

# Lustre data migration among heterogeneous storage platforms ( Archive Lustre cold data to alicloud)

---

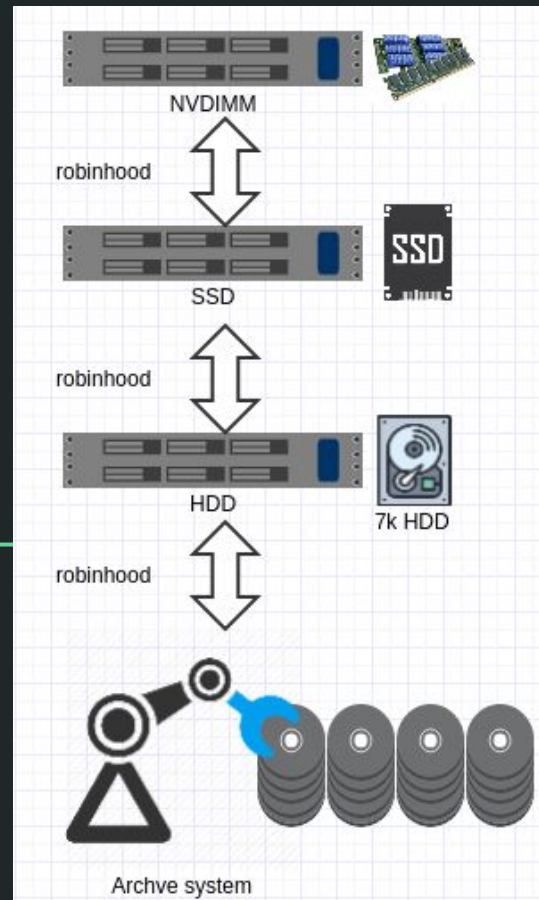
-- Homer Li (Beijing Genomics Institute)

- Background

- Sequencing platforms output a lot of cold data
  - 3~4TiB raw data per day
    - Create more project data
  - Data could not delete
  - A huge data pool archive cold data
    - Huge pool (Object storage)
  - A policy engine
    - Automatic
    - Easy to manage
    - Robinhood like pipe

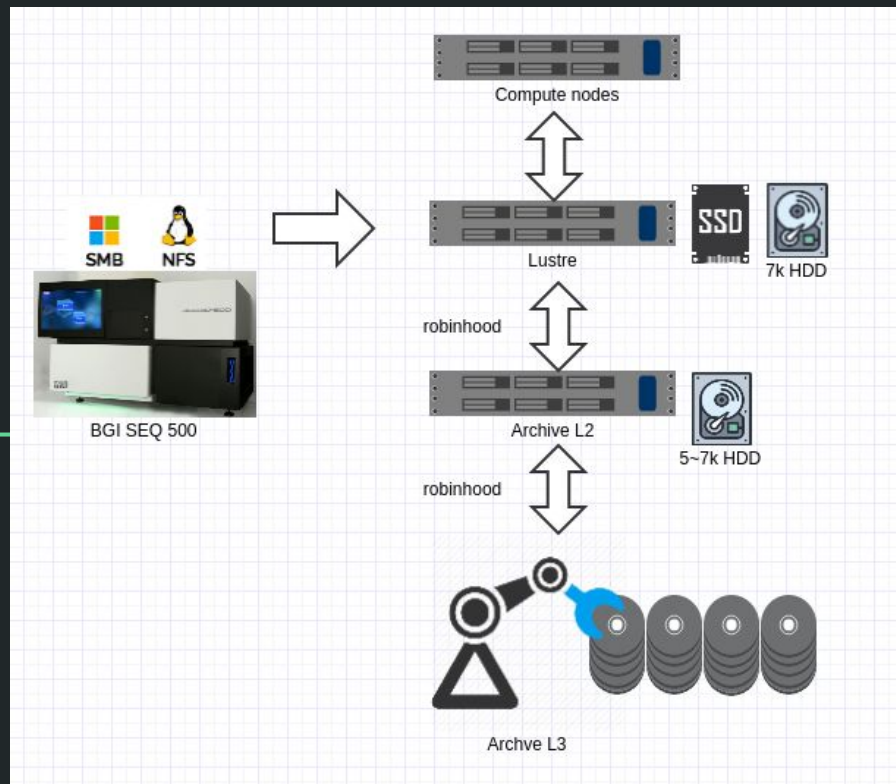


- Why data classification
  - Different types of storage media
    - Performance
    - Cost
    - Capacity
    - Density
    - Lower power

