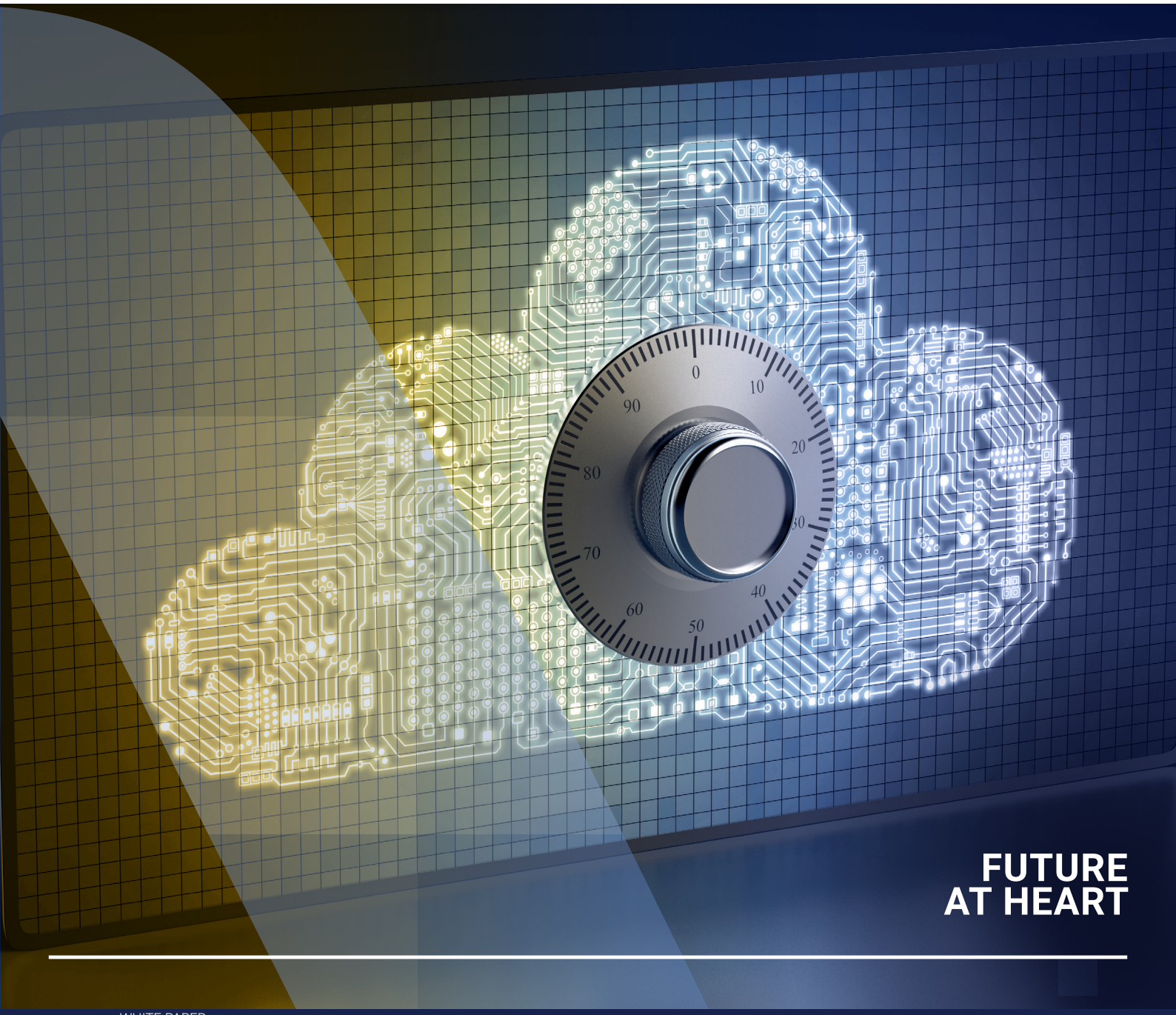
NTT DATA




WHITE PAPER

# CLOUD, BANKING & FUTURE CHALLENGES

## A FRESH APPROACH FOR REGULATORY REPORTING PROCESS MODERNIZATION

It's a time of particularly significant change in the banking sector. Digital transformation is opening fresh opportunities worldwide, and as a result, the adoption and integration of cloud platforms has become a pivotal component in a company's technology strategy. It's easy to see the advantages too: new environments are bringing with them greater security and reliability. That, and a huge range of PaaS and SaaS options continue to evolve. There's opportunity for real, reachable modernisation, and for cost savings too, and that's what we're going to explore here.

