

Why the time is right
to implement a
cloud data lakehouse



Gold
Microsoft Partner

Executive summary.

Human beings are great at gathering information. According to IDC, the world's store of digital info will grow by more than 60 per cent by 2025 — reaching 175 zettabytes. That's an unfathomable amount of data, and at least half of it will live in the cloud.

As good as we are at capturing, creating, and storing it, our ability to find the correct data and get it to the people who need it – when they need it – is stalling.

We have too much data in too many places in too many formats and environments. Hybrid cloud environments make it harder to access it when you want it; even harder to apply the necessary security controls and governance policies that regulators and customers demand.

- Gartner says that by 2025, 80% of organisations seeking to scale digital business will fail because they haven't modernised their approach to cloud data management and analytics.
- A survey by Palantir found that more than 95 per cent of businesses still struggle to manage unstructured data.

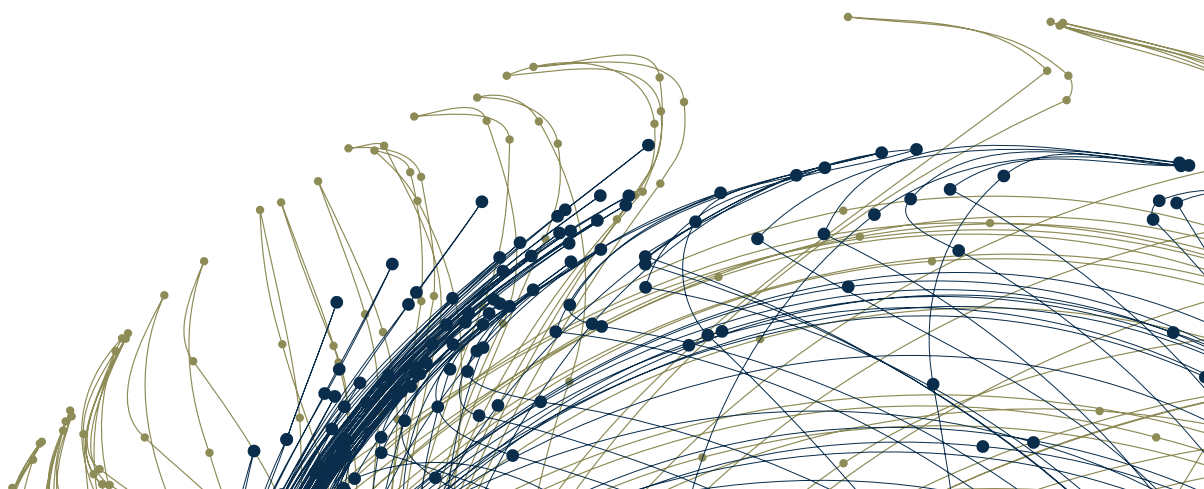
Organisations need to be able to pull value from the exploding array of information held across their systems. For many, that will mean changing how they manage data in the cloud.

It needs an approach that's nimble enough to handle the workloads created by expanding data volume and complexity, reflect the requirements of analytics, and ensure everything captured is delivering value back to the business.

Integrating data, knowing where it sits, and understanding its provenance requires numerous checks and data quality rules. Using the capabilities of world-leading cloud data platforms to automate and execute those requirements at scale is more critical than ever.

This paper will explain how a cloud data lakehouse, enabled by solutions from Microsoft Azure and Informatica, can improve data management and produce the kind of analytics that help businesses make better decisions faster.

We'll also explain our own phased methodology for making the migration of data to the cloud seamless.



Making the case for the cloud data lakehouse.

By the end of next year, Gartner says that over half of enterprise data will be created and processed outside of data centres — and outside of the cloud. It's no mystery then why organisations have begun to struggle with discovering, understanding, and trusting their data quality.

Legacy IT infrastructure often lacks the ability to sort through data and turn around insights quickly. To overcome that, many organisations have invested in cloud data warehouses or created data lakes.

But those models have begun to buckle under the strain of exponential data growth and complexity.

- Cloud data warehouses are optimised for structured data but struggle to handle unstructured and semi-structured data.
- Data lakes were designed to overcome those limitations and accommodate data in various formats. Still, lack of support for transactions and an inability to enforce data quality means data lakes can't keep up with modern analytics requirements.

