

The High Cost of Hand Coding

Why Intelligent, Automated and Modern Cloud Data
Management Matters Now More Than Ever

About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead—in any sector, category or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities or create new inventions. With 100% focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

Table of Contents

Introduction	4
The Choices Facing IT.....	4
The Challenges of Data Management in a Multi-Cloud and Hybrid World	5
Why Some Organizations Still Consider Hand Coding.....	5
4 Drawbacks to Hand Coding for IT	6
3 Drawbacks to Hand Coding for Business	7
4 Elements of Intelligent, Automated Cloud Lakehouse Data Management	8
The Informatica Cloud Lakehouse Data Management Solution	10
Take a Comprehensive Approach to Intelligent, Automated, Modern Cloud Data Management	10
Next Steps	11

“To manage all of this [data complexity] will need an army of people with hand coding and spreadsheets.”

— Director of Enterprise Information Management and Governance, Healthcare Organization

Introduction

Today, organizations are investing heavily in new cloud-based processes, platforms, and environments in order to achieve benefits like scalability, elasticity, agility, and cost efficiencies. At the same time, organizations also understand that data is foundational to successful digital transformation, and cannot be an afterthought.

As these trends converge, IT departments are tasked with helping the business become cloud ready or cloud first, specifically as they modernize analytics. Enterprises are modernizing or standing up new data warehouses and data lakes in the cloud, or what they are starting to refer to as a lakehouse. In one cloud data platform, you have a combined solution for both historical and predictive analytics.

However, when it comes to managing the data to accelerate first time to value and deliver ROI with an investment in cloud data warehouses, data lakes, and lakehouses, the typical approach that IT departments tend to first take can have significant implications such as increased cost, project overruns, and maintenance complexity—wiping out any benefits of modernizing analytics in the cloud.

The Choices Facing IT

Consider the case of a financial services firm that wants to consolidate and modernize its many on-premises data warehouses and data lake into a cloud lakehouse as part of an ongoing customer experience initiative. The first business initiative is to target customers who have shown interest in products from different categories online but never purchased a product. However, there are digital signals (for example, web click-stream, social, transactions, etc.) that may help the firm influence or better engage with its customers to increase share of wallet.

IT must move all data from on-premises databases, files, or CRM and ERP data to the cloud and be ready to support various advanced analytics and AI projects. The perceived quickest path to success—and the trap many IT teams fall into—is to quickly prototype a custom home-grown integration solution in the cloud to solve the first business use case. However, business initiatives consist of multiple projects and workstreams, and so it doesn't end with a one-off project, and the solution doesn't scale to meet today's complex data management challenges.

New data types (such as social, machine sensor and log files, unstructured text, third-party data, and so on) have led organizations to adopt new technologies such as open-source data processing frameworks like Apache Spark to process and store data in cloud environments such as Amazon Web Services, Microsoft Azure, or Google Cloud Platform. And many organizations must manage a complex, multi-cloud landscape—it's common for enterprises to have two or more cloud environments. IT faces the real possibility of not being able to integrate on-premises systems and cloud applications in order to leverage the data effectively and efficiently.

“If you’re the one going to hand code and build these scripts to ingest the data directly ... you’re the one we’re calling. ... The last thing I want is for the development teams ... to also then be on call to support them. I really need code reusability and code maintainability.”

— Director of Data Architecture, Manufacturer

Many IT departments are required to address several new challenges for these new requirements:

- How does IT get siloed data to decision makers?
- Should IT managers rely on in-house developers to create custom data integration?
- Should IT leaders consider data integration tools?

This white paper is intended to provide IT leaders guidance on implementing an intelligent, automated data management solution versus spending time building custom hand-coded solutions.

The Challenges of Data Management in a Multi-Cloud and Hybrid World

As IT organizations begin supporting cloud and analytics or AI initiatives, the temptation is to charge their technical developers with designing, developing and deploying a solution. However, they quickly run into data challenges if they go down the hand-coding path. In many cases, these difficulties are the same that have bedeviled on-premises data warehouses and data lakes:

- **Varied and siloed data:** Most organizations have many various data types residing in many different systems and storage formats, either on-premises or in the cloud. The data is often distributed across siloed data warehouses, data lakes, cloud applications, or third-party resources. However, more data is increasingly originating from online transactions systems and interactions such as web and machine log files, and social media. For example, in a retail environment, data is distributed across many different systems. These include point of sale (POS) systems containing in-store transaction data, customer data in a CRM and MDM system, social and web click-stream data stored in a cloud data lake, and more.
- **Lack of data governance and quality:** With varied and siloed data, there are often varying degrees of data quality and governance. Policies (if they exist) are rarely enforced consistently. Data is dumped into data lakes producing swamps where data is difficult to search, understand, manage, and protect whereas a result only a few data scientists have access. Even worse is dirty data making its way into a cloud data warehouse where hundreds of business analysts and other data consumers depend on it for decision making, predictive analytics and AI.
- **Too many emerging and changing technologies:** With more data, there are new vendors, technologies, and open source projects that change the IT landscape. There are legacy, new and evolving technologies for storage, compute, databases, applications, analytics, and more recently AI and machine learning. Developers may struggle to keep track of this shifting landscape, making it difficult to standardize or implement a methodology.

Why Some Organizations Still Consider Hand Coding

Some organizations choose hand coding, believing that it’s an easier path than procuring a data integration tool, which may have a short learning curve. Additionally, developers may feel that integration tools restrict their creativity for a given use case. In most cases, these are short-sighted objections to an intelligent, automated solution. However, hand coding may be appropriate for quick proof-of-concepts (POC) with a low cost of entry.

4 Drawbacks to Hand Coding for IT

Hand-coded data integrations may appeal to IT departments initially as a quick, inexpensive way to build data pipelines. But there are significant downsides to consider.

Hand Coding Is Expensive

Over time, hand coding is expensive to deploy, operate, and maintain in production. Hand coding needs to be tuned and optimized from development to deployment. And with the bulk of IT budgets in operations and maintenance, the price of hand coding rises over time. Furthermore, the price of a homegrown solution rises as complexity increases in the form of more sources, more targets, more advanced data transformations, or simply event scheduling. By the third or fourth iteration, IT realizes that payroll is escalating dramatically, changing the economics of the solution. By then it's too late, because a substantial commitment has been made and no one wants to explain to management that the millions spent on hand coding must be thrown out in favor of a new investment in a vendor's data management solution, which goes beyond basic data ingestion and integration.

Hand Coding Isn't Future Proof

With rapidly new and emerging technologies, developers have to reengineer and recode anytime there is a technology change, an upgrade, or even a change to the underlying processing engine. Not long ago, we had Hadoop, today it's Spark, and tomorrow there will be another version of Spark or even better technology. As soon as your hand-coded data integration is up and running it's obsolete. What's more, a hand-coded solution doesn't address today's multi-cloud reality. Most organizations end up with multiple clouds across departments or project types or as a result of a merger or acquisition.

Hand Coding Lacks Automation

Hand-coding doesn't scale for data-driven organizations and can't keep pace with enterprise demands. There are simply too many requests for data integration pipelines for IT teams to accommodate. The only way to scale the delivery of production data integration jobs is through automation, and this requires AI and machine learning.

Hand coding means that development, testing, deployment, and source code management are mostly manual. Documentation is laborious and often can't keep up. With hand coding, you must use different tools to debug different environments, schedule, and monitor, and hope that it can recover gracefully from failed jobs. Furthermore, hand-coded data integration processes are limited in the transformations involved. For example, to address data quality, a data engineer will need to hand-code a process to standardize codes or cleanse addresses for a specific region.

Hand Coding Lacks Enterprise Breadth

It's taken many years for traditional data integration hand coders to realize how important and necessary data quality and governance are to ensure the business has trusted data. This is even more important today with data-driven companies and the rise of AI and machine learning. Hand coding can't offer enterprise breadth for data integration, data quality, and metadata management.

3 Drawbacks to Hand Coding for Business

Hand coding's limitations aren't just limited to IT. Ultimately, hand coding affects business outcomes. Here are three key areas where hand coding can have a negative business impact:

Higher Costs

Hand coding and scripting labor costs can be significant, and skilled resources are limited. These costs increase over time. While an individual project of limited scope may be quickly and inexpensively delivered, those economies are lost at enterprise scale. There is no concept of reusability with hand coding—sometimes data engineers will end up writing scripts that already exist without realizing it. Skills are not repeatable for future efforts. For example, modern data integration techniques include multi-latency data ingestion, complex ETL/ELT transformations, streaming, cleansing, and parsing data. With a custom hand-coded solution, these techniques are difficult to develop and aren't reusable with newer data or projects. Therefore, the IT organization will need to retrain developers at great expense and with limited future use.