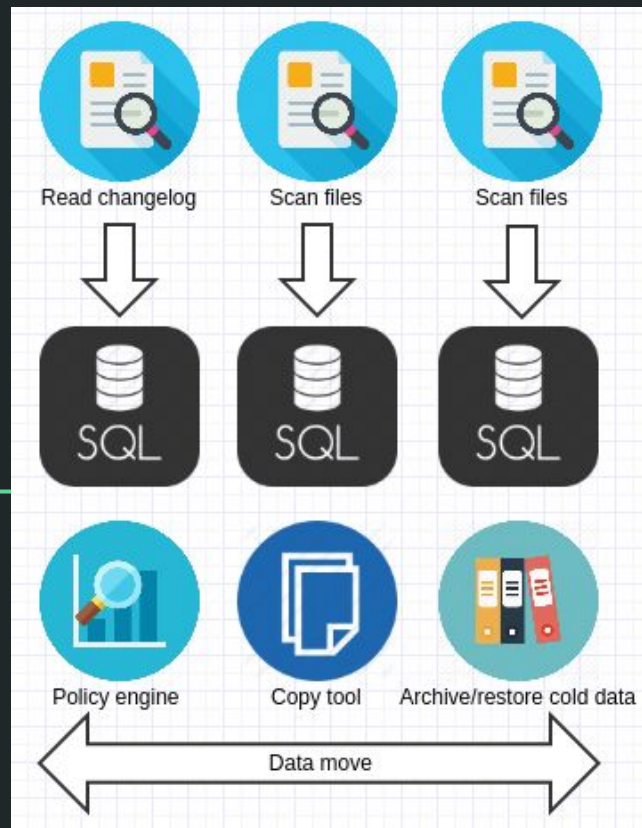


# ● Sequencing platforms workflow

- NFS/SMB -> Lustre
- Compute <-> Lustre
- Lustre -> Archive System
- Archive system
  - Object storage
  - Lustre Level 2
  - Tape
  - Blue disc (evaluating)
    - Offline



- Robinhood workflow
  - Read Lustre change log
  - Scan files
  - Policy engine trigger rules
  - Copy tool move data