So a new model has emerged that combines the best aspects of both: the cloud data lakehouse.

The lakehouse combines the best elements of data lakes and data warehouses, blending data structures and data management features similar to what you'd get in a data warehouse, and applies them in the same low-cost storage used for data lakes. The result is a single repository for data of all kinds that can be read and augmented by machine learning and AI.

The lakehouse offers enterprise-scale data integration, data quality, and metadata management suited to today's analytics. It also enables today's governance requirements by managing discovery, cleansing, integration, protection, and reporting across cloud environments.

Successful implementation of a cloud data lakehouse can deliver the kind of insight-led, data-driven decision making modern digital businesses need to compete and survive — all the more reason to trust data migration and lakehouse cloud infrastructure to the right vendor.



Key considerations for a cloud data lakehouse implementation.

What is the current architecture? First, you'll need to look at your existing architecture and understand its current pain points.

If your analytics team is spending too much time processing and prepping data and not enough conducting value-add analytics, that's one sure sign that you need a new approach to data management.

What are the limitations of the current system? Some legacy environments make it difficult to extract data and understand it, which complicates the creation of effective data models for analytics.

What are the use cases you need to enable? It's essential to define problem statements and be clear on the tangible business benefits you want to achieve. Resolving some data management issues can involve a lot of effort and budget. Will it help increase sales? It's important to know which of the use cases are worth investing in and prioritise them according to expected short- medium- and long-term ROI.

How do you select the right vendor?

✓ Price

You want the cost of implementation to be low, as well as cloud consumption costs.

✓ Agility

Ask if the vendor can ingest large data sets quickly and get you to a minimum viable product so you can prototype analytics and see it demonstrated on the platform.

✓ Flexibility

Can the vendor's platform store and catalogue structured, semi-structured, and unstructured data types? To maximise the uptake and value of analytics, it's also important that data sets be shared easily across the organisation.

✓ Scalability

Data storage needs to be elastic and accommodate rapid scaling at the lowest possible cost.