High development costs lead to a high cost of maintenance. If a code change is required, data engineers need to determine the impact of the changes, rewrite code, retest, and redeploy the code and any impacted artifacts such as BI reports. If there are errors in data processing, a hand-coded solution is difficult to debug. As a result, developers are chained to their code in a constantly changing environment.

Higher Risk

Frequent upgrades to the underlying infrastructure require complex coordination. When you hand code, you no longer develop at the pace of innovation. Any time there is a change in the data landscape (for example, an Oracle upgrade, a switch from an on-premises Teradata data warehouse to Amazon Redshift, or a move from SAP to Salesforce CRM), developers must stop what they are doing and maintain their code.

These dependencies increase overall risk when skilled resources are limited. Depending on the skills required, organizations can spend a significant amount of money finding and training the right resource to support home-grown solutions. In many regions it can be difficult to find highly skilled hand-coding developers proficient in the latest technologies, and competition for skilled developers is fierce.

For example, a skilled developer at a global pharmaceutical created an ingestion framework to get data into the company's data lake leveraging Sqoop (an open source component of Apache Hadoop). The process was entirely dependent on one person to maintain and operate the ingestion framework. The developer moved on, leaving no one to maintain or operate the existing framework. No new data was able to make it into the data lake because no one else understood or could maintain the code. The company realized its mistake and switched to an enterprise data integration solution.

“[With Informatica]
We can make timely
system updates
available when needed
without hand coding,
and easily connect
disparate sources
in different locations
without limitations on
the data type
or format.”

— Senior IT Director,
Travel Retailer

Slower Time to Value

Organizations run the risk of delays when they rely on a handful of developers to develop, implement, and maintain the home-grown custom solution. In some cases, these dependencies can significantly impact time to value and result in lost opportunities.

A large energy company had a Hadoop environment that was not returning the expected value. It turned out that, despite searching for six months, the lead developer was not able to find skilled resources to build the required Spark jobs. The company decided to invest in an enterprise data integration tool to get more out of their investments and help the lead developer with the backlog of requests.

A lack of data governance, data quality, and end-to-end data visibility decreases the trusted data available to the enterprise for any kind of digital transformation initiative. Whenever there is a lack of trust or governance, business puts the brakes on, lowering user adoption and thereby delaying any expected business value.

4 Elements of Intelligent, Automated Cloud Lakehouse Data Management

As organizations consolidate and modernize their on-premises data warehouses and data lakes in the cloud or stand up new ones in the cloud, it's more important than ever to avoid the pitfalls of hand coding. Particularly today—with the emergence of lakehouses promising the best of data warehouses and data lakes with cloud agility and scalability—it's essential to leverage metadata-driven intelligence and automation to build efficient data pipelines.

While many IT departments focus just on data integration, a broader solution is needed to solve today's enterprise needs across the entire lifecycle of data management. Here are the key elements.

Data Integration

A best-of-breed intelligent, automated data integration solution is essential for managing cloud data warehouses, data lakes, and lakehouses. The following capabilities enable you to quickly and efficiently build data pipelines to feed your cloud data warehouse, data lake, and lakehouse:

- Codeless integration with templates and AI-recommended next-best transformations
- Mass ingestion for files, databases, changed data, and streaming
- Pushdown optimization to databases, cloud data warehouses, and PaaS lakehouses
- Serverless and elastic scaling
- Spark-based processing in the cloud
- Broad and native connectivity
- Stream processing
- AI and machine learning augmentation to handle schema drift and complex file parsing
- Support for data and machine learning operations (DataOps and MLOps)

Data Quality

In today's world of cloud lakehouses, it's not enough to have world-class data integration. You also need best-of-breed data quality. Intelligent, automated data quality capabilities ensure that data is cleansed standardized, trusted, and consistent across the enterprise.