## Introduction & context

# EXECUTIVE SUMMARY

The banking industry remains incredibly complex, an environment dominated by low return scenarios in a very fast-evolving market where customers demand fast, cheap and innovative services. Services that are gradually displacing the business of traditional banks. Customers are talking about:

**Low interest rates** for global banking. That mean banks need to be even more efficient to be able to offer low rates.

Peace of mind, and reliable governance: **new regulations** and requirements around data and its usage.

**COVID-19**, the ongoing impact of the pandemic, and continued uncertainty around the global economy as a result of it.

Then there's the emergence of new names to the fintech market: global consolidated tech giants such as Apple and Google. Companies leveraging their **state-of-art technology**, and disrupting the banking sector. Against that, traditional big banking corporations are quickly shifting their strategy to respond.

**This document addresses these issues, arming you with a view of how cloud adoption can help the banking industry to overcome the challenges ahead.**

A mix of **cost reduction and technological innovation** is evidently now key for the banking sector, especially focused on processes with no direct income stream to them.

In particular, we've detected real challenges around **regulatory reporting processes**, and the way information is being managed. There's a lack of coherence and integration here, and that's leading to limited scalability, and duplicated processes. Costs are also being driven up by outdated technology that delivers slow responses at a time when the industry needs to be at its most nimble.

## Challenges to overcome

# CURRENT SITUATION

With regulatory reporting, the challenges the industry faces are:

- **Processing information as a whole, regardless of the objective of use.**
- **Providing consistency and coherence.**
- **Gaining efficiency in processes and data management.**
- **Consolidating and consuming information in a homogeneous way.**

## Cloud Datalake as a Service approach

# ADDRESSING CHALLENGES

To overcome these issues, our regulatory reporting 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.**

Which leads us to DLaaS: a **delta data architecture implemented on Azure** platform, that relies on Microsoft's stack + Databricks.

