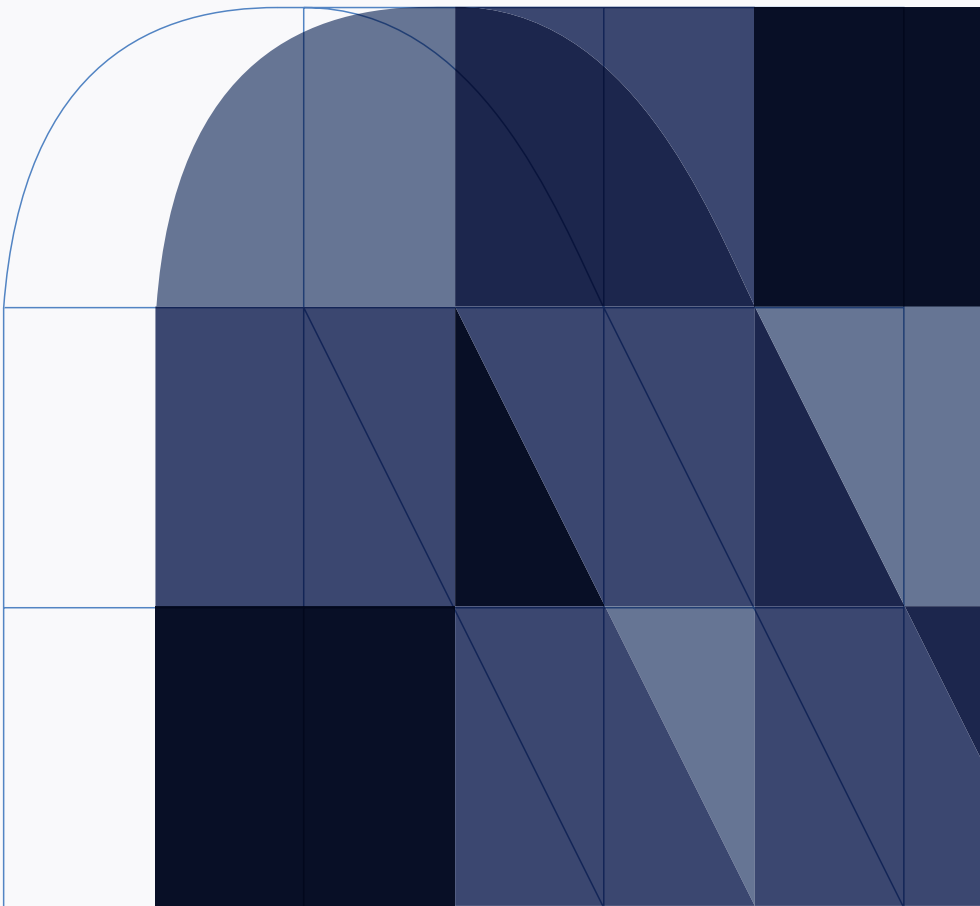
NTT DATA

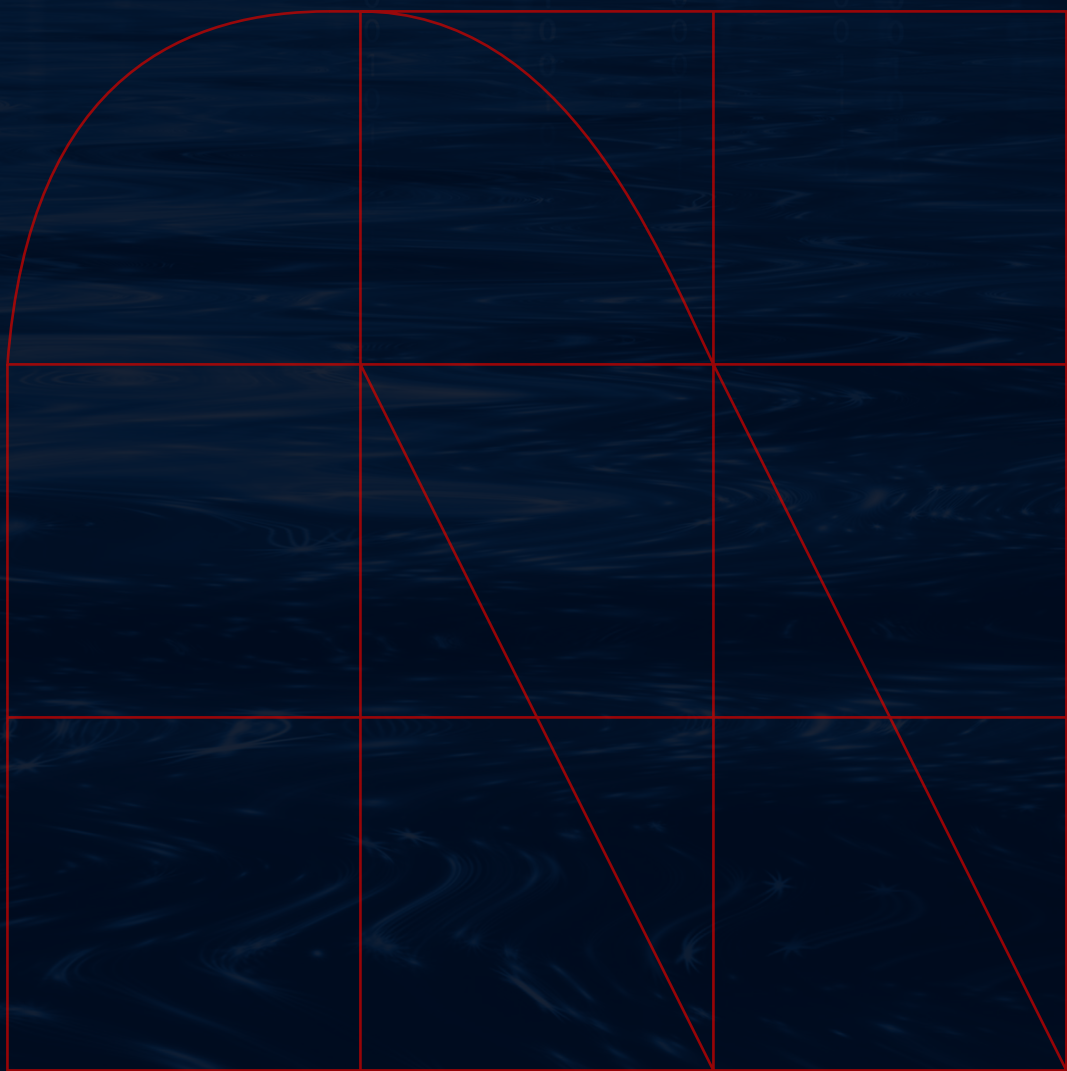
Cloud Datalake as a Service



An approach to regulatory reporting process modernization

Digital transformation is now developing and expanding in all economic and social areas. The adoption and integration of cloud platforms has subsequently become an essential part of companies' technological strategy. These new environments are not only more secure and reliable than traditional On-Premises systems but also offer a wide range of PaaS and SaaS services in continuous growth and evolution, which help companies to face new challenges with guarantees updating their business processes while reducing costs.

Moreover, Real-time monitoring helps organizations to identify the actual times an incident occurs, the reporting time, and the resolution time accurately. By identifying these times, they can get more proactive with their response methods and deal with recurring problems efficiently.



Executive summary

Nowadays, the banking industry is a very complex environment, dominated by **low return** scenarios in a rapidly evolving market. Customers are demanding **fast, cheap and innovative services** which are gradually displacing traditional banks' business.

Low interest rates for banking activities on a global scale force banks to put pressure on achieving as much efficiency as possible.

Demand for governance processes related to **new regulations** and requirements around data and its usage.

COVID-19 pandemic will have an impact on the global economy that is yet to be fully assessed.

On top of this, new players are emerging. Brand-new fintech companies and consolidated tech giants, such as Apple and Google, are postulating as candidates to provide these new services. And they're well aided by **state-of-the-art technology.** The market that once belonged immovably to traditional big banking corporations is changing indefinitely, pushing them to shift their strategy accordingly.

In this context, **cost reduction** and **technological innovation** are the new mantras for the sector, especially focusing on those processes that do not generate any direct returns or incomes in terms of business. Specifically, looking into **regulatory reporting processes,** we have detected that the banking industry faces information silos in this area with little coherence and integration. There's also poor scalability, as banks are duplicating processes and performing calculations over outdated technology that prevents a rapid response to changes and causes high technological costs.

This document addresses these issues, aiming to provide an overall view of how cloud adoption can help the banking industry to overcome the most pressing challenges of today.

