin@ 3 TB ---
documentation bucket dcsadmin@ 715 MB ---
dpflieger bucket dpflieger@tesseract.ibmp.unis... 16 TB 5 TB
drouard-1 bucket dcsadmin@ --- ---
giege-1 bucket dcsadmin@ 25 GB ---
gmarais bucket gmarais 214 MB ---
heinlein-1 bucket dcsadmin@ 433 GB ---
Afficher un menu

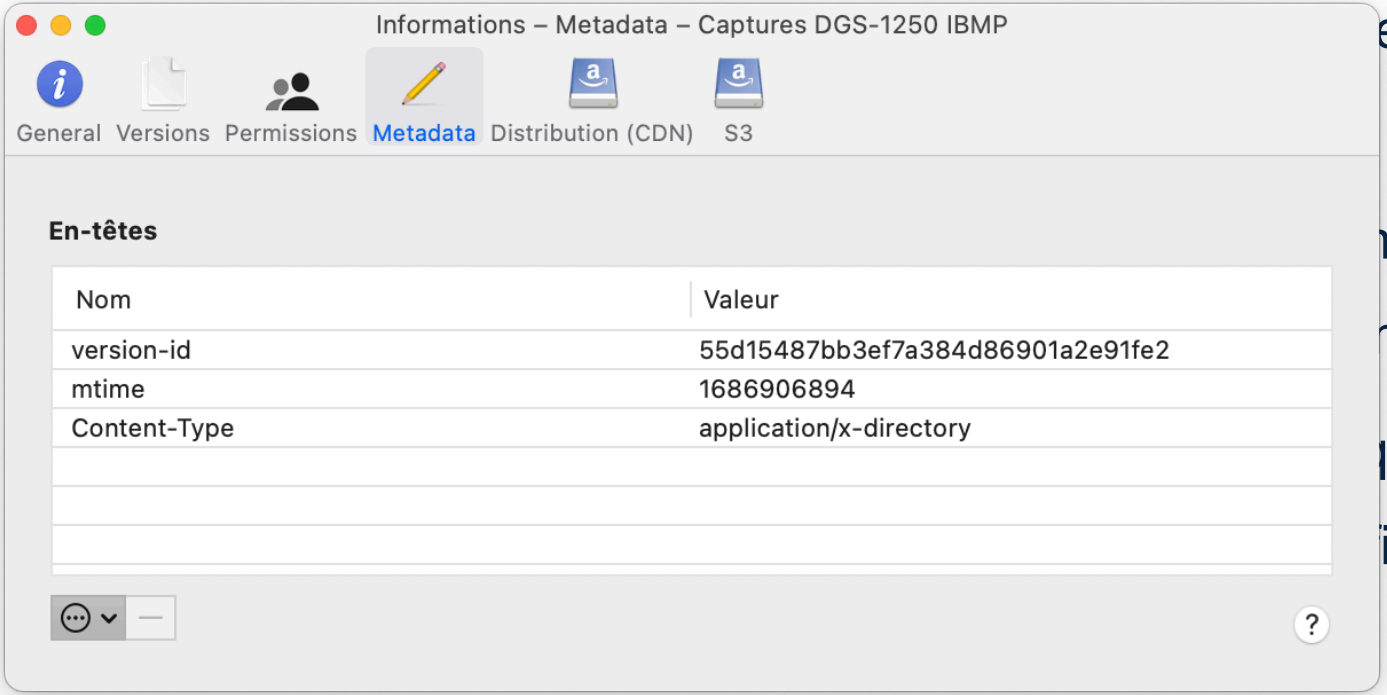
7


production
Tenant
tesseract...
Domain
jle-demo
Bucket
dcsadmin
Objects 15
+ Add
Uploads
Storage 62 GB raw | Bandwidth 0 bytes past month
Charts
Storage Used
1 GB
500 MB
0 bytes
Bandwidth Used
1 GB
500 MB
0 bytes
Filter Objects
Actions
Search
Refresh
Name
Owner
Storage Date
Size
Type
Captures DGS-1250 IBMP
application/x-directory
Tropiques Criminels S4
video/mp4
Ah ! Par les dieux inspirés où va la jeune hindoue.m4a
jlevrard@tesseract.ibmp.unist...
2023-06-16 9:19:24 AM
21.21 MB
application/octet-stream
Bandit.mp4
jlevrard
2023-01-18 2:24:58 PM
8.87 GB
video/mp4
Belle musique.aac
jlevrard@tesseract.ibmp.unist...
2023-06-06 2:29:56 PM
5.42 MB
audio/x-aac
CDI Sturm 2.pdf
jlevrard
2023-01-18 2:06:31 PM
948.92 KB
application/pdf
Capture d'écran 2023-06-15 à 16.17.46.png
jlevrard
2023-06-16 9:14:47 AM
115.55 KB
image/png
Cuisine. Paris-Brest : la recette facile d'un dessert inc...
jlevrard
2023-01-18 2:15:41 PM
117.56 KB
application/pdf
How to get old versions of macOS - Apple Support.we...
jlevrard
2023-01-18 2:03:52 PM
93 bytes
application/octet-stream
Invoice Notion 6.pdf
jlevrard
2023-01-18 2:06:11 PM
62.18 KB
application/pdf
Négociation.mp4
jlevrard
2023-01-18 2:15:50 PM
6.16 MB
video/mp4

8


Bandit.mp4
Share
Edit Metadata
Cancel
+ Add
Update
dcsadmin
Create Collection
Name
content-type
Value
