

Brought to you by:



Modern Data Architectures

for
dummies[®]
A Wiley Brand

Uncover emerging
data architecture trends

—
Deliver data-driven
decisions faster

—
Use AI to make
meaningful connections



Mike McCallister

Informatica Special Edition

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™.

Visit www.informatica.com for more information.

Connect with Informatica on LinkedIn, Twitter and Facebook.



<https://linkedin.com/company/informatica>



<https://twitter.com/Informatica>



<https://facebook.com/InformaticaLLC>



Modern Data Architectures

Informatica Special Edition

by Mike McCallister

for
dummies[®]
A Wiley Brand

Modern Data Architectures For Dummies®, Informatica Special Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2023 by John Wiley & Sons, Inc., Hoboken, New Jersey

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, For Dummies, the Dummies Man logo, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. Informatica and the Informatica logo are registered trademarks of Informatica. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: WHILE THE PUBLISHER AND AUTHORS HAVE USED THEIR BEST EFFORTS IN PREPARING THIS WORK, THEY MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES REPRESENTATIVES, WRITTEN SALES MATERIALS OR PROMOTIONAL STATEMENTS FOR THIS WORK. THE FACT THAT AN ORGANIZATION, WEBSITE, OR PRODUCT IS REFERRED TO IN THIS WORK AS A CITATION AND/OR POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE PUBLISHER AND AUTHORS ENDORSE THE INFORMATION OR SERVICES THE ORGANIZATION, WEBSITE, OR PRODUCT MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING PROFESSIONAL SERVICES. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR YOUR SITUATION. YOU SHOULD CONSULT WITH A SPECIALIST WHERE APPROPRIATE. FURTHER, READERS SHOULD BE AWARE THAT WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ. NEITHER THE PUBLISHER NOR AUTHORS SHALL BE LIABLE FOR ANY LOSS OF PROFIT OR ANY OTHER COMMERCIAL DAMAGES, INCLUDING BUT NOT LIMITED TO SPECIAL, INCIDENTAL, CONSEQUENTIAL, OR OTHER DAMAGES.

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

ISBN: 978-1-119-98246-3 (pbk); ISBN: 978-1-119-98247-0 (ebk). Some blank pages in the print version may not be included in the ePDF version.

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

Project Manager:
Carrie Burchfield-Leighton
Sr. Managing Editor: Rev Mengle

Managing Editor: Camille Graves
Sr. Client Account Manager:
Matt Cox

Special Help from Informatica:
Vishwanath Belur, Rudra Ray,
Michelle Schallhorn

Table of Contents

INTRODUCTION	1
About This Book	1
Icons Used in This Book.....	2
Beyond the Book.....	2
CHAPTER 1: Making Sense of Modern Data Architectures	3
Looking at the Reasons for Modernizing	4
Listing the Characteristics of a Modern Data Architecture	5
CHAPTER 2: Emerging Patterns in Modern Data Architecture	7
Building a Modern Data Stack	7
Benefiting with MDS	8
Getting started with MDS.....	8
Weaving a Data Fabric	9
The magic of active metadata	9
The benefits of the data fabric.....	10
Embracing the Ubiquity of a Data Mesh.....	12
Benefiting from a data mesh.....	13
Evaluating a data mesh	13
CHAPTER 3: Managing Data in a Modern Architecture with IDMC	15
Demonstrating Architectural Flexibility	16
Modern data stack.....	16
Data fabric	17
Data mesh.....	18
Automating Data Management with CLAIRE	18
Learning From Proven Success.....	20
AXA XL.....	20
Shire Pharmaceuticals.....	21
Railinc	21

CHAPTER 4: Ten Benefits of a Modern Data Architecture..... 23

- More Effectively Use Your Cloud-Based and Hybrid Data..... 23
- Break Down Your Data Silos 24
- Choose Patterns that Work Best 24
- Get the Information You Need from a Variety of Tools and Services..... 25
- Unite Disparate Data Sources to Identify Unique Solutions 25
- Leverage Decentralized Teams to Manage Domain-Oriented Data 26
- Use AI to Connect the Dots among All Your Data Sources and Build a Metadata Knowledge Graph 26
- Democratize Your Data 26
- Drive Data-Driven Decisions Faster 27
- Provide Flexibility to Data Architecture with a Holistic Data Management Platform..... 27

Introduction

Much has changed in the world of data since the days when every bit of it was stored on a mainframe and accessed through a terminal. Today, most businesses receive data from network-based services and process it in the cloud. The questions are, “Are you managing that data effectively? And, if you had more accurate insights based on that data, could you serve your customers better?”

What if we had data systems that let more people in our organizations generate better insights? What if those people could use artificial intelligence to do amazing things with data? Can we get our data-driven decisions made faster?

Proponents of modern data architecture believe there’s a better way to do it.

About This Book

Welcome to *Modern Data Architectures For Dummies*, Informatica Special Edition. This book helps you understand the goals and objectives of modern data architecture and the benefits of migrating your cloud-based data to a modern setup.

This book covers several topics, including

- » Why you need to modernize
- » What a modern data architecture looks like and what it does
- » Three flavors of modern data architecture: modern data stack (MDS), data fabric, and data mesh
- » Understanding the magic of metadata in a modern data architecture
- » How Informatica’s Intelligent Data Management Cloud helps you get the most from your data, regardless of what flavor you favor

Icons Used in This Book

Throughout this book, different icons highlight important information. Here's what they mean:



REMEMBER

The Remember icon marks a generally interesting and useful fact about modern data architectures — something you may want to remember for later use.



