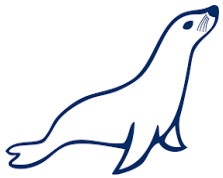
Xtra Large Data Platform

AIDEL Osman

Who I am ?

Database service manager at CC-IN2P3

- Mariadb



- PostgreSQL

- MySQL

- Oracle

ORACLE

- MongoDB

- ElasticSearch

- XLDP

Plan

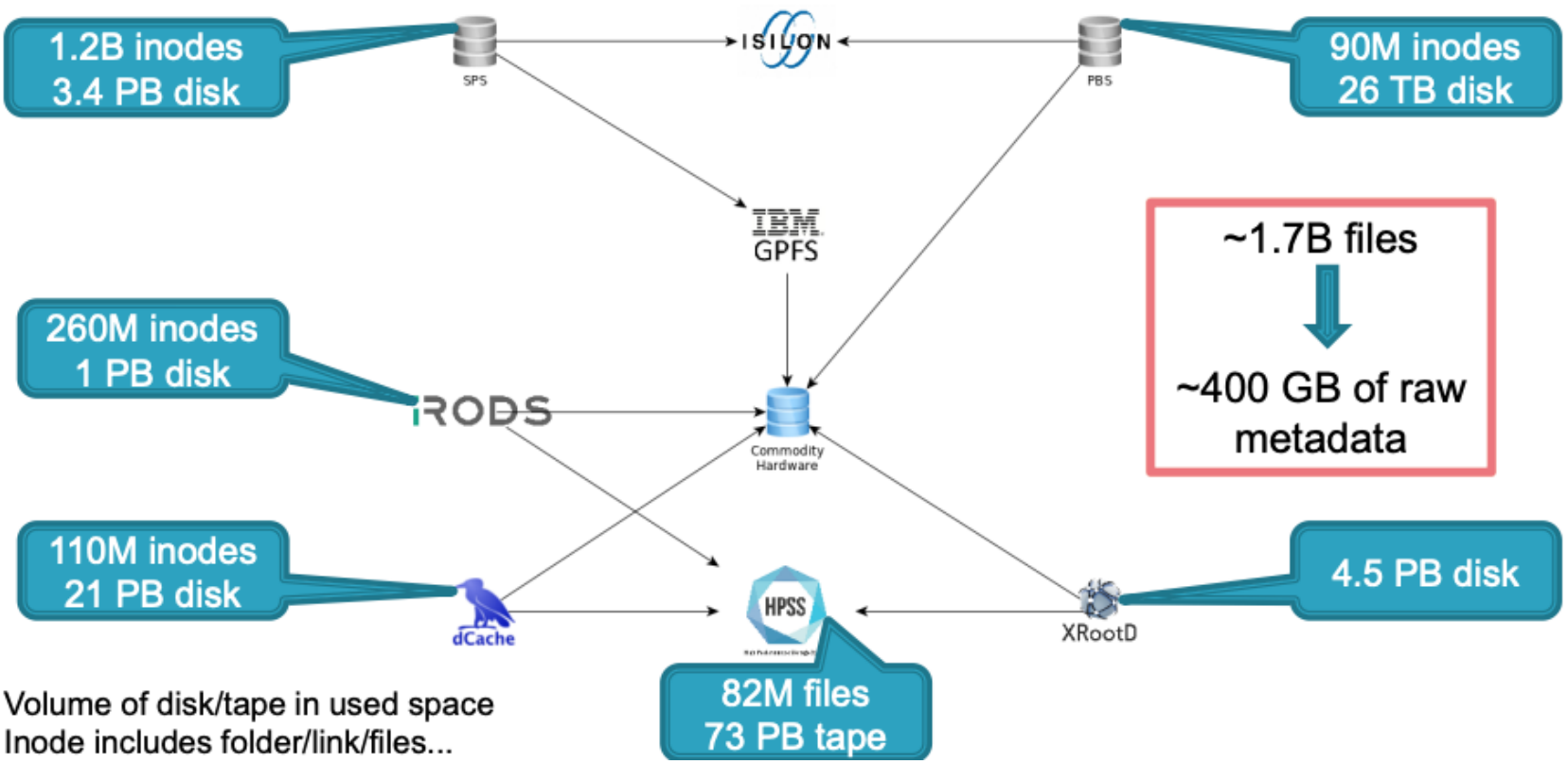
Context

Objectives

XLDP overview

Prospectives

Context

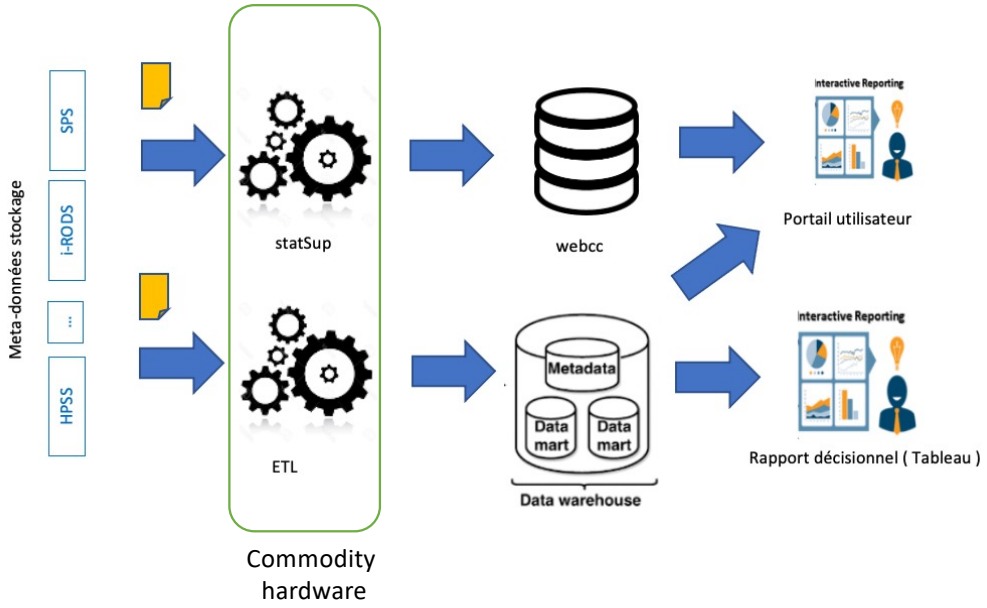


Volume of disk/tape in used space
Inode includes folder/link/files...

Context

Storage systems provide metadata such as file size, owner, last access time, last modification time ...

Those metadata are processed by some internal tools in order to build storage statistics.