video/mp4
Name
x-owner-meta
Value
jlevrard
Name
x-acl-meta
Value
F:U:jlevrard@tesseract.ibmp.unistra.fr
Name
x-amz-storage-class-meta
Value
STANDARD
Name
x-composite-md5-meta
Actions
X-Mtime-Meta
1674051592152

9 POUR FACILITER LA VIE DES UTILISATEURS "DE BASE"...



es

nple :

mple…

que

fier les

10 (R)ÉVOLUTIONS

- En ce qui conc rm, la conteneurisation est en route : il e pour les petites volumétries (10 nsporter)
- Une interface W rain d'être développée : elle permettra à space de gestion buckets/objets/
- Les interfaces W rnisées et faciliter la gestion des mé "de masse" en GUI

11 (R)ÉVOLUTIONS À VENIR

- L'IBMP vient de déployer le CLE ElabFTW *on premise* avec dépôt des données dans un bucket dédié
- ElabFTW va être capable sous peu (?) de présenter aux utilisateurs leur bucket afin de lier les données au CLE
- Mais il reste un gros travail sur la propagation des métadonnées "pertinentes" définies par un groupe de travail interne "IBMP"

12 CONCLUSION

- La solution Swarm est une vraie solution objet/S3 et s'avère techniquement parfaitement robuste et compatible avec le S3 d'Amazon
- La partie GUI pour les administrateurs est suffisante, même si pas "extraordinaire"
- La partie GUI pour les utilisateurs est "dans les tuyaux" et devrait apporter du confort pour ces derniers
- Il existe une solution de transition "*filesystem* vers objets" automatique potentiellement intéressante : FileFly
- La gestion des métas est le point dur du stockage objet…

ibmp
institut
de biologie moléculaire des plantes

cnrs

Amazon Simple Storage Service

Application in bioinformatics

Bioinformatics Plateform (BiP)