TIP

The Tip icon points out helpful suggestions and useful information for a better way to do something, save time, or save money.



WARNING

Information with Warning icons alert you to issues you may want to avoid or highlight challenges you may face.

Beyond the Book

This book can help you discover more about modern data architecture, but if you want resources beyond what this book offers, Informatica can give you additional insight:

- » www.informatica.com/platform.html: This site gives you everything you need to know about Informatica Intelligent Data Management Cloud.
- » www.informatica.com/enterprise-architecture.html: Reference architectures, solution diagrams, templates, best practices for common data management patterns and more — get the details on how to efficiently build your enterprise data infrastructure using Informatica solutions.
- » www.informatica.com/blogs/data-platform-ai.html: Get the latest perspectives on Informatica Intelligent Data Management Cloud, the intelligent data platform that builds off CLAIRE, Informatica's artificial intelligence (AI) engine that supports data for analytics. Whether you're a chief data officer or simply have a role in data management, you'll find clear insights in these blogs.
- » www.informatica.com/resources.html: Stay up to date with the latest marketing content.

- » Understanding why to modernize
- » Listing the features of a modern data architecture

Chapter 1

Making Sense of Modern Data Architectures

Businesses run on data, and the most successful businesses just may be the ones that mine the most strategic information from the raw data their machines collect each millisecond. This fact is especially true when you store your data in the cloud. Whether you store your data on Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP), Snowflake, Databricks, Salesforce, or another provider, combining data from different data sources is difficult. But you can leverage business data in an attractive new way, and it's open to nearly everyone.

Modern data architectures have many of the right answers and could be the solution to the problems with data today. In this chapter, you discover the basic characteristics of modern data architecture and explore three different types:

- » The modern data stack (MDS), the most accessible way to modernize
- » Data fabric, focusing on generating new insights from multiple data sources
- » Data mesh, centered on data products, decentralized data governance, and agile leadership

Looking at the Reasons for Modernizing

Monolithic architectures and centralized data platforms may have worked years ago. Today, they can hinder business agility, making it difficult to quickly adjust to the constantly changing data landscape that perpetually demands new views, new aggregations, and new projections of the data (the basis of *data products*).

Now, there's nothing wrong with the standard way of processing data: extract, transform, load (ETL). For that matter, those who like loading before the transform (a process called *extract, load, transform* [ELT]) aren't wrong either.



WARNING

But when you start to think through these processes, you may realize that your organization struggles with the following challenges:

- » **Data proliferation and complexity:** This includes increasing data domains (such as product, supply chain, marketing), data consumers (that is, your staff and customers), and data sources (that stuff you're trying to gain insights from).
- » **Lack of agility:** You can't meet data demands in a timely manner.
- » **Lack of collaboration:** Your data developers are stuck in a silo far from where the data originates or, even worse, where the decision makers define what data they use.
- » **Lack of trust in decisions made:** This is based on people using different copies of varying instances of the same data, making it difficult to identify the point of truth.

If these examples sound like your workplace, perhaps you've tried one or more of these approaches:

- » Semantic layers
- » Data virtualization
- » Data as a service (a data management strategy aiming to leverage data as a business asset for greater business agility)

Maybe you've made some progress, but none of these approaches became an obvious solution by itself.

Listing the Characteristics of a Modern Data Architecture

Data strategy insider Wayne Eckerson outlined ten characteristics of modern data architectures and what a modern data architecture should look like. In the following list, I give you his characteristics, but the commentary is mine:

- » **Customer-centric:** Don't focus on how you process the data. Your customers, internal and external, have to be able to make sense of the data you're presenting. It wouldn't hurt to give them some control over how to define the data they need.
- » **Adaptable:** Manage the data flow by building pipelines that serve a variety of business needs. Build the pipelines with basic data objects.
- » **Automates everything:** Connect disparate data elements by tagging and profiling everything in sight while you're pulling it into the system.
- » **Smart:** While your automated system is tagging and profiling, consider that the best systems use artificial intelligence (AI) and machine learning (ML) to identify commonalities among different terms. Make your people more efficient and effective. Legacy systems used brute force to make those connections, and that may not have gone so well.
- » **Flexible:** Flexible isn't the same as adaptable. Your system has to deal with massive varieties of people and processes. Other systems have different refresh rates, slow networks, processing engines, query options, and pipelines. You have to be able to deal with it all.
- » **Collaborative:** Modern architectures may run on SQL queries, but the data belongs to the business users. IT needs to work with the business to understand the best ways to deliver data to them. This is where data catalogs and other tools come in. (I cover data catalogs more in Chapter 2.)
- » **Governed:** For Eckerson, "governance is the key to self-service." You need to define access points for every user type to make things go.



REMEMBER

Modern data architecture is based on the idea of decentralized governance based on teams of users.