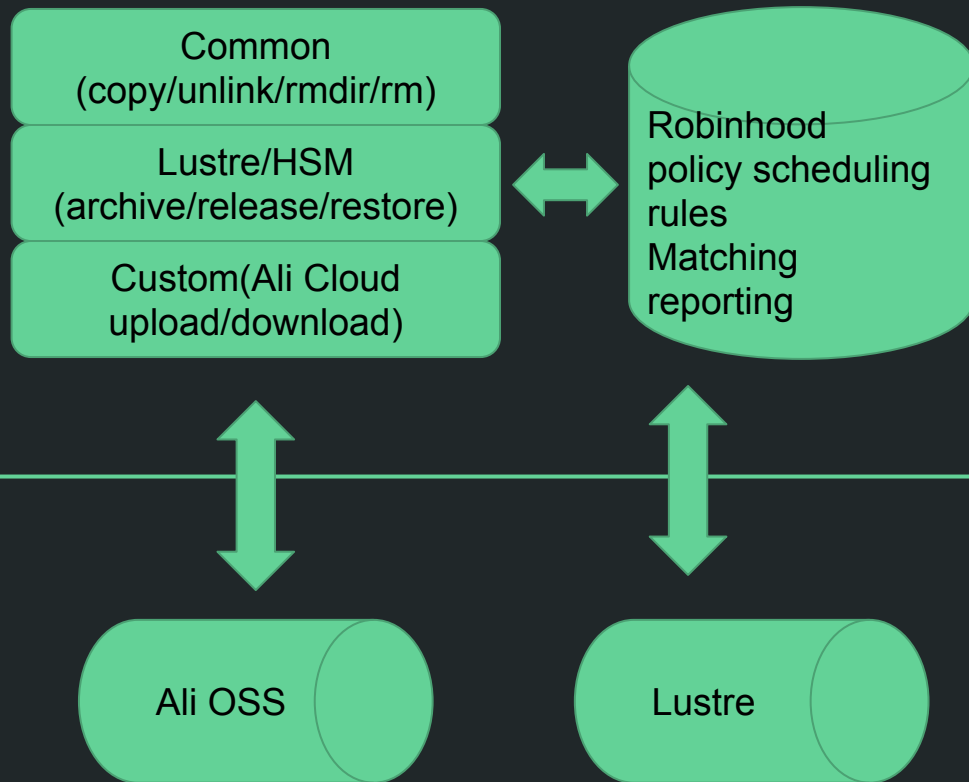


- Custom plugins

- lshmtool\_cmd

- Lustre

- Archive
    - Release
    - Restore

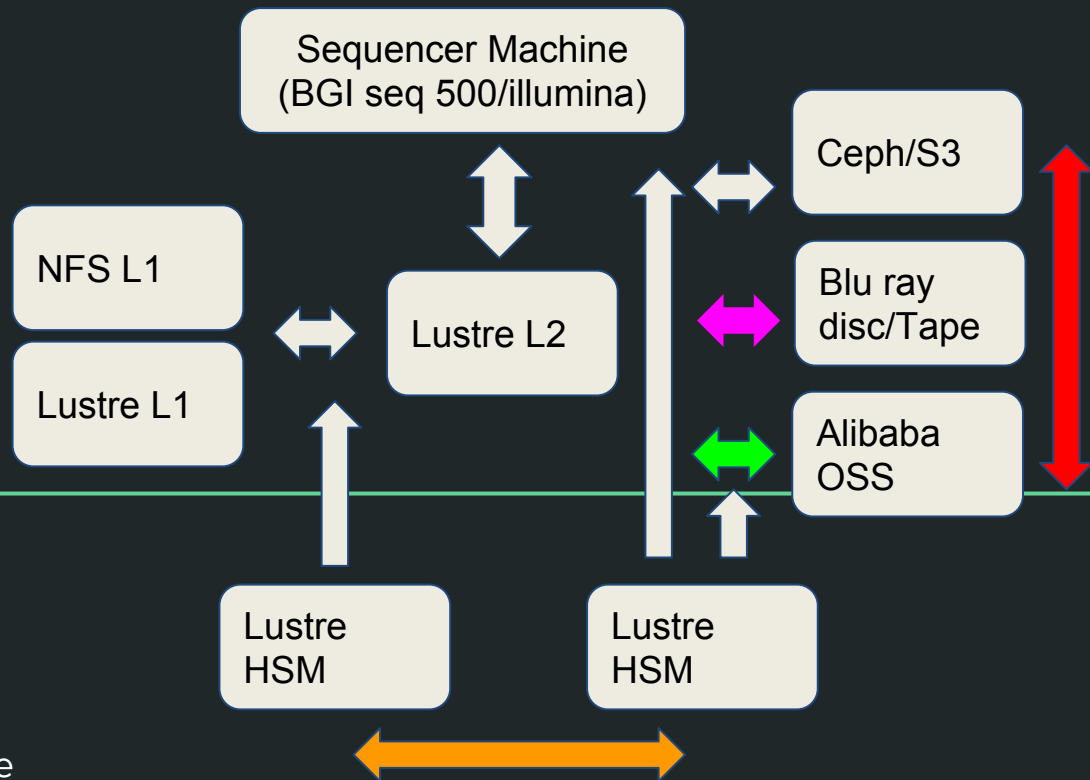


- Ali Cloud OSS

- Upload
- Download

## ● Architecture problems

- In process (swift)
- Ali OSS (that what I do)
- Could not move data from cloud to cloud, only lustre to another
- HSM database Isolation
- No Distributing DB
- No Distributing scans
- No lfs remove function



## ● Resolve our problems

- A lot of cold data automatic archive/restore to the cheaper storage media
- Analyse and group data
  - Found some of unreasonable use cases
- Replace du/find by rbh-du/find
  - MDT performance

Thanks  
2016-Shanghai

---