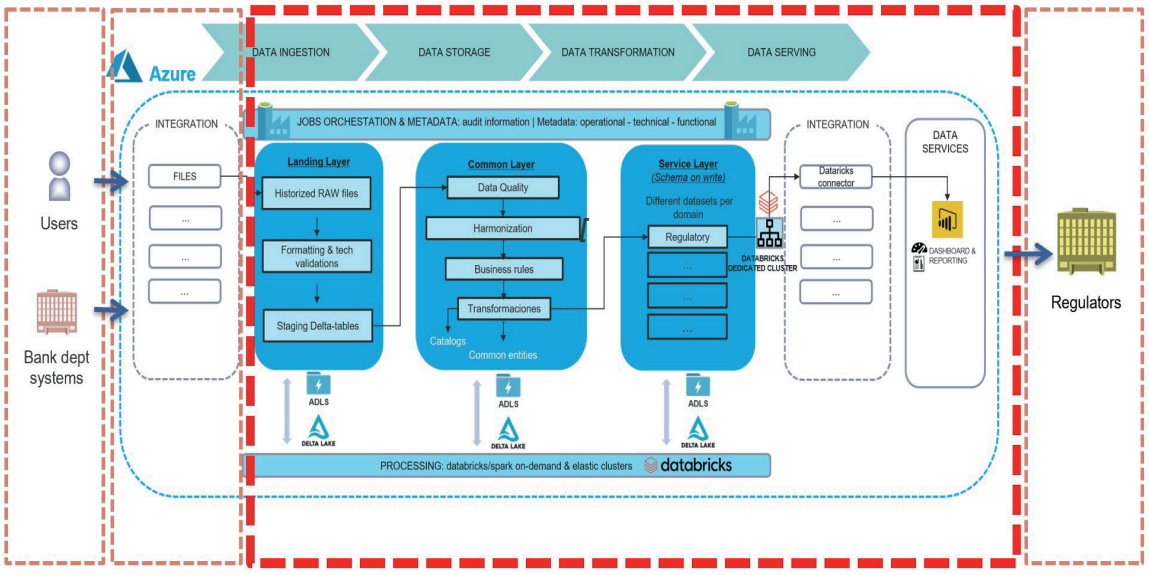



FIG 1: Regulatory reporting architecture building blocks.

The main characteristics of DLaaS are:

### Reduced Technological Scope
To keep management of the solution simple, the DLaaS proposal is based on open source and standardized enterprise solutions, such as Databricks and Power BI. These allow for high integration with other systems.

### Centralized data architecture
Layer-oriented architecture designed around data purpose and usage, allowing flexibility and full governance integration.

**To keep management of the solution simple, the DLaaS proposal is based on open source and standardized enterprise solution.**

### Architectural components full integration
All the architecture components have been designed to take on segregated responsibilities. That, and they can be integrated in a highly secure manner. It allows for easier evolution of the technology, and also the support of practically any use case and additional components.

### Efficient cost and Cloud benefit oriented
Our approach means costs are directly related to usage. Additionally, the architecture is horizontally scalable, and more flexible than commodity hardware. It allows a broad range of configurations as well.

### Use case compatibility
Whilst focused on regulatory reporting, the architecture provides core features that can integrate with other components and case typologies (such as real time and advance analytics).

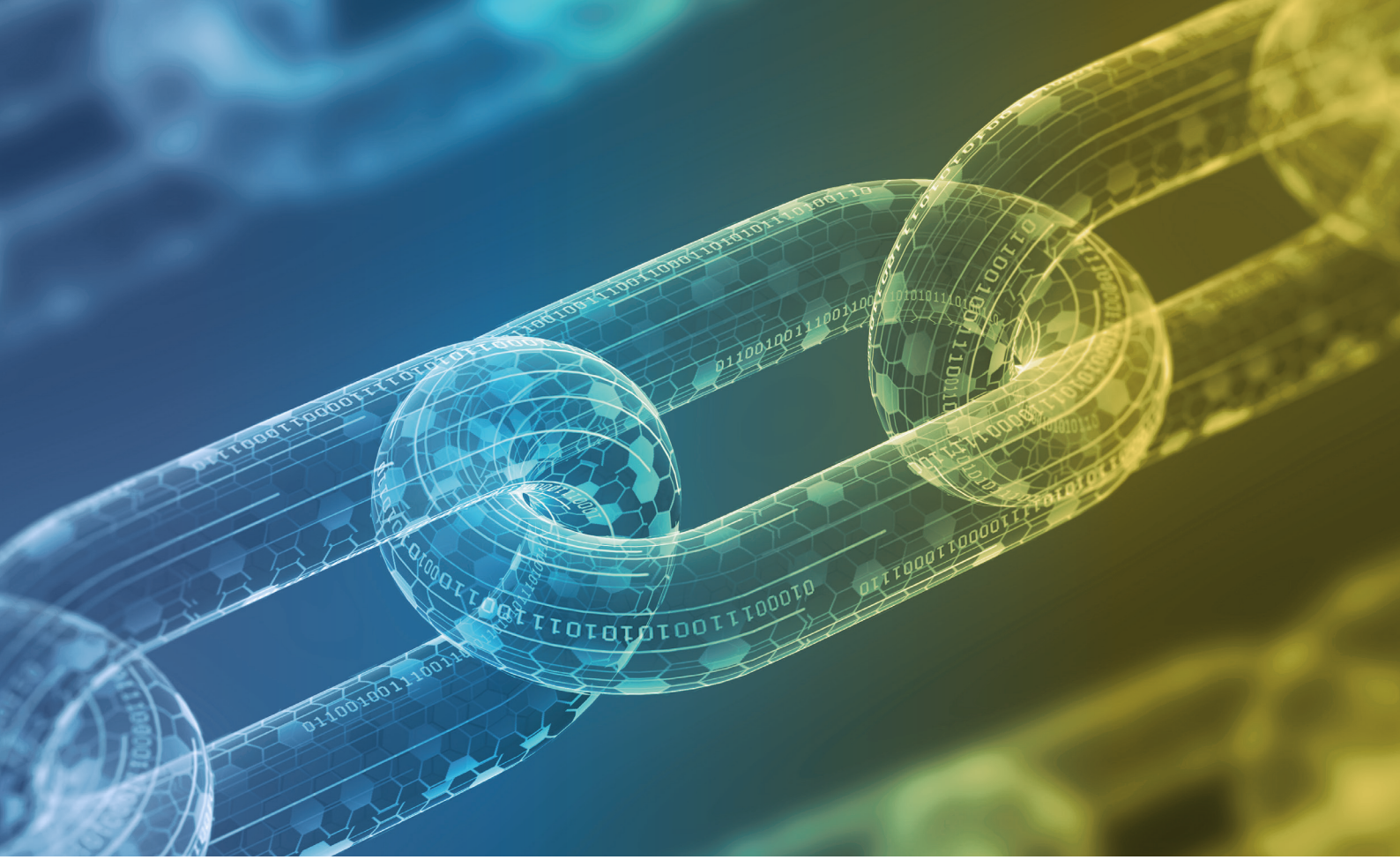
### Low code & user-oriented solution
Component parts with a focus on visually aided applications. These are designed so that non-expert users can still develop their own algorithms and analysis.

