

Qlik Talend Data Integration

Copyright © 2018-2024 QlikTech International AB. All rights reserved.
Published: August 2024

1 Qlik Cloud overview	3
1.1 Architecture overview	3
Qlik Cloud Analytics	4
Qlik Talend Data Integration	4
1.2 Qlik Cloud Editions	5
2 Qlik Talend Data Integration	6
2.1 Data pipeline	7
Data pipelines, ETL & other approaches	8
2.2 Data movement	9
Qlik Data Gateway - Data Movement	10
From on-premise to cloud data warehousing	10
Delivering your data to Qlik Cloud	11
From cloud sources to cloud data warehouses or Qlik Cloud	11
Data Architecture patterns	12
Data projects	13
2.3 Data transformation	15
Visual transformations with transformation flows	17
SQL transformations	18
AI-assisted SQL transformations	18
Third-party data transformation	19
2.4 Security and governance	19
Connection and perimeter security	20
Data storage	20
Metadata storage	21
Data encryption	21
2.5 Data Products	23
Data Marketplace	24
2.6 Qlik Application Automation	25
Architecture	25
Reliability	25
Automations	26
Run modes of automations	30
Use cases	31
Security model	31
Governance	32
Data privacy and compliance	32
3 About Qlik Evaluation Guides	33
3.1 Document history	33
3.2 Changelog	33
Changelog — Qlik Talend Data Integration evaluation guide	33

1 Qlik Cloud overview

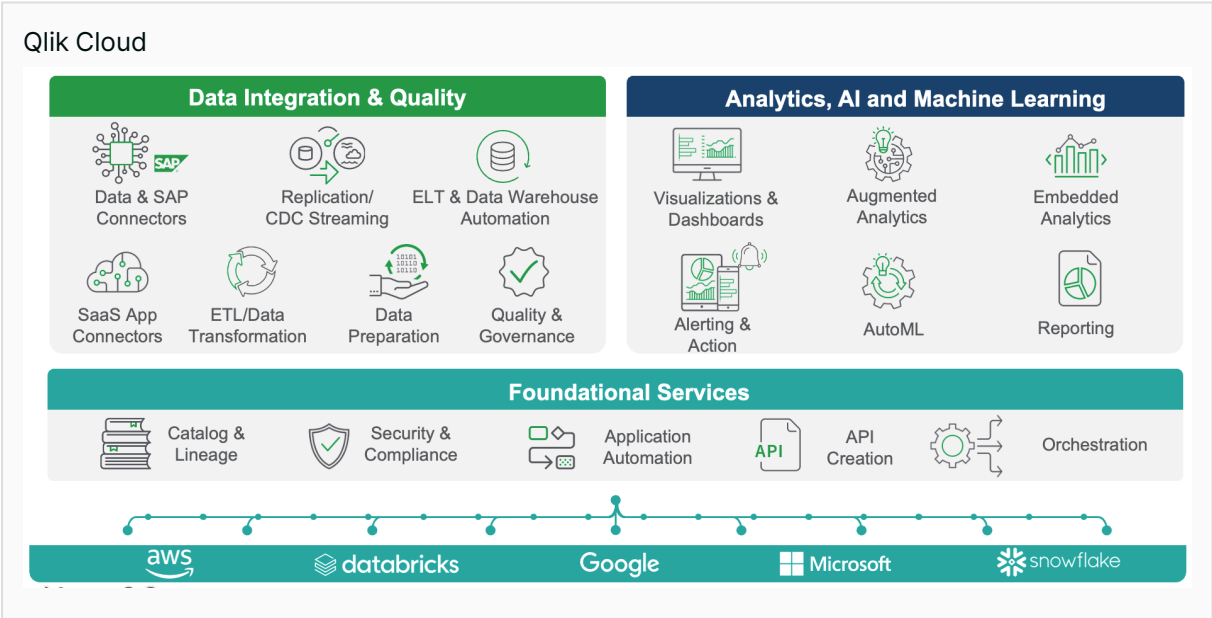
Qlik is a leader in data and analytics with a core mission to provide solutions that ensure organizations can work smarter and use data as a competitive edge. Qlik Cloud is a powerful end-to-end solution for data and analytics services. Our platform empowers curiosity-driven exploration offering everyone – at any skill level – the ability to use data to make transformative change for their organization.

Through several data-focused services, the Qlik Cloud platform supports a full range of users and use-cases across the lifecycle from data integration to insight generation. These services include change data capture, transformation, data cataloging, application automation, self-service analytics and dashboards, conversational analytics, custom and embedded analytics, and alerting.

This document highlights key aspects of the Qlik Cloud platform, including architecture, security, governance, and reliability. It is designed to complement the technical documents for the Qlik solutions that run on the Qlik Cloud platform.

1.1 Architecture overview

All of Qlik’s SaaS offerings and services, known collectively as the Qlik Active Intelligence Platform, run on the Qlik Cloud platform. The platform delivers the underlying compute, storage security, and governance features to provide services to our customers. The Qlik Active Intelligence Platform enables the creation of the analytics data pipeline. Powered by Qlik Cloud and a rich set of foundational services, it provides all the data integration and analytics services you need to transform raw data into informed action.



A customer’s instance of the Qlik Cloud platform is called a tenant. It is logically separated from other tenants by using unique encryption keys. Access to the platform is controlled by the customer’s configured identity provider and any access to functions within the platform is based on the entitlements the customer has assigned across roles and users. A number of services are available on the Qlik Cloud platform:

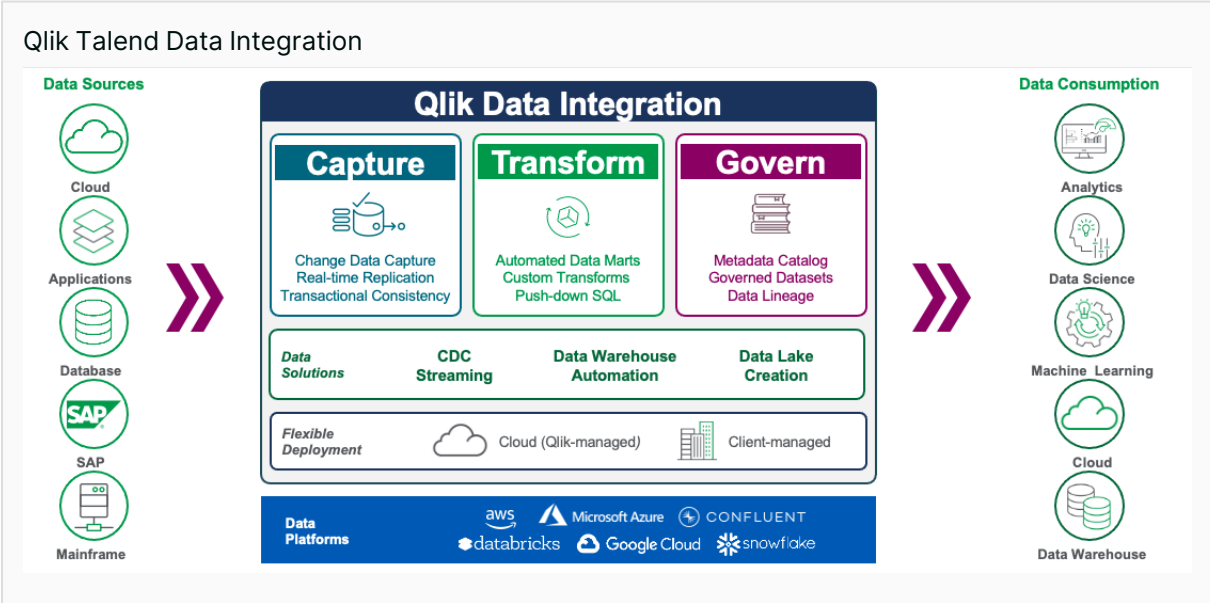
- Analytics – provides a complete third-generation analytics solution including Qlik Sense Enterprise SaaS
- Data Integration – provides the ability to manage your data assets and utilize change data capture to provide real-time access to your data, as well as application automation to automate integrations between cloud applications.

Qlik Cloud Analytics

Incorporating our premier offering Qlik Sense, Qlik Cloud Analytics sets the benchmark for third-generation analytics platforms, empowering everyone in your organization to make data-driven decisions. Built on our unique Associative Engine, it supports a full range of users and use-cases across the lifecycle from data to insight: self-service analytics, interactive dashboards, conversational analytics, custom and embedded analytics, mobile analytics, reporting, and alerting. It augments and enhances human intuition with AI-powered insight suggestions, automation, and natural language interaction.

Qlik Talend Data Integration

Qlik Talend Data Integration is Qlik’s hosted and managed data integration platform as a service (iPaaS). Our vision is to provide a broad variety of data integration services aimed at helping you move from passive to active BI. Qlik Talend Data Integration is architected for real-time data capture, transformation, and analytics-ready data delivery leveraging a variety of methods in our unique change data capture approach.



Real-time data movement

Replicate data from on-premises or cloud sources into Qlik Cloud and other leading cloud data platforms. Automatically and continuously ingest data without the need for job scheduling or scripting. Your data is updated without manual intervention to drive insights and actions of important business moments.

Data transformation

Quickly turn raw transaction records into consumption-ready data via auto-generated, push-down SQL. Our no-code interface helps you create reusable transformation pipelines that intelligently conform data to dimensional models or custom formats.