Challenges to overcome

Current situation

The detected challenges are:

- Providing consistency and coherence.
- Gaining efficiency in processes and data management.
- Real-Time Monitorization of processes status and data validation.
- Consolidating and consuming information homogeneously.

Cloud Datalake as a Service approach

Adressing the challenges

To solve these issues, our proposal is a data platform based on the following guidelines:

- Public Cloud implementation.
- Flexibility first.
- Centralized data hub for all use cases.
- Use-Case driven implementation.
- Governance by design.

This has led to the DLaaS concept: a **delta data architecture implemented on the Azure** platform, relying on Microsoft's stack + Databricks.

For RT monitorization Stream Analytics and Azure Event Hub have been used.

To provide user Data Consolidation, the proposed solution is based on Microsoft's Power Platform.

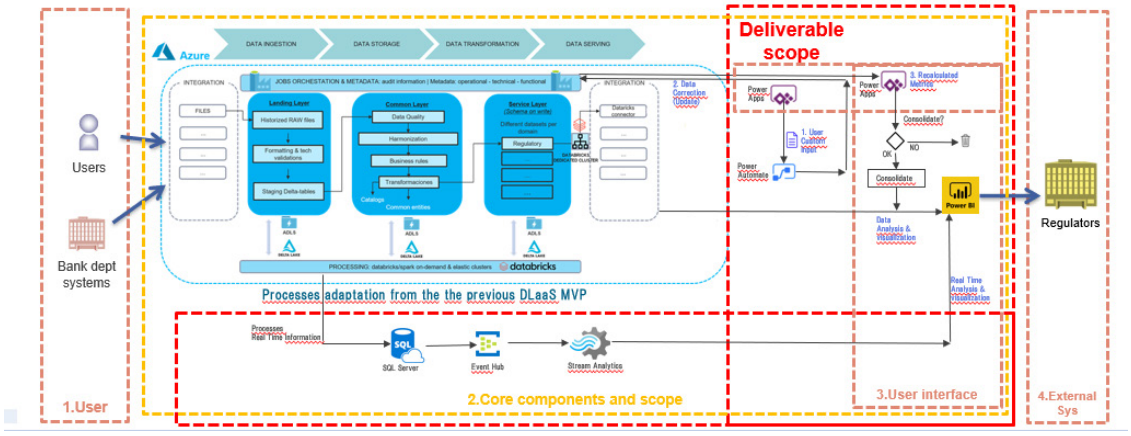
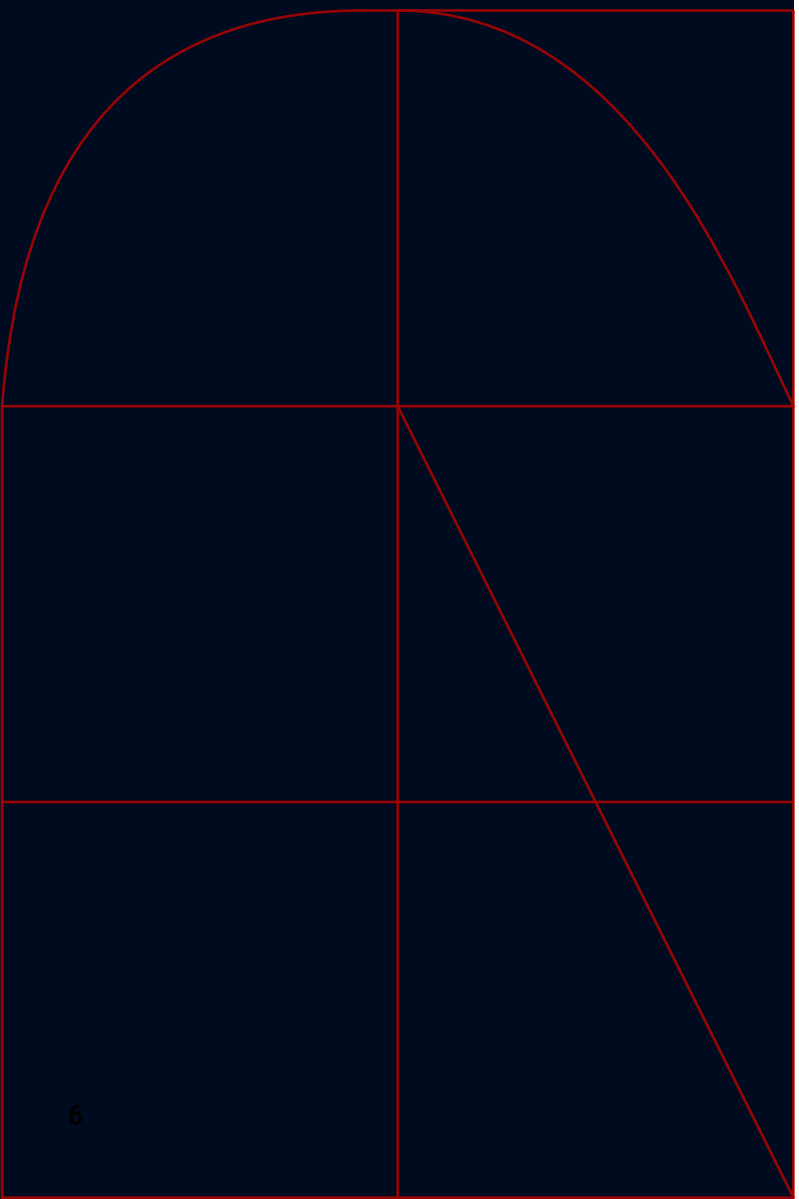