- » **Simple:** Don't overcomplicate the system. Give the people what they want and no more.
- » **Elastic:** You have to be in the cloud. Your data will grow. Your data sources will multiply. They all have to connect with each other. You have to be able to adapt to changing requirements. The cloud may not make all this possible, but it makes it all affordable (for now, at least).
- » **Secure:** Eckerson calls the modern data architecture "a freedom fortress." If you're an authorized user, you can gain ready access to data while keeping hackers and intruders at bay. You don't have to be constrained by the European General Data Protection Regulation (GDPR) or the United States Health Insurance Portability and Accountability Act (HIPAA) to understand how important privacy is to your users. Everything should be encrypted in transit and at rest.

- » Defining a modern data stack
- » Using a data fabric to weave together connections
- » Understanding and better accessing data with a data mesh

Chapter 2

Emerging Patterns in Modern Data Architecture

Getting started with modern data architecture often starts with moving some or all your data from on-premises to the cloud. The cloud offers important opportunities to break down silos and see how you can connect data in creative ways. In this chapter, you get the basics on modern data architecture patterns.

Building a Modern Data Stack

Years ago, business data was collected and stored on a mainframe or in relational databases such as Oracle or PostgreSQL. Tools were developed to extract, transform, and load (ETL) data in a format that non-technical people could view and understand.

Along with traditional sources of data, open-source database management systems will never disappear. Now, data often comes from a software-as-a-service (SaaS) specialized product that produces analyzable data in your industry. Data streams from a

variety of sources, including clickstreams, location, and smart-factory data from internet of things (IoT) devices, and your phone. These traditional ETL tools don't easily connect to new sources or deliver information in the same way. New tools are needed.

Take a look at Figure 2-1. Modern data stack (MDS) refers to the tools and technology that are used to collect, process, and store data faster and in more scalable and accessible ways than traditional data stack. MDS moves all data gathered from a variety of cloud-based and on-premises sources into a more centralized, analyzable space in the cloud, with better analytics than the legacy data stack.

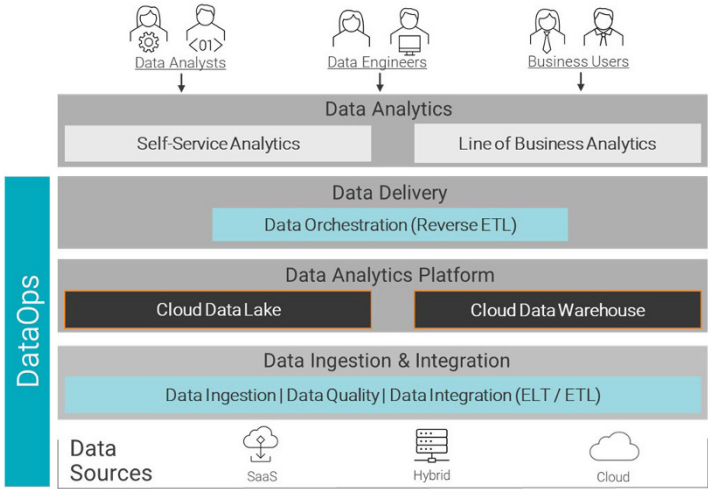


FIGURE 2-1: The modern data stack.

Benefiting with MDS

MDS has a scalable and flexible framework. You can choose the services and tools you use to assemble and work with your data. There is little data engineering overhead. You don't need a massive team of data professionals to implement MDS, either. Even small shops can benefit. You can start and execute your data strategy quickly and get valuable performance metrics to test your strategies as you go.

Getting started with MDS

The MDS is built with non-technical business users in mind. MDS removes the technical barriers that left company data firmly in the

hands of those who could recite the incantations of the *structured query language* (SQL) in just the right ways to deliver answers to the waiting masses. MDS tools require no to very little knowledge of coding, such as SQL. To run MDS, you just need an ingestion tool, a storage platform, a transformation tool, and a business intelligence tool.

Weaving a Data Fabric

After you've built an MDS, you can extend its power with a data fabric. By harnessing the magic of metadata through artificial intelligence (AI) and machine learning (ML), a data fabric can generate previously undreamed of connections between disparate data sources. A data fabric can mean different things, depending on who you're talking to, but here's a useful starting point:

- » It integrates and connects your organization's data intelligently and efficiently by abstracting underlying complexity.
- » It minimizes disruption by enabling an adaptable data management strategy with augmented data integration and management.
- » It doesn't matter what deployment platforms, data processing methods, data delivery methods, locations, and architectural approaches you use. It's independent and neutral regarding the ecosystem platform.
- » It can manage large volumes of data at any latency that may be structured, unstructured, or semi-structured.

The magic of active metadata

Every database includes some metadata that allows humans and machines to know what this thing is. Structured data systems with data in rows and columns have labels to help identify the characteristics that identify the data in this particular context.



REMEMBER

Active metadata is data about everything that happens to the data and is done to the data. Active metadata helps metadata to flow effortlessly and quickly across the entire data stack.

Take this example: A financial institution maintains its personal banking and mortgage departments on separate cloud platforms.

Bridging the gap between the clouds to identify and sell mortgages to existing personal banking clients would be time-consuming and expensive. A data fabric can provide connectivity across a hybrid multi-cloud environment as a data network.

Mergers and acquisitions happen every day, and one of the biggest headaches in that process is pulling together the data from each company and making sense of the combined results. New challenges arise with piecemeal data quality and duplicate master data. Data fabric's ability to develop active metadata across disparate databases can ease the pain of merging two companies, but it can also make the combined company much more empowered to learn lessons that the data can teach. AI-powered matching helps in identifying duplicate master data such as customer, product, supplier, and so on.

