

Bringing the Promise of Data to Life

A three-part strategy to realize more data value at petabyte scale





INTRODUCTION

The Business Need for Cloud-Based Data Warehousing

It has become more than clear—business is moving to the cloud.

The number of companies using the cloud for data warehousing or business intelligence (BI) has increased nearly 50% since 2013.¹ And those companies are reaping the benefits.

The cloud provides better insights and analytics through easier access to—and management of—data. In the cloud, companies are also able to quickly meet fluctuating business demands and lower costs by instantaneously scaling resources up or down without wasting money on excess capacity. It's no wonder that this kind of performance is fast becoming table stakes for business decision-makers.

Even with all these advantages, in a recent *Harvard Business Review* survey of global executives, 80% of respondents say their organizations are struggling to become mature users of data and analytics, and 45% say their organizations operate on-premises IT systems that are at or near performance capacity.²

To stay competitive, companies need a way to simplify and accelerate the process of migrating their existing data warehouses to the cloud.

¹ "The Total Economic Impact™ of Microsoft Azure Analytics with Power BI," a Forrester Total Economic Impact™ Study, commissioned by Microsoft, October 2019.

² "Understanding Why Analytics Strategies Fall Short for Some, but Not for Others," a Harvard Business Review Analytic Services Pulse Survey, commissioned by Microsoft, 2019.



So, how do you dismantle data silos and capitalize on the massive computing capacity of the cloud—without interrupting critical business data flows?

Simply shifting your data warehouse “as-is” to the cloud leaves you with inefficient redundancies and legacy systems that may soon be outdated. That translates to thousands of tables and files that IT must manage without any hope of return on the effort.

Instead of an all-at-once approach, take a well-governed, incremental route to migration that focuses first on high-value data that can deliver business results in priority areas, with minimal disruption to existing applications and analytics workloads.

In this e-book, we’ll present a three-part strategy to launch a new, petabyte-scale data warehouse on Azure (in your choice of regions) in the fastest and most effective way possible.

The key parts of the strategy are:

- 1. Discovery**—Understand the business objective and the intent of data—and ensure that you maintain data lineage so that you don’t break anything in the migration process.
- 2. Integration**—Take an intelligent, automated approach to connecting and migrating data from a variety of sources.
- 3. Platform**—Choose a modern data-warehouse environment that fully enables and complements your analytics-processing activities.

PART I

Discover and Understand: the Modern Approach to Discovering the Right Data

Start by prioritizing resources for the most valuable workloads.

The goal is to find the fastest, most cost-effective, and least disruptive path to extending or migrating your on-premises data warehouse (such as Netezza or Teradata) to a world-leading cloud environment like [Microsoft Azure Synapse Analytics](#) (formerly known as SQL Data Warehouse or Azure SQL Data Warehouse).

To that end, an intelligent data catalog like [Informatica Enterprise Data Catalog](#) helps you **cut the time it takes to find and prioritize the right data for cloud data-warehouse and analytics projects by as much as 90%**. Its downstream-dependency identification and lineage tracking enable you to identify sensitive data and determine if there's any risk in moving it, while minimizing disruption.

Given the dependency requirements, even incremental migration is not simply a copy-and-paste job. It can be time consuming unless you can automate it through intelligent data discovery that catalogs and identifies all types of data and data relationships.

Make sure your intelligent data tool uses AI to:

- ✓ **Discover and catalog any data asset** with automatic scanning across on-premises, cloud, big data, databases, data warehouses, BI platforms, applications, and structured and unstructured data types.
- ✓ **Curate data using AI-powered automation**, including automatic identification of data domains and entities (such as orders and invoices), data similarity, and intelligent recommendations.
- ✓ **Establish end-to-end data lineage** in order to identify and prioritize what data to migrate from a legacy data warehouse to a modern cloud data warehouse.
- ✓ **Enable collaboration** across the organization to select the best data by harnessing the combined power of human expertise, social curation (such as data-set ratings and reviews), and AI-driven insights.
- ✓ **Provide an enterprise-scale solution.** Look for a machine-learning-based data catalog that can scale to classify and organize tens of millions of data assets across any environment throughout the enterprise.
- ✓ **Support open APIs** and leverage metadata APIs to easily integrate into your environment and consume catalog content anywhere.
- ✓ **Work seamlessly with cloud data-integration tools** to automate data-provisioning workstreams to provide you with a more comprehensive view and control of all your data sources.