Many views :

- Services
- Experiments
- Users

For many customers :

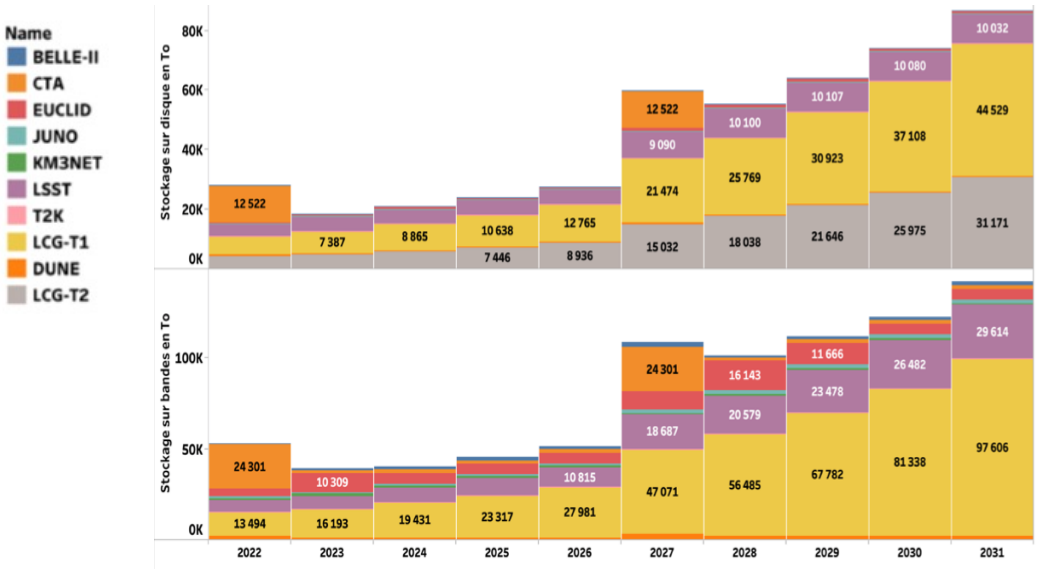
- CC-Users
- Czar : Experiment manager
- Service expert
- Team manager / Direction

Context

Current metadata volume by storage service

	Compressed size in GB	Files number in million
Dcache	9	119
SPS	13	1569
PBS	1,5	150
HPSS	2,9	109
Xrootd	5,3	
Total	33,2 (uncompressed ~200)	~ 1947

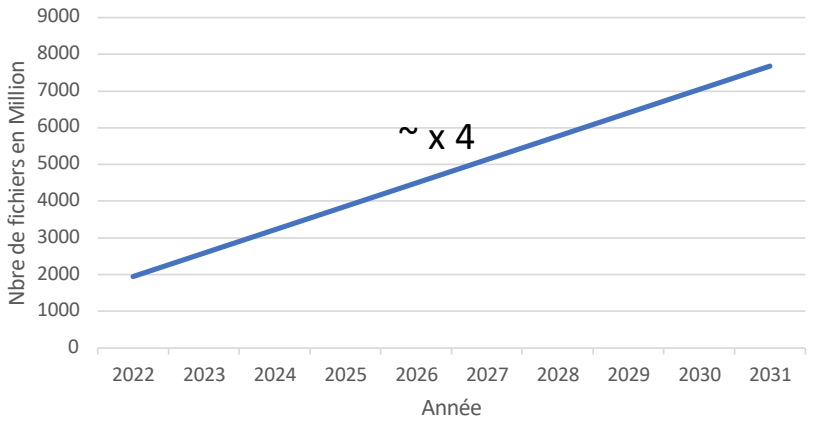
Estimated volumetry at 10 years



~ x 4

~ x 3

Estimated number of files at 10 years



HOW TO PROCESS STORAGE STATISTICS IN 2030 ?