Valérie Cognat
Stéfanie Graindorge
David Pflieger

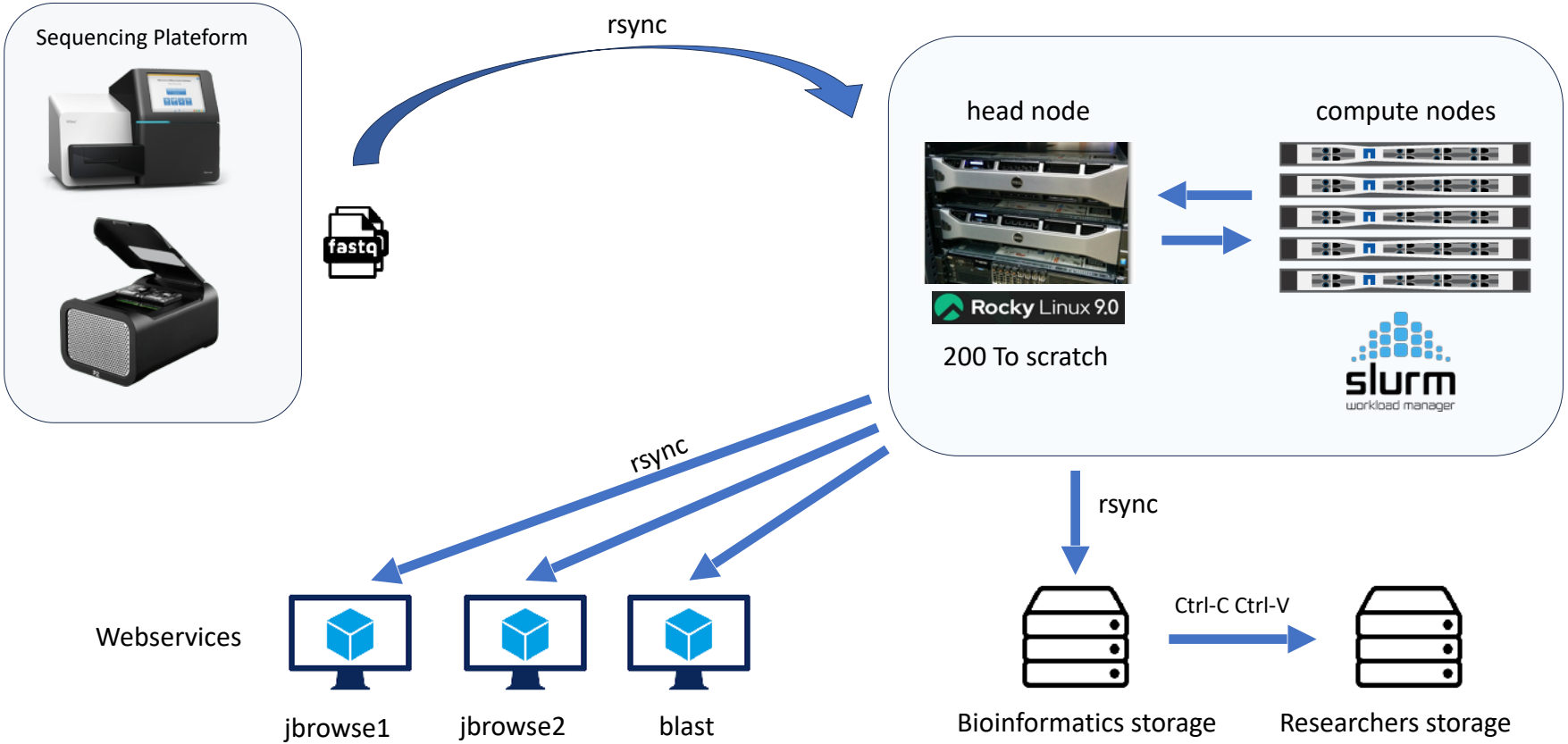
Bioinformatics projects

- Developing workflows for Next-Generation Sequencing (NGS) data analyses
- 1-8To of data per project