Qlik Application Automation

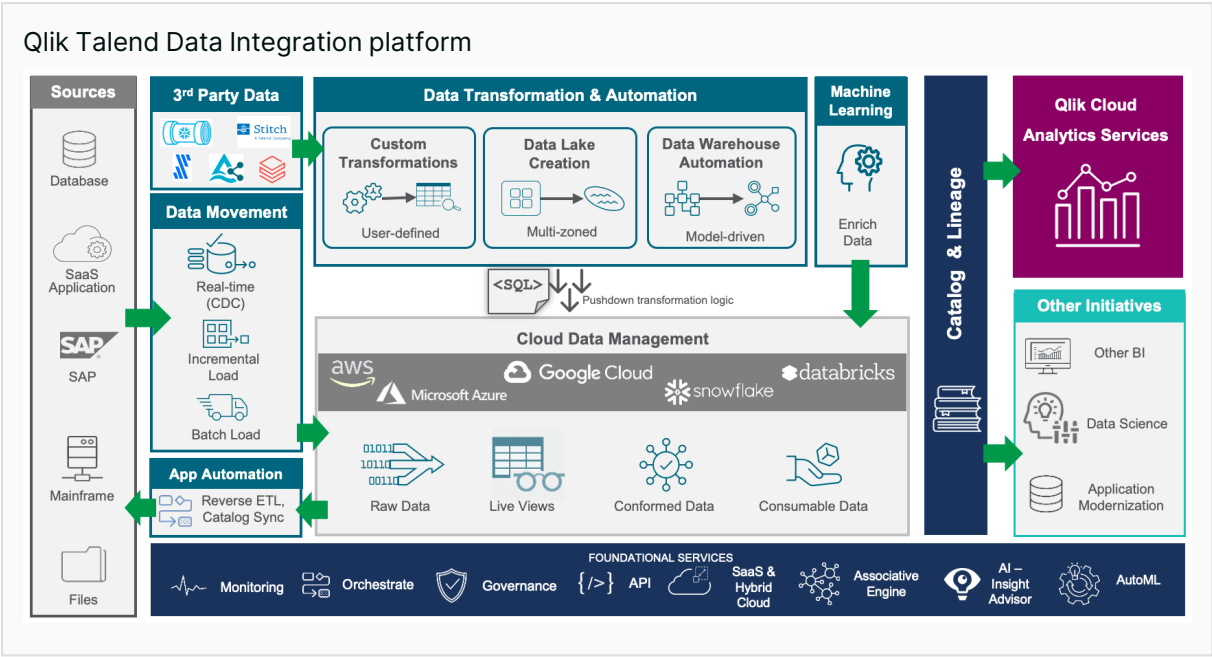
Qlik Application Automation is an integration platform to build integrations and automation flows between cloud applications. Closely integrated with the other Qlik Cloud services, Qlik Application Automation is able to build workflows between your cloud applications using a no-code approach by connecting data sources, applying conditions, calling webhooks, adding loops, scheduling runs, and setting up triggers. For example, a webhook in your CRM system could initialize a reload of your sales performance Qlik Sense application.

1.2 Qlik Cloud Editions

There are several Editions of Qlik Cloud to meet the varied needs of our customers. The Evaluation Guides are designed to cover all the functionality of Qlik Cloud Analytics, Qlik Talend Data Integration and the Qlik Cloud Platform. There is also a special version of Qlik Cloud for Government organizations in the US, known as Qlik Cloud Government. Information provided in the Evaluation Guides does not imply that those features are available in all Editions.

For information on your edition, see the relevant link below:

- [Qlik Cloud Analytics user-based subscriptions](#)
- [Qlik Cloud Analytics capacity-based subscriptions](#)
- [Qlik Talend Data Integration capacity-based subscriptions](#)
- [Qlik Cloud Government](#)



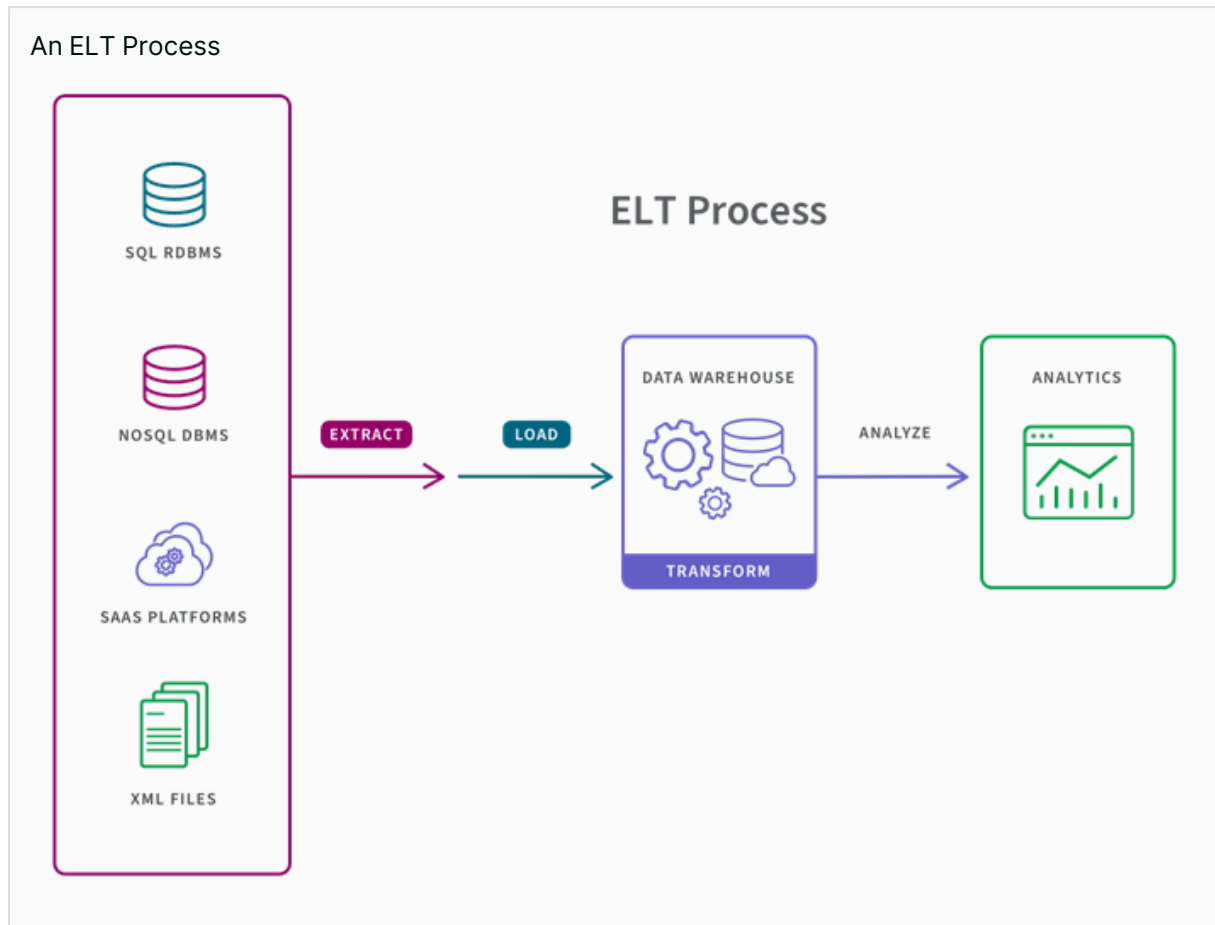
These services rely on the underlying platform services provided by the Qlik Cloud platform. For more information, see the Qlik Cloud platform Evaluation Guide.

2.1 Data pipeline

A data pipeline is an end-to-end process to move data from source to target, including any required transformations. A pipeline can be as simple as a straight mirroring of data from source to target, or as complex as a complete enterprise data warehouse solution including multiple data marts serving a diverse range of requirements.

A data pipeline takes your source data, often from multiple operational systems, and moves the data to where you need it. It transforms, merges, and validates the data into a consumption-ready format.

Data pipelines, ETL & other approaches



A data pipeline is a high-level concept. The actual implementation of the data pipeline will be based on one or more of the following techniques:

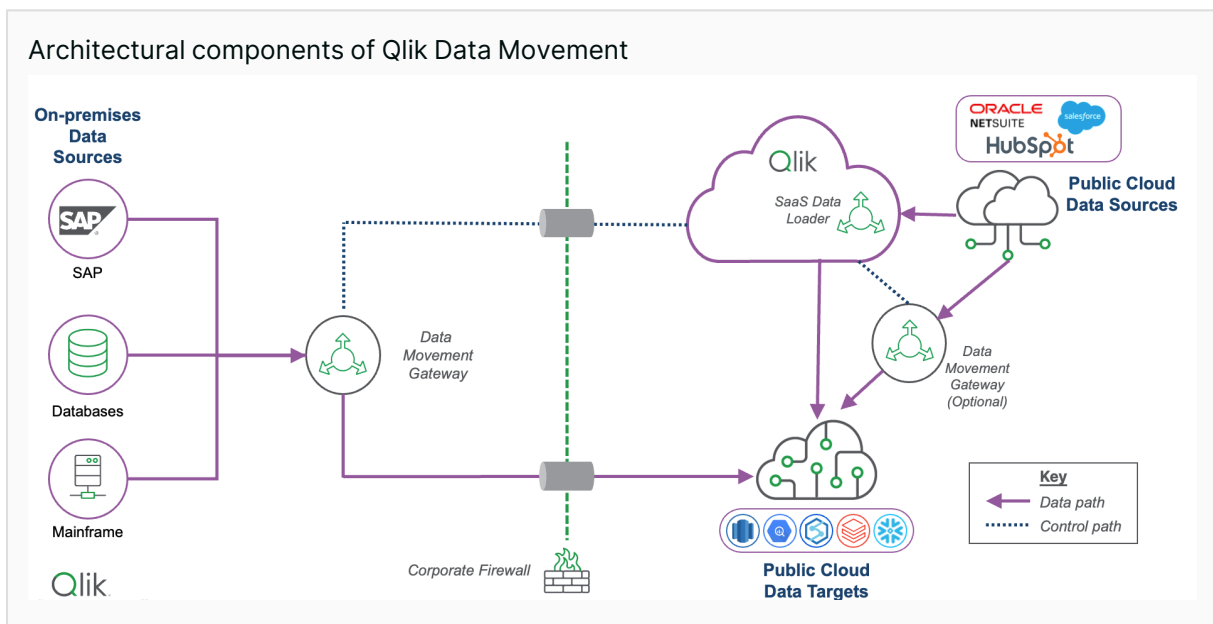
- **ETL: Extract, Transform & Load.** ETL is one of the oldest techniques in data movement and has been widely used since the 1970s. Data is extracted from the source system, then processing and transformation occurs within the ETL tool before the transformed data is loaded into the target system. Talend Studio is an example of a popular tool that supports developing ETL pipelines.
- **ELT: Extract, Load & Transform:** ELT is an alternative technique to ETL, where the data is loaded into the destination system before being transformed. SQL pushed down from the ELT tool performs the actual transformations, although this is usually transparent to the user and is generated by the ELT tool. ELT is generally seen as performing better due to this, however it is limited by the capabilities of the destination system. Qlik Talend Data Integration is largely based on ELT, however where necessary can use some ETL techniques.
- **Event Streams and Message Queues:** Both variants of the message broker architecture, these techniques deal with data that is in motion rather than read from in a static data source. Operations on these technologies cannot rely on having access to the full data set

and often will process a subset rather than the full data set. These technologies are often used in combination with traditional databases and will be integrated into a ETL/ELT based solution. Talend Studio is an example of a tool that supports transformations against these technologies.

Qlik understands that there is not a single type of data pipeline that meets every need and our focus in developing Qlik Talend Cloud is to provide solutions that meet our customers needs.

2.2 Data movement

Qlik Data Movement helps customers onboard their data rapidly and securely from their on-premise and cloud-based data sources to cloud data warehouses and data lakes. An end-to-end solution for data movement, this service securely accesses data sources, automatically replicates in real-time to cloud targets, and catalogs data sets without manual scripting.



