Informatica®

10 Critical Factors for Cloud Analytics Success

Cloud-native, intelligent and automated data management strategies to accelerate time to value and ROI for cloud data warehouses, data lakes and lakehouses

Contents

Tip: Click to jump straight to any section.

Introduction

Today's Cloud-Native and Multi-Cloud Reality

Today's Cloud-Native and Multi-Cloud Reality

Enterprises today recognize that cloud is integral to their digital transformation.

Cloud adoption is growing exponentially, and it is a key enabler to every digital transformation. As organizations move to the cloud for agility, scalability and flexibility, they are adopting cloud-native, intelligent and automated solutions and multi-cloud platforms. According to a recent survey, 89% of respondents report having a multi-cloud strategy.[1] And the focus today is less on justifying the move to the cloud and more on how to get there faster and with less difficulty—and how to realize greater business value and ROI.[2]

At the same time, organizations moving to the cloud are consolidating and modernizing their traditional on-premises data warehouses and data lakes to take advantage of cost savings and operational efficiencies of the cloud. Cloud and analytics modernization go hand-in-hand.

A recent Accenture survey found that 90% of business and IT executives agree that to be agile and resilient, their organizations need to fast forward their digital transformation with cloud at its core.[3]

To support their cloud analytics initiatives, organizations are adopting cloud data warehouses, data lakes and lakehouses (which combine the benefits of cloud data warehouses and data lakes into one data platform). But common data management mistakes can hinder organizations on their journey to analytics modernization and ROI. This is also true for organizations starting new data warehouses and data lakes in the cloud, whether for the first time or for new departments.

The good news is that proven, cloud-native and automated cloud data management strategies for modernizing analytics in the cloud are available to accelerate your journey and finally realize the expected ROI from your investments. Let this ebook serve as your guide to modernizing analytics in the cloud.
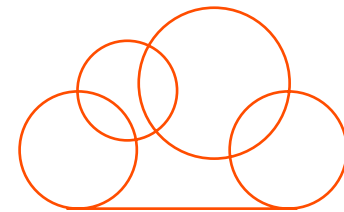
Section One

The Evolution of Cloud Analytics

The Evolution of Cloud Analytics

The data and analytics landscape has recently undergone tremendous changes.

For many years, organizations relied on data marts and on-premises data warehouses to power business intelligence (BI) and reporting systems.

With the advent of big data, enterprises added Hadoop-based data lakes. New technologies— such as EMR, Hive and Spark—eventually followed. These changes paved the way for faster analytics and insights. But operational complexity, maintenance and cost considerations remained challenges.

Businesses that depended on analytics coming from on-premises data warehouses and data lakes went through an arduous phase of expectations versus reality when it came to time to value and ROI. Millions of dollars were spent trying to address serious data quality and data management challenges. Data lakes very quickly became known as data swamps. Organizations that succeeded were those that decided to tackle these challenges head-on with intelligent, automated data management.

Now with the promise of agility, scalability and cost savings, organizations are turning to cloud-native data lakes, data warehouses and lakehouses. Yet the data management and data quality issues that affected on-premises data warehouses and data lakes remain barriers to success today.

Enterprises can avoid these barriers by taking a cloud-native, intelligent and automated approach to cloud data management. Those that do so will be well-positioned to realize the full possibilities of their cloud data warehouses, data lakes and lakehouses—gaining analytics insights into past performance as well as predictions for the future.

Three Common Data Management Mistakes

In an earlier era, organizations often struggled to accelerate time to value and maximize ROI from their on-premises data warehouses and data lakes. As you modernize analytics in the cloud, it's essential to avoid three common mistakes from the past.

- **Using manual hand coding to hydrate and process data in cloud data warehouses, data lakes and lakehouses:** Organizations may turn to hand coding for prototyping, but in the long run hand coding is insufficient to address enterprise requirements of scale and maintainability. Hand coding falls short when it comes to data integration capabilities for building high-performance data pipelines to ingest and prepare data for analysis. Code isn't reusable as the underlying technology changes. This means you have to reengineer and recode every time there's a change to technology, platform or processing engine.

Hand coding also fails to ensure data quality and doesn't provide metadata management to help you discover, catalog and understand how your data moves through the organization. Over time, hand coding is more expensive, time consuming and riskier than using an intelligent and automated solution that doesn't require coding.