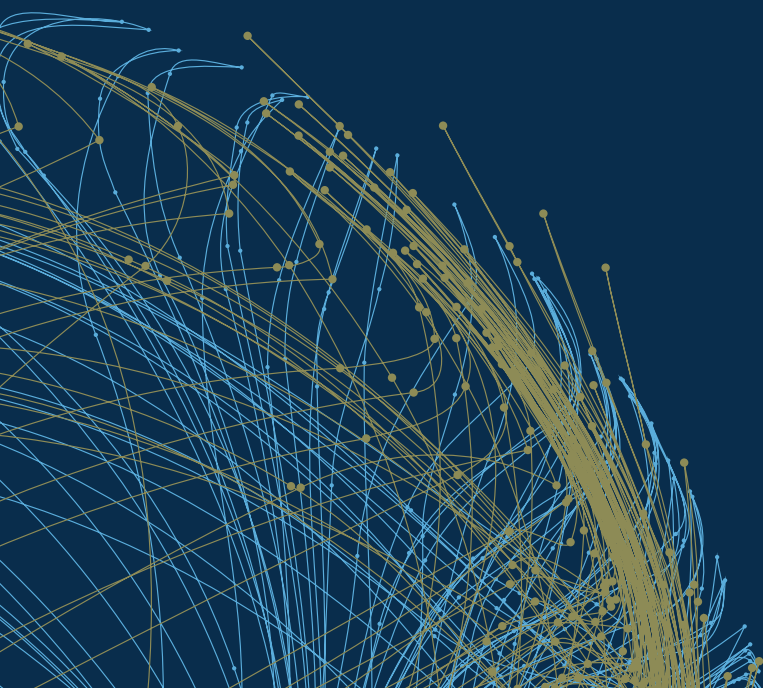
✓ Functionality

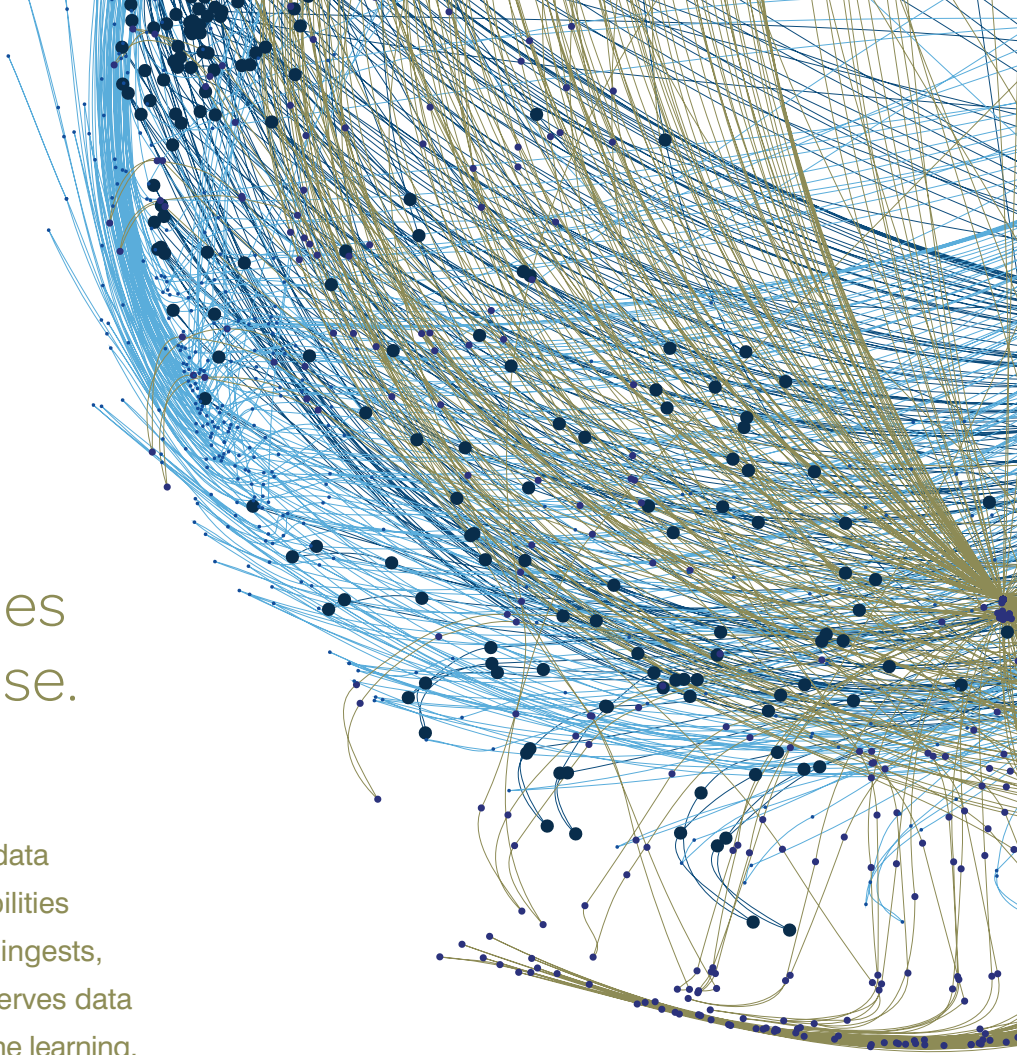
How well will the cloud platform integrate with existing IT architecture, and will it make data governance requirements easier to deliver.

✓ Future-proof

Ensure the vendor's platform has a roadmap that includes new automation options and improved integration with machine learning tools.

One vendor consistently ticks all these boxes: Microsoft Azure.





How Azure enables the data lakehouse.

With Azure, enterprises can bring data warehouses and data lakes' capabilities together for a unified platform that ingests, stores, processes, enriches, and serves data for business intelligence and machine learning.

Creating a data lakehouse and implementing applications like Microsoft Azure Synapse Analytics and Microsoft Power BI delivers on essential business requirements — the first being cost. Azure Synapse can deliver price-performance that's 94 per cent less than other vendors.

It also brings together data integration, enterprise data warehousing, and big data analytics at cloud scale. Unifying those workloads can significantly reduce development time and speed up access to usable insights.

Looking to the future, Synapse has already integrated machine learning capabilities. Data engineers working in Synapse can use the templates in Microsoft Azure Machine Learning's central registry to build new models and start using predictive analytics faster.

Of course, before these benefits can be realised, the data has to be migrated from current repositories into the new cloud environment. Agile Solutions recommends a phased approach based on our proven methodology.