The data movement process in Qlik Talend Data Integration is managed from the Qlik Cloud hub. It initializes and monitors the process of capturing data from Enterprise and Cloud application data sources.

Data from on-premise systems or running in a customer's cloud does not pass through, nor is it stored, in Qlik Cloud, unless Qlik Cloud is the chosen destination for the data. SaaS application source data is captured by Qlik Cloud and stored transiently while data flows from source to target via Qlik Data Gateway - Data Movement. In the case of Cloud data sources, data can be pulled directly from the source without the need for a gateway.



A note on Qlik Data Gateway - Direct Access This paper will not detail the functionality of Qlik Data Gateway - Direct Access. This gateway has a different purpose and helps solve different use cases than does the Qlik Data Gateway - Data Movement. The Direct Access gateway is considered a Qlik Cloud Analytics component, where users can connect directly to on-premise data sources from an app in Qlik Cloud Analytics and load data from there. It is not technically speaking a data integration tool, which means it will not be covered here.

Qlik Data Gateway - Data Movement

A challenge for many customers when moving to SaaS is providing access to their on-premise and private cloud data sources without compromising security. Qlik's solution to this is Qlik Data Gateway - Data Movement. This allows customers to access data sources in their data center and private cloud, without exposing them to the public internet.

Qlik Data Gateway - Data Movement is a component controlled from Qlik Cloud, but physically located near your data. It initiates connections to your source and target systems, orchestrating both full loads and change data capture (CDC). For simplicity we will refer to this simply as the Data Movement Gateway going forward.

Source data is onboarded directly into and persisted to the target cloud platform by the Qlik Data Movement Gateway, removing the need to expose data sources to the internet.

When started, the Data Movement Gateway makes an outbound connection to Qlik Cloud, which then initiates a reverse tunnel back to the gateway for command and control.

From on-premise to cloud data warehousing

Delivering data from on-premise data sources to cloud data warehouses is achieved with the Data Movement Gateway:

- **Source** — The data source types available in the Data Movement Gateway will govern the sources of data that can be delivered from. The sources are divided into two types:
 - Database sources: At the time of writing (April 2024) over 20 source database platforms were supported. see [Connecting to databases](#) in help for an up to date list of supported databases.
 - SaaS Sources: At the time of writing, 60 SaaS applications were supported. See [Supported SaaS applications](#) in help.
- **Target** — Many target data warehouse platforms are supported by Qlik Talend Data Integration (see [Connecting to cloud data platforms in your data projects](#) in the help). Currently, those targets are:
 - Snowflake[®]
 - Azure Synapse Analytics[®]
 - Databricks[®]
 - Google BigQuery[®]

- Amazon Redshift[®]
- Microsoft SQL Server[®]
- Microsoft Fabric[®]

Both staging and storage will happen in the target system when data is delivered via the pipeline. It is possible to use targets in a private cloud; these connections will be proxied via the Data Movement Gateway.

Delivering your data to Qlik Cloud

You can deliver data from on-premise and cloud data sources directly to Qlik Cloud and store as QVD files (Qlik's proprietary file format, designed for fast loading into memory) with the Data Movement Gateway.

- **Source** — The data source types available in the Data Movement Gateway will govern what sources data can be delivered from. We are regularly adding new sources. See [Data sources](#) in the help for details on the latest available sources.
- **Target** — There are two options for target storage of these files, Qlik-managed storage and Customer-managed storage:
 - The Qlik-managed storage option requires customers to bring their own Amazon S3 bucket for the staging area. This storage is configured, maintained, and financed by the customer. Qlik will however provide storage for the storage area once the files have passed staging and are stored at rest. This is recommended if your goal is to make the data available for Qlik Cloud Analytics.
 - The Customer-managed storage option means the customer brings their own Amazon S3 bucket for both the staging area and the storage area, which means configuring, maintaining, and financing them. This is recommended if you need to make the data available to sources in addition to Qlik Cloud Analytics.

From cloud sources to cloud data warehouses or Qlik Cloud

Delivering data from cloud sources and storing it directly in cloud data warehouses is also possible with Qlik Talend Data Integration. This still requires the Data Movement Gateway for some scenarios. This allows us to support public and private cloud data warehouses and data lakes as source and target.

SaaS Data Loader

For Many SaaS platforms and some cloud database platforms, we are able to move data without a customer-deployed gateway. This SaaS data loading capability uses cloud-managed infrastructure and supports:

- Full and incremental loads
- No gateway or driver management required
- All third party SaaS targets
- Periodic Change Data Capture (frequency varies by edition, increasing in higher tiers)
- Data Movement only (no transformation or pipeline tasks)

Use-cases not covered in the above list require a data gateway to be deployed. The SaaS data loader capability is the only method supported in Starter edition of Qlik Talend Data Integration. For other editions this is supported in addition to the Data Gateway - Data Movement.

Data Architecture patterns

A cloud data warehouse built by Qlik Talend Data Integration creates artifacts to support a number of key data warehousing patterns. These include:

- **Landing:** Landing contains an up-to-date copy of the raw data "landed" from your sources by the gateway. This could be kept up to date using either CDC or full loads. The Landing zone is designed for internal use, not consumption, and Qlik recommends against using landing tables for any downstream tasks.
- **Storage:** Storage contains tables and external views based on the data in landing. When consuming data, the best practice is to use views which provide improved data concurrency.
- **Views:** Live views include data from change tables that are not yet applied to the current or prior tables. This lets you see data with lower latency and reduces the costs of processing requirements in the target platform. You can access both current data (ODS) and historical data (HDS) using live views. Depending on your project settings, the following views will be available:
 - Current view
 - Live view
 - Changes view
 - History view
 - History live view
- **Tables:** The following tables are created:
 - Current table (ODS): This table contains the replica of the data source updated with changes during the latest apply interval.
 - Prior table (HDS): This table contains type 2 historical data. It is only generated if History is enabled in the data task settings.
 - Changes table: This table contains all changes that are not yet applied to the current table. It is only generated if the landing mode Full load and CDC is used.



The benefits of delayed merge

In a cloud data warehouse architecture, data is not available for analysis until it has been physically loaded into the Data Warehouse or Data Mart. This is often an expensive (in terms of resources, if not costs) and time-consuming operation and may prevent the latency required by end users. Cloud data warehouses typically have multiple charge metrics including compute up-time, data scan and processing. This often leads to a trade-off between latency, cost and business need which can lead to the expense of running these loads regularly through the day.

Qlik Talend Data Integration however provides an alternative solution for this challenge. The combination of delayed merge with the live views created by Qlik Talend Data Integration provides a solution which supports the best of both worlds; where a use-case demands real-time access, a live view provides this by joining the change tables with the prior loaded tables, for batch use cases they can use the current views which are highly performant. Delayed merge allows users to compress data changes at wider intervals while also providing an updated copy along with a full type 2 history.

Data projects

The first step to create a data pipeline in Qlik Talend Data Integration is to create a project. A project defines the use-case, source, and onboarding of the data. This is sufficient for a complete pipeline for some use cases. Qlik Talend Data Integration supports two use cases, Replication and Data Pipeline.

Replication

A data replication task uses change data capture (CDC) to efficiently and securely move data between a source and target system. Replication can be to an RDBMS, Data Lake, or Qlik Cloud target.

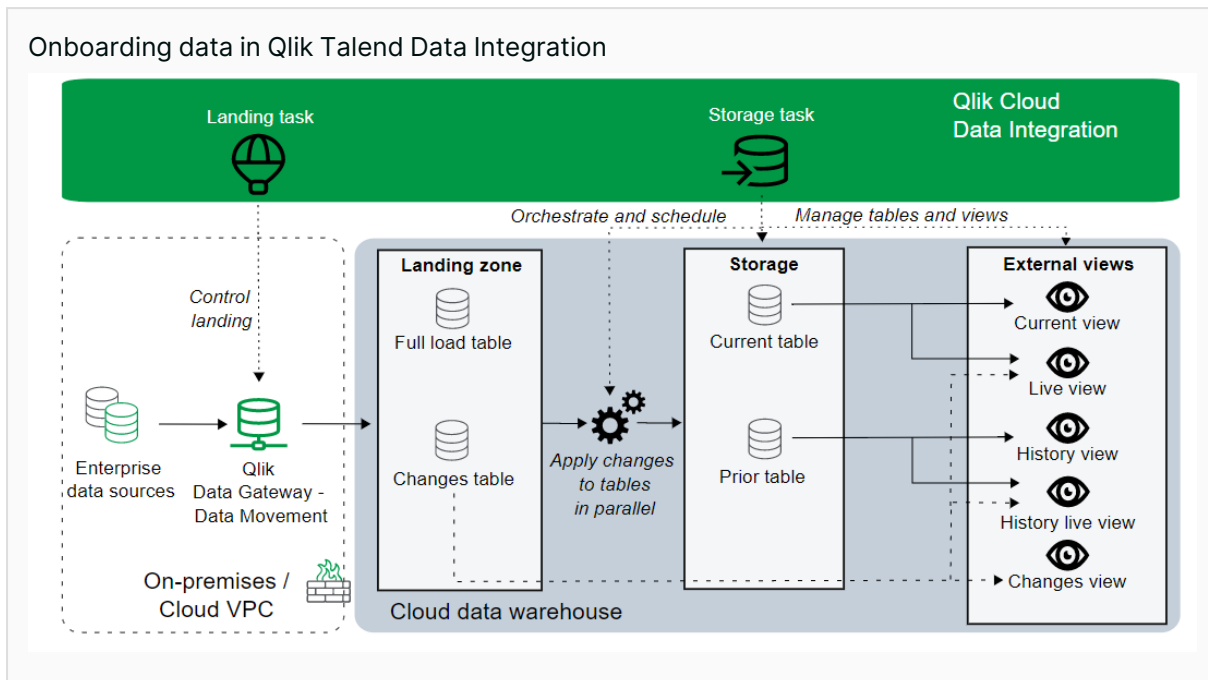
Data Pipeline

Data Pipeline projects allow you to modify and enrich the data, join disparate data sets, and transform the data into the formats you need such as Snowflake schemas or Data vaults. Data Pipelines can be created with our visual no-code designer, with custom SQL tasks, or created for you by generative AI.

Onboarding Data

The initial focus of a data project is the onboarding the data. This involves transferring the data continuously from the on-premise or cloud data source and generating datasets in read-optimized format. Onboarding involves two steps: landing and storing

- **Landing** – The data involves transferring the data continuously from the data source to a landing area, using a Landing data task.
- **Storing** – The data involves generating datasets based on the landing data, using a Storage data task.