- **Depending on disjointed point products to achieve end-to-end data management:** Using multiple, non-integrated products increases complexity as well as cost. It can take up to 10 separate products to achieve the end-to-end data management you need for modern cloud data management. And stitching together disjointed products means that you must constantly deal with changing roadmaps, cost and time overruns and—most significantly— inconsistent data governance and quality.

- **Relying on limited solutions from cloud vendors that only offer basic data integration or ingestion and don't work across clouds:** Although offerings from platform-as-a-service (PaaS) or infrastructure-as-a-service (IaaS) vendors are designed for the cloud, they tend to have both of the above downsides. They typically offer basic data integration and ingestion, are reliant on hand-coded development and provide capabilities that extend only as far as their own platforms, making them inadequate for today's multi-cloud environments. Cloud data management requirements for modern enterprises must extend beyond any single PaaS to a multi-cloud strategy and deployment model.

What Does a Modern Approach to Cloud Analytics Look Like?

As you look to modernize analytics in the cloud, you need industry-leading, best-of-breed data management capabilities optimized and built for cloud-native environments. Your solution should include three essential pillars built on a foundation of artificial intelligence (AI) and machine learning (ML):

- **Cloud-Native Data Integration:** Intelligent, automated, cloud-native data integration capabilities—such as a codeless visual interface, prebuilt mappings and elastic scaling—enable you to quickly and efficiently build data pipelines to feed your cloud data warehouse, data lake and lakehouse.

- **Cloud-Native Data Quality:** An enterprise data governance program with intelligent, automated, cloud-native data quality ensures that the data in your cloud data warehouse, data lake and lakehouse is cleansed, standardized, trusted and secure.

- **Cloud-Native Metadata Management:** A common enterprise metadata foundation enables intelligent, automated, end-to-end workstreams across your data environment, facilitates collaboration, provides visibility into end-to-end data lineage and promotes efficiencies.

In addition, a modern solution for cloud data management should provide intelligent, automated, end-to-end cloud data management that works across any cloud. That way, you're not locked in by hand coding or limited functionality. Instead, you're able to reuse and leverage existing investments, ultimately future-proofing your cloud or multi-cloud solutions.
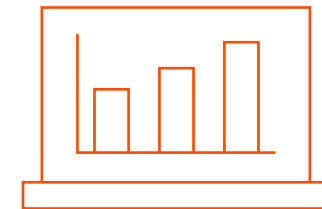
Cloud Analytics Modernization Benefits by Industry

Organizations of all types are using intelligent, automated cloud data management to deliver cloud analytics that accelerate insights and drive innovation. Here are a few examples:

Financial Services

Banks, investment companies and other financial services firms rely on data to detect fraud, assess risk for credit and loans, predict market movements, evaluate investments and manage billions of transactions and interactions each day. Intelligent, automated cloud data management gives them deeper insight into customer needs and behavior to ensure that they can compete on customer service, connect customers to the most appropriate information and products, and protect customer accounts and sensitive data effectively.

To combat fraud, financial services firms are modernizing their existing fraud monitoring systems, investing in cloud data lake solutions, and adopting AI and machine learning-powered predictive analytics to detect and combat fraudulent activities faster than ever before. Intelligent, automated cloud data management also helps commercial banks win new customers by integrating and analyzing huge volumes of structured, semi-structured big data.

Healthcare

Healthcare organizations are moving their analytics platforms to the cloud in order to improve patient care, increase operational efficiencies and provide superior customer service. By ingesting and processing large volumes of data, cloud-based solutions enable healthcare providers to better predict deteriorating health conditions and suggest preventative measures with the help of sensor data, patient history and cohort data.

Intelligent, automated cloud data management enables healthcare providers to make data available to be served to more users, including patients themselves via patient portals. For example, healthcare providers can make data available on a patient portal in near real time. As a result, it's easier for patients to access their health information, contact their doctors, understand their bill and receive financial assistance.

Government

Federal agencies create, collect and access huge volumes of data—on-premises and in the cloud. Intelligent, automated cloud data management enables government agencies to break down data silos, unify resources and apply data governance and data quality policies to improve operational effectiveness, uncover efficiencies, better serve citizens and strengthen data privacy protections.