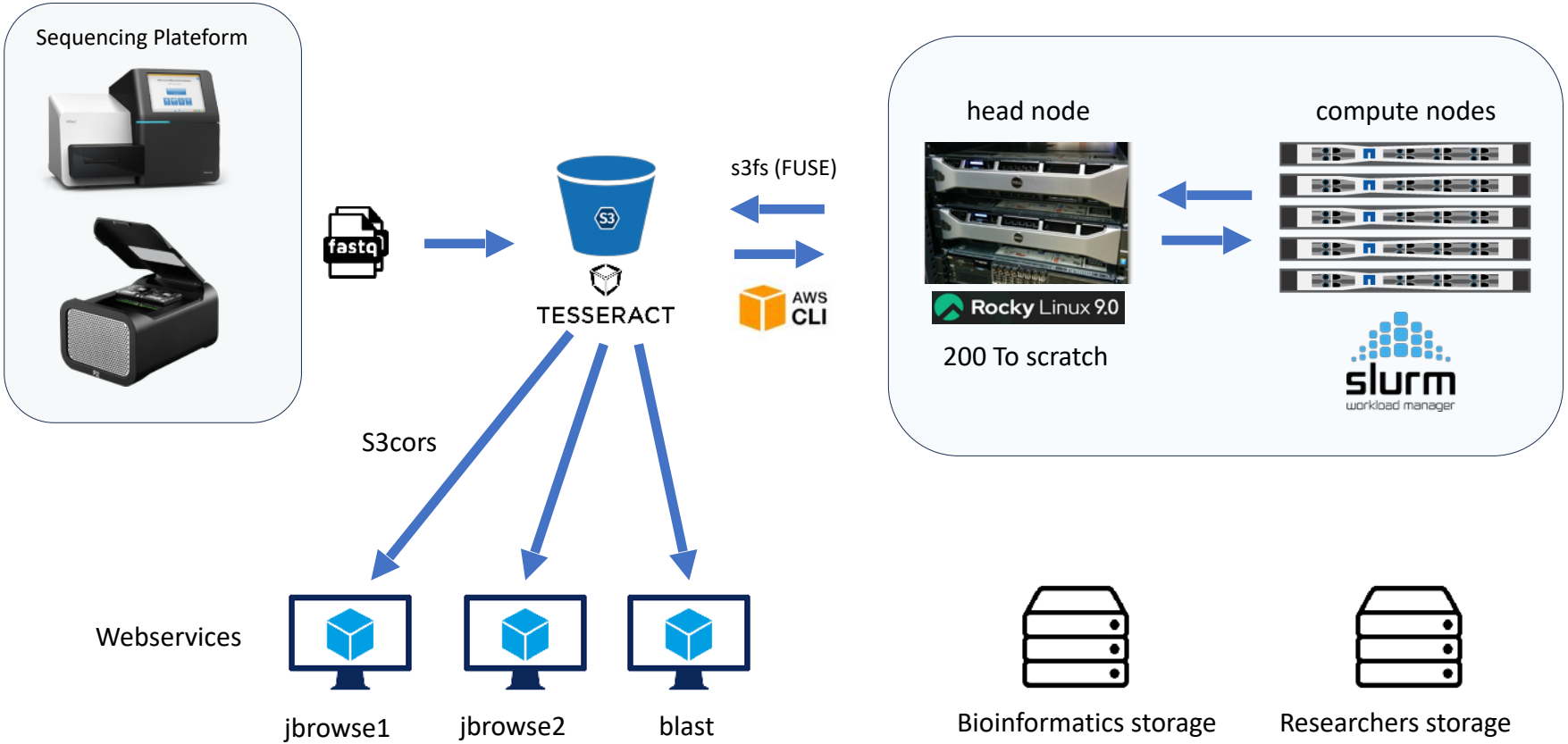
```
URT1_KO_1  URT1_KO_2  URT1_KO_3  URT1_WT_1  URT1_WT_2  URT1_WT_3
dpflieger@babel:~/Nanopore_DRS/test_snakemake2/1_Basecalled$ du -sh */
383G    URT1_KO_1/
325G    URT1_KO_2/
340G    URT1_KO_3/
379G    URT1_WT_1/
287G    URT1_WT_2/
220G    URT1_WT_3/
dpflieger@babel:~/Nanopore_DRS/test_snakemake2/1_Basecalled$ du -sh
1,9T    .
```

- Storage for raw data
- Storage for data analysis
- Storage for data results

Schema infrastructure



Schema infrastructure



Amazon S3 API usage with CLI

```
# Example of command for our S3
aws s3 --endpoint-url https://tesseract.ibmp.unistra.fr sync Gagliardi s3://DATA/Gagliardi

# We can now add an alias in the .bashrc with those options by default:
alias aws='aws --endpoint-url https://tesseract.ibmp.unistra.fr'

# Synchronize a folder to a bucket (similar to rsync and useful when you made some changes)
aws s3 sync TAIR10 s3://jbrowse/Genomes/TAIR10

# Copy a folder to a bucket
aws s3 cp TAIR10.1 s3://jbrowse/Genomes/TAIR10.1


# Command on windows powershell
 aws --endpoint-url https://tesseract.ibmp.unistra.fr s3 sync X:\Résultats\AEG\Bioinfo\Nanopore\TGS081_SCFLR_NN2_06092023 s3://dpflieger/SEQUENCI
```

```
dpflieger@pangloss:~$ alias aws
alias aws='aws --endpoint-url https://tesseract.ibmp.unistra.fr'
dpflieger@pangloss:~$ aws s3 ls s3://jbrowse
                           PRE Genomes/
                           PRE data_achard/
                           PRE data_blevins/
                           PRE data_gagliardi/
                           PRE data_giege/
                           PRE data_molinier/
                           PRE data_ryabova/
dpflieger@pangloss:~$
```

```
aws s3api put-bucket-cors \
    --profile my-aws-profile \
    --bucket jbrowse \
    --cors-configuration file://s3-cors.json
```

13 QUESTIONS ?