

Introduction to Analytics Hub

Preview

This feature is covered by the [Pre-GA Offerings Terms \(/terms/service-terms#1\)](/terms/service-terms#1) of the Google Cloud Terms of Service. Pre-GA features might have limited support, and changes to pre-GA features might not be compatible with other pre-GA versions. For more information, see the [launch stage descriptions \(/products#product-launch-stages\)](/products#product-launch-stages).

Analytics Hub is a data exchange platform that enables you to share data and insights at scale across organizational boundaries with a robust security and privacy framework. With Analytics Hub, you can discover and access a data library curated by various data providers. This data library also includes Google-provided datasets.

For example, by using Analytics Hub you can augment your analytics and ML initiatives with third-party and Google datasets.

As an Analytics Hub user, you can perform the following tasks:

- As an Analytics Hub publisher, you can monetize data by sharing it with your partner network or within your own organization in real time. [Listings \(#listings\)](#) let you share data without replicating the shared data. You can build a catalog of analytics-ready data sources with granular permissions that let you deliver data to the right audiences.
- As an Analytics Hub subscriber, you can discover the data that you are looking for, combine shared data with your existing data, and leverage the [built-in features of BigQuery \(/bigquery/docs/introduction#explore-bigquery\)](#). When you subscribe to a listing, a [linked dataset \(#linked_datasets\)](#) is created in your project.
- As an Analytics Hub viewer, you can browse through the datasets that you have access to in Analytics Hub and request the publisher to access the shared data.
- As an Analytics Hub administrator, you can create [data exchanges \(#data_exchanges\)](#) that enable data sharing, and then give permissions to data publishers and subscribers to access these data exchanges.

Analytics Hub in a minute



For more information about Analytics Hub user roles, see [Configure Analytics Hub roles](#) (/bigquery/docs/analytics-hub-grant-roles).

Architecture

Analytics Hub is built on a publish and subscribe model of BigQuery datasets. The separation of compute and storage in BigQuery's architecture enables data publishers to share data with as many subscribers as they want without having to make multiple copies of the data. Publishers are only charged for data storage, whereas subscribers only pay for queries that run against the shared data. The publisher and subscriber workflows in Analytics Hub are explained in detail in the following sections.

Publisher workflow

The following diagram describes how publishers interact with Analytics Hub:

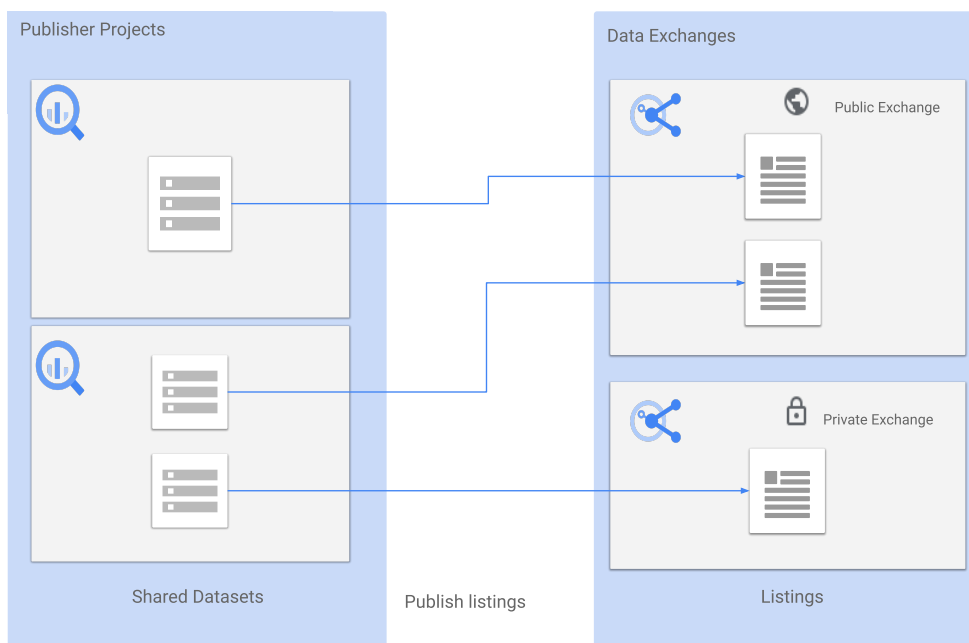


Figure 1. Analytics Hub Publisher workflow.

In figure 1, the following features are labeled: Shared dataset (#shared_datasets), Data exchange (#data_exchanges), and Listing (#listings).