State and local governments are increasingly leveraging data from new sources, such as Internet of Things (IoT) sensors and emergency and social services. Using trusted data to drive cloud analytics, local government agencies can improve city planning, reduce traffic congestion and accidents, monitor and track community health and more—all with an objective to make better decisions on resource allocation and to better serve their constituents.
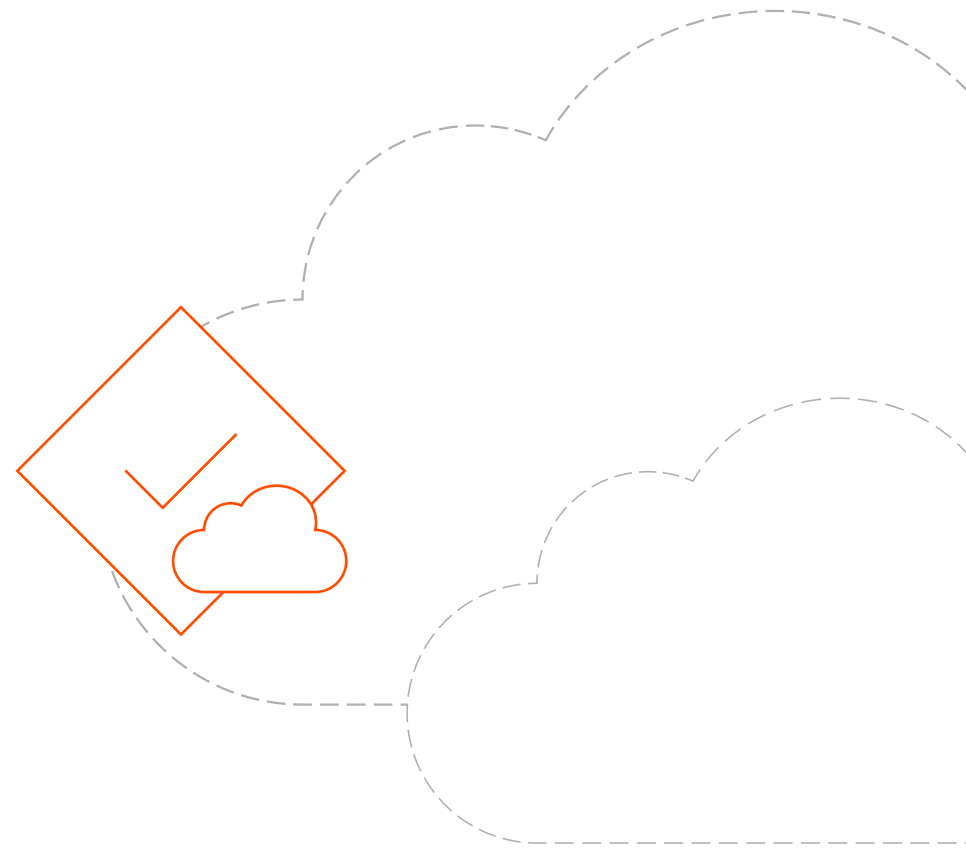
Section Two

Pathways to Analytics Modernization in the Cloud

Pathways to Analytics Modernization in the Cloud

Modernizing analytics in the cloud enables you to achieve greater agility, increase scalability and optimize costs. But organizations take different paths to achieving their goals. The best way to proceed will depend on your data environment and business objectives.

You'll find two common paths to modernizing analytics in the cloud:

- **Cloud Data Warehouse, Data Lake and Lakehouse Modernization:** Strategically move your data warehouse and data lake to the cloud over time and adopt a modern, end-to-end data infrastructure for advanced analytics, data science, AI and machine learning projects.

- **New Cloud Data Warehouse and Data Lake:** Start small and fast and grow as needed by spinning up a new cloud data warehouse or cloud data lake. The same guidance applies whether you are implementing new data warehouses and data lakes in the cloud for the first time, or doing so for an individual department or line of business.

We'll examine each of these paths in more detail. Keep in mind that you have options for how to proceed. You might start on one pathway and eventually move to another.