Key Concepts in a data integration project

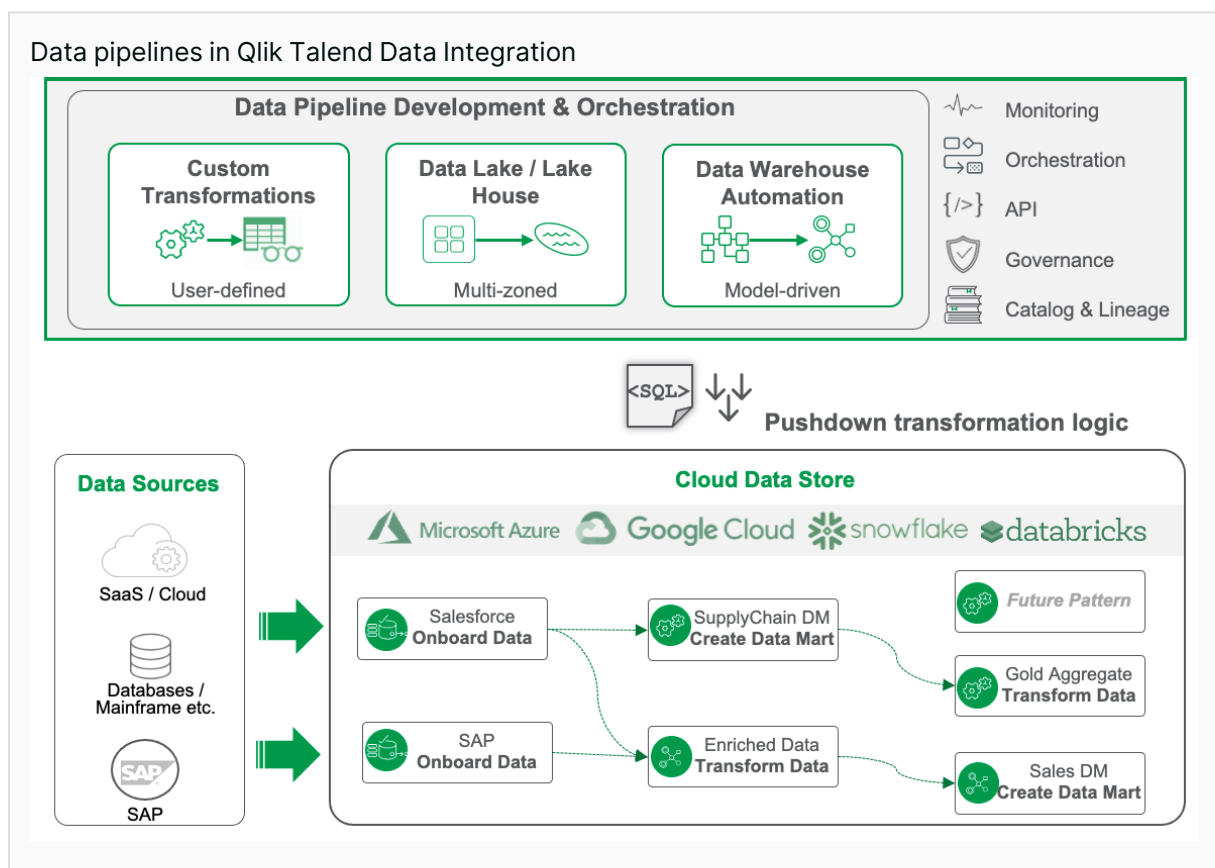
A data integration project is how we build, run, and monitor data pipelines.

Concept	Relationship to project	Description
Data tasks	Component	Data tasks are a fit-for-purpose collection of tables or files and an associated operation on those files. It is the main unit of work within a project in the data project. Examples of data tasks include transform and data mart.
Data spaces	Dependency	Data spaces are governed areas of your Qlik Cloud tenant that are used to manage projects and their data assets. Access to a data space is determined by membership to the space. Access to projects and their data assets inside a data space is determined by roles assigned to members of the space. This means that a user must first be a member of the data space, and second, have the required roles to create, manage, or monitor data assets and resources in a data space. Members with the roles to consume data assets can also use data assets from a data space when building apps in personal, shared, and managed spaces.
Data Gateway	Dependency	Data Gateway is used by the landing data asset for associating a replication task with it, as well as for control and basic monitoring of this task.

Concept	Relationship to project	Description
Data connection	Dependency	A data connection is used by the storage data asset for connecting to AWS S3 buckets or cloud data warehouses, for the purpose of either reading from the staging area or writing to the customer-managed storage area.
Registered data	Component	Registered data is similar to a data task, however it does not perform and actions against the data directly. It is designed to expose data landed outside of Qlik Cloud to the data project, so it can then be used in the data pipeline.

2.3 Data transformation

The data transformation service in Qlik Talend Data Integration provides ELT capabilities to cloud data warehouses and data lakes. It is a key part of Qlik's vision to provide our customers with a raw-to-ready data pipeline.

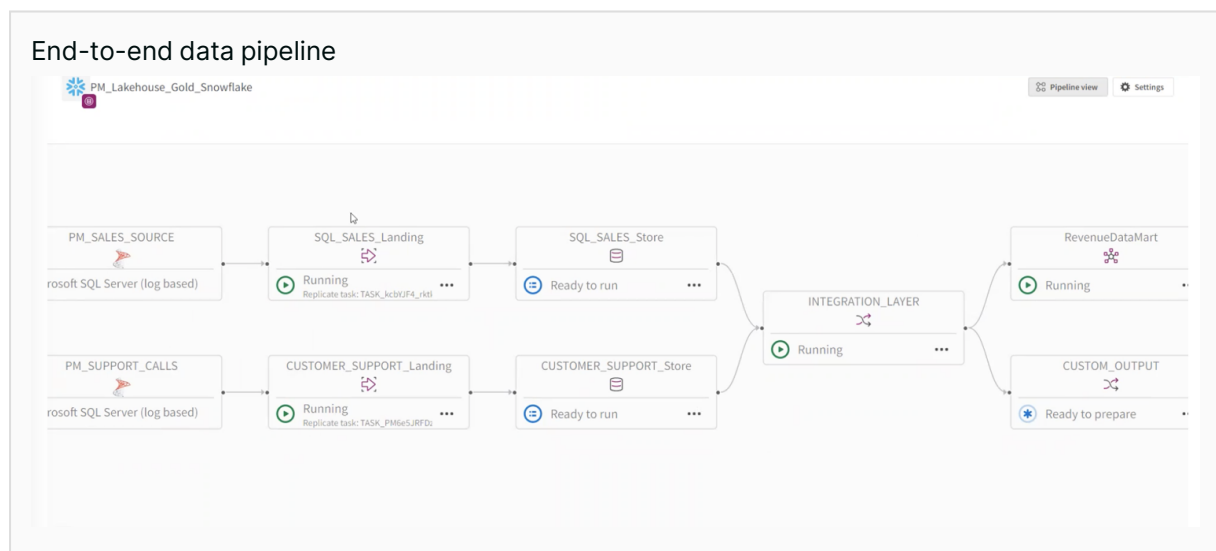


The data transformation service is:

- **Cloud-based:** Customers design, deploy, and monitor data pipelines in Qlik Cloud with transformations executed via SQL in popular cloud targets.
- **Flexible, template-driven approach:** Customers can define reusable transformations, set policies, and create custom design templates to simplify and accelerate the development of their data projects.
- **Automated:** Automation of common methodologies and DataOps best practices allow customers to operationalize their transformation workloads quickly and reliably.
- **Integrated:** Pipeline tasks are integrated with the data movement service to transform data in near real-time to analytics-ready data in your target cloud platform.

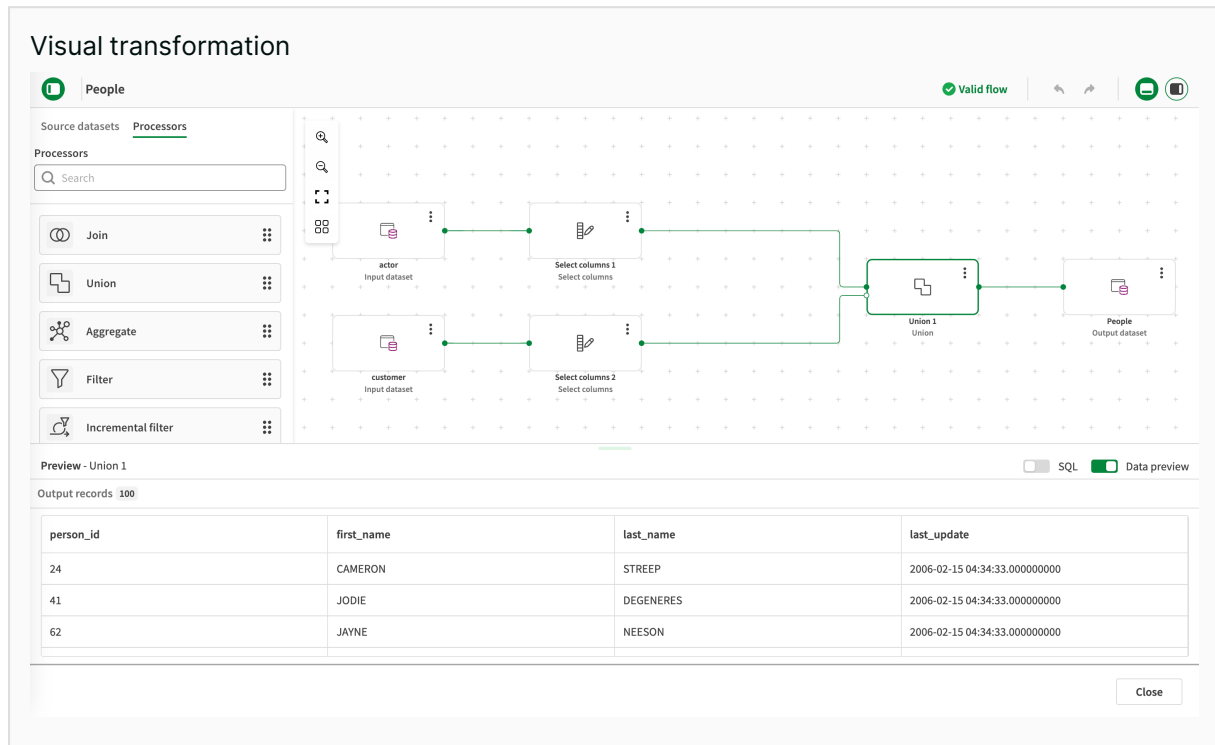
The data transformation service contains the following functionality:

- Creation of flexible, fit-for-purpose data pipelines
- Rule-based, row-level transformations
- Creation of new derived objects based on:
 - Source to target mappings
 - Visual transformation designer
 - AI-assisted transformations
 - Custom SQL for more complex logic
 - Automated generation of star schema data marts
 - Defined logical relationships between data sets
- Choice to materialize data sets as tables or create as views
- CDC (change data capture) support for low latency and incremental data movement
- Push-down SQL execution to cloud DW platforms (Snowflake, Azure Synapse, Google BigQuery, Microsoft SQL Server)



Visual transformations with transformation flows

The transformation flow designer allows you to create a transformation flow visually, using sources, processors, and targets to define complex or simple transformations. Processors allow you to perform common ELT operations in a visual no-code environment.