Here's what to look for:

- Data profiling that is integrated to data governance
- Data quality rules and automated rule generation
- Data dictionaries to manage values lists
- Cleansing, standardization, parsing, verification, and deduplication/consolidation processes
- Integration with your data integration solution
- Data quality analytics
- Spark-based processing in the cloud

Metadata Management

A common enterprise metadata foundation enables intelligent, automated, end-to-end visibility and lineage across your environment. Broad metadata connectivity across disparate data types and sources ensures that you have visibility into and can access data locked in heterogeneous transactional applications, data stores and systems, SaaS applications, and proprietary legacy systems. A common enterprise metadata foundation enables intelligent, automated:

- Data discovery
- End-to-end lineage
- Asset tagging and data curation
- Understanding of technical, business, operational and usage metadata
- Connectivity across on-premises and cloud for databases (data warehouse and data lakes), apps, ETL, BI tools, and others

Cloud-Native Capabilities Built on a Foundation of AI and Machine Learning

The fourth element is foundational and runs beneath the other three. Data integration, data quality, and metadata management should be built on a foundation of AI and machine learning in order to handle the exponential growth in enterprise data. Look for a cloud-native solution that is multi-cloud, API-driven, and microservices-based in addition to offering the following capabilities:

- AI/ML-powered automation, such as next best transformation recommendations, data pipeline similarity, operational alerting, and auto-tuning
- Containerization
- Serverless architecture
- Minimal install and setup
- Auto-upgrades
- Usage-based pricing
- Trust certifications
- Integrated full-stack high availability and advanced security (i.e., platform, network, infrastructure)

“Informatica’s Intelligent Data Platform enables us to leverage the best of open source technology for optimal execution and faster deployment. We can now manage our data in a systematic, repeatable, and sustainable manner.”

— Senior Data Engineer,
Legal Company

The Informatica Cloud Lakehouse Data Management Solution

Informatica® Cloud Lakehouse Data Management is the only enterprise-class, cloud-native, end-to-end data management solution for lakehouses—as well as data warehouses and data lakes—in the cloud.

The solution is built on the industry-leading Informatica Intelligent Cloud ServicesSM (IICS), the industry’s most advanced enterprise iPaaS. Informatica Cloud Lakehouse Data Management combines best-of-breed data integration, data quality, and metadata management in a completely automated, cloud-native solution with foundational metadata-driven AI and machine learning capabilities.

The solution addresses the many complex data management challenges facing businesses that have moved—or are moving—their data to the cloud. With it, you can:

- Lower TCO by increasing efficiency across the entire software development lifecycle (build, deploy, maintain, operate)
- Leverage out-of-the-box connectivity to hundreds of data sources to take advantage of graphical, codeless integration and built-in complex transformations
- Eliminate the risks of using hand coding and limited point solutions for data management by hiding complexity and adapting to data and technology changes
- Achieve all the benefits that as-a-service cloud solutions offer: scale, agility, minimal install and setup, automatic upgrades, high availability, and advanced security
- Get faster time to value by accelerating your efforts to migrate data lakes and data warehouses to the cloud with AI/ML-powered automation that eliminates thousands of manual tasks
- Future proof your data analytics initiatives from the ever-changing public cloud ecosystem and avoid vendor lock-in in a multi-cloud world
- Establish a modern data management infrastructure with comprehensive data integration, data quality, and metadata management

Take a Comprehensive Approach to Intelligent, Automated, Modern Cloud Data Management

Many organizations need data to effectively understand, operate, and grow their business, but data complexity is an impediment. IT organizations are looking for an intelligent, automated data management solution that bridges the gap between on-premises and cloud deployments without having to rebuild everything from scratch before they can reap the benefits of a successful implementation.

Without a unified and comprehensive data platform, organizations are forced to cobble together disparate point solutions that were never designed to work together in the first place. Integrating these systems is time consuming, costly, risky, and inflexible to change. If one point solution changes, then you need to redo and retest all the system integrations.

Taking an enterprise approach does not require a “big bang” implementation. One of the benefits of intelligent and automated data management is that organizations can roll out common methodologies, processes, and technologies incrementally, starting with one or two projects.

Using a high-productivity enterprise data management platform, IT can accelerate initial projects to deliver immediate business value. As the IT organization executes additional projects, it can leverage and reuse existing assets, significantly decreasing the cost and time to deliver new functionality to the business and improving consistency and control.

With the industry’s leading, metadata-driven cloud lakehouse data management solution, you unleash the full potential of your cloud data warehouse and data lake across a multi-cloud, hybrid environment. You gain efficiencies and cost savings and can start small and scale with best-of-breed data integration, data quality and governance, and metadata management—built for the cloud, on an AI-powered, intelligent data platform.

Next Steps

Experience Informatica intelligent, automated cloud lakehouse data management free for 30 days. Sign up for a [free trial of Informatica Cloud Data Integration](#).