Data fabrics rely on active metadata, knowledge graphs, ML, and other metadata-driven capabilities. You then use these capabilities to make recommendations for data integration and analytics, and over time, an intelligent data fabric can then become independent.

ML and graph analysis (GA) form the smart core of the metadata discovery by running deep data analysis on existing and newly arriving data. They can profile the new data, extract value and structure, and run comparisons with the available information to look for similarities and then map users and their usage patterns to that data. ML enhances data fabric with learning models that can train by observing the functions of data engineers, enabling them to prescribe the best next actions and use the outcome to calibrate the next decisions into a continuous improvement practice.



TIP

For more on the importance of active metadata, read “Why a Metadata Knowledge Graph is Essential to an Intelligent Data Fabric.” Visit www.informatica.com/blogs/why-a-metadata-knowledge-graph-is-essential-to-an-intelligent-data-fabric.html.



TIP

The benefits of the data fabric

The data fabric has many benefits:

- » **Analytics productivity:** Increase productivity and accelerate data-driven decision-making by abstracting the complexity of

the distributed data landscape while leveraging the existing investments in data warehouses, data lakes, data hubs, and more.

- » **Cloud migration and modernization:** Accelerate cloud transformation by significantly reducing risks and optimizing costs while easing disruption to data consumers.
- » **Complete view of business data:** Drive faster, more accurate decisions with a complete view of data by accelerating unification and enrichment of disparate data.
- » **Data and AI governance:** Activate data and AI governance by accelerating data stewardship, data classification, business glossary, and data quality with AI and ML-powered automation.
- » **Data privacy, risk, and compliance:** Enable automated discovery and classification of sensitive data assets at scale, while promoting responsible sharing of personal data.
- » **Operational efficiency:** Improve efficiency by automating the operationalization and management of both data and AI model life cycle. This enables a bird's-eye, multilevel view of key operational insights.
- » **Self-service data search and discovery:** Enable fast, easy discovery of business-critical data for a wide range of data consumers by breaking down data and organizational silos.

Figure 2-2 shows you the key components of data fabric architecture.

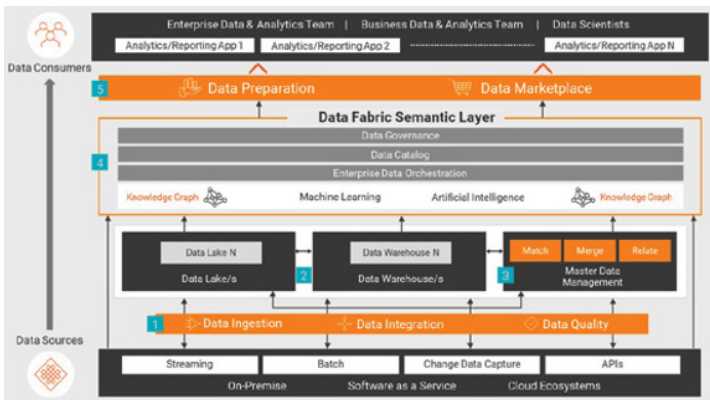


FIGURE 2-2: The key components of a data fabric.

Embracing the Ubiquity of a Data Mesh

A *data mesh* is a data platform architecture that embraces the thought that data is everywhere in the enterprise. It leverages a domain-oriented, self-serve design and aims to democratize data to help data consumers discover, understand, trust, and use data and data products to drive data-driven decisions and initiatives.

A data mesh focuses on organizational change and enables domain teams to own the delivery of data products with the understanding that the domain teams are closer to their data so they understand their data better. Data teams view a data mesh as a prime opportunity to transition from monolithic data platforms to modern data platforms that are cloud-native, microservices-based, API-driven, and multi-cloud and hybrid.

Core to a data mesh is the concept of breaking apart the monolithic architecture and custodianship or ownership of the data around domains in an organization. Data warehouses and data lakes can still exist in the mesh architecture, but they become just another node in the mesh (see Figure 2-3).

A data mesh advocates distributed, domain-based ownership, custodianship of data, and building data products that are self-described and atomic, more easily managed, and delivered at the domain level. These data products are sharable with other domains and interoperable with other data products that form the data mesh. A data mesh manages data as a distributed network of self-describing data products.

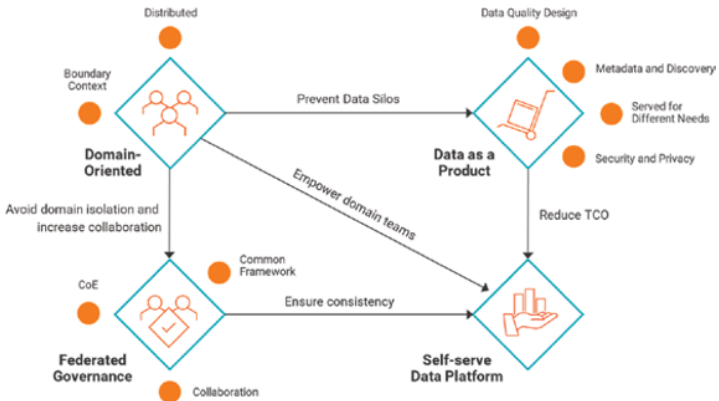


FIGURE 2-3: The key principles of a data mesh.