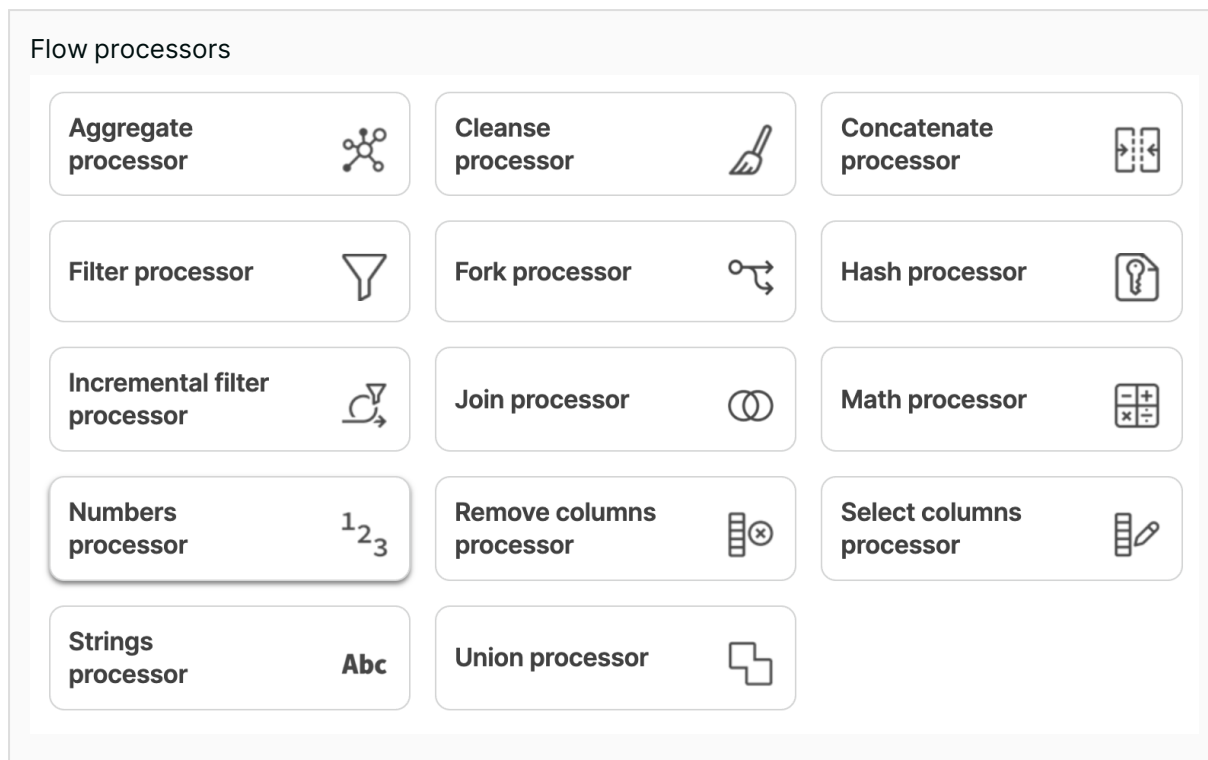
The screenshot displays the 'Visual transformation' interface in Qlik Talend. The main workspace shows a flow starting with two input datasets: 'actor' and 'customer'. Each dataset is followed by a 'Select columns' processor. The outputs of these processors are combined in a 'Union 1' processor, which then feeds into the final 'People' output dataset. The interface includes a 'Processors' sidebar with options like Join, Union, Aggregate, Filter, and Incremental filter. A 'Preview' window at the bottom shows the output records for 'Union 1'.

person_id	first_name	last_name	last_update
24	CAMERON	STREEP	2006-02-15 04:34:33.000000000
41	JODIE	DEGENERES	2006-02-15 04:34:33.000000000
62	JAYNE	NEESON	2006-02-15 04:34:33.000000000

Flow processors allow you to perform common transformation tasks such as Joins, unions and aggregation, filtering and many more.

The transformation capabilities also offer advanced data cleansing capabilities. Businesses can identify and correct data quality issues such as missing values, duplicate entries, and inconsistencies. This helps ensure that their data is accurate and reliable, leading to better insights and decision-making.

If you need to do something that a visual transformation can't do, you can feed one transformation as a source to another, which could add advanced logic without to the need to hand-code all the logic.

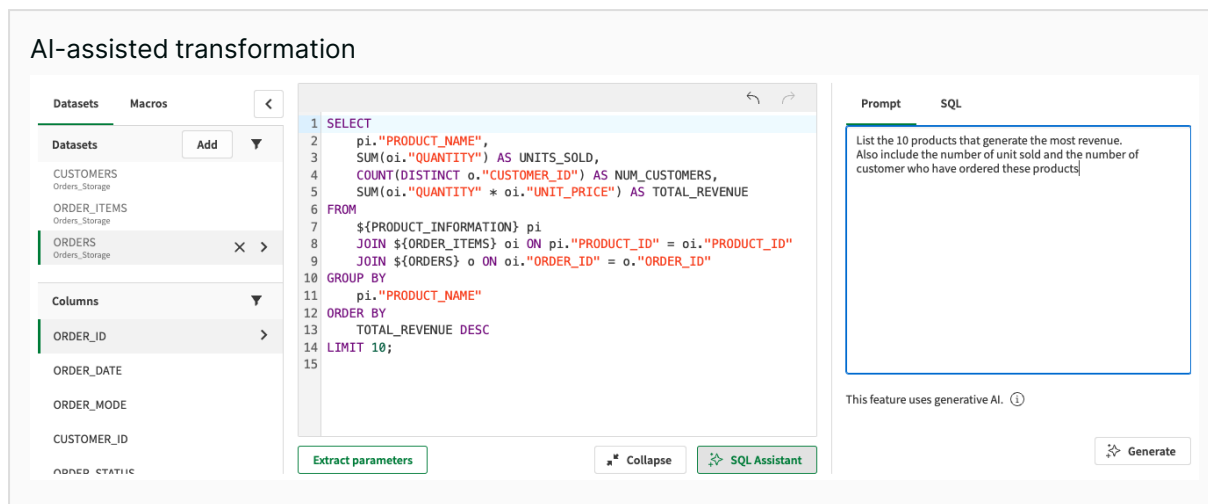


SQL transformations

Where you need to perform operations not supported by visual transformations, SQL transformations can be used. SQL transformations use select statements to transform the data into your required output. The SQL syntax is determined by the target platform, so exact functionality varies by platform.

AI-assisted SQL transformations

AI-assisted transformations use the metadata from your dataset, combined with your plain english request, to generate a SQL transformation with generative AI (Gen-AI). This feature is disabled by default and is dependent of the availability of Gen-AI features in your chosen region. You can then choose to use the generated transformation as-is, or modify it. It is important to review the generated SQL to ensure Gen-AI has correctly interpreted your request. AI-assisted transformations are aware of the underlying target so will generate SQL syntax specific to that target.



Third-party data transformation

Third-party data transformation in Qlik Talend Data Integration refers to the process of registering already existing data that has been landed in the chosen cloud platform by external tools (including Qlik Replicate and Talend Studio).

This means that customers can build workflows on top of and incorporate existing data into data pipelines without having to duplicate existing processes and consume the data twice. This includes to create transformation tasks, data cleansing, and data warehouse automation.

Use-cases for third party transformation include:

- A temporary process during migration from legacy tools to Qlik Talend Data Integration
- Leveraging an existing cloud data warehouse or data lake for a new requirement
- Allowing Qlik Talend Data Integration to work with a propriety solution where connectivity is not available

Third-party transformation supports key principles of Qlik: leave the data where it is, register it, understand it, make it qualitative, and start delivering it.

2.4 Security and governance

A Qlik Talend Data Integration deployment is made up of three key components:

- Qlik Talend Data Integration
- Cloud storage:
 - File stores (e.g. S3)
 - data warehouse tables
- Data Gateway - Data Movement

There are several aspects of how we secure different data delivery strategies, which we will discuss here.

Connection and perimeter security

All the connections from on-premise to the cloud are outbound. This means there is no need to open any port through the corporate firewall, nor to provide a publicly resolvable IP address. The connection is initiated by the data movement gateway and promoted to a web socket secure (WSS) connection, which allows Qlik Cloud to send command and control to initiate the data synchronization. All data in transit is secured by the TLS 1.2 and 1.3 protocol. Command and control plane communication is additionally encrypted with a random channel key which is changed every time the gateway connects to Qlik Cloud.

Data storage

Qlik Talend Data Integration does not transfer data into Qlik Cloud for the purposes of running a data pipeline, nor does it cache this data in Qlik Cloud. Data is only stored in Qlik Cloud if one of the following is true:

- Qlik Cloud is the target platform
- Connection to source is direct (i.e. Data Movement gateway is not used)
- It is metadata

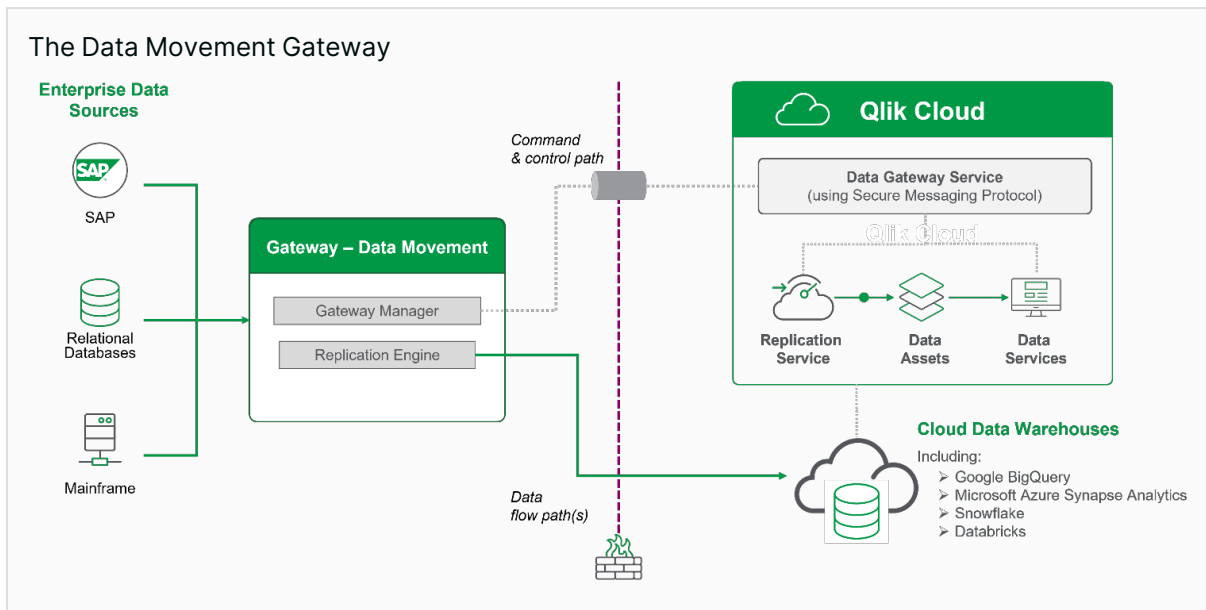
Qlik Cloud as a target

When choosing Qlik Cloud as the target, data from source systems can be stored either in Qlik-managed or customer-managed storage. In these workflows, the only time data is stored in Qlik Cloud is when choosing Qlik Cloud – Qlik-managed storage as the data platform (target) for a data project. This means that the landed data will reside in a Qlik data space encrypted with the customer's unique encryption keys. Customer must still provide a separate S3 storage area for the landing of the data (staging).