FIG 1: Regulatory reporting architecture building blocks.

NTT DATA



The main characteristics of DLaaS are:

REDUCED TECHNOLOGICAL SCOPE

To manage the solution simply, the proposal is based on open-source and well-standardized enterprise solutions such as Databricks and Power BI. These solutions provide high integration capabilities with other systems.

CENTRALIZED DATA ARCHITECTURE

Architecture with layer oriented based on data purpose and usage, allowing flexibility and full governance integration.

ARCHITECTURAL COMPONENTS FULL INTEGRATION

All the architecture components have been designed to perform segregated responsibilities and can be integrated in a highly secure manner and prepared for evolving and supporting practically any use case and additional components.

EFFICIENT COST AND CLOUD BENEFIT-ORIENTED

With this approach, costs are directly related to usage. Additionally, the architecture is horizontally scalable and elastic over commodity hardware allowing a wide range of configurations.

USE CASE COMPATIBILITY

The solution is mainly oriented to regulatory reporting, but the architecture provides core features that can integrate other components and use case typologies like real-time and advanced analytics.

LOW CODE & USER-ORIENTED SOLUTION

Components are oriented to provide visually aided applications designed for assisting non-expert users with the development of their own algorithms and analysis.

Components and tools

Regarding the technology, there are some technical features to highlight over the implemented architecture.

POWER AUTOMATE

Power Automate is a tool that allows the user to save time by automating repetitive manual tasks. Some of its capabilities include:

- Automating business processes.
- Sending automatic reminders about overdue tasks.
- Moving business data between systems on a schedule.
- Connecting to 500+ data sources or any publicly available API.
- Automating tasks on users' local computers, like calculating data in Excel.

POWER APP

Power Apps is a service for building and using custom business apps that connect to your data and work across the web and mobile - without the time and expense of custom software development.

Using Power Apps, you can quickly build custom business apps that connect to your data stored either in the underlying data platform or in various online and on-premises data sources.

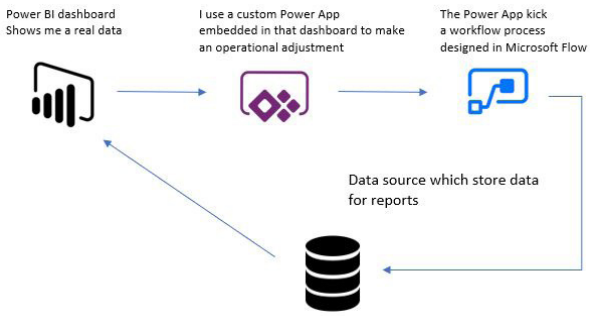


FIG 2: Power Platform example

STREAM ANALYTICS

Azure Stream Analytics is a real-time analytics and complex event processing engine that is designed to analyze and process large volumes of fast streaming data for multiple sources simultaneously.

Data patterns can be used to trigger actions and initiate workflows, such as creating alerts, providing information to a reporting tool, or storing transformed data for later use.

AZURE EVENT HUB

Azure Event Hubs is a Big Data streaming platform and event ingestion service that can receive and process millions of events per second.

Event Hubs can process, and store events, data, or telemetry produced by distributed software and devices.

Data sent to an event hub can be transformed and stored using any real-time analytics provider or batching/storage adapters.