Data Integration as a Service—A Strategic Approach to Migration

Once you identify the data you want to move and its dependencies and linkages, you can use a modern integration platform as a service (iPaaS) to connect and integrate the data sources and help you move the data to [Azure Synapse](#).

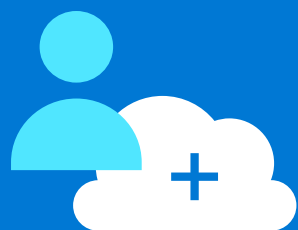
[Informatica Intelligent Cloud Services \(IICS\)](#), the next-generation iPaaS, facilitates easy ingestion of a variety of data sources to derive the maximum benefit.³ This is best enabled by a single solution for self-service data integration, synchronization, and management.

Combining all of these capabilities increases the efficiency of managing data-integration tasks.

Key questions to consider when evaluating an iPaaS solution include:

- ✓ Does it easily integrate cloud, on-premises, and third-party data systems such as Oracle, Netezza, and Teradata, and applications such as Salesforce and Marketo?
- ✓ Does it provide preconfigured mapping templates for common integration use cases, with the ability to scale data-integration tasks?
- ✓ Does it include prebuilt task-based wizards to simplify replication and synchronization tasks?
- ✓ Will it use intelligent data-structure discovery to automatically parse complex files?
- ✓ Does it combine batch and real-time data integrations with task-flow orchestration?
- ✓ Will it provide mass data ingestion for high-volume streaming data, files, and databases with change data capture?
- ✓ And finally, does it deliver high-quality, clean, trusted, and secure data?

³ "Cloud Data Warehousing with Microsoft Azure Workbook," Informatica and Microsoft, 2017.



The high-level benefits of Informatica Intelligent Cloud Services (IICS)

With IICS you get:

- **All your data integration, application integration, and data-management needs managed** with a single, integrated iPaaS solution.
- **Ingestion of large data volumes** from stream, file, or database sources with change data capture and enterprise data integration.
- **Rich out-of-the-box application integration** with connectivity and event-driven integration and a single-canvas design tool.
- **Fast identification, repair, and monitoring** of data quality with prebuilt business rules.
- **Integration with a data catalog** to enable data search and discovery within IICS, and to leverage data-source connectivity information to simplify data mappings.

Connecting the dots from day one

Right out of the box, a solution like IICS easily connects all your data to Azure Synapse sources, whether they're in the cloud or on-premises.

This immediately replaces cumbersome manual management with high-performance automation, giving you a well-integrated data-discovery and data-management platform. It offers:

- **A metadata-driven solution** with role-based tools that provides data visibility and a single point of control for end-to-end production data flows.
- **Development agility** to accelerate time to market with easy-to-use ingestion and integrations wizards, without the need for any coding.
- **Deployment scalability and operational confidence** that includes support for both ETL and ELT patterns for end-to-end high-performance data pipelines, as well as support for Spark serverless processing for on-demand scalability.
- **Delivery of high-quality clean, trusted, and secure data** that's scalable, with enterprise-class data quality, data security, and governance solutions to address any size and format of data, platform, or technology.

These capabilities accelerate the deployment and delivery of trusted data in the cloud, on-premises, and across hybrid environments.

Add the Power of Azure Synapse

When considering the environment in which to establish your enterprise data warehouse, we use and recommend Azure Synapse for its best-in-class price performance, which provides:

- A solution that's up to 14x faster and costs 94% less than other cloud providers.
- Massively parallel processing to aid initial data load and reduce load time.
- Elastic cloud computing resources and modern infrastructure to fully enable and complement your analytics-processing activities.⁴
- Combined enterprise data warehousing and big-data analytics in a single service.
- A choice of either serverless query-as-a-service or provisioned resources for data warehousing to empower data professionals to query data on their terms.



