Top Cloud Data Warehouses for the Enterprise

Amazon vs. Azure vs. Google vs. Snowflake

Qlik Q®

The Rise of Cloud Data Warehousing

Data warehouses have been staples of enterprise analytics and reporting for decades. But they weren't designed to handle today's explosive data growth or keep pace with end users' ever-changing needs. All that changed when cloud data warehousing emerged.

Cloud data warehousing provides businesses of all sizes with benefits and flexibility they couldn't enjoy before. No longer constrained by physical data centers, companies can now dynamically grow or shrink their data warehouses to rapidly meet changing business budgets and requirements.

Modern cloud architectures combine three essentials: the power of data warehousing, flexibility of Big Data platforms, and elasticity of cloud at a fraction of the cost to traditional solution users. This eBook describes leading cloud data warehouses with noteworthy differences and a proven approach to make them accessible, effective, and efficient for all your data users.

MASSIVELY PARALLEL PROCESSING (MPP)

Data warehouses that support big data projects use massively parallel processing (MPP) architectures to provide high-performance queries on large data volumes. MPP architectures consist of many servers running in parallel to distribute processing and input/output (I/O) loads.

COLUMNAR DATA STORES

MPP data warehouses are typically columnar stores — the most flexible and economical for analytics. Columnar databases store and process data by columns instead of rows and make aggregate queries, the type often used for reporting, run dramatically faster.

Amazon Redshift

Redshift is a fully managed, petabyte-scale data warehouse service in the cloud. You can start with as little as a few gigabytes of data and scale to petabytes. This empowers you to acquire new insights from your business and customer data.

The first step to creating a Redshift data warehouse is to launch a set of nodes, called an Amazon Redshift cluster. After you provision your cluster, you upload your data set and then perform data analysis queries. Regardless of the size of your data set, Amazon Redshift delivers fast query performance using familiar SQL-based tools and business intelligence applications.
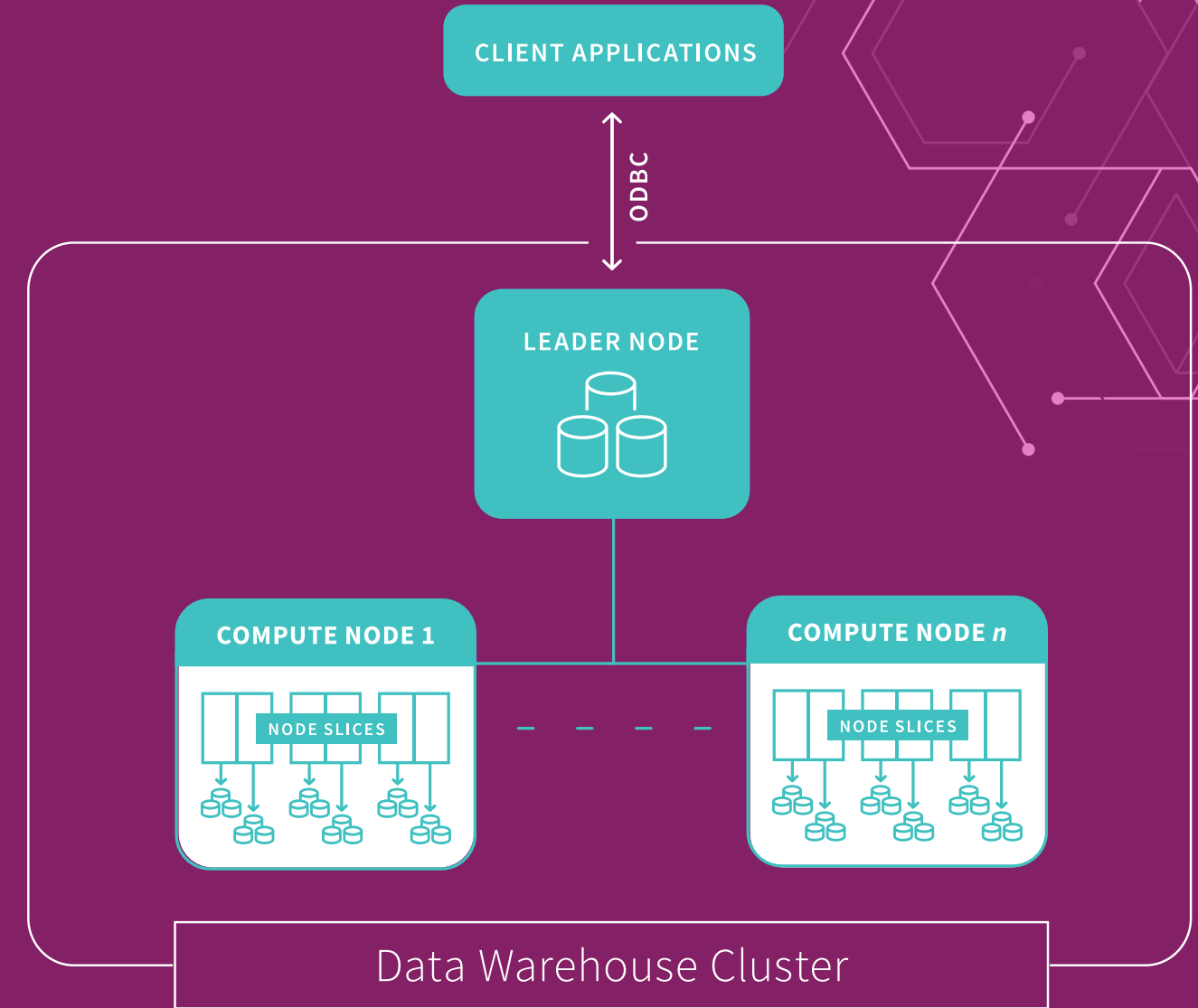