TIP

Benefiting from a data mesh

A data mesh has many benefits:

- » Enables delivery of customized data products to meet business demands by linking strategic business objectives to an ecosystem of data products to drive value.
- » Scales delivery of data products via decentralized ownership and domain-specific expertise, transforming organizational culture to a data product mindset.
- » Improves agility by abstracting complexity and breaking down monolithic architectures that can be bottlenecks to meeting business demands for data.
- » Facilitates a flexible data governance operating model with federated governance, allowing organizations to enhance the model to meet their unique needs.

Evaluating a data mesh

If you're considering a data mesh architecture, you want to ask questions such as

- » What is a good data product?
- » Who owns the data?
- » Will data standards be different for each domain?
- » Will it lower or increase costs?
- » How do I onboard new data products?
- » What does the data governance operating model look like?

Even if you feel comfortable with your answers to the above questions, consider factors for the following dimensions:

- » **People:** Data mesh can potentially create a burden for domain teams, so an assessment of their skills, roles, responsibilities, and availability is necessary for establishing the appropriate team structure that is a prerequisite for success. An organization's culture is also an important consideration in determining if decentralized domain-based decision making is the optimal strategy.

- » **Process:** Creating data products doesn't imply that domain teams own the data itself. Rather, the teams assume the role of data custodians. Governance operating models, workflows, and key performance indicators (KPIs), for example — data product quality, for both local and federated governance — should be clearly defined and require buy-in.
- » **Technology:** Enterprises need to carefully evaluate how to provision the self-serve data infrastructure at the domain level and the potential costs associated with it. Considerations such as governance policies, security standards, interoperability, risk assessments, DataOps, data lineage, data provisioning, cloud-native elastic and serverless deployments, and more will be necessary.

**Is your company in the right place to implement data mesh now?
See if your company matches this profile:**

- » You manage many data domains, data sources or data team members.
- » You have issues with scaling data delivery and meeting data demands.
- » You aim to create or maintain a decentralized decision-making culture.
- » You have budget for enabling domain-level, self-serve data infrastructure.
- » You need to break the bottleneck that centralization of data (for example, data warehouse and data lake) creates to meet data delivery requirements.
- » Your workforce at the domain level has the skills and resources to define and build data products.

If you answered yes to most of these profile characteristics, you should be ready for a data mesh.

- » Supporting flexible modern data architecture
- » Using AI and ML to automate and augment data management
- » Looking at customer success stories

Chapter 3

Managing Data in a Modern Architecture with IDMC

The Informatica Intelligent Data Management Cloud (IDMC) is a cloud-native, end-to-end data management platform powered by artificial intelligence (AI) that promotes agility, reduces costs, and improves operational efficiencies via intelligent automation of data management capabilities. IDMC provides comprehensive data management capabilities, including data cataloging, data ingestion, data replication, data integration, application programming interface (API) and application integration, data prep, data quality, master data management, data governance and privacy, and a data marketplace, all on an enterprise metadata foundation.

IDMC enables your modern data architecture across a distributed data landscape with an active metadata-based AI and machine learning (ML) engine, CLAIRE, that utilizes a breadth of metadata to automate data management tasks. CLAIRE is optimized for intelligence and automation and is built on a modern, elastic, serverless microservices stack that connects data consumers to the data sources they need.



TIP

With IDMC, you and your staff can

- »» Discover and understand all the data inside and outside of the enterprise.
- »» Access, ingest, and integrate all types of data wherever and whenever you want.
- »» Curate and prepare data in a self-service fashion so it's fit for use.
- »» Deliver an authoritative and trusted single view of all your data.

In this chapter, you discover how IDMC can serve your company's modern data architecture.

Demonstrating Architectural Flexibility

IDMC supports almost any modern data architecture because of its overall flexibility, but there are particular features that support each architecture.

Modern data stack

IDMC supports every layer of the modern data stack (MDS):

- »» Data ingestion and integration from just about any pattern, user, and data
- »» Data analytics for every major cloud provider, including Amazon Web Services Redshift, Snowflake, Azure Synapse, Databricks Delta Lake, Google BigQuery, and Oracle Cloud
- »» Data Orchestration
- »» DataOps

After the data is ingested, you can run whatever analytics tool you want on it. Check out Figure 3-1 to see how this process works.

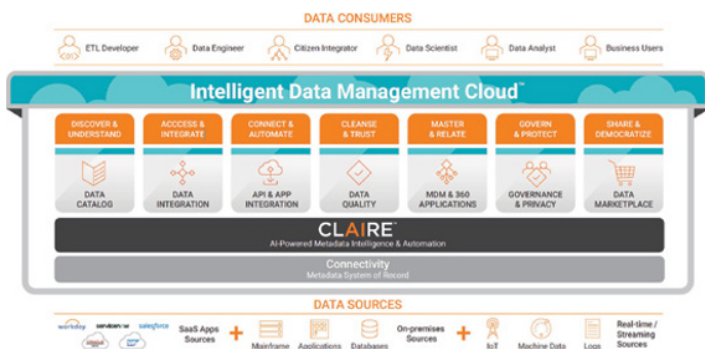


FIGURE 3-1: AI-powered data management capabilities with IDMC.