Cloud Data Warehouse, Data Lake and Lakehouse Modernization

As cloud adoption grows, most organizations will eventually want to modernize their enterprise analytics infrastructure entirely in the cloud. With the modernization pathway, you rebuild everything to take advantage of the most modern cloud-based enterprise data warehouse, data lake and lakehouse technology to end up in the strongest position long term. But you migrate data and workloads from your existing on-premises enterprise data warehouse and data lake to the cloud incrementally, over time.

This approach allows you to be strategic while minimizing disruption. Your organization can take the time to carefully evaluate your data and bring over only what you need, which makes this a less risky approach. It also enables more complex analysis of your data, using artificial intelligence, machine learning and data science. The combination of a cloud data warehouse and data lake (or lakehouse) allows you to manage the big data necessary for analytics by providing economical scalability across compute and storage that is not possible with an on-premises infrastructure. And it enables you to incorporate new types of data, from IoT sensors, social media, text and more, into your analysis to gain new insights

For this pathway, you'll need an intelligent, automated data platform that delivers a number of critical capabilities. It should handle new data sources, accommodate AI and machine learning projects, support new processing engines, deliver performance at a massive scale and offer serverless scale up/scale down capabilities. As with a brand-new cloud data warehouse or data lake, you'll need cloud-native, best-of-breed data integration, data quality and metadata management to ensure that you're maximizing the value of your cloud analytics.

Once the data is in the cloud, your organization can provide users with self-service access so they can more easily and seamlessly create reports or answer questions. Following this modernization pathway gives organizations an end-to-end modern infrastructure for next-generation cloud analytics.

New Cloud Data Warehouse and Data Lake

Lines of business increasingly rely on analytics to improve processes and business impact. For example, sales and marketing no longer ask, "How many leads did we generate?" They want to know how many sales-ready leads we gathered from Global 500 accounts as evidenced by user time spent consuming content on the web. But individual lines of business may not have the time or resources to create and maintain an on-premises data warehouse to answer these questions.

With a new cloud data warehouse and data lake, departments can get analytics projects off the ground quickly and cost effectively. Departments simply spin up their own cloud data warehouses, populate them with data and make sure they're connected to analytics and BI tools. For data science projects, a team may want to quickly add a cloud data lake. In some cases, this approach enables the team to respond to requests for sophisticated analysis faster than centralized teams can normally handle.

Whatever the purpose of your new cloud data warehouse and data lake, you'll need intelligent, automated cloud data management with best-of-breed, cloud-native data integration, data quality and metadata management all built on a cloud-native platform in order to deliver value and drive ROI.

While this approach allows you to start small and scale as needed, the downside is that your data warehouse and data lake may only benefit a particular department inside the enterprise. If you take this path, consider how to position your cloud data warehouse and data lake for the future.

Avoiding "Lift & Shift"

Some organizations with significant investments in on-premises enterprise data warehouses and data lakes are looking to simply replicate their existing systems to the cloud. By lifting and shifting their data warehouse or data lake "as is" to the cloud, they seek to improve flexibility, increase scalability and lower data center costs while migrating quickly to minimize disruption.

Lifting and shifting an on-premises system to the cloud may seem fast and safe. But in reality, it's an inefficient approach, one that's like throwing everything you own into a moving van instead of packing strategically for a plane trip. In the long run, reducing baggage and traveling by air delivers greater agility and faster results because you are not weighed down by unnecessary clutter.