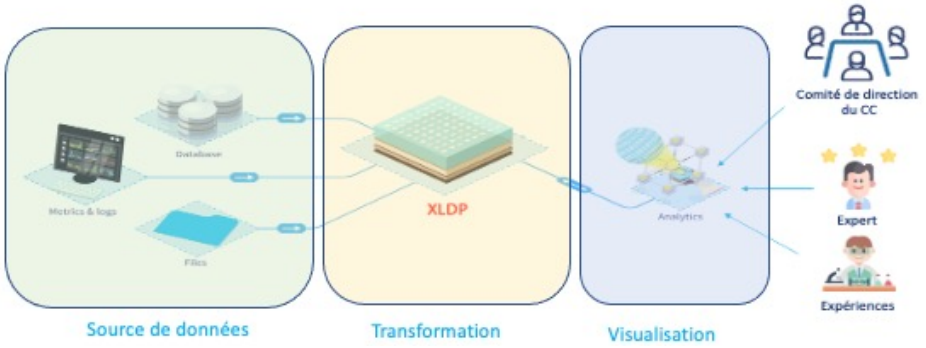
Objectives

Reproducing storage statistics regardless of the metadata volume.

Collecting and centralizing heterogeneous data sources.

Standardizing and centralizing metadata processing :

- Data cleaning
- Data consolidation with OpenIDM



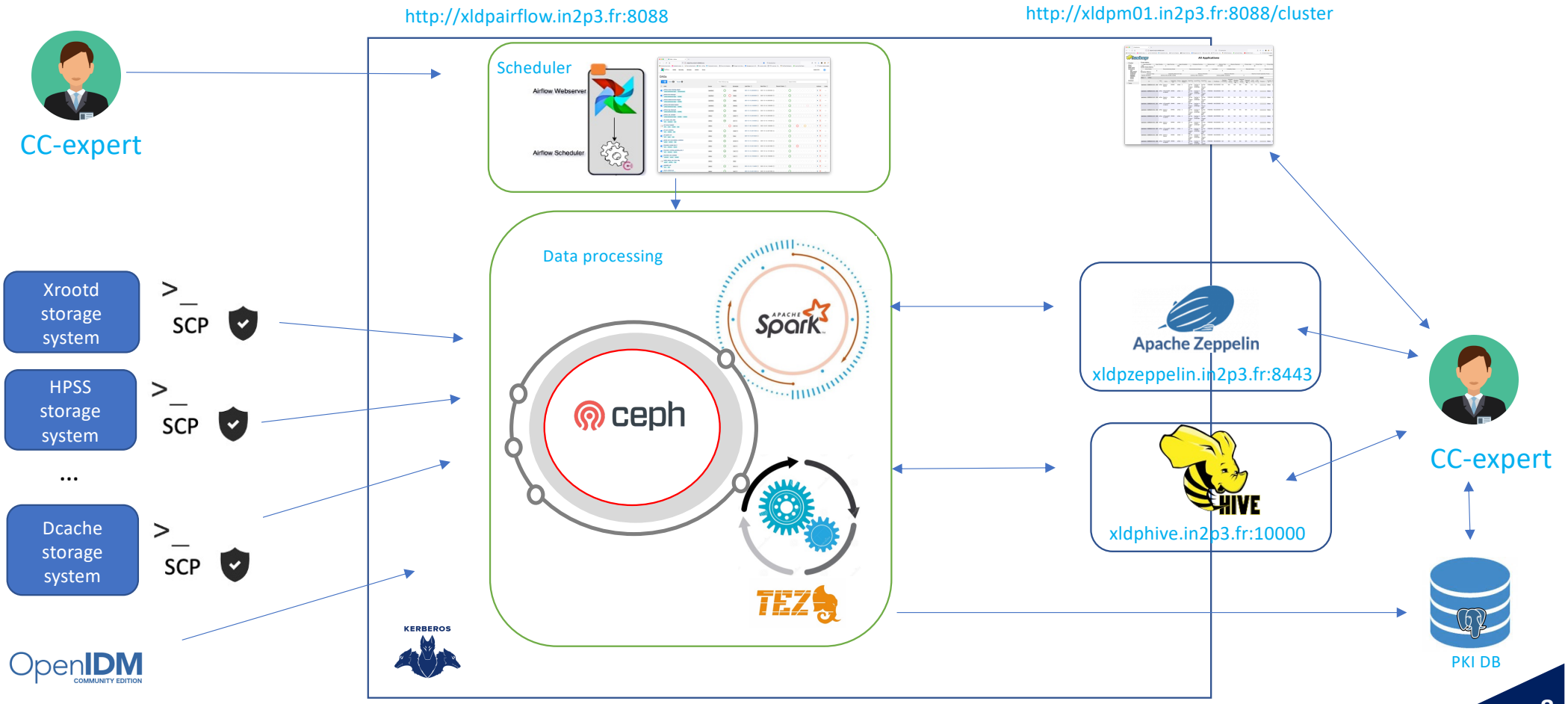
Performing simple and complex processing.

Secure access to the platform: Compliance with General Data Protection Regulation.