A phased approach to cloud data migration.

Our cloud data migration methodology enables customers to break free from the inertia that technical complexity can bring and accelerate the move to a data lakehouse.

It's based on learnings obtained from hundreds of implementations but flexible enough to be tailored to each organisation's individual needs and objectives.

This methodology has been proven effective whether businesses need to migrate a single workload or thousands. It aims to set in place a data framework that suits the lakehouse model and enables AI-driven analytics.

Agile's methodology has four key phases.

01



Assess

Are you really ready to undertake cloud data migration? This is the foundational phase where we consider current readiness for operating in the cloud, looking at existing applications, data, infrastructure, and licensing. Then we look at the business objective for the migration, identify the workloads that underpin it, and decide what their priority should be.

02



Mobilise

Once we've clarified where an organisation is starting from and where it wants to get to, we move to the second phase and decide how we're going to achieve our goals. The mobilise phase provides us with an opportunity to lay the foundation for tooling, processes, and any cultural changes required to accelerate the migration at scale.

03



Migrate and modernise

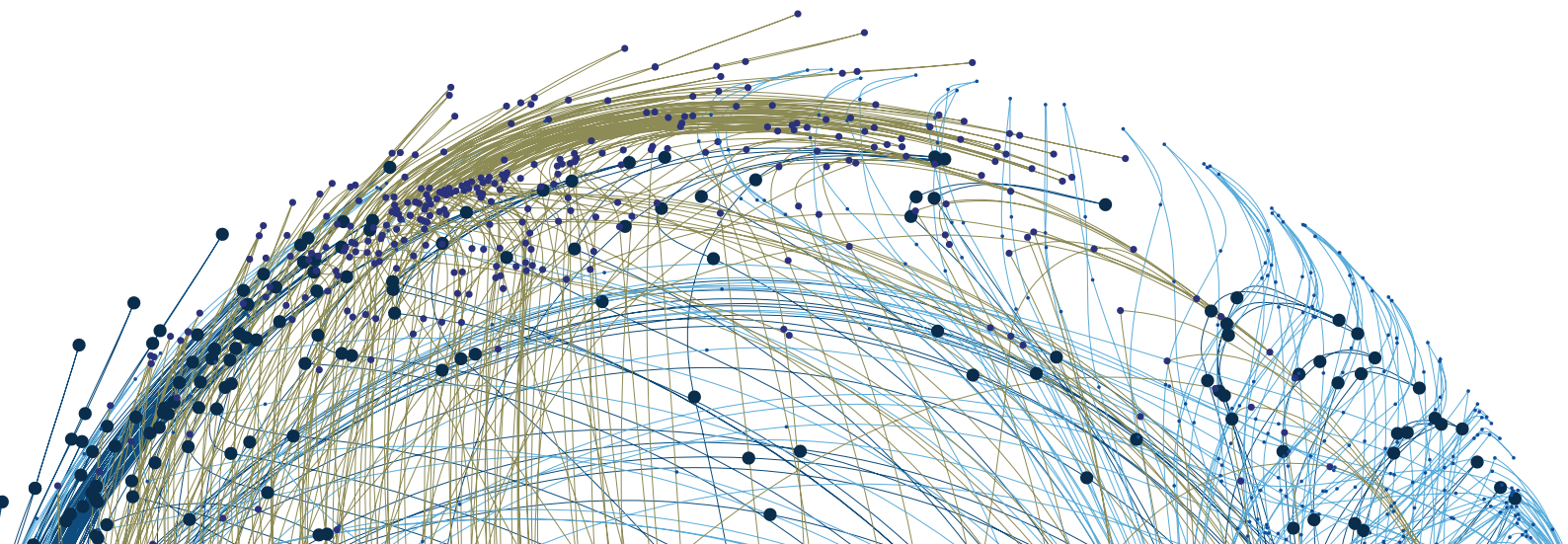
Now we can put what we've learned in phases 1 and 2 to work. We work with our customers in the migration phase to execute the migration and implementation plan developed during the earlier phases. This covers data, infrastructure and applications. Of course, each migration is different in detail, but they tend to follow well-known patterns, which allows us to leverage best practices.

04



Manage and monetise

Once the lakehouse has been established, and data quality is where it should be, there may be an opportunity to turn your business information into revenue. Agile will help you assess the value of your cloud data assets and identify cloud platform services that can create opportunities.