SQL SERVER

SQL Server is a relational database management system, or RDBMS, developed and marketed by Microsoft.

Like other RDBMS software, SQL Server is built on top of SQL, a standard programming language for interacting with relational databases. SQL server is tied to Trans-act-SQL, or T-SQL, Microsoft's implementation of SQL that adds a set of proprietary programming constructs.

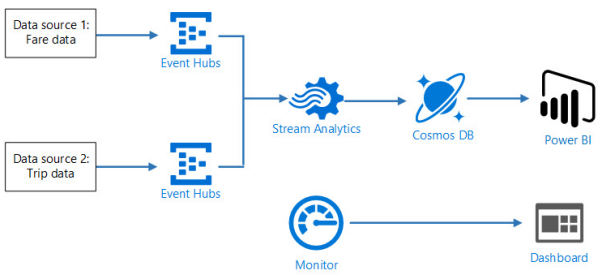


FIG 3: Stream Processing architecture example

Roadmap and evolution

Next steps

Although proposed cloud architecture covers most of the requirements, some evolutions may be of interest to evolve and provide **new features** for regulatory reporting processes modernization:

1 **Data storage optimizations** to manage the regulatory processes that use large amounts of information such as file management (bin-packing, data skipping, z-ordering, file sizing tunning), auto-optimize, caches, DFP, Bloom filters, and optimization of queries and transformation.

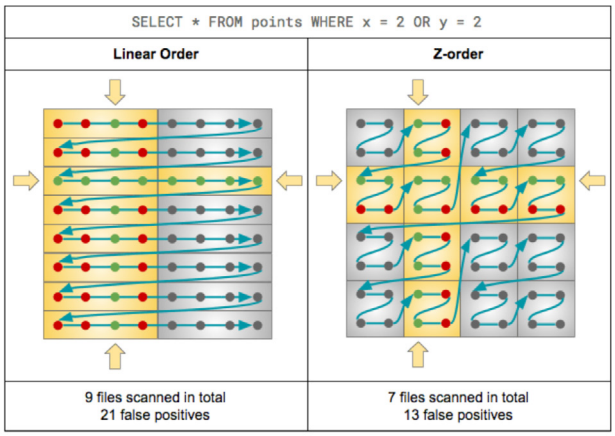


FIG 4: Z- ordering with Databricks example.

2 Databricks' specific DWH solution, as part of the Lake House approach, implements a new vectorized queries engine written in very high-performance C++, called Photon. This might be applicable for those **heavy processing calculation** cases. Thanks to Cloud platform scalability features, this capability can be easily added to the architecture.

3 **Automated application security** to keep pace with dynamic clouds and rapid software development practices. The solution could provide automated runtime vulnerability detection and risk assessment for cloud-native applications across the entire software development lifecycle and in every operating environment, including dynamic multiloads and clusters.

4 Application of **artificial intelligence and machine learning,** for the detection of problems and anomalies in the data, and the application of automated solutions in some cases, or the generation of reports for the supervision of data engineers or business analysts.

5 Application of the **Data Mesh paradigm,** under the premise of the existence of different functional domains, and the construction of data products based on three axes: code, data, and infrastructure. All of this is designed to take advantage of the scaling capabilities, flexibility, efficiency and multi-tenant solutions that the Cloud offers.

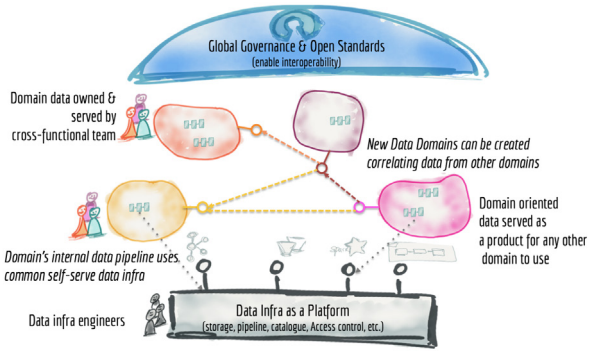


FIG 5: Data Mesh strategy example.

NTT DATA

NTT DATA Corporation

NTT DATA is a leading IT services provider and global innovation partner headquartered in Tokyo, with business operations in over 50 countries. Our emphasis is on long-term commitment combining global reach with local intimacy to provide premier professional services varying from consulting and systems development to outsourcing.