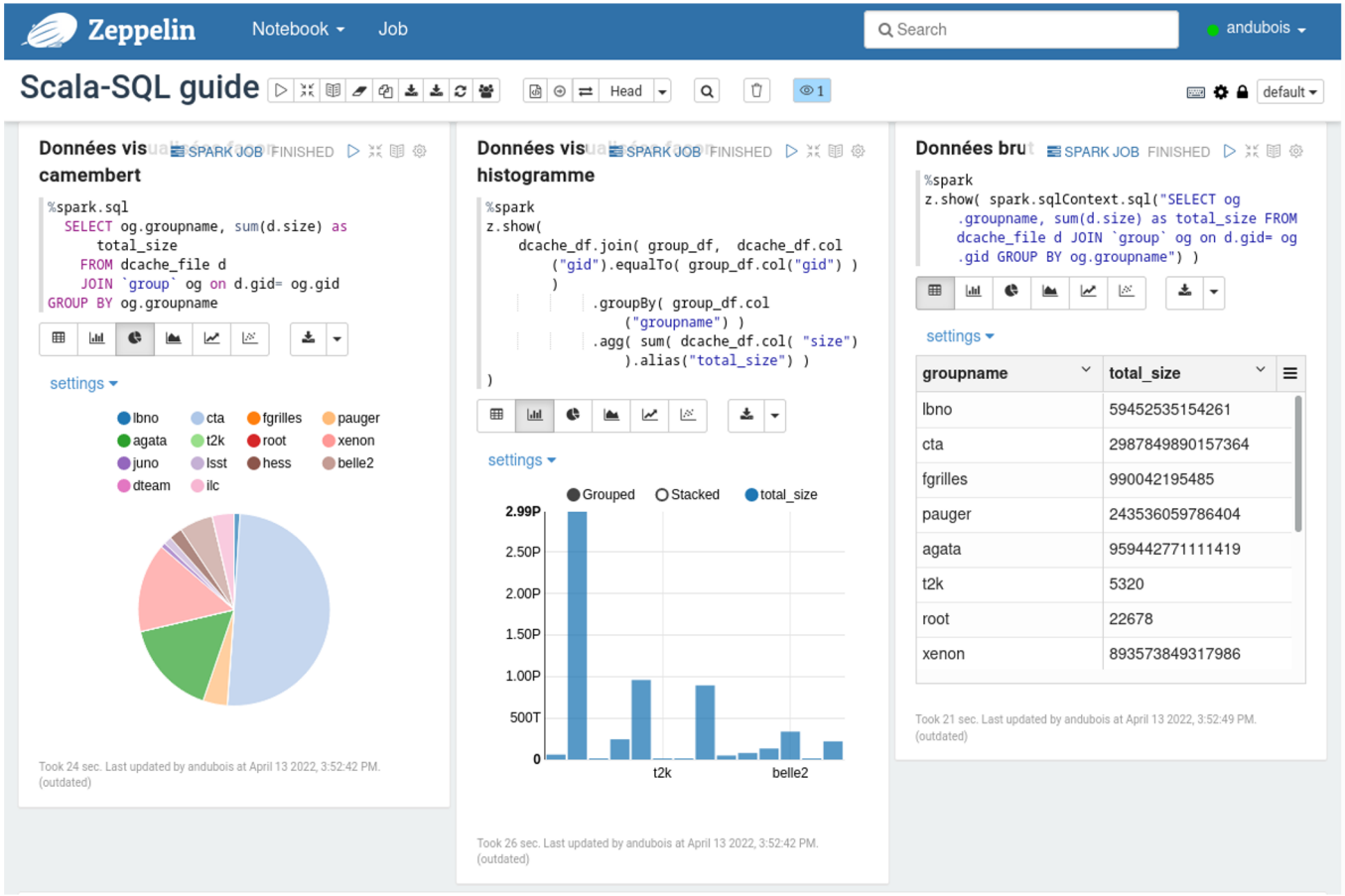
Facilitate the integration of new Key Performance Indicators.

Providing access to data sources if needed.

