

3 Ways to Increase Your Data Lake ROI

Use real-time data pipeline automation to get better insights, faster.



A new normal.

In today's rapidly evolving digital landscape, multiple technological innovations, market changes and business demands have coincided to change how data needs to be moved, stored, processed and analyzed:



REAL-TIME INSIGHTS

To respond to the breakneck pace of business and intensifying competition, enterprises need to uncover real-time, actionable insights without sacrificing accuracy or quality.



DATA SPEED AND SCALE

As data from all sources (including streaming and transactional) and formats (such as sensor, image, text, video and audio) pours in, businesses have to find a way to process and monetize it – quickly.



ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

AI and ML have become critical to quickly deriving real-time, actionable insights from massive volumes of continuously streaming data.



FLEXIBILITY AND AGILITY

With both the data and analytics infrastructures constantly growing and changing, enterprises need the ability to efficiently scale and conform to shifting requirements.



BUILDING THE BUSINESS

All of the above has to happen with respect for the usual business imperatives – lower costs, higher productivity and faster time-to-market.

Enter the data lake.

To help enterprises adapt and succeed in this new environment, data lakes have emerged as valuable architectural constructs. Designed to efficiently and cost-effectively store and process massive volumes of diverse data sets, they can provide IT organizations with the scalable platform they need to harness all data – and enable analytics for data scientists, analysts and decision-makers.

No longer hampered by all Hadoop-based implementations, data lakes are now more elastic, manageable and cost-effective, leveraging cloud-based platforms, embracing Spark-based real-time processing and more.

Evolving Data Lake Architecture

Legacy Hadoop-Based Architecture

On-prem – In Corporate Data Center

Combined Storage & Compute

Hadoop Distributed File System (HDFS)

MapReduce Batch Processing

Physical Hardware/Servers

Modern Data Lake Architecture

Cloud Platforms (Amazon, MS Azure, GCP)

Decoupled Storage & Compute

Cloud-Based Object Store + File Store
(Amazon S3, ADLS, GCS)

Spark-Based Real-Time Processing + Batch

Serverless/Scalable

Where's the ROI?

However, challenges remain. Despite all the investment in updating data lake architectures, enterprises are still struggling to:

- **Access and ingest data from a wide variety of enterprise sources**
- **Process and refine data to quickly create analytics-ready datasets for AI, ML and data science programs**
- **Establish data integrity and trust with rich metadata and traceable lineage**
- **Provide data consumers access to secure, governed data so they can self-provision data in a timely manner for analysis**

As a result, the ROI isn't materializing quickly enough – or at all.

The culprit: Traditional data integration processes.

What's behind all these difficulties? Outdated data integration processes. Practices like scheduling batch updates and performing manual design and transformations are slow and coding-intensive. They often result in error-prone data pipelines, data integrity and trust issues, and delayed time-to-insight. And they're expensive, requiring significant engineering and data science resources.





3 ways Qlik® accelerates the ROI from your data lakes.

Every enterprise needs real-time, actionable insights based on a single, trusted view of all their data. To make that possible, Qlik's solution for data lake creation automates the entire data pipeline, from real-time ingestion to the creation and provisioning of analytics-ready data sets. In the process, it delivers a continuous stream of the most up-to-date data to power AI, ML and data science initiatives.

Qlik's solution enables you to:

1 Universally ingest and replicate data in real time – for up-to-date data at every moment.

With a market-leading change data capture capability, you can ingest massive volumes of transactional data in real time from virtually any source – databases, data warehouses, legacy mainframe systems, enterprise applications like SAP and more – directly into your data lake.