Data fabric

IDMC helps to build a data fabric architecture by

- » Providing comprehensive metadata connectivity for enterprise data connecting to practically any source through 50,000+ metadata-aware connections and scanners
- » Linking and enriching semantic metadata through an enterprise knowledge graph that puts data in context to deliver intelligence to data management functions such as data cataloging, data governance, data integration, data quality, and master data management
- » Enabling automation with CLAIRE through recommendations and insights that allow you to scale your data management to meet your business needs
- » Simplifying and speeding up data preparation through enterprise data preparation with advanced ML-based automation and data cataloging
- » Enabling automatic data delivery flows through enterprise orchestration and operationalization by employing DataOps, MLOps, and FinOps in support of continuous analysis and monitoring



TIP

With IDMC's data fabric support, you can reap the following benefits:

- » **Expedited time to value:** Industry-leading innovations in active metadata and AI/ML-powered data management accelerate time to value.

- » **Optimized costs while maintaining elasticity:** Flexible consumption-based pricing optimizes your data management costs by enabling you to dynamically scale up or down and leverage capabilities that fit your needs.
- » **Supported multi-cloud, hybrid environments:** Multi-hybrid, multi-cloud enables you to run, interoperate, and support practically any combination of cloud and hybrid infrastructures.

Data mesh

IDMC facilitates data-driven decision-making. Automated data governance and privacy consistently enforce governance and compliance across all data. A data mesh increases transparency and collaboration while reducing compliance risks.

In addition, IDMC supports data mesh by

- » Integrating data within or across domains to create a data product with data integration and application integration services
- » Managing local and federated governance by using data governance and data cataloging services
- » Helping with self-service data discovery and data preparation by using data marketplace, data cataloging, data quality, and profiling services

Automating Data Management with CLAIRE

If you can connect the right data to the right consumers with ease and trust, you'll accelerate your digital transformation. At the heart of IDMC is CLAIRE, the embedded AI engine.

Consider what CLAIRE can do to maximize metadata across your data sets to provide more analysis, making your data smarter. CLAIRE taps into four major categories of metadata:

- » **Technical metadata:** Database schemas, mappings and code, transformations, and quality checks

- » **Business metadata:** Glossary terms, governance processes, and application and business context
- » **Operational and infrastructure metadata:** Run-time stats, time stamps, volume metrics, log information, and system and location information
- » **Usage metadata:** User ratings, comments, and access patterns

Metadata in these four categories becomes the basis for a common metadata foundation. Informatica metadata management uses a rich set of capabilities to create this shared foundation:

- » **Collect:** Scan metadata from all enterprise data systems, such as data lakes and data warehouses, across the cloud and on-premises. This includes databases, filesystems, data integration tools, and processes that help fuel your data analytics and data science tools with a high level of trustworthiness toward data fitness.
- » **Curate:** Document the business view of data with glossary terms, concepts, relationships, and processes. Augment the collected metadata with this business context. Gather user input in the form of ratings, reviews, and certifications. This helps to assess the usefulness of data assets to other users.
- » **Infer:** Apply intelligence to derive relationships not obvious in the collected metadata. This includes data provenance and data lineage, data similarity, and ranking the most useful data sets for different types of users and purposes.

All this metadata is used to create a knowledge graph of an enterprise's data assets and their relationships. CLAIRE then activates the knowledge graph by applying AI and ML and integrates it with data management capabilities. Active metadata serves as the unifying foundation for IDMC and fuels data intelligence back to the CLAIRE AI engine. This accelerates and automates core data management and governance processes.

CLAIRE uses metadata, AI, and machine learning to automate thousands of tasks, including

- » Automatically discovering data domains
- » Classifying data

- » Identifying similar data and other data relationships
- » Recommending next-best actions
- » Associating business terms with physical datasets



TIP

Making an intelligent data catalog a core part of your data infrastructure brings additional benefits. It ensures that active metadata is integrated into your data management processes. IDMC helps you capture enterprise-wide metadata and turn it into active metadata. This is done using extensive metadata connectors. These scan and index metadata augmented with intelligence from CLAIRE.

Active metadata adds automation to make data easier and more efficient to use. Users can better build, deploy, and operate data management applications. These apply to data analytics, data science, data governance, and nearly any other data-driven business priority.

Learning From Proven Success

Many companies have turned to Informatica for help managing their data. In this section, I give you a few examples.

Beyond the customer success stories in this section, you can read more at www.informatica.com/about-us/customers/customer-success-stories.html.

AXA XL

Global insurance and reinsurance company AXA XL, based in Hamilton, Bermuda, and Stamford, Connecticut, has more than 100 offices across six continents. The company needed to deploy a big data infrastructure entirely in the cloud that fed a Microsoft Azure-based data lake. AXA XL wanted to standardize on a single data management platform with integrated products to manage all its metadata.



REMEMBER

AXA XL chose Informatica's data cataloging. With this solution, the company attained more profitable growth through cross-selling and upselling of insurance policies by making data actionable and easy to find. Then, by using Informatica data preparation, it

democratized data discovery and preparation to give data scientists, analysts, and actuaries faster and deeper insights that drove faster policy introductions by the company.

Shire Pharmaceuticals

Shire Pharmaceuticals is a leading global biotechnology company focused on serving people affected by rare diseases and highly specialized conditions. These diseases are often misunderstood, under-diagnosed, and potentially life-threatening. Headquartered in Dublin, Ireland, Shire maintains operations in 65 countries and makes its therapies available in over 100 countries.