THE FIRST WIDELY ADOPTED CLOUD DATA WAREHOUSE

For many years, data warehousing was only available as an on-premise solution. Then in November 2012 Amazon Web Services (AWS) launched Redshift. Although not the first cloud data warehouse, it was the first to gain market share through adoption. Redshift's SQL dialect is based on PostgreSQL, which is well understood by analysts worldwide, and uses an architecture familiar to many on-premises data warehouses users.

Microsoft Azure Synapse Analytics

Azure Synapse Analytics is a newer analytics service that brings together enterprise data warehousing and Big Data analytics. It gives you the freedom to query data using either serverless on-demand or provisioned resources. Azure Synapse offers a unified experience to ingest, prepare, manage, and serve data for your business intelligence (BI) and machine learning (ML) needs.

At the heart of Azure Synapse is a cloud-native, distributed SQL processing engine. It's built on the foundation of SQL Server to drive your most demanding enterprise data warehousing workloads. Similar to other cloud MPP solutions, Azure SQL Data Warehouse (SQL DW) separates storage and compute, billing for each separately. Azure Synapse saves relational tables data with columnar storage and abstracts physical machines by representing compute power in the form of data warehouse units (DWUs). This allows your users to easily and seamlessly scale compute resources at will.

POWER BI

ECOSYSTEM

AZURE SYNAPSE ANALYTICS

AZURE MACHINE LEARNING
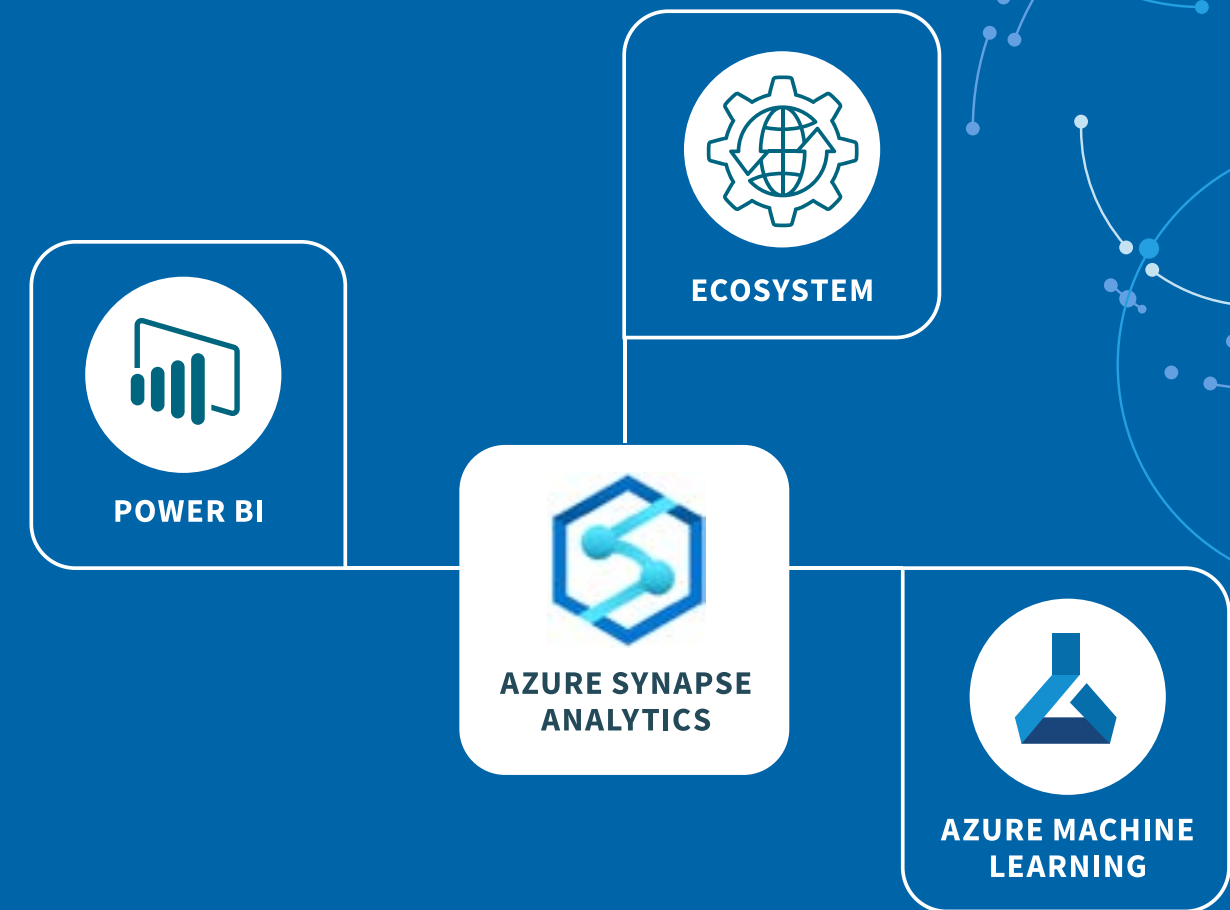
TAKING SQL BEYOND DATA WAREHOUSING

Synapse Analytics aims to unify a range of analytics workloads, such as data warehouses, data lakes, and ML, in a singular user interface (UI). The combination of an SQL Engine, Apache Spark with Azure Data Lake Storage (ADLS), and Azure Data Factory gives users the option to control both data warehouse/data lakes and data preparation for ML tasks. Azure Synapse allows for both vertical and horizontal scaling of the data warehouse. Vertically by changing the service tier or placing the database in an elastic pool. Horizontally by adding more data warehouse units.

Google BigQuery

BigQuery is a fully managed, serverless data warehouse that automatically scales to match storage and computing power needs. With BigQuery, you get a columnar and ANSI SQL database that can analyze terabytes to petabytes of data at incredible speeds. BigQuery also lets you do geospatial data analysis using familiar SQL with BigQuery GIS. In addition, you can quickly build and operationalize ML models on large-scale structured or semi-structured data using simple SQL with BigQuery ML. And you can support real-time interactive dashboarding with BigQuery BI Engine.