2 Fully automate your data pipeline – for continuous streams of consumption-ready data.

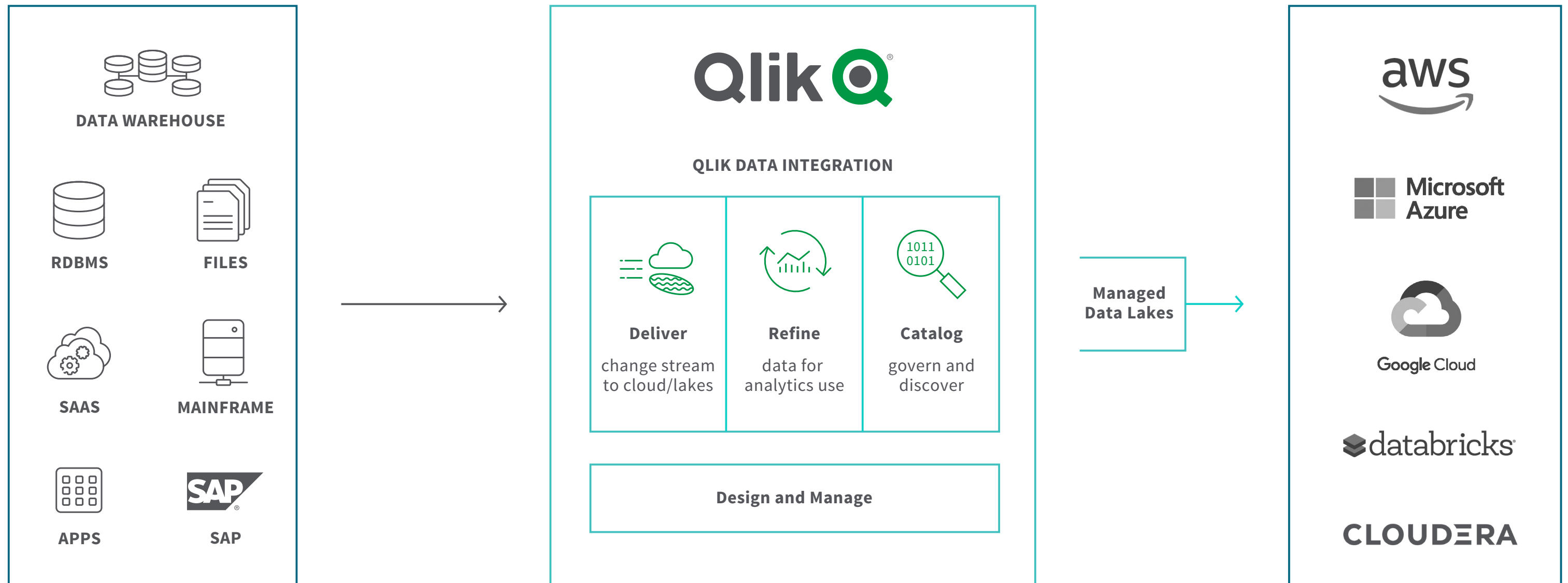
Fully automate data ingestion, transformation and creation of analytics-ready data to drive agility. And reduce risk by minimizing complex, code-intensive ETL modeling and scripting processes.

3 Take advantage of a smart, integrated data catalog – for trusted, governed data.

Provide data science teams and other data consumers visibility into all data – not just in your data lake but across all enterprise sources – through a fully governed enterprise-scale catalog. Give users the ability to easily find, understand and self-provision data for their AI and machine learning initiatives.

The Qlik solution for data lake creation at a glance.

A faster, lower-hassle approach – at scale.



Qlik automates the complete data lake pipeline.

The solution:



Captures and lands real-time change data into data lake of choice



Standardizes, merges and refines data – and subsets it into analytics-ready data sets



Includes a data catalog to auto-generate rich metadata and ensure data security and governance



Provides an orchestration layer for easy setup, monitoring and management of data pipelines



Supports all major cloud providers, source and target endpoints and analytic tools of choice



Universally ingest and replicate data in real time – for up-to-date data at every moment.

Qlik provides real-time change data capture to ingest data directly into your data lake. With continuous capture of the incremental changes to both data and metadata, updates are applied at near-zero latency.

The solution supports out-of-the-box connectivity to the widest set of enterprise data sources and directly loads into cloud object storage on your desired platform. No other solution offers this breadth and depth of source coverage – including databases, data warehouses and enterprise systems like SAP and mainframe.

