

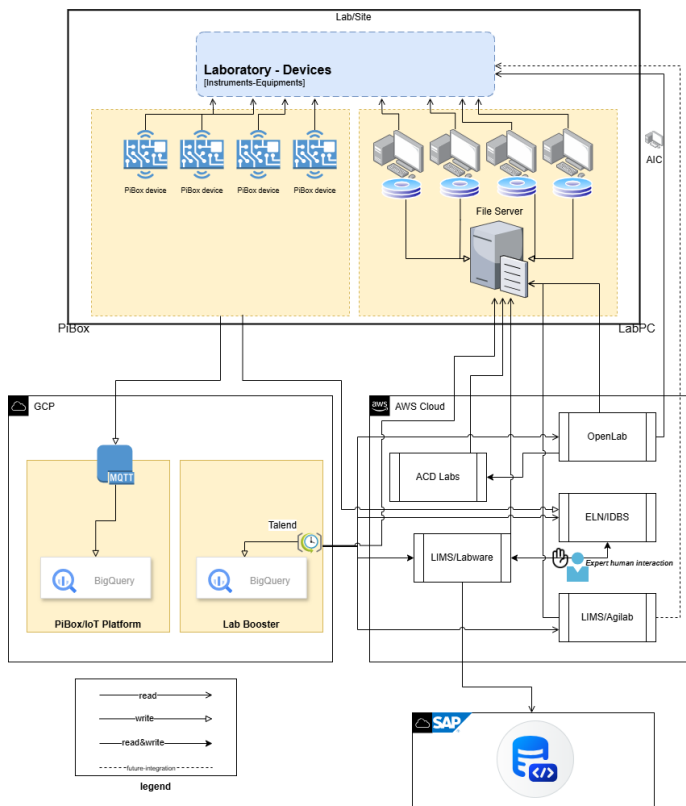
[WiP] Design Proposal for Evolution - LabPC - Storage for lab-local data at scale

Status	WIP
Stakeholders	
Outcome	
Due Date	
Owner	SAUSSOL, Olivier marie.goavec@syensqo.com
Solution/Domain/Data Architect	TRIFFAUX, Eric ERIC.PASQUALI@syensqo.com , Laurent alexandre.mathe@syensqo.com

Most of the lab sites are equipped with File Server atop iOMega NAS, which is in some cases if malfunctioning also are running at the limit for adding more hard disks to accommodate new demands for storage and been compliant with data storage regulations.

From the producer/consumer data and data logistic perspectives, LabPC disks and Local File Storage are just different levels of data buffering. They are not the place to store data for long term neither to serve with proper interoperability and scalability data to its consumers.

It is therefore essential to develop a good classification of the data captured from the instruments in terms of country regulation, access restrictions, data lifecycle, experience in data access - relationship between latency and data availability - in order to design the most economically efficient solution that guarantees business continuity effectively.



- Use cases
- Questions and Concerns
- Architectural Significant Requirements
- Impediments and Blockers
- Tradeoff analysis
- Quick-Wins
- Design Solution Proposal
- Target Architecture - R&I
- Meeting Notes
- References

Use cases

1. Lyon/RICL
 - a. OpenLab instruments generate files at hundreds GB scale for some techniques which demands local computer power to load in stand alone applications for expert analysis, this magnitude size also makes it impossible to load these files to cloud repository as well as creating difficulties in being compliant with regulatory terms for storage.
2. Shanghai
 - a. Waters instruments demands 12TB space to store application data. This storage magnitude reaches the local limitation for LabPC disk and also for local network in the iOMega NAS.
3. Bristol
 - a. Waters instruments demands 12TB space to store application data. This storage magnitude reaches the local limitation for LabPC disk and also for local network in the iOMega NAS.
4. Nuclear Magnetic Resonance:
5. Mass spectrometry:

Questions and Concerns

Architectural Significant Requirements

- Data rotation
- Retention
- Export Control/Cyber Sec
- Access Control
- Data transfer latency SLA
- Data Consumers (data structure/data model)

Impediments and Blockers

Tradeoff analysis

- Alternatives
 - XYZ
 - YZX
 - Sensitivity Points
 - Risks
 - Non Risks
 - Architectural Approaches

Quick-Wins

The criticality of the problem demands that there be at least a temporary solution as an alternative to mitigate the immediate impact of losing sensitive data for the business due to the local storage limit at the current date.

Some questions that can facilitate the analysis:

1. What type of search criteria is used on historical data?
2. What retrieval is done on this found data?
3. What type of processing is done on the found and retrieved data?
4. Is it possible to parse this data?

Possible alternatives to consider in advance:

1. Increase the storage capacity in Lab File Server - then move LabPC disk data to File Server
2. Promote historical data close to the (AWS Landing Zone) ACD Labs domain so that it can be ingested for later analysis in reports
3. Promote historical data close to the (GCP) Lab-Booster domain so that it can be ingested for later analysis in reports
4. Promote historical data to the Azure Fabric Lakehouse so that it can be ingested for later analysis in reports