The BigQuery architecture is composed of several components. Borg is the compute. Colossus is the distributed storage. Jupiter is the network. And Dremel is the execution engine.
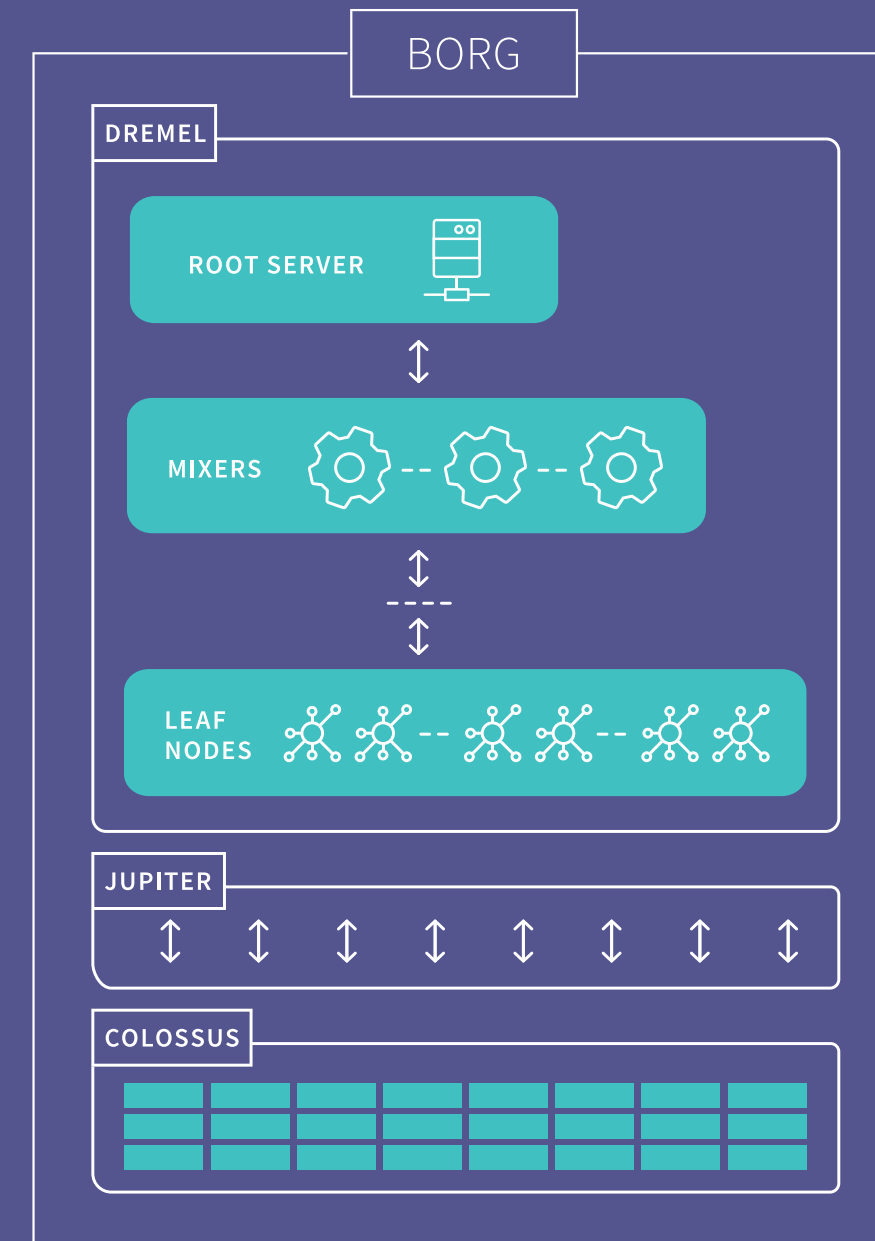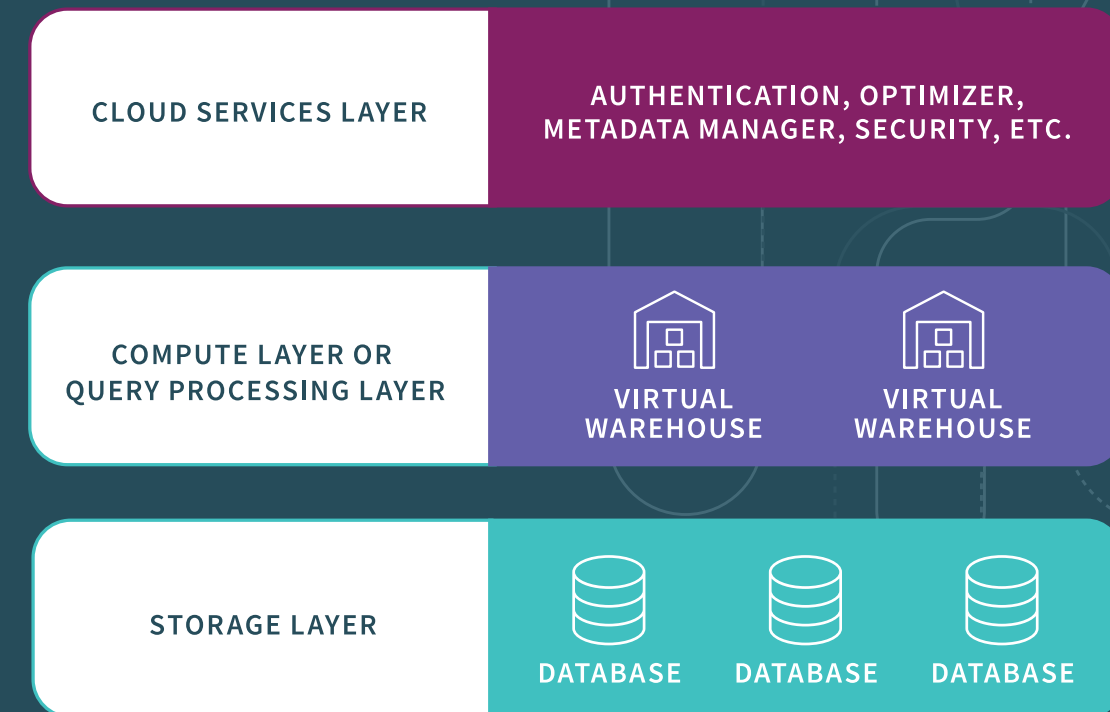