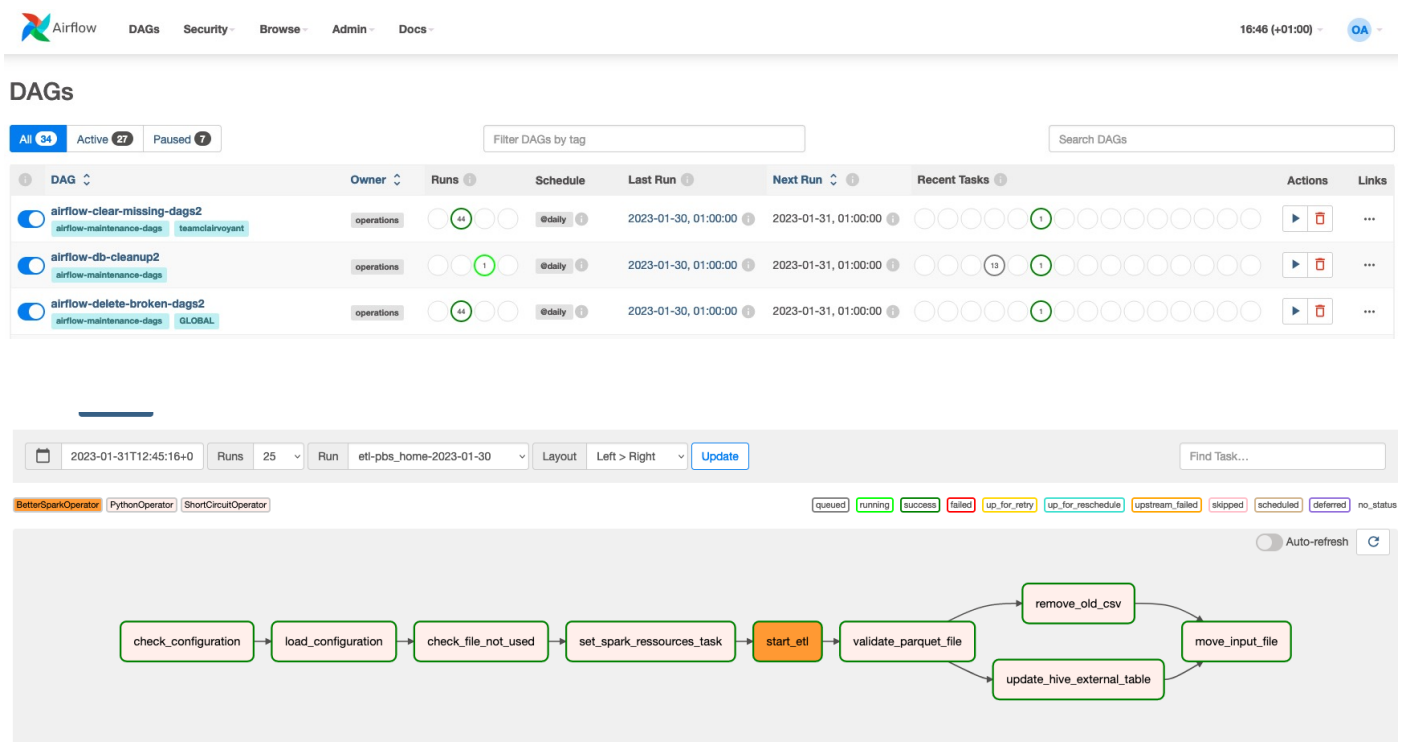
XLDP


XLDP
http://xldpairflow.in2p3.fr:8088
http://xldpm01.in2p3.fr:8088/cluster
CC-expert
Scheduler
Airflow Webserver
Airflow Scheduler
Data processing
Xrootd storage system
HPSS storage system
...
Dcache storage system
SCP
ceph
Spark
TEZ
KERBEROS
OpenIDM
COMMUNITY EDITION
Apache Zeppelin
xldpzeppelin.in2p3.fr:8443
HIVE
xldphive.in2p3.fr:10000
CC-expert
PKI DB