Shire had the following business needs:

- » Use research data to fight rare diseases and specialized conditions, and to deliver breakthrough therapies faster.
- » Enable quick, easy access to analytics tools to speed research and development.
- » Create an integrated, central repository for enterprise data.

The company used Informatica's data integration capabilities that were scalable for large data volumes and Microsoft Azure to consolidate disparate data sources and create a single version of the truth. As a result, Shire leveraged Informatica synergies with Microsoft Azure for faster deployment to allow more value to be extracted from enterprise data.

In addition, Shire brought together data ingestion, integration, and visualization tools to support analytics, reduce research and development time, and reduce costs for data acquisition, data integration, and IT support.

Railinc

Based in Cary, North Carolina, Railinc serves the North American freight railroad industry and provides innovative, reliable resources for rail data, IT, and information services. The company wanted to

- » Provide business users with easier search and discovery of interline rail data in a diverse, distributed environment.

- » Improve rail data quality and automatically capture data changes in critical systems as they occur.
- » Increase users' confidence in data by providing greater visibility into its end-to-end lineage.



WARNING

On the company's quest to fulfill these needs, it faced these requirements:

- » Use a simple, intuitive search interface for enhanced usability.
- » Ability to scale a large volume of metadata with room to grow as the business expands.
- » Boost collaboration among business and IT teams for near real-time data delivery.

By implementing Informatica's data cataloging, Railinc enjoyed the following results:

- » Indexed metadata and added business context to rail data
- » Built users' confidence in data by displaying data history to when data was first created
- » Gave business users a better experience with self-service interline rail data discovery while decreasing reliance on IT
- » Delivered accurate, trusted, and timely rail data to the business by automating data quality and change data capture

IN THIS CHAPTER

- » Breaking down data silos
- » Finding the patterns that work for you
- » Using AI to make connections
- » Managing domain-oriented data
- » Getting data-driven decisions faster
- » Discovering datasets of interest

Chapter 4

Ten Benefits of a Modern Data Architecture

If you've been reading this book from start to finish, you may be ready to start your modern data architecture journey, but just in case you needed a few more reasons to make the switch, welcome to the Part of Tens chapter. In this chapter, I give you ten benefits of a modern data architecture. Feel free to take this list to your higher-ups to help make your case.

More Effectively Use Your Cloud-Based and Hybrid Data

Enabling timely access to trusted data for all data consumers is critical. This statement is true whether you're automating front- and back-office processes, analyzing customer sentiments, digitizing product catalogs, or creating hyper-personalized email campaigns across multi-cloud and hybrid environments.

Some examples include departmental use cases such as customer acquisition and profitability analysis, sales forecasting, process optimization, employee retention, and supplier performance.

Similarly, timely access to data is critical for enterprise-wide initiatives such as sustainability/ESG, operational excellence, privacy/compliance and procurement cost analysis across multi-cloud and hybrid environments.

Modern data architecture is all about managing data in the cloud. You get better insights from your data, and build speed, scalability, flexibility, and simplicity with a cloud-native solution.

Break Down Your Data Silos

Organizations that rely on legacy architectures struggle to scale data and analytics due, in part, to developers siloing data from the domains where the data originates or is used for decision making.

As organizations modernize their data landscapes, they still need to access legacy data along with new sources and applications in hybrid and cloud. This problem involves literally thousands of data silos, ranging from newer applications such as Salesforce and Workday, and cloud ecosystems such as AWS, Azure, Snowflake, Databricks, Google Cloud, and more, to legacy mainframe, PeopleSoft, Teradata, and MicroStrategy. The landscape is constantly changing as new data silos and sources emerge every day.



TIP

Modern data architecture can parse data from wherever it comes: your website analytics, network, Salesforce, HubSpot, customer feedback, and so on. You don't have to run queries on a half-dozen different data sources to generate the insights needed to be effective.

Choose Patterns that Work Best

Your business is different from everyone else's, and your data architecture needs to make sense to you and your staff.

All the modern data architectures empower both your technical people and your businesspeople to master the insights your data can provide. You choose the pattern that works for your team. For example:

- » If you're a small operation pulling data from a variety of commercial cloud services, borrowing from the modern data stack (MDS) toolset may be just what you need.
- » If you're inundated with data or acquiring multiple datasets (or companies) that don't know how to talk to each other, much less share common data, weaving a data fabric could be the ticket to solving a variety of problems.
- » If you have disparate teams with widely differing data needs, decentralizing your data and giving each team the ability to draw its own insights from the data, look at building a data mesh.

Get the Information You Need from a Variety of Tools and Services

While MDS is a framework that highlights the ability to use different tools to access, process, and analyze data, all these architectures are designed to pull in data from just about everywhere. To get the information you need, you can use a variety of tools and services:

- » Build data pipelines (ETL or ELT) to move data at any latency from its source into an analytics-focused environment.
- » Manage a target data warehouse or data lake.
- » Use an analytics tool for creating business value out of the data.

Unite Disparate Data Sources to Identify Unique Solutions

Data fabric taps into both human and machine capabilities in using data processes to read, capture, integrate, and deliver data based on the user, context of usage, and prevalent and changing usage patterns. It continuously identifies and weaves data from disparate sources to discover unique, and relevant relationships — in the form of metadata — between the available data points.