## Putting technology to work
# COMPONENTS AND TOOLS

Here, we'll highlight some of the key technical features that make up the implemented architecture.

### DataFactory
This is a tool used both as an orchestrator of the processes and also an integrator of the tools used on the platform. It's been chosen given its integration with the rest of the native Azure ecosystem - and some of the main tools on the market too - is simple and complete. As such, it effectively manages all data flows. In addition, it efficiently facilitates the development of processes, reducing the necessary code demands and increasing the use of easier to follow graphical tools for maintaining processes. The cost

**DataFactory is a tool used both as an orchestrator of the processes and also an integrator of the tools used on the platform.**



## Databricks

By deploying ADLS Gen2 in hierarchical mode (for file storage) with the Delta Lake format, it means that ACID transactional capabilities are made available. This allows users to manage and modify data in a similar way to familiar SQL-like database systems. It pulls out details of data distribution and consistency, adding audit capabilities (historicization too, with time-travelling and schema-evolution).

Focusing on clusters, we need to keep in mind that Databricks parallelizes its processes by delegating tasks to available workers. It does not have a YARN-type resource manager to optimize processes.

This is important, as it changes the way of working with respect to typical On-Premises architectures where there are resource negotiators available.

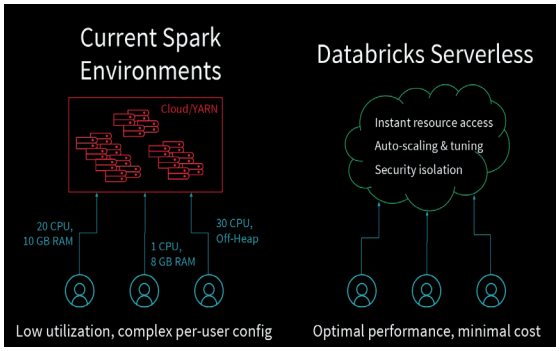



FIG 2: databricks.com

It's therefore recommended to use job clusters for batch executions and RT processes: that way, each process lifts the cluster it needs during its execution. The use of all-purpose persisted clusters is also recommended exclusively for analytical or data science use, by running notebooks and other tools that have Spark / Databricks connectors.

**The use of all-purpose persisted clusters is also recommended exclusively for analytical or data science use.**


## Power BI

Power BI Desktop is a user-friendly reporting tool that provides a straightforward link to Azure Databricks clusters using the built-in Azure Databricks connector.