Shared datasets

A shared dataset is a BigQuery dataset that is the unit of data sharing in Analytics Hub. As a publisher, you create or use an existing BigQuery dataset in your project

with the collection of objects, such as tables and views, that you want to deliver to your subscribers.

Data exchanges

A data exchange is a container that enables self-service data sharing. It contains listings that reference shared datasets. With Analytics Hub, publishers and administrators can grant access to subscribers at the exchange and the listing level. This method helps to avoid granting access on the underlying shared datasets explicitly. An Analytics Hub subscriber can browse through data exchanges, discover data that they can access, and subscribe to shared datasets. A data exchange can be of the following types:

- **Private data exchange.** By default, a data exchange is private and only users or groups that have access to that exchange can view or subscribe to the data.
- **Public data exchange.** By default, a data exchange is private and only users or groups that have access to that exchange can view or subscribe to its listings. However, you can choose to make a data exchange public. Listings in public data exchanges can be discovered (/bigquery/docs/analytics-hub-view-subscribe-listings#view_listings) and subscribed (/bigquery/docs/analytics-hub-view-subscribe-listings#subscribe_to_a_listing) by Google Cloud users (allauthenticated users) (</iam/docs/overview#all-authenticated-users>). For more information about public data exchanges, see Make a data exchange public (</bigquery/docs/analytics-hub-manage-exchanges#make-data-exchange-public>).

As an Analytics Hub administrator, you can create multiple data exchanges in Analytics Hub, and manage other Analytics Hub users.

Listings

A listing is a reference to a shared dataset that a publisher lists in a data exchange. As a publisher, you can create a listing and specify the dataset description, sample queries to run on the dataset, links to any relevant documentation, and any additional information that can help subscribers to use your dataset. For more information, see Manage listings (</bigquery/docs/analytics-hub-manage-listings>). A listing can be of the following two types based on the Identity and Access Management (IAM) policy that is set for the listing and the type of data exchange that contains the listing:

- **Public listing.** It is shared with all Google Cloud users (allauthenticated users) (</iam/docs/overview#all-authenticated-users>). Listings in a public data exchange are public listings. These listings can be references of a *free public*

dataset or a *commercial dataset*. If the listing is of a commercial dataset, subscribers can request access to the listing and the data provider contacts those subscribers directly.

- **Private listing.** It is shared directly with individuals or groups. For example, a private listing can reference marketing metrics dataset that you share with other internal teams within your organization.

Subscriber workflow

The following diagram describes how subscribers interact with Analytics Hub:

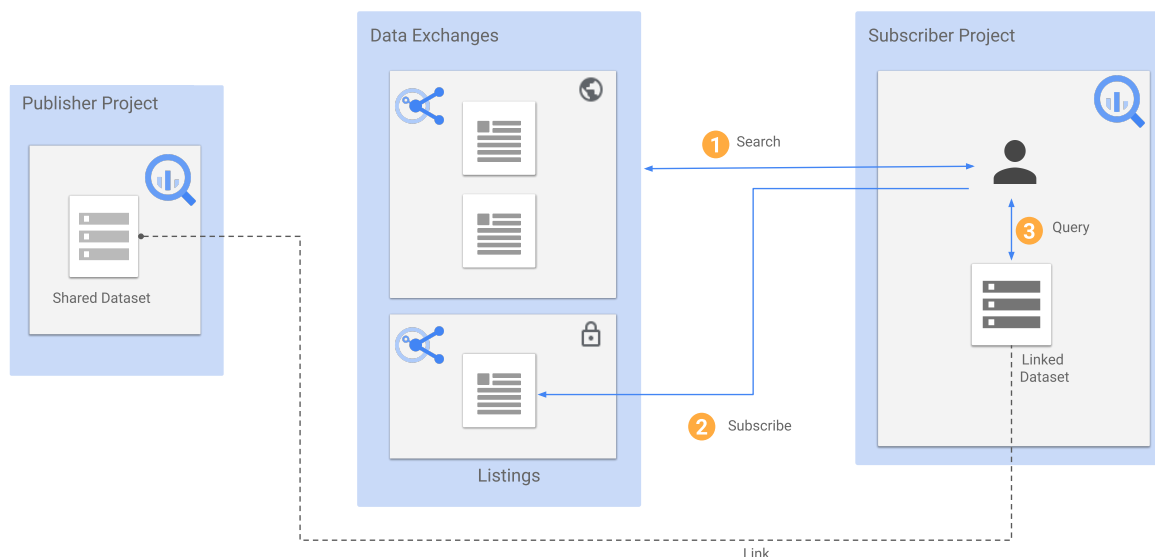


Figure 2. Analytics Hub Subscriber workflow.