Simply put, Azure Synapse is the next evolution of Azure SQL Data Warehouse.

Microsoft has taken the same industry-leading data warehouse to a whole new level of performance and capabilities. In fact, it's the first and only analytics system to have run all TPC-H queries **at petabyte-scale**.

For intelligent workload management, it separates storage and processing, providing greater flexibility and scalability to support multiple workloads and use cases. It also uses adaptive caching to accelerate query performance and query concurrency.

In addition to speed, elasticity, and efficiency, Azure provides industry-leading, defense-in-depth security and a financially backed 99.9% availability SLA. It keeps you in control with a comprehensive suite of security capabilities, including data protection, fine-grained access control with column- and row-level security, and built-in threat detection.



Azure provides industry-leading, defense-in-depth security and a **financially backed 99.9% availability SLA**.

⁴ "Cloud Data Warehousing with Microsoft Azure Workbook," Informatica and Microsoft, 2017.

Putting It All Together



A successful enterprise data-warehouse migration to Azure includes three key elements:

- 1. Intelligent data cataloging** that can reduce the time it takes to find and prioritize the right data for cloud data-warehouse and analytics projects by as much as 90%. You get machine-learning-based data discovery to help create a manageable data-migration roadmap while ensuring holistic visibility into data lineage.
- 2. Integration platform as a service (iPaaS)** that enables more efficient management of data ingestion and data integration. This hybrid data access and processing model allows you to run data integration jobs on-premises, in the cloud, or in a hybrid environment.
- 3. Scalable, agile, high-performance data warehouse** with big-data analytics in a single service that offers faster, more elastic data capabilities, as well as a more convenient end-user experience.

You also get a best-of-breed, industry-leading solution on Azure, the platform used by 95% of Fortune 500 brands. Informatica is the only enterprise cloud data management leader in five Gartner Magic Quadrants in 2019.⁵

All this adds up to disruption-free, incremental migration without high costs. And prebuilt integration means simplified scheduling, loading, transforming, and managing of trusted data.⁶

⁵ 2019 Magic Quadrant Reports, Gartner.

⁶ "Changing Landscape: Microsoft Adopts an Open Approach," Nidhi Singal, Business Today, May 8, 2019.

Data Warehouse Due Diligence

Make sure your partner does more than check the boxes

Take a closer look at your data warehouse options. If they can't provide all these critical elements, look into Informatica:

- ✓ **Integrated AI-driven self-service data cataloging**, cloud data integration, data warehousing, and visualization.
- ✓ **Simplified transformation**, synchronization, and scheduling of data to migrate to the data warehouse in a few easy steps with easy-to-use self-service tools.
- ✓ **A hybrid data management platform** to take advantage of existing on-premises investments and skill sets in the cloud.
- ✓ **Native portal integration** with streamlined migration to Azure Synapse.
- ✓ **End-to-end self-service data management** with governance solutions.
- ✓ **Connection of all Azure data services** to 200+ external cloud and on-premises data sources rapidly, with prebuilt native connectivity.
- ✓ **Enterprise-scale performance** with advanced capabilities and simple interface.
- ✓ **Fully supported deployment** on Azure Synapse.



Data Warehousing on Your Terms

Discover first-hand how our strong and long-standing partnership with Microsoft accelerates your digital transformation. We're prepared to help you intelligently lead—in any sector, category, or niche, with a selection of industry-leading offers.

The next step is yours to choose:

Where do you stand? Find out from Microsoft and Informatica.

Take off for the cloud with a free evaluation, free code conversion, and free proof of value.

SIGN UP NOW

Get the tools to envision your cloud future.

Explore your path to accelerated digital transformation with Informatica and Azure.

LEARN MORE

Experience Azure now with a **FREE** trial.

Accelerate and optimize your analytics with a fully managed data warehouse including 12 months of free services on Azure.

START TRIAL



Informatica™



Microsoft Azure