All other targets

If the data platform is any one of the cloud targets, such as Snowflake, Azure Synapse Analytics, Microsoft SQL Server, Databricks, or Google BigQuery, the data will be both staged and landed there, and never in Qlik Cloud.

The Data Movement Gateway's replication engine will directly push source data from on-premise / VPC source systems to the cloud target, only involving Qlik Cloud if the source is a SaaS application (or Qlik Cloud is the target) . The gateway manager will be responsible for metadata and command and control, but never data transfer.



Generative AI

Some optional features of Qlik Talend Data Integration use generative AI. These features are disabled by default and it is up to each customer to decide if using these features meets the customer's policies. When using generative AI features in Qlik Talend Data Integration, a customer's metadata only will be sent to our generative AI shared services which operate within the same region as the customer's tenant. This communication is via a private network and at no time is this exposed to public endpoints. At no time is customer data sent to the generative AI services.

Metadata storage

Metadata about the data source and connections are stored in Qlik Talend Data Integration and are not persisted in the Data Movement Gateway. Metadata required to run a job is transferred securely to the gateway, but is kept in-memory and not persisted to the gateway server.

Data encryption

The Data Movement Gateway communicates with the source and target endpoints (typically databases) using either the vendor-provided client package or via a standard ODBC driver. For this reason, the Data Movement Gateway relies on the source/target endpoint's vendors for the in-transit encryption. While the majority of vendors all support encrypted connections, it is out of scope for this document to cover the specifics of other vendor software.

Databases are assumed to sit on a private high-secure network with limited access from other applications. The Data Movement Gateway requires access to the source databases. Data landing on the client-managed cloud storage can be encrypted using server-side encryption.

Qlik Cloud offers data at rest encrypted with the tenant private key or optionally with the customer's own managed keys. For more information about encryption, see the Qlik Cloud platform evaluation guide.

Service-to-service authentication and authorization

The Data Movement Gateway uses a JWT-based token to communicate with Qlik Cloud. Communication between the Data Movement Gateway and Qlik Cloud is always encrypted with TLS 1.2 or higher.

Client-managed cloud storage (S3 bucket) requires a key and secret and is subject to IAM control. This allows customers to set the read/write permissions for different buckets. See [What is IAM?](#) for more details.

User authentication and authorization

Authentication

Authentication of users on the Qlik Cloud platform is handled by the configured identity provider:

- Qlik Account — This is the default mechanism and is managed by Qlik. It provides a secure but basic authentication solution for customers, however is not configurable and does not integrate with a customer's existing solutions.
- Third party identity provider — Qlik Cloud supports configuring third party identity providers that support the OpenID Connect (OIDC) protocol.

For more information about authentication, see the Qlik Cloud platform evaluation guide.

Authorization

Users of Qlik Talend Data Integration are managed using the same role based access control system as is used for the rest of Qlik Cloud. The available space based roles are described in the following table:

Data space role	Summary
Is owner	Full permissions on the space including the ability to grant others access
Can view	Monitor, but not alter the data pipeline
Can consume	Consume data from data tasks in the data space
Can manage	Manage the space details and members
Can operate	View data tasks with basic details and perform actions, such as run, stop, and resume
Can edit	View and edit data tasks in this space, as well as create new data assets

In addition, there are several system-wide data integration roles available:

System role	Summary
Data Admin	Administrator with limited permissions to manage only data spaces
Data Product Manager	A user who can create data products
Data Space Creator	A user who can create data spaces
Data Services Contributor	A user who has access to Data Integration services

For more information about platform level authentication and authorization, see the Qlik Cloud platform evaluation guide.

2.5 Data Products

A data product is an entity that allows you to gather, and package trusted datasets related to the same domain, or any grouping that makes sense for your business. Data product teams can quickly produce high-quality, curated, domain-centric data assets, that can be easily accessed by data consumers and leveraged in several business use cases across the organization. Data Products build on top of and compliment existing dataset functionality.

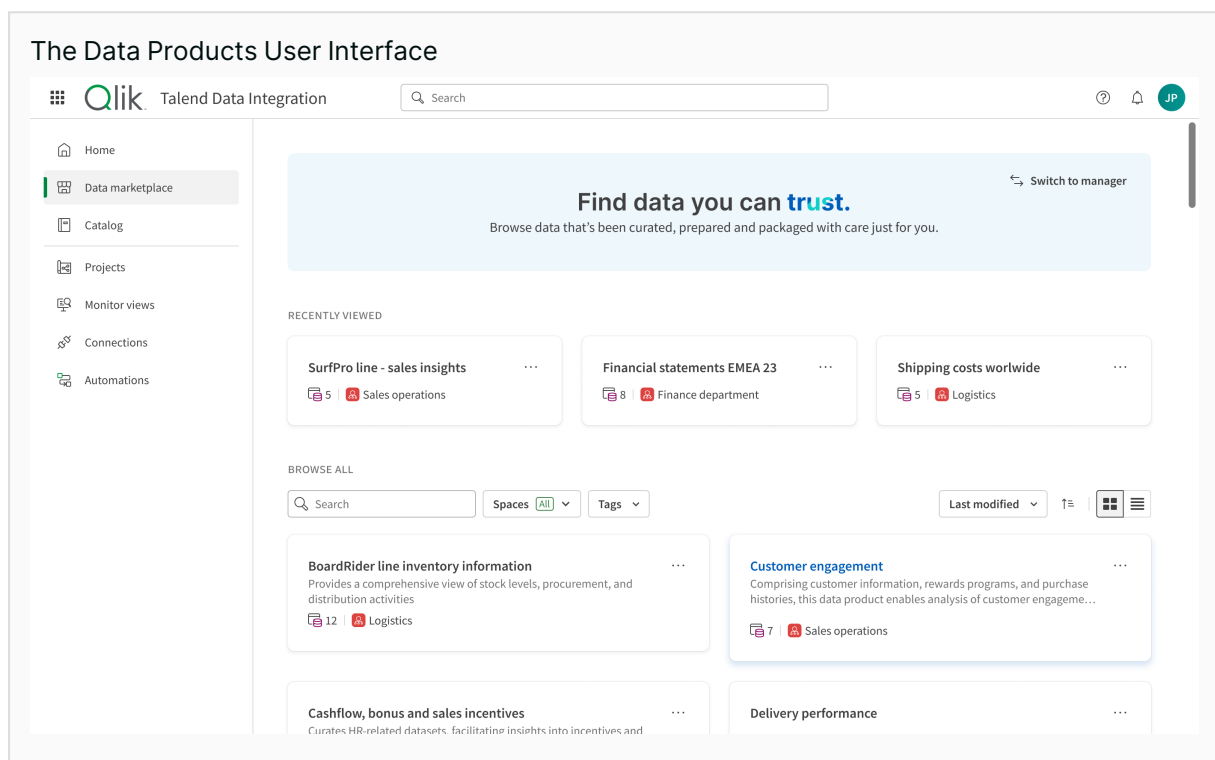
The data product catalog in Qlik Talend Cloud offers several key capabilities to manage the data product lifecycle end-to-end:

- **Data marketplace:** This is where data products are readily available for data consumers to shop and leverage according to their use cases. Data consumers can get a preview of the data products, including quality information, descriptions, and relevant tags.
- **Data product quality and profiling:** Data Product Quality and profiles are calculated for entities in a Data Product. Depending on the Data source, this information will be calculated in the data source (push down) or within your Qlik Talend Data Integration tenant (pull-up).
- **Data product lineage and impact analysis:** This allows users to trace the origins and transformations of data products. Understanding where the data came from and how it changed along the data pipeline enhances data trust and brings more transparency for data consumers.
- **Integrated data catalog experience:** Providing an integrated experience means that data products can be composed of datasets that are registered in the Qlik Talend Data Integration catalog, streamlining the workflow for data product teams without requiring any additional tools.
- **Data product access control:** As in the rest of Qlik Cloud, user access is controlled through spaces. This simplifies access control for data products, emphasizing easy data sharing with data consumers within the Qlik Talend Data Integration environment.
- **Data product lifecycle management:** Lifecycle management enables data product teams to treat data products just like real products, composing them from relevant datasets, evaluating data quality, assigning ownership, managing data product attributes and publishing them for consumption. If a data product no longer meets business needs, it can be deactivated. The Data product manager role is required for this.

- **Data product consumption for Qlik Cloud Analytics:** Data products can be consumed directly by Qlik Sense apps. This allows data consumers to easily load data products into Qlik Sense applications, enabling a seamless transition from data discovery to actionable insights with just a few clicks.

Data Marketplace

The Data Marketplace is a central location within Qlik Talend Data Integration where Data products are both managed and available to consumers. It shows data products from all spaces the user is allowed to see. Data products are also available in the Catalog view, however less information about the data products are shown there.



For a user to be able to consume data products they require the Can Consume role in the space that data product resides as well as access to the space where the datasets themselves reside. To create and manage Data products requires the Data product manager role.

When a user finds a Data product they are interested in, they are able to look at the datasets it contains, documentation on the data Product, relevant contacts in your organization for the data product and metadata about the datasets. If the data product meets the users needs, they can directly create an analytics app from it.

2.6 Qlik Application Automation

Qlik Application Automation is a no-code visual design tool for automating manual work. Instead of writing code, users combine configurable blocks that they drag and drop onto a canvas, forming a sequence of action steps from a variety of SaaS applications, including Qlik Cloud.

These steps run like a program to automate business processes, using analytics to direct the flow of data through a multitude of applications.

