QoS for LOFAR

Pandey & Yan
(March 3rd, 2021)

Disclaimer

- Use cases have much more implications than QoS alone
- This presentation is a zeroth order starting point for discussion

USE CASES:-1 (LTA)

- **LOFAR-Long Term Archive (LTA)**

  - **Ingest** (Observatory => Upload Staging Area)

  - **Archive** (Upload Staging Area => Long Term Archive)

  - **Egest** (Long Term Archive => Download Staging Area)

  - **Retrieval** (Download Staging Area => End User / EOSC)

- **PROJECT/KSP/USER-LTA**

  - Steps Similar to LOFAR LTA except that Specific projects and users can initiate the first two steps Ingest, Archive. The last two steps are by definition initiated by users in general.

USE CASE:-1 (LTA)

Name	Usage	Example Media	Latency	Throughput	Comments
Ingest	Observatory => Upload Staging Area	spinning HDD (10K RPM?)	normal	high (~25Gbps)	Bulk users but not many users; Raid; Good Reliability
Archive	Upload Staging Area => Archive	Tape	High	normal	High Reliability
Egest	Archive => Download Staging Area	Spinning HDD (7200 RPM; less expensive)	normal	normal	Raid; Normal reliability
Retrieval	Download Staging Area => EndUser/EOSC	Spinning hard disk	normal	high, multiple transfers	Large number of users with a few bulk users; Raid

Upload and Download Staging Areas have slightly different requirements and thus may be separate

USE CASES:-2 (Visualisation)

- Visualisation use case:

  - Surveys - access to all stored files (e.g. Images; Source Catalogues)
  - Visualization of specific metadata properties (e.g. auto-correlation plots for each observation, RFI stats plots, Ionospheric quality plots etc.)

  Non visual cases:

  - Streaming Processing ?

  - Retrieval of log files, files containing data characteristics (station health monitoring, noise, Observation Pointing, Array configuration etc.).

USE CASE:-2 (Visulization)

Name	Usage	Example Media	Latency	Throughput	Comments
Visualisation (Survey & Source Catalog)	Datalake => EndUser/EOSC/ESAP	spinning HDD (10K RPM?)/SSD	low	high	Many users; Interactive; cross correlate surveys; Database; high reliability; Raid
Visulaisation (Metadata)	Datalake => EndUser/EOSC/ESAP	Spinning HDD (7K RPM)	HIGH	normal/low	plots with png, non-interactive; high reliability; not many users

USE CASES:-3 (tentative for discussion)

- Direct data lake **APPLICATION READ** access for processing (caching - remote, compute next to data lake location)
  - Linear access: pretty much all processing (Source finding, Spectra etc.)
  - Random access: some use cases (Pulsars?)
- Direct data lake **APPLICATION REPLACEMENT** access for processing (caching - remote, compute next to data lake location): Data size changes ?
  - Linear access: pretty much all processing (RFI flagging etc.)
  - Random access: some use cases (Pulsars/Transients?)

(those use cases could be using xcache, webdav)