A SERVERLESS SOLUTION

Google doesn't expect you to manage your data warehouse infrastructure which is why BigQuery hides many of the underlying hardware, database, nodes, and configuration details. Its elasticity automatically works out of the box. And getting started is simply a matter of creating an account with Google Cloud Platform (GCP), loading a table, and running a query. Google takes care of the rest.

Snowflake Cloud Data Platform

Snowflake is a fully managed MPP cloud data warehouse that runs on AWS, GCP, and Azure. When you're a Snowflake user, you can spin up as many virtual warehouses as you need to parallelize and isolate the performance of individual queries. Snowflake enables very high concurrency by separating storage and compute to ensure that many warehouses can simultaneously access the same data source.

You interact with Snowflake's data warehouse through a web browser, the command line, an analytics platform, or via Snowflake's ODBC, JDBC, or other supported drivers. The platform supports ACID-compliant relational processing and has native support for document store formats such as JSON, Avro, ORC (Optimized Row Columnar), Parquet, and XML.

Snowflake's hybrid architecture is separated into three distinct layers:

CLOUD SERVICES LAYER	AUTHENTICATION, OPTIMIZER, METADATA MANAGER, SECURITY, ETC.

COMPUTE LAYER OR QUERY PROCESSING LAYER	VIRTUAL WAREHOUSE	VIRTUAL WAREHOUSE

STORAGE LAYER	DATABASE	DATABASE	DATABASE

## THE FIRST MULTI-CLOUD DATA WAREHOUSE

Snowflake, unlike the other data warehouses we've profiled, is the only solution that doesn't run on its own cloud. It's the first multi-cloud data warehouse available globally on AWS, GCP, and Azure. With a common and interchangeable code base, Snowflake features global data replication, which means you can move your data to any cloud, in any region — without having to re-code your applications or learn new skills.

Top Cloud Data Warehouses at a Glance

	Amazon Redshift	Microsoft Azure Synapse	Google BigQuery	Snowflake Cloud Data Platform
Initial Release	2012	2016	2010	2014
Separates Storage and Compute	No	Yes	Yes	Yes
Multi-Cloud	No	No	No	Yes
Query Language	Amazon Redshift SQL	TSQL	Standard SQL 2011 & BigQuery SQL	Snowflake SQL
Elasticity	Yes - Manual	Yes – Manual and Automatic	Yes – Automatic	Yes – Automatic
MPP	Yes	Yes	Yes	Yes
Columnar	Yes	Yes	Yes	Yes
Foreign Keys	Yes	Yes	No	Yes
Transaction	ACID	ACID	ACID	ACID
Concurrency	Yes	Yes	Yes	Yes
Durability	Yes	Yes	Yes	Yes
Automation	No	No	No	No
Website	Link	Link	Link	Link
Free Trial	Yes	Yes	Yes	Yes