Some organizations may need to do a lift and shift, but most will find it's not the best course of action because it simply persists outdated or inefficient legacy systems and offers little in the way of innovation.
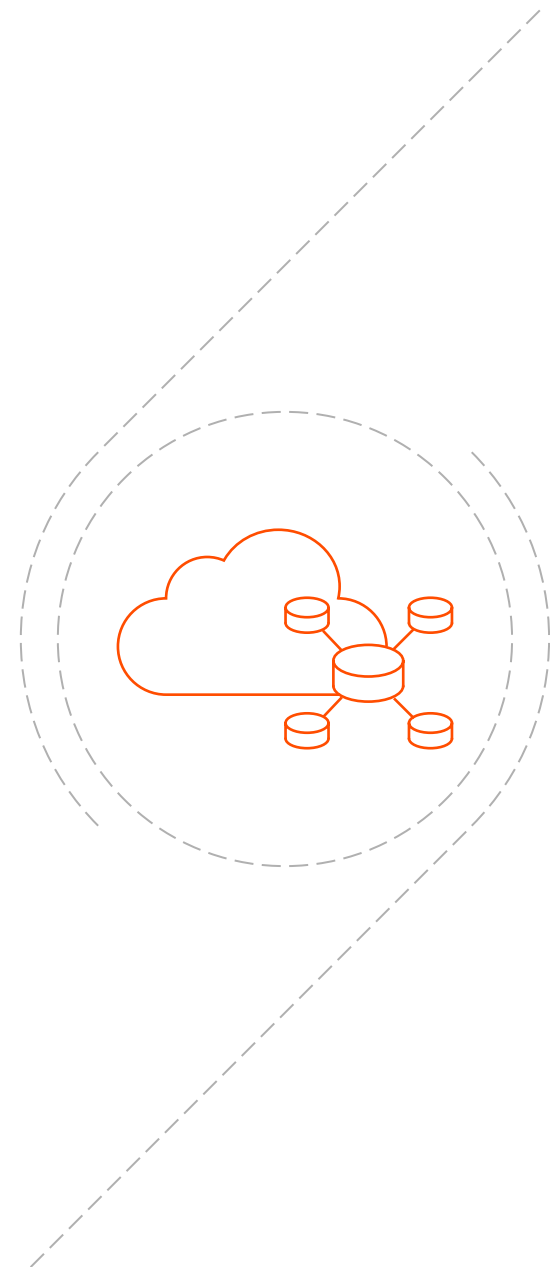
Home Point Financial Modernizes Data Analytics in the Cloud

Home Point Financial realized that to grow its business and become a top ten leading residential mortgage provider, it needed to become data driven. But its legacy infrastructure was not built for scale, agility or modernization. The firm wanted to deploy an agile infrastructure that could lower labor costs, was future proof and could scale to accommodate its vision for growth.
To achieve its vision for market disruption, Home Point Financial committed to a cloud-first data strategy running on Microsoft Azure.
It partnered with data warehouse and integration vendors, including Snowflake for hosting and storing the data, Tableau for presenting data, and Informatica's Intelligent Data Management Cloud™ (IDMC) for delivering and integrating data with high availability.

The solution has reduced costs while at the same time enabling Home Point Financial to scale. The platform also allows the firm to better organize and deliver data, which has translated into better data quality.

READ MORE

CUSTOMER STORY

AXA XL Simplifies Advanced Analytics

AXA XL, a commercial property and casualty and specialty insurance company, was looking to differentiate itself in a commoditized market by adopting a data-driven strategy. But because the company had grown through mergers and acquisitions, AXA XL's data existed in silos, which limited its ability to leverage data to gain insights. AXA XL needed a more effective way to support advanced analytics and identify new sales opportunities. It also wanted to allow data analysts and actuaries to prepare and operationalize data themselves for faster, trusted insights into sales opportunities and risks.

To meet these demands, AXA XL completely redesigned and consolidated its data architecture with a cloud-based data lake, advanced analytics and AI/ML tools. In order to integrate, govern and cleanse the data from various cloud and on-premises solutions and make it easy to find and prepare for analysts, they used an integrated suite of cloud-based data management products from Informatica to scan and catalog data across the enterprise and allow users to prepare data for analysis in a self-service manner. The platform simplified advanced analytics, helping the company build innovative data assets faster and at a lower cost.

READ MORE

Section Three

10 Considerations for Success

10 Considerations for Success

Here are 10 considerations for successful cloud data management that accelerates time to value and drives ROI from cloud data warehouses, data lakes and lakehouses:

1. **Intelligence and automation**—Are you using metadata-driven artificial intelligence and machine learning to automate and simplify tasks related to data management across your enterprise? AI and ML can automate and   scale complex cloud data management tasks, accelerating time to value from your cloud data warehouse, data lake and lakehouse initiatives.

2. **Data discovery**—Do you have a way to find data assets across your enterprise and prioritize which data you need to migrate first? Are your data engineers and data scientists easily able to find the data they need? Do you have visibility into end-to-end data lineage for your enterprise data assets?

3. **Data integration**—Do you have a reliable, scalable, high-performing cloud-native integration tool to help you rapidly build data pipelines and increase performance and efficiency? Can your data engineers leverage a zero-code, serverless solution for streamlined cloud-based data integration processing?