How Informatica's iPaaS solution complements Azure.

While many companies choose Azure as their cloud platform, 93% of enterprises are on a multi-cloud footing. Since Informatica is platform-agnostic, as a tool to manage cloud data migration, it pairs neatly with a varied infrastructure approach. In fact, it's commonly requested by Agile clients for demos and proof-of-concept work.

Informatica is the leading integration platform as a service (iPaaS) vendor. It has been rated number one by Gartner for seven years running, and its cloud data integration offering is one of the most modern and comprehensive available.

The Informatica iPaaS platform is cloud-native, microservices-based, API-driven and AI-powered. Informatica also uses an AI engine called 'Claire', which uses metadata and machine learning to accelerate time to value. Large enterprises, businesses with hybrid environments, and organisations who may have used Informatica's PowerCenter solution in the past will find that migrating cloud data with Informatica will accelerate the move to Azure.

Globally, the Informatica cloud supports more than 15 trillion transactions every month and holds the industry's highest number of security certifications. It offers a broader range of connectors than any other iPaaS vendor and works seamlessly with Microsoft Azure solutions.

A unified infostructure for digital business.

To take advantage of data and turn it into actionable insights, today's organisations need a unified repository that can manage data across environments and then act as a mechanism for delivering insights and information to users and applications.

A cloud data lakehouse delivers fast and contextualised access to the data businesses need. That way, they can generate analytics that point to changes in customer behaviour, investment opportunities, activities with a higher propensity for risk, and emerging market trends.

By pairing the tools offered by Microsoft Azure and the Informatica cloud, enterprises can achieve an immersive analytics experience from pipeline to visualisation.



About Agile Solutions

Agile Solutions GB is an independent consultancy focused on deriving value from data. We work with cloud software leaders like Microsoft with our Gold/Silver partner status and Informatica where we have Platinum partnership status.

Our mission is to help our clients achieve their IT and business objectives and become more agile as they react to the changing needs of a digitally-driven economy. Our aim is to create tangible benefits for your business.

From our experience, we recognise that the quality, order, accessibility and presentation of data is key to any forward-looking business in a globalised market. We make data work harder for you by helping you manage it, monetize it, leverage it and make better use of it.

Some of the world's largest companies rely on us for information management applications, technology, and support services. Want to learn more?

Get in touch:

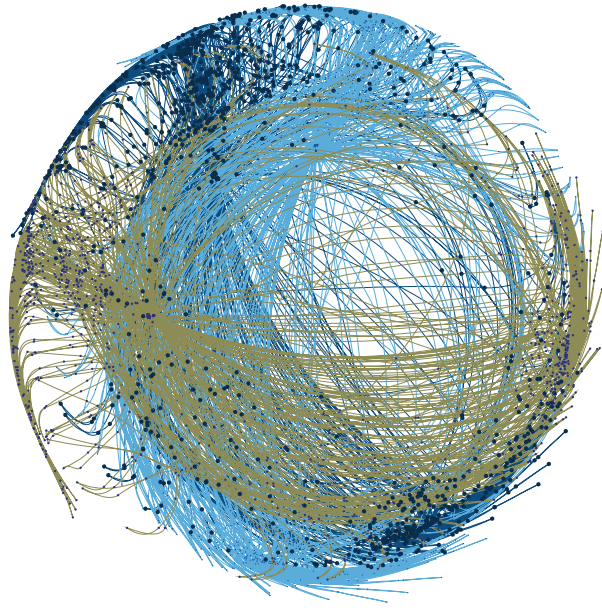
+44 (0)190 801 0618

info@agilsolutions.co.uk

www.agilesolutions.co.uk

agile
solutions

Breaking ground
for a cloud data
lakehouse



Top tips

Don't boil the ocean. Given the breadth of capability available in MS Azure and Informatica, find the critical use case and deliver it first. You can springboard onto new use cases later and add value incrementally.

Collaborate. Capitalise on the collaborative environment inside Azure Synapse Studio by bringing together your engineers, analysts and data scientists to harmonise operational models

Don't gamble. Execute your cloud data migration using a proven, repeatable customisation process

The right tools for the job. Use the proper tooling and apply it according to best practice.

Automate. Automate. Automate.