It is important to choose the right data granularity for the performance and ease of use of Power BI reports. Creating a streamlined data model will help users better understand the data, and also make it easier to create useful Power BI reports.

Also, we need to take into account that nine out of ten times, poor performance is the direct result of incorrect data models or analysis expressions (DAXs), or a combination of both. As such, this is a core point to address and be aware of.

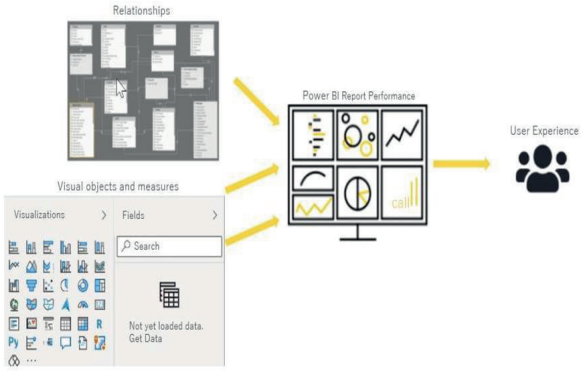



FIG 3: microsoft.com

**Power BI Desktop is a user-friendly reporting tool that provides a straightforward link to Azure Databricks clusters using the built-in Azure Databricks connector.**

## Roadmap and evolution

# NEXT STEPS

**Although our proposed cloud architecture covers most requirements, there are some further evolutions to consider. These may be of interest to those looking to evolve and provide new features for regulatory reporting processes modernization:**

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

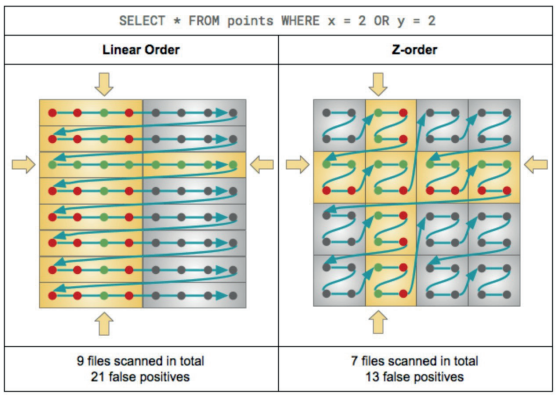



FIG 4: databricks.com

**2** Databricks' specific DWH solution - which is 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 **heavy processing calculations** cases. Thanks to the flexibility of Cloud platform scalability features, this capability can be easily added into the architecture.

**3** As part of the use of Delta Lake, Real-Time and Batch flows can be easily integrated in the current cloud architecture to enrich regulatory reporting processes. The advantage is that **there is no need to include additional tools**, nor to implement separate process and governance depending on the nature of the data.

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

**5** Application of the **Data Mesh paradigm**: This is under the premise of different functional domains existing, and the construction of data products based on three axes: code, data, and infrastructure. All the while taking advantage of the scaling capabilities, flexibility, efficiencies and multi-tenant solutions that the Cloud offers.

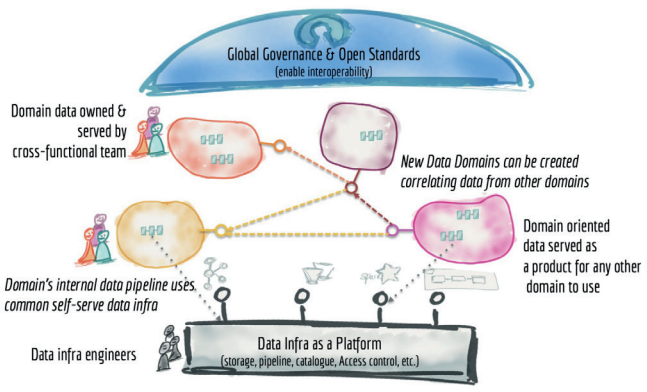



FIG 5: martinfowler.com


## For More Information

To find out more about how we can help your organization, contact your NTT DATA representative or visit:
**www.nttdata.com**