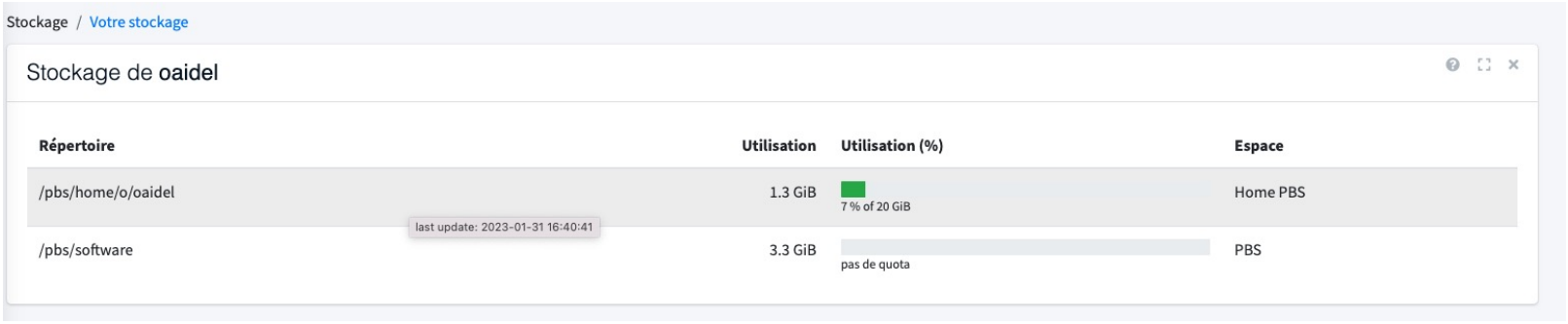
Zeppelin

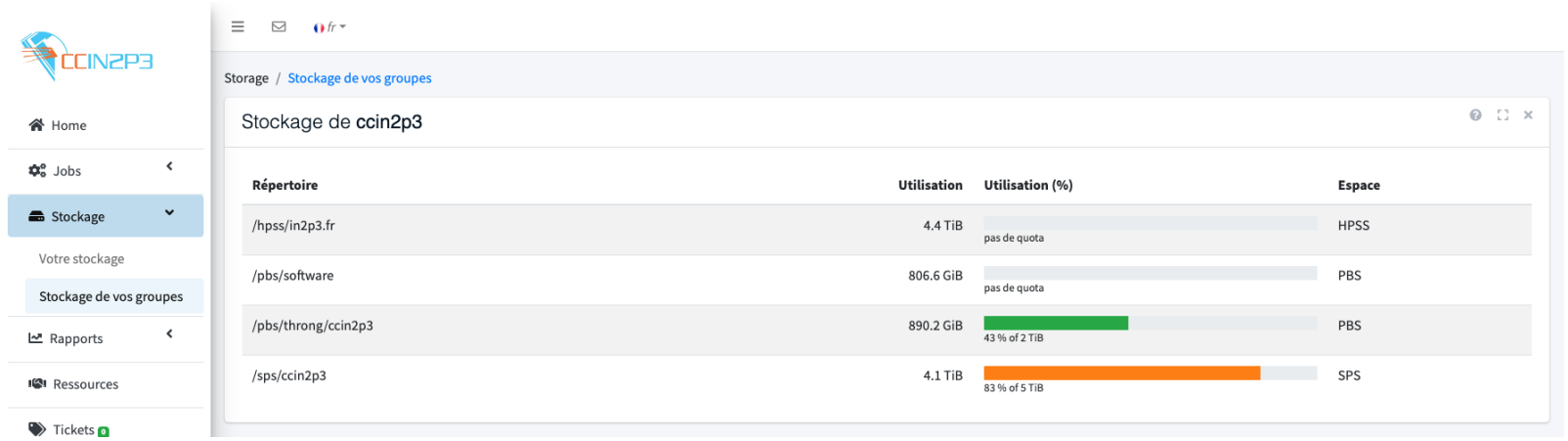

Zeppelin
Notebook
Job
Search
andubois
Scala-SQL guide
Head
default
Données vis
SPARK JOB FINISHED
camembert
%spark.sql
SELECT og.groupname, sum(d.size) as
total_size
FROM dcache_file d
JOIN `group` og on d.gid= og.gid
GROUP BY og.groupname
settings
lbno
cta
fgrilles
pauger
agata
t2k
root
xenon
juno
lsst
hess
belle2
dteam
ilc
Took 24 sec. Last updated by andubois at April 13 2022, 3:52:42 PM.
(outdated)
Données vis
SPARK JOB FINISHED
histogramme
%spark
z.show(
dcache_df.join(group_df, dcache_df.col
("gid").equalTo(group_df.col("gid"))
)
.groupBy(group_df.col
("groupname"))
.agg(sum(dcache_df.col("size")
).alias("total_size"))
)
settings
Grouped
Stacked
total_size
2.99P
2.50P
2.00P
1.50P
1.00P
500T
0
t2k
belle2
Took 26 sec. Last updated by andubois at April 13 2022, 3:52:42 PM.
(outdated)
Données bru
SPARK JOB FINISHED
%spark
z.show(spark.sqlContext.sql("SELECT og
.groupname, sum(d.size) as total_size FROM
dcache_file d JOIN `group` og on d.gid= og
.gid GROUP BY og.groupname"))
settings
groupname
total_size
lbno
59452535154261
cta
2987849890157364
fgrilles
990042195485
pauger
243536059786404
agata
959442771111419
t2k
5320
root
22678
xenon
893573849317986
Took 21 sec. Last updated by andubois at April 13 2022, 3:52:49 PM.
(outdated)