Leverage Decentralized Teams to Manage Domain-Oriented Data

Data mesh focuses on decentralized teams. You can get powerful insights from empowered teams. Decentralized ownership transforms your organizational culture to a data product mindset. To manage domain-oriented data, you can break down monolithic, centralized architectures that can be bottlenecks to meeting business demands for data and become more agile in the process.

Use AI to Connect the Dots among All Your Data Sources and Build a Metadata Knowledge Graph

You can use artificial intelligence (AI) to make connections between all your data sources and to build a metadata knowledge graph. Tapping into the magic of metadata — technical, business, operational, and usage — gives you less obvious insights. Knowledge graphs enable exploration and navigation of networks of relationships to uncover value of new connections between existing data such as identifying duplicates through synthesis matching, automated data quality rule recommendations, intelligent structure discovery from logs and other semi-structured files, and many more.



REMEMBER

An intelligent data catalog helps you import business glossary assets such as terms, policies, and classifications from data governance tools. Behind the scenes, the metadata knowledge graph automatically associates business terms to technical metadata.

Democratize Your Data

Managing data isn't the exclusive province of the SQL society. It includes huge volumes, high velocity, and an extensible variety of data such as structured, un-structured, semi-structured and machine data. Modern data architectures empower every staff member to find, understand, trust, and access needed data

through a self-service data marketplace. This data democratization helps everyone better serve their internal and external clients.

Drive Data-Driven Decisions Faster

As businesses evolve in the era of hyperautomation and distributed environments, faster data-driven decisions are critical to digital transformation by including digital processes into all interactions and transactions across your organization. Organizations need to integrate hundreds to thousands of applications and orchestrate the data efficiently across the enterprise through APIs, application integration, business process management, and so on. Automating data management enables faster decisions by data sharing and optimizing data integration and data preparation to improve productivity in a cost-effective manner.

Provide Flexibility to Data Architecture with a Holistic Data Management Platform

Using a data management platform (like the Informatica Intelligent Data Management Cloud) offers your modern data architecture multiple advantages:

- » **Data cataloging** to discover, organize, and curate data assets
- » **Data integration** with just about any pattern, streaming data, batch data, ETL, ELT, data engineering, and hybrid and multi-cloud integration
- » **Application integration** to connect business applications and built-in API management functionality
- » **Data preparation** to enrich and prepare data
- » **Data quality** to profile, cleanse, and improve the quality of data
- » **Master data management** and tailored solutions that provide a 360-degree view of your business (including Customer 360, Product 360, Patient 360, and Supplier 360)

- » **Data marketplace** and data services to democratize data and facilitate data sharing
- » **Data privacy** to discover, classify, protect, and tag sensitive data
- » **Data governance** to build business glossaries for data standardization, to manage and enforce policies, to maximize the value of data, and to enable a foundation of trusted data
- » **A metadata knowledge graph** to deliver consistent data intelligence across practically all data management applications to provide a unified metadata foundation

IDMC enables end-to-end data management and a connected data strategy with a fully integrated, complete data life cycle that's essential for enabling numerous business use cases, all in a single platform.



How One Platform Helps Build Modern Data Architectures

In today's rapidly changing landscape, organizations want to adopt modern data management architectures that are comprehensive, agile and flexible. To address the needs of the intelligent data enterprise, these architectures must be cloud-native, operate at petabyte scale, support an increasing number of non-technical users and provide AI-powered automation.

We can help. The Informatica Intelligent Data Management Cloud™ (IDMC) is a neutral and complete solution for a multi-cloud, multi-hybrid world designed to handle complex challenges — like dispersed and fragmented data — to innovate with data on any platform and any cloud.

Informatica IDMC supports all emerging patterns in today's modern data architecture and provides AI-powered data management capabilities that enable you to:

- ▶ Democratize trusted insights for AI and analytics
- ▶ Orchestrate and automate applications and processes
- ▶ Connect data and digital experiences
- ▶ Discover, catalog, govern and share data

Powered by our AI and machine learning (ML) engine, CLAIRE®, IDMC enables you to discover, understand and access your data, ingest just about any type of data, and curate and prepare data in a self-service fashion.

For more information on Informatica Intelligent Data Management Cloud, please visit informatica.com/platform

© Copyright 2022 Informatica LLC. Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and other countries

Optimize data architectures with one platform

Remember when data used to be stored on a mainframe and accessed through a terminal? Today, most businesses receive data from network-based services and process it in the cloud. But are you managing all that data effectively? And could you better serve your customers if you knew more about their needs? We think so! This book will help you understand the ins and outs of modern architecture patterns and how they can give you a competitive edge in today's demanding market.

Inside...

- Make sense of modern data architecture
- Discover patterns in modern data architecture
- Learn to manage data in a holistic platform
- Find out ten benefits of modern data architecture



Mike McCallister is a technical writer based in Milwaukee, Wisconsin. He's been empowering people to master their tech tools in print, online, and in-person for 25 years. Check out his portfolio at www.michaelmccallister.com.

Go to **Dummies.com**[™]
for videos, step-by-step photos,
how-to articles, or to shop!

ISBN: 978-1-119-98246-3

Not For Resale



for
dummies[®]
A Wiley Brand

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.