A fully automated interface eliminates manual coding and gives your data engineers the ability to configure, control and monitor thousands of streaming data pipelines through a single console. Setup is quick and easy; administrators have the power to monitor for peace of mind, control and auditing; and industry-standard SSL encryption ensures secure data delivery.

Fully automate your data pipeline – for continuous streams of consumption-ready data.

The Qlik Data Integration Platform accelerates data delivery by fully automating data pipelines, from the ingestion of data streams all the way through to the creation of analytics-ready data sets – including data transformation and refinement tasks along the way. And the platform scales with your organization’s ever-growing needs, both on-premises and in a multi-cloud environment.

With automated ETL scripting to transform data into an analytics-ready state, your data scientists can focus on high-value modeling tasks. Data engineers can quickly create reusable, automated data pipelines to rapidly add new data sources and streamline the delivery of analytics-ready data sets to data consumers, without any coding.

The platform persists the complete change history through the entire “raw-to-ready” preparation process, for end-to-end data lineage. By capturing changes to data and metadata, it ensures that schema changes are automatically propagated and that data consistency and integrity are maintained.

Take advantage of a smart, integrated data catalog – for trusted, governed data.

Qlik's integrated catalog automatically profiles data and generates rich metadata for all data in the collection. And with metadata provided across all sources – not just your data lake – your users can leverage all enterprise data.

Via a data marketplace, Qlik delivers a modern online shopping experience with powerful search, preview and selection capabilities. As a result, data is easy to access, find and understand. Your data consumers can self-provision the data and act on it within the analytics tool of their choice.

With enterprise-scale access controls and obfuscation capabilities, all data is protected and secure. Via a centralized platform, all security and governance measures are consistently applied, from end to end. This removes potential failure points, tightens security and reduces risk. Finally, data protection features are easy to administer, even in large settings with many users, data sources and/or complex infrastructure.



Automate and accelerate the entire data pipeline.

Customers using the Qlik solution for data lake creation are seeing the following:

✓ **Faster insights and increased ROI**

The accelerated delivery of analytics-ready data into AI, ML and analytics platforms is fast-tracking insights.

✓ **Increased agility and flexibility**

Organizations are gaining the ability to deploy in any cloud configuration; rapidly integrate with an ever-growing array of data sources, targets and platforms; and consume data in the analytic tools of choice.

✓ **Increased efficiency and reduced risk**

The automation of coding-intensive tasks is minimizing error-prone processes and freeing up data specialists to focus on higher-value tasks.

✓ **Improved IT and business alignment**

IT can quickly build scalable data pipelines and enforce centralized security and governance – while giving data consumers the self-sufficiency to easily find, understand and take action on data.

Qlik for data lakes.

Qlik's Data Integration Platform can help you realize a faster return on your data lake investment by continuously delivering accurate, timely and trusted data sets for business insights. The solution provides an unparalleled ability to automate data streams from any source – including mainframes, SAP, databases and data warehouses – into your data lake, either on-premises or in the cloud. You can now unlock the value of all your data by accelerating delivery of analytics-ready data sets for AI, machine learning and analytics programs to maximize efficiency, agility and flexibility.

How will this solution impact data delivery in your business? Start with a free trial.

[Test Drive](#)

And of course, for any questions or a deeper dive, we're here to help.

[Get in Touch](#)



Inteligencia

Contact details

<http://www.inteligencia.co>

+853 6616 1033

support@inteligencia.co



ABOUT QLIK

Qlik's vision is a data-literate world, where everyone can use data and analytics to improve decision-making and solve their most challenging problems. Qlik offers real-time data integration and analytics solutions, powered by Qlik Cloud, to close the gaps between data, insights and action. By transforming data into Active Intelligence, businesses can drive better decisions, improve revenue and profitability, and optimize customer relationships. Qlik serves more than 38,000 active customers in over 100 countries.

qlik.com