4. **Data ingestion**—Do you have a scalable, automated data ingestion tool that can bring all the data types and sources you need (including streaming, files, databases and changed data) into your cloud data warehouse, data lake and lakehouse?

5. **Data cleansing, standardization, an transformation—**Are you able to discover data quality issues and correct them on an ongoing basis to ensure the delivery of trusted data for decisions and key business processes? Do you have appropriate transformation capabilities to cleanse and aggregate data to prepare it for analysis?

6. **Data quality**—Are you able to ensure the quality, availability, usability, integrity and security of the data in your cloud data warehouse, data lake and lakehouse?

7. **Data preparation**—Are your data scientists spending their time on data prep instead of using data to discover new insights?

Are you using a solution with an intuitive user interface, built-in prep functions and data quality rules with the ability to make recommendations about which datasets a user should use for their analysis?

8. **Data lineage**—Are your teams able to see end-to-end data lineage to understand how data moves through your organization? Can your analysts see data lineage to ensure that your reporting tools such as Microsoft Power BI or Tableau are using the right data?

9. **Operationalization**—Does your solution operationalize data engineering pipelines to create standard data ingestion processes that leverage automation as much as possible? Are you easily able to provision trusted data to data consumers and enable self-service?

10. **Reusability**—Are you able to leverage existing investments as you modernize in the cloud? Look for a solution that enables reuse with the ability to abstract the data pipelines that feed into your cloud data warehouses and data lakes.

Conclusion

Start Your Journey Today

Whether you are building a new data warehouse, data lake or lakehouse in the cloud, or consolidating and modernizing existing on-premises data warehouses and data lakes in the cloud over time, it's essential to make sure you have the tools and best practices in place to manage your data during migration and once it's in the cloud.

If yours is like most organizations, your ultimate destination will be a cloud data warehouse with a cloud data lake or lakehouse that can enable your organization to incorporate all types of data and analytics to drive your digital transformation.

To gain efficiencies, cost savings and scale, look for intelligent, automated, cloud-native data management solutions with best-of-breed data integration, data quality and metadata management—built for the cloud on an AI-powered, intelligent data platform.

Further Reading

Learn about the Informatica solution for cloud data warehouses and data lakes in these resources.
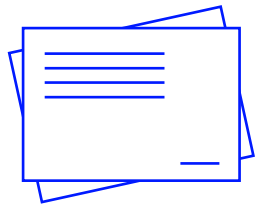
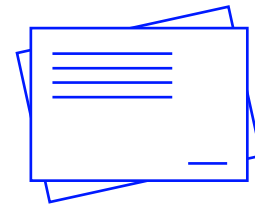*Modernize Your Data Warehouse and Data Lake in the Cloud*

READ MORE
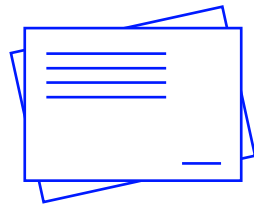
Intelligent Cloud Lakehouse Data Management for Cloud Analytics

READ MORE

3 Ways Manual Coding Is Killing Your Business Productivity (And How You Can Fix It)

READ MORE

Sources

1   Flexera, 2022 State of the Cloud Report

2   TDWI Best Practices: Modernizing Data
    and Information Integration for Business
    Innovation, David Stodder, Q4, 2021

3   Accenture, Technology Vision 2021

About Informatica

At Informatica (NYSE: INFA), we believe data is the soul of business transformation. That's why we help you transform it from simply binary information to extraordinary innovation with our Informatica Intelligent Data Management Cloud™. Powered by AI, it's the only cloud dedicated to managing data of any type, pattern, complexity, or workload across any location — all on a single platform. Whether you're driving next-gen analytics, delivering perfectly timed customer experiences, or ensuring governance and privacy, you can always know your data is accurate, your insights are actionable and your possibilities are limitless. Informatica. Cloud First. Data Always.™

Worldwide Headquarters
2100 Seaport Blvd, Redwood City, CA 94063, USA
Phone: 650.385.5000
Fax: 650.385.5500
Toll-free in the US: 1.800.653.3871

informatica.com
linkedin.com/company/informatica
twitter.com/Informatica

CONTACT US

Informatica®