Automations can be executed manually, run on a schedule, listen to webhook events, and be triggered using an API endpoint. Automations are compiled into native code that runs in Kubernetes pods on Qlik Cloud to automatically provide the scalability you need to scale and automate your business processes.

Qlik Application Automation provides blocks to implement conditions, work with loops, use variables, and a wide range of other building blocks to implement automation flows.

Architecture

Qlik Application Automation is an integral part of Qlik Cloud and is available in a Qlik Cloud tenant alongside other capabilities such as analytics applications, connectivity to your data sources, and notes.

The elastic nature of the underlying technology allows Qlik Application Automation, as well as the other Qlik Cloud services, to scale horizontally as needed.

Qlik Application Automation shares access to common services, such as authentication services and file storage, to ensure an integrated experience with the other data analytics and data integration capabilities within Qlik Cloud.

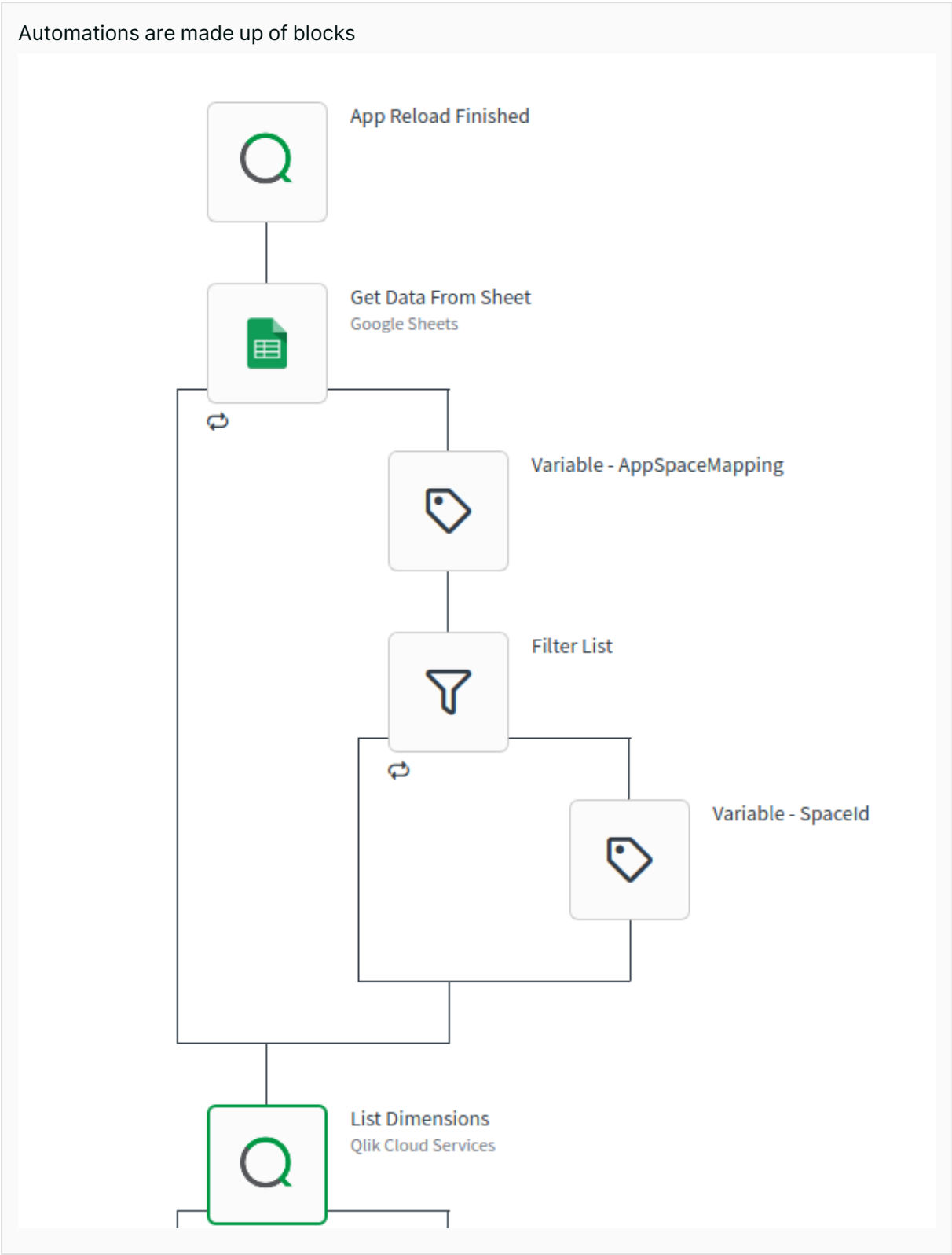
For more information about the Qlik Cloud architecture, see the [The Qlik Cloud platform technical paper evaluation guide](#).

Reliability

As part of the Qlik Cloud platform, Qlik Application Automation is built on a resilient, autoscaling architecture to handle any customer workload. Qlik Application Automation is subject to the same uptime commitment policy reported in the [Service Level Agreement](#) (Section 7, point 7.2).

Automations

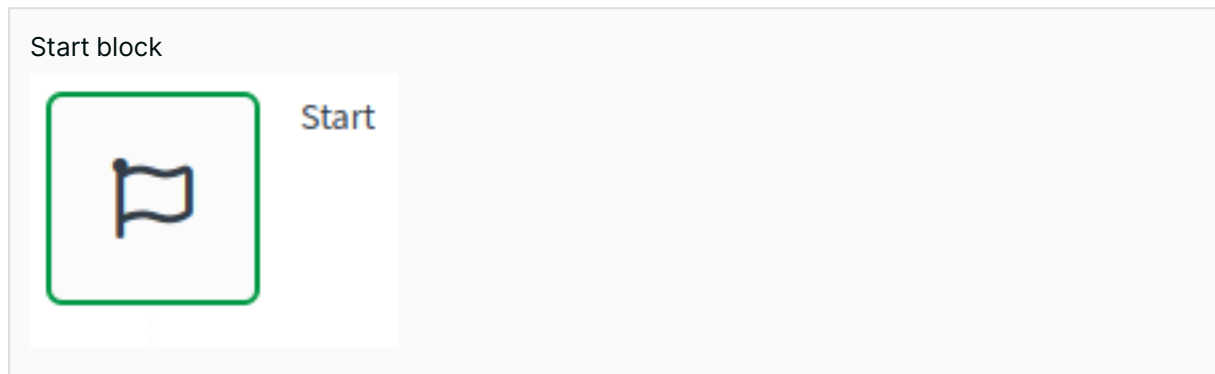
Blocks



Every automation is made of different building blocks chained together to build a flow. There are three types of building blocks:

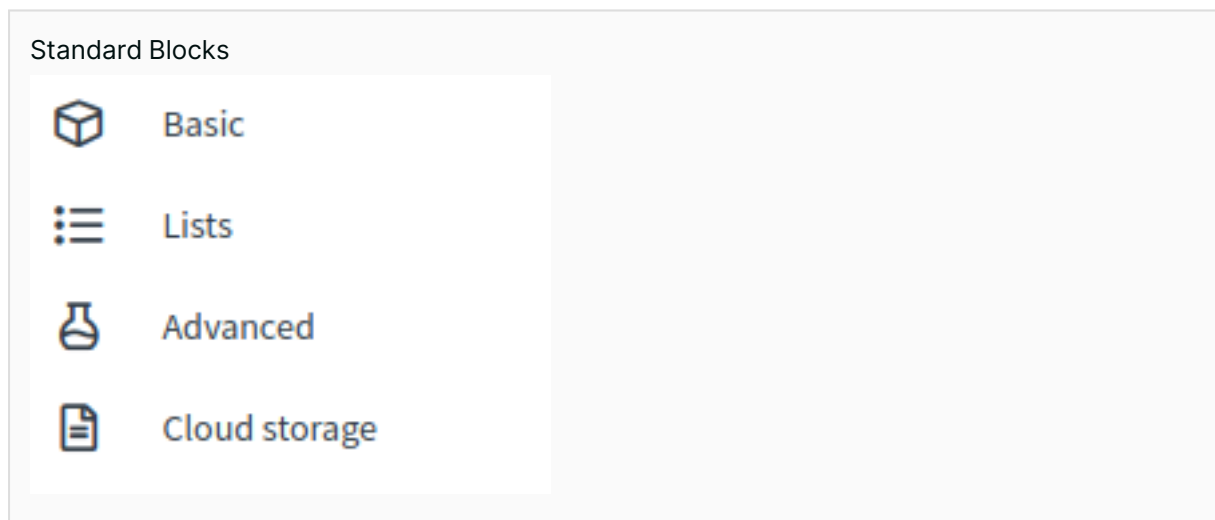
- Start block
- Standard blocks
- Connector blocks

Start blocks



The Start block is your starting point and mandatory in each automation. It is already present on the canvas and defines the start of the automation process; any other block in the automation must be connected to the start block chain.

Standard blocks

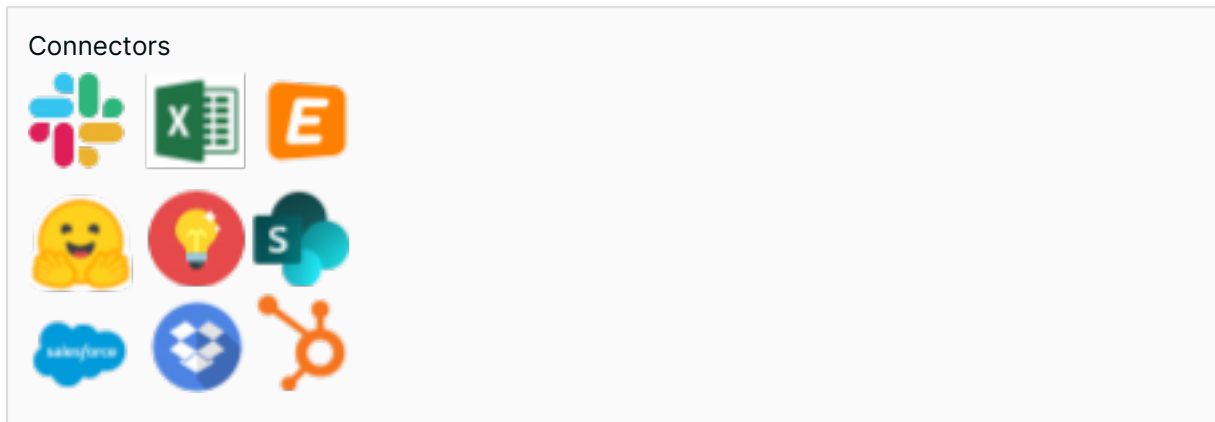


Standard blocks fall into four categories: Basic, List, Advanced, and Cloud storage. Some example uses are:

- Loop over data and conditional block
- Working with variable and user input block

- Merge, filter, lookup or transform lists block

Connectors



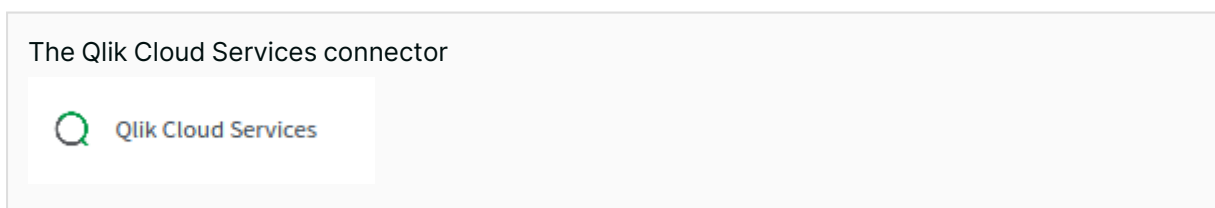
Qlik Application Automation offers a range of connectors. Connectors are the bridge between third-party applications and the automation itself. They can be used to get data from or write data to a third-party application.