Airflow

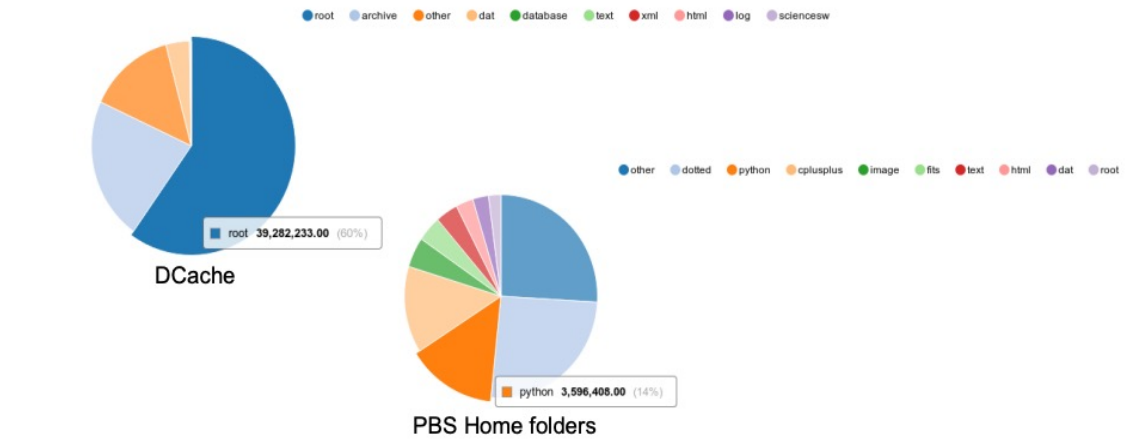

Airflow
DAGs
Security
Browse
Admin
Docs
16:46 (+01:00)
OA
DAGs
All 34
Active 27
Paused 7
Filter DAGs by tag
Search DAGs
DAG
Owner
Runs
Schedule
Last Run
Next Run
Recent Tasks
Actions
Links
airflow-clear-missing-dags2
airflow-maintenance-dags
teamclairvoyant
operations
@daily
2023-01-30, 01:00:00
2023-01-31, 01:00:00
airflow-db-cleanup2
airflow-maintenance-dags
operations
@daily
2023-01-30, 01:00:00
2023-01-31, 01:00:00
airflow-delete-broken-dags2
airflow-maintenance-dags
GLOBAL
operations
@daily
2023-01-30, 01:00:00
2023-01-31, 01:00:00
2023-01-31T12:45:16+0
Runs
25
Run
etl-pbs_home-2023-01-30
Layout
Left > Right
Update
Find Task...
BetterSparkOperator
PythonOperator
ShortCircuitOperator
queued
running
success
failed
up_for_retry
up_for_reschedule
upstream_failed
skipped
scheduled
deferred
no_status
Auto-refresh
check_configuration
load_configuration
check_file_not_used
set_spark_ressources_task
start_etl
validate_parquet_file
remove_old_csv
update_hive_external_table
move_input_file