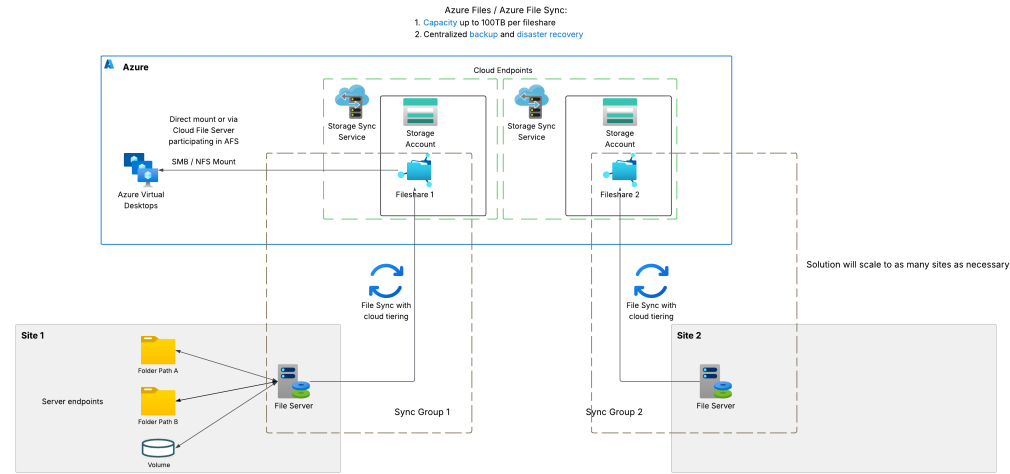
Having Azure as cloud destination for the data, a PoC could also be evaluated for a LabPC virtualization (with no connection with instruments), where only Analytics capabilities from the software vendor would be leveraged on the loaded historical data.

Design Solution Proposal

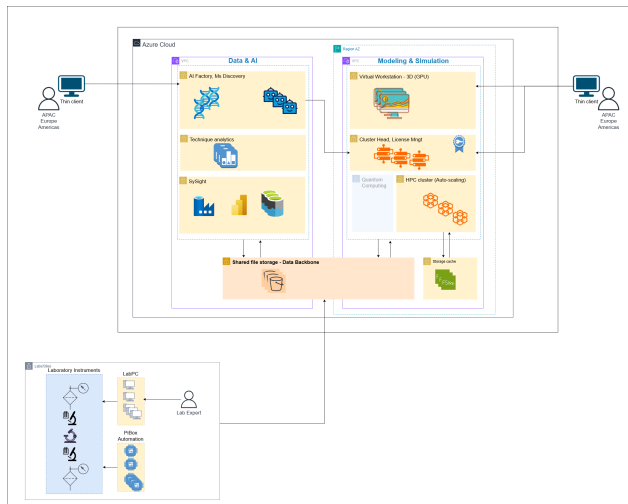
The building blocks in the following diagram are mostly from the Azure Tech-Stack for the sake of exercising concepts.

Apart from the convenience to use Azure Tech-Stack in this technical assessment, given the tight dependency among producers and consumers from the data perspective, the cloud provider has to be chosen strategically.

Azure File Sync: key component for synchronizing file from the Site to the Cloud (Azure)



Target Architecture - R&I



Meeting Notes

[SoW] LabPC - Storage for lab-local data at scale 26 Feb 2025 (Olivier SAUSSOL, Mijajlovic, Julie, Tiago Oliveira)

1. LabPC Storage Study

a. SoW

i. Assess the issue

1. Scenarios (business needs?)

- a. *Pictures* for real case situation (disks at the floor...)
- b. Worst case scenarios
- c. Application data, storage for long term, later data combination cross apps
- d. Instruments categories (data volume, SLA,...)
- e. (inventory for storage) LABPC Storage needs consolidation https://docs.google.com/spreadsheets/d/1U-d6W2LEGV9mz4XK9hBmHg1FTZY_-oYxlkXKkukgSNQ/edit?gid=0#gid=0
- f. Total storage needed for now and the forecast?

2. Impact

- a. Reaching limit of storage
- b. User storing data on inappropriate devices (NAS, - shadow it)

3. Risks

- a. 1.Data loss

- b. 2.Data steal
 - c. 3.Shadow IT
 - d. 4.Export data control/legal regulations concerns
 - e. 5.Business continuity
 - 4. Cost
 - a. Getting expensive to maintain/extend disks
 - 5. Impediments
 - a. Hard to standardize instruments
 - b. Blocked to onboard new instrument
 - 6. User complains
 - a. Bolate: 77 pcs, storage not appropriate
 - b. Shanghai:
 - c. "Mark's use case: large file generated
- ii. Evaluate Alternatives
 - 1. Meeting Stakeholders
 - 2. Business strategy
 - 3. Skills demanded for the assessment
 - 4. Solution Vendor providers
 - 5. Short Term solution (quick-wins)
 - a. Already some in place - *highlight quick-wins ongoing
 - i. Distributed storage on LabPC (disks): store on LabPC by convenience to avoid losing data
 - ii. Leveraging local File Server (*Shanghai - non standard storage)
 - iii. Cloud storage for historical LabPC data - it depends on application readiness for Virtualization deployment
 - 1. Recall AWS gateway for data - PoC (ask Khemaies)
 - 2. PoC for instrument application virtualization (pick that one with larger data volumetry; ask Olivier)
 - 3. Consider ACD Lab (as Mark Kwasnik) as the interface to run analytics - in case instrument application not ready for virtualization
 - 6. Long Term solution
- iii. Outcome
 - 1. Presentation
 - a. (1st phase)structure scenario (deadline 31 Mar 2025) [Julie, Olivier, Tiago]
 - i. touch points weekly basis meeting (30min) March (Wednesday 10:30)
 - b. (2nd phase)engage expert per domain for the solution alternatives
 - i. Decision to be taken
 - ii. Business Case
 - iii. Project organization
 - iv. Solution Architecture Proposal

References

LABPC Storage needs consolidation https://docs.google.com/spreadsheets/d/1U-d6W2LEGV9mz4XK9hBmHg1FTZY_-oYxlkXKkukgSNQ/edit?pli=1&gid=0#gid=0

Presentation https://docs.google.com/presentation/d/1UE4-b8e0GxHoxIVHJomt54JpcWwCv5Uh48Tw3qLA41o/edit#slide=id.g33d7e03e36c_0_0