In figure 2, the following Analytics Hub features are labeled: Shared dataset (#shared_datasets), Data exchange (#data_exchanges), Listing (#listings), and Linked dataset (#linked_datasets).

Linked datasets

A linked dataset is a *read-only* BigQuery dataset that serves as a symbolic link to a shared dataset. Subscribing to a listing creates a linked dataset in your project and not a copy of the dataset, so subscribers can read the data but cannot add or update objects within it. When you query objects such as tables and views through a linked dataset, the data from the shared dataset is returned. For more information about linked datasets, see [View and subscribe to listings](https://cloud.google.com/bigquery/docs/analytics-hub-view-subscribe-listings) (/bigquery/docs/analytics-hub-view-subscribe-listings). Linked datasets are authorized to

access tables and views of a shared dataset. Subscribers with linked datasets access tables and views of a shared dataset without any additional Identity and Access Management authorization.

Limitations

Analytics Hub has the following limitations:

- Owners of shared datasets and data exchanges cannot see subscription metrics.
- If a project is deleted, then the data exchanges within it are not deleted. You need to manually [delete these data exchanges](#) (/bigquery/docs/analytics-hub-manage-exchanges#delete_a_data_exchange) before deleting the project.
- If you delete a shared dataset that has subscribers, then the linked datasets are not deleted. Subscribers need to manually [delete these linked datasets](#) (/bigquery/docs/analytics-hub-view-subscribe-listings#delete_a_linked_dataset) from their projects.
- If you [create a listing](#) (/bigquery/docs/analytics-hub-manage-listings) for a shared dataset that uses a [customer-managed encryption key](#) (/storage/docs/encryption/customer-managed-keys), subscribers won't have access to the Cloud KMS key required to access the dataset.
- A shared dataset can have a maximum of 1,000 linked datasets. All subscribers, combined, can have a maximum of 1,000 linked datasets per shared dataset.
- The following BigQuery objects can be shared using Analytics Hub:
 - [Authorized views](#) (/bigquery/docs/authorized-views)
 - [Authorized datasets](#) (/bigquery/docs/authorized-datasets)
 - [BigQuery ML models](#) (/bigquery-ml/docs/introduction)
 - [External tables](#) (/bigquery/docs/external-tables)
 - [Materialized views](#) (/bigquery/docs/materialized-views-intro)
 - [Tables](#) (/bigquery/docs/tables-intro)
 - [Table snapshots](#) (/bigquery/docs/table-snapshots-intro)
 - [Views](#) (/bigquery/docs/views-intro)

A dataset with unsupported resources cannot be selected as a shared dataset when you are creating a listing (/bigquery/docs/analytics-hub-manage-listings#create_a_listing).

- If you are a publisher, the following BigQuery interoperability applies to you:
 - Using streaming inserts (/bigquery/docs/streaming-data-into-bigquery) or the BigQuery Storage Write API (/bigquery/docs/write-api) to stream data into a shared dataset might cause issues for the linked datasets (#linked_datasets).
 - Shared datasets (#shared_datasets) support column-level (/bigquery/docs/column-level-security-intro) and row-level security (/bigquery/docs/row-level-security-intro).
 - If a view (/bigquery/docs/views) in the shared dataset doesn't contain fully qualified URI references to its source data, then subscribers won't get the correct result when querying that dataset. To avoid this issue, use a fully qualified reference—for example, *PROJECT_NAME.DATASET_NAME.TABLE_NAME*.
 - Shared datasets are indexed in Data Catalog (/data-catalog/docs/concepts/overview). Updates on a shared dataset, such as adding tables or views, are made available to subscribers without any delay. However, in certain scenarios, for example, when there are more than one hundred subscribers or tables in a shared dataset, the updates might take up to 18 hours to get indexed in Data Catalog. Due to the delay in indexing, subscribers cannot search for these updated resources in the Google Cloud console immediately.
- If you are a subscriber, the following BigQuery interoperability applies to you:
 - Using the BigQuery Storage Read API (/bigquery/docs/reference/storage) on linked dataset resources is not supported.
 - Time travel on linked dataset resources is not supported.
 - Materialized views that refer to tables in the linked dataset are not supported.
 - Taking snapshots (/bigquery/docs/table-snapshots-intro) of linked dataset tables is not supported.
- If linked datasets are not colocated with the shared dataset, then read operations to linked dataset tables with a query size of more than 5 GiB might fail. This error might resolve automatically. You can also contact support (/bigquery/docs/support) to resolve this issue.
- The following **INFORMATION_SCHEMA** views don't support dataset qualifiers (/bigquery/docs/information-schema-intro#dataset_qualifier) for linked datasets:



- [INFORMATION_SCHEMA.TABLES](/bigquery/docs/information-schema-tables) (/bigquery/docs/information-schema-tables)
- [INFORMATION_SCHEMA.TABLE_OPTIONS](/bigquery/docs/information-schema-table-options) (/bigquery/docs/information-schema-table-options)
- [INFORMATION_SCHEMA.COLUMNS](/bigquery/docs/information-schema-columns) (/bigquery/docs/information-schema-columns)
- [INFORMATION_SCHEMA.COLUMN_FIELD_PATHS](/bigquery/docs/information-schema-column-field-paths) (/bigquery/docs/information-schema-column-field-paths)
- [INFORMATION_SCHEMA.VIEWS](/bigquery/docs/information-schema-views) (/bigquery/docs/information-schema-views)
- [INFORMATION_SCHEMA.PARTITIONS](/bigquery/docs/information-schema-partitions) (/bigquery/docs/information-schema-partitions)
- You cannot use [region qualifiers](/bigquery/docs/information-schema-intro#region_qualifier) (/bigquery/docs/information-schema-intro#region_qualifier) with `INFORMATION_SCHEMA` views to [view metadata for your linked dataset](/bigquery/docs/analytics-hub-view-subscribe-listings#view_table_metadata) (/bigquery/docs/analytics-hub-view-subscribe-listings#view_table_metadata).



Supported regions




Analytics Hub is supported in the following regions and multi-regions.

Regions

The following table lists the regions in the Americas where Analytics Hub is available.

Region description	Region name	Details
Iowa	us-central1	 <u>Low CO₂</u> (/sustainability/region-carbon)
Las Vegas	us-west4	
Los Angeles	us-west2	
Montréal	northamerica-northeast1	 <u>Low CO₂</u> (/sustainability/region-carbon)
Northern Virginia	us-east4	

Region description	Region name	Details
Oregon	us-west1	 Low CO₂ (/sustainability/region-carbon)
Salt Lake City	us-west3	
São Paulo	southamerica-east1	 Low CO₂ (/sustainability/region-carbon)
Santiago	southamerica-west1	
South Carolina	us-east1	
Toronto	northamerica-northeast2	
The following table lists the regions in Asia Pacific where Analytics Hub is available.		
Region description	Region name	Details
Delhi	asia-south2	
Hong Kong	asia-east2	
Jakarta	asia-southeast2	
Melbourne	australia-southeast2	
Mumbai	asia-south1	
Osaka	asia-northeast2	
Seoul	asia-northeast3	
Singapore	asia-southeast1	

Region description	Region name	Details
Sydney	australia-southeast1	
Taiwan	asia-east1	
Tokyo	asia-northeast1	
The following table lists the regions in Europe where Analytics Hub is available.		
Region description	Region name	Details
Belgium	europe-west1	 Low CO₂ (/sustainability/region-carbon)
Finland	europe-north1	 Low CO₂ (/sustainability/region-carbon)
Frankfurt	europe-west3	
London	europe-west2	
Netherlands	europe-west4	
Warsaw	europe-central2	
Zürich	europe-west6	 Low CO₂ (/sustainability/region-carbon)
Multi-regions		
The following table lists the multi-regions where Analytics Hub is available.		
Multi-region description	Multi-region name	

Multi-region description	Multi-region name
Data centers within <u>member states</u> (https://europa.eu/european-union/about-eu/countries_en) of the European Union ¹	EU
Data centers in the United States	US

¹ Data located in the EU multi-region is not stored in the `europa-west2` (London) or `europa-west6` (Zürich) data centers.

Example use case

This section shows an example of how you can use Analytics Hub.

Suppose you are a retailer and your organization has real-time demand forecasting data in a Google Cloud project named **Forecasting**. You want to share this demand forecasting data with hundreds of vendors in your supply-chain system. Here's how you can share your data with vendors through Analytics Hub:

Analytics Hub administrators

As the owner of the **Forecasting** project, you must first enable the Analytics Hub API and then assign the Analytics Hub Admin role (</bigquery/docs/analytics-hub-grant-roles#ah-admin-role>) to a user who administers the data exchange in the project. Users with the Analytics Hub Admin role are called the *Analytics Hub administrators*.

An Analytics Hub administrator can perform the following tasks:

- Create, update, delete, and share