KPI : File extensions



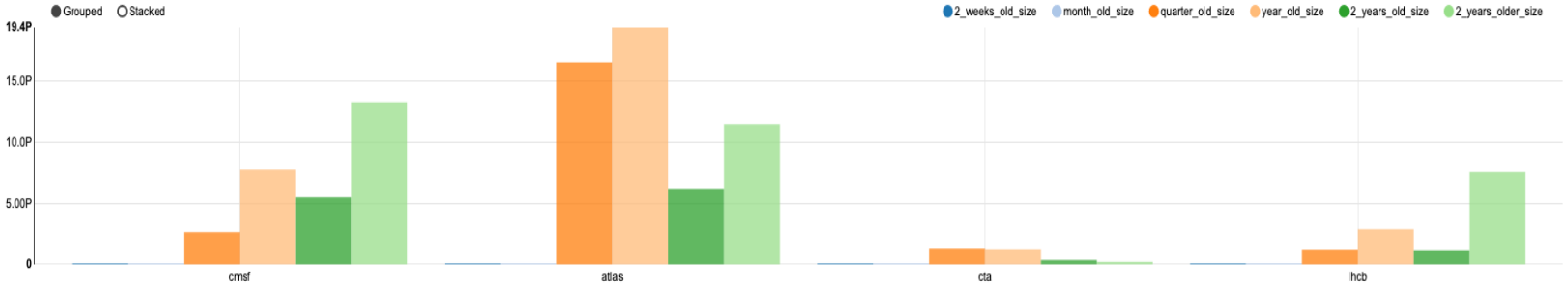


KPI example



File extensions repartition in # of file

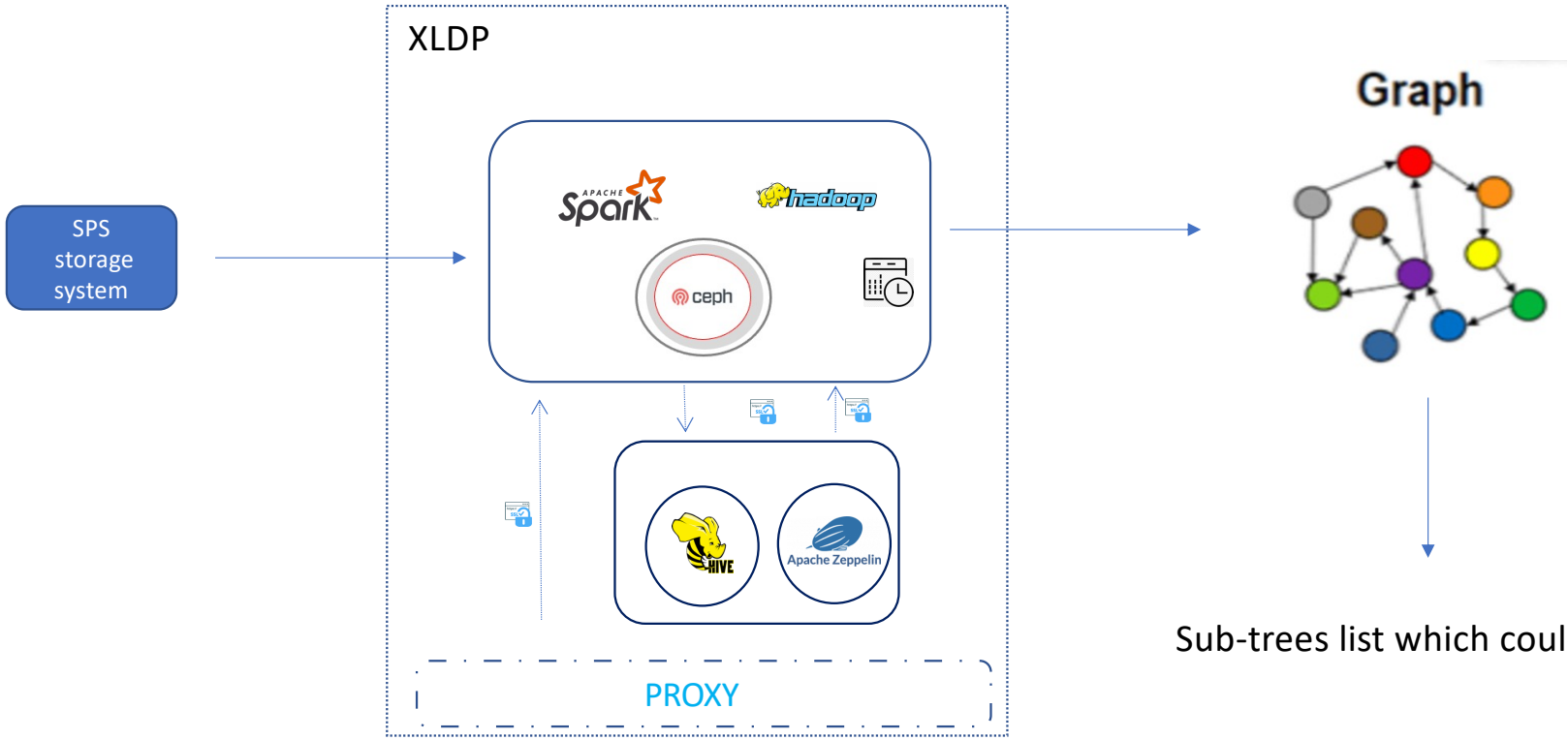
File age per access time



Prospectives

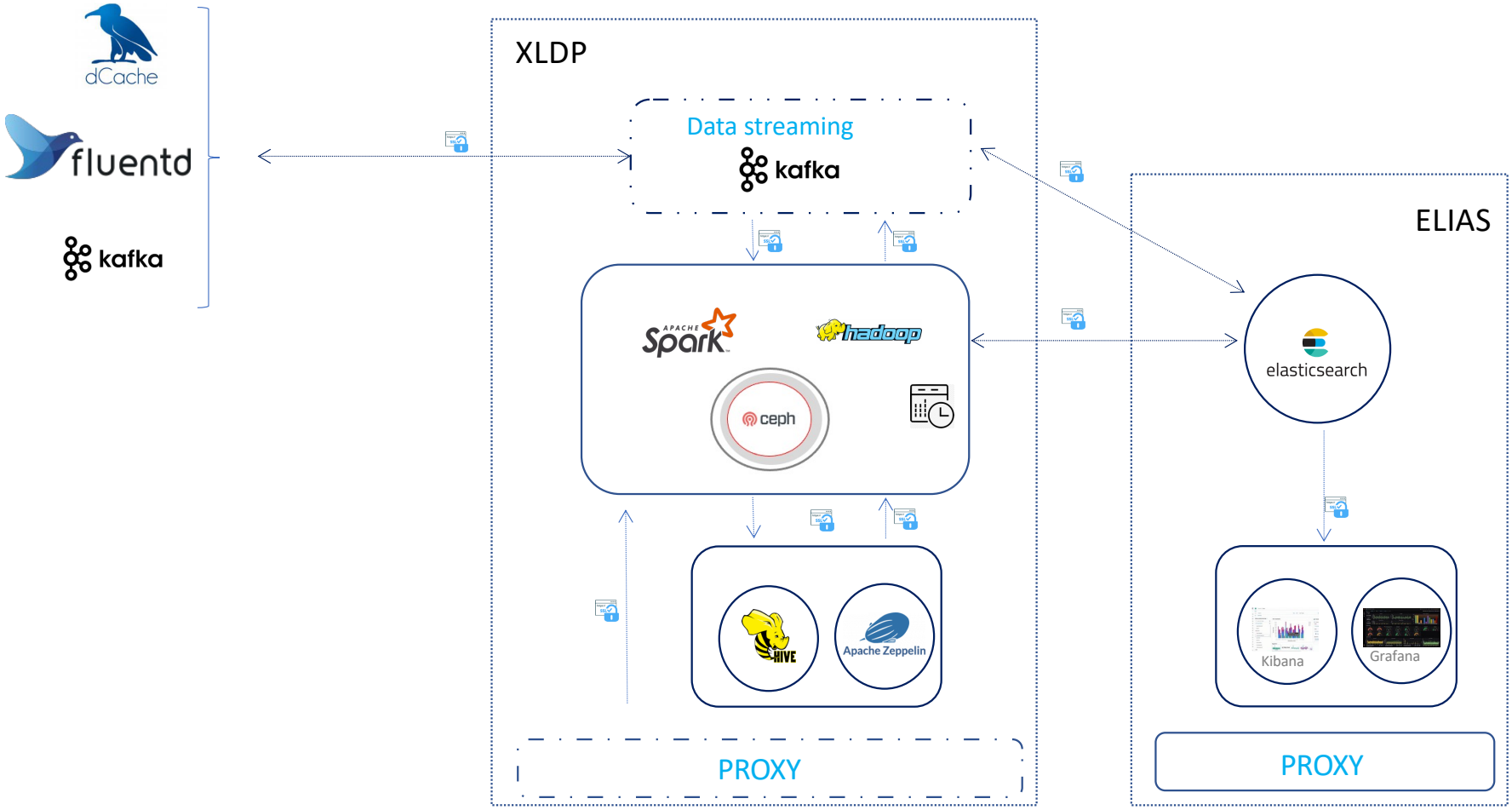
Tapes are cheaper than disks.

To save money, we try to identify dead sub-trees no more accessed on disks for a long time.



Sub-trees list which could be moved to Tape

Prospectives



Questions