There are too many connectors to cover them all here, and we regularly add new connectors to the platform. For information on all connectors available see [Application Automation connectors](#) in the help.

Qlik Cloud connectors

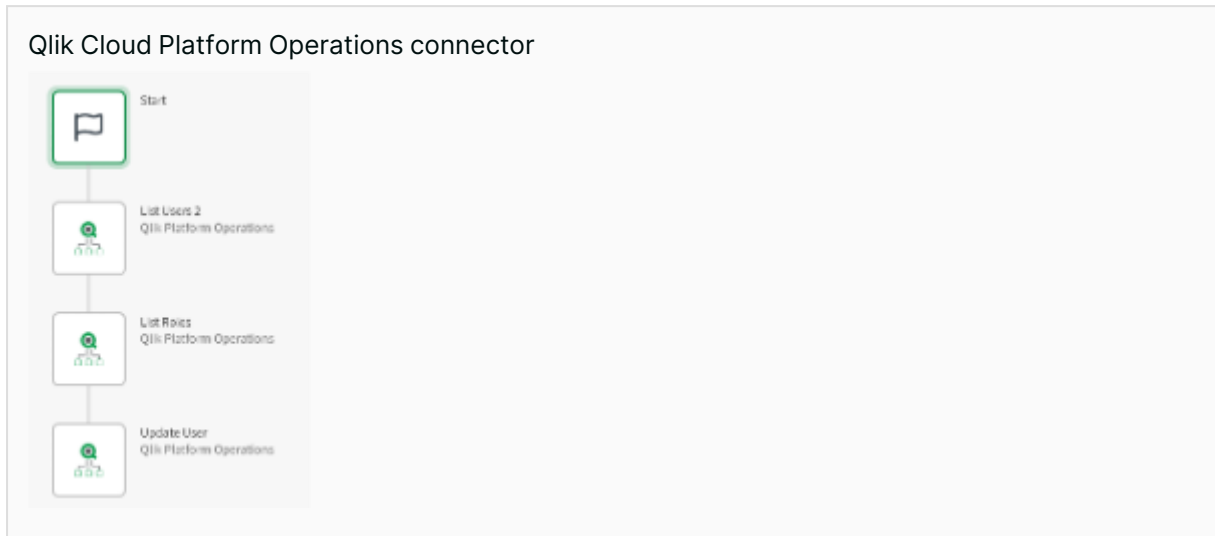
As well as connectors to other platforms and services, Qlik provides connectors to perform actions on Qlik Cloud itself. These cover the full life-cycle of Qlik Cloud from tenant provisioning to reporting.

Qlik Cloud Services connector



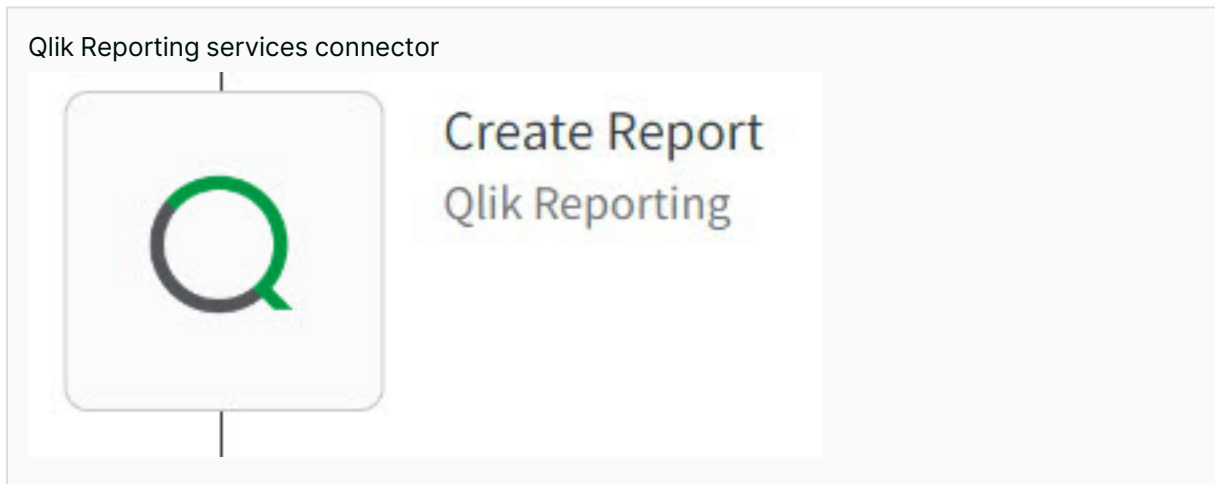
One connector of note is the Qlik Cloud Service connector, it allows the connection to the user's current tenant. It does not require authentication and comes with the same logged-in user's authorization.

Qlik Cloud Platform Operations connector



The Qlik Cloud Platform Operations connector can orchestrate one or many Qlik Cloud tenants using either tenant or subscription OAuth clients.

Qlik Reporting services connector



Qlik Reporting is a Qlik-native connector that lets you build customized, multi-page reports based on app sheets. Reports can be delivered via e-mail as PDF or PowerPoint documents.

Run modes of automations

Automations are built by combining blocks that represent API endpoints of connected applications. Each Automation execution is called a job or run. Automations can be run either manually, or automatically. Manually an automation can be run either from the console, or directly within the automations editor.

Automations can be executed in three ways systematically:

- On a scheduled basis - from every 30 seconds, up until monthly.
- Triggered from an external webhook
- Exposed as an API Call

Use cases

Given the flexibility and wide range of connectors supported by Qlik Application Automation, several common types of use cases are possible:

- Operational automations in service of Qlik Cloud: Qlik Application Automation provides a direct visual access to various Qlik Cloud APIs to facilitate operations such as:
 - Task chaining (e.g., Qlik Cloud Services app reloads)
 - App versioning
 - Assign licenses, roles, spaces, rights to users
 - Alerting and messaging: send messages using e-mail, Slack, and MS Teams triggered by events
 - Active intelligence: Reverse ETL and write-back in business applications (such as CRM, SCM, ERP etc.) based on analytical insights

These use cases can be combined to create complex automations to support business processes, implemented in the same automation or by chaining different automations together.

Security model

Connector authorization and authentication

In general, connectors allow two types of authentication, and this is dependent on the third-party provider service:

- Access token authentication
- OAuth authentication

The access token authentication method requires users to provide an access token (e.g., API key or secret) to the third-party application and provide it in the connector settings.

With OAuth authorization, users are forwarded to the application login page which requires user permission access to grant access to Qlik Application Automation, on behalf of Qlik Cloud. If the user agrees, Qlik Application Automation will store and refresh the access token for accessing data in the automations.

Qlik Application Automation will request both read and write permissions in OAuth flows, due to the nature of the features.

Automation owners can export automation in a JSON file format and subsequently import them in a different tenant or share with other users on the same tenant. Authentication tokens are not part of the export; thus, the import requires a re-authentication of all the connectors involved in the automation.

Governance

The Qlik Cloud platform's management console allows any administrator to manage and govern all the automations. An administrator is entitled to list all the tenant's automations, and sort and filter them based on their name, status, last run, run mode, owner or completed time. Each automation can further show the list of executions and their states.

A Qlik Cloud administrator may enable or disable automations, change owner, or delete them.

Data privacy and compliance

An automation can make API calls to connected SaaS applications (e.g. Salesforce), retrieving data for processing within the automation. For example, a user could build an automation that fetches contacts/leads from their CRM and send it on to another SaaS application. Each automation job will have a history log that shows the API calls made by each block.

Where an automation processes personally identifiable information (PII), or other private data, part of this data will also be present in the history log of the automation. History logs can be viewed in the UI for each automation. History logs have a retention of 30 days and are automatically deleted afterwards. These logs are only visible to the automation owner and can not be viewed by other users or administrators.

In cases where there is a requirement not to log this type of data, it is possible to disable the logging of information on a per-block basis within the automation. This ensures this data is never written to the logs of the automation.

3 About Qlik Evaluation Guides

The content provided herein is provided for informational purposes. Due to Qlik Cloud's continuous release process, at times the content herein may differ from actual platform functionality. Please refer to [Qlik Cloud Help](#) for the product documentation for Qlik Cloud.

Any statement about future plans or intentions for the Qlik Cloud platform contained herein is not a commitment to deliver those features or functionalities, as the development, release, and timing of any features or functionality described for Qlik's products remain at our sole discretion.

For additional information regarding Qlik Cloud, please see [Qlik Cloud](#) or contact your Qlik representative.

3.1 Document history

This content has been developed to assist customers and prospective customers to understand and evaluate the Qlik Cloud platform and its related services. Traditionally this content has been published in document format only as a PDF; however, it is now primarily published as web content with PDF files available if required.

Over its history, this content has been known by the following names:

- Qlik Technical papers
- Qlik White papers
- Qlik Technical overview

This documentation supersedes the above documents.

3.2 Changelog

The PDF documents are generated from the evaluation guides at [Qlik Help](#). The changelog for this evaluation guide is shown below.

Changelog — Qlik Talend Data Integration evaluation guide

July 2024

Qlik Talend Data Integration

- Updated to reflect the name change from Qlik Cloud Data Integration to Qlik Talend Data Integration

July Updates

- Many updates to Data Transformations: SaaS Data Loader, Gen-AI

New Content

- Data Products

October 2023

October Updates

- Cloud sources in *Data movement*
- Minor updates

July 2023

Initial release to help.qlik.com

- Initial release



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. Our cloud-based Qlik Active Intelligence Platform® delivers end-to-end, real-time data integration and analytics cloud solutions 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 does business in more than 100 countries and serves over 38,000 active customers around the world.

[qlik.com](https://www.qlik.com)