Achieve an Agile Data Warehouse

Our Qlik® Data Integration Platform (formerly Attunity) automates the entire data warehouse lifecycle to accelerate the availability of your analytics-ready data. Our model-driven approach helps your data engineers to design, deploy, manage, and catalog purpose-built, cloud data warehouses faster than traditional solutions. Add Qlik to any cloud data warehouse you choose and achieve the cost and efficiency promises of agile data warehousing.
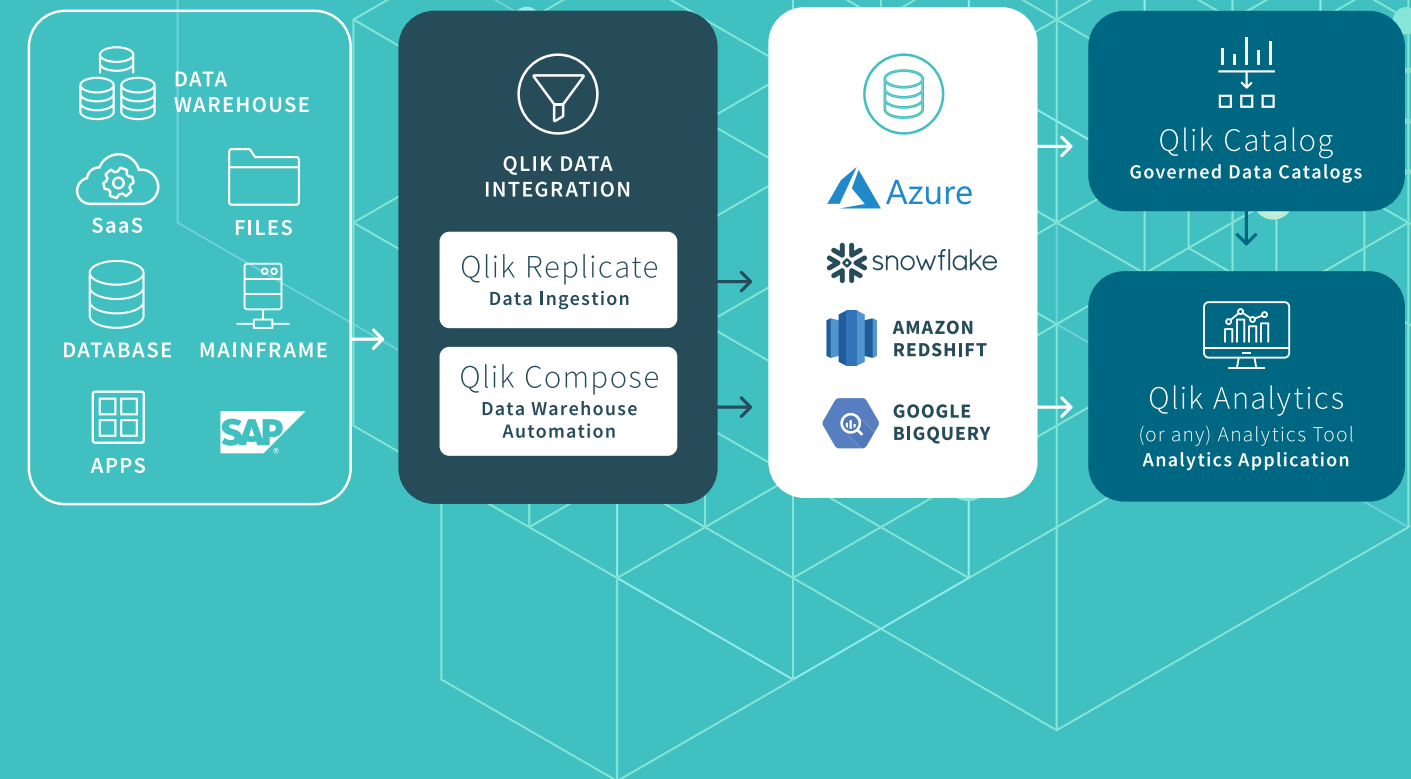
OUR QLIK PRODUCTIVITY DRIVERS

- **Real-time data ingestion and updates** – A simple and universal solution for continually ingesting your enterprise data into popular cloud data warehouses in real time.

- **Automated workflow** – A model-driven approach for continually refining your data warehouse operations.

- **Trusted, enterprise-ready data** – A smart, enterprise-scale data catalog to securely share your data marts.

Choose Cloud Data Warehousing and Innovate with Qlik

The cloud is now the go-to platform for modern analytics. That's why your enterprise needs approaches and technologies that enable analytics in the cloud, in a solution that delivers more value with faster iterations and fewer resources. With Qlik, you automate your data warehouse, optimize your data pipeline, deliver a secure data catalog, and cap it all off with industry-leading analytics.

Qlik is a robust, comprehensive, and innovative solution for enabling world-class data architecture, integration, delivery, and analytics in your business.

Ready to move to agile data warehousing? We're ready to help.

For more information, visit
qlik.com/us/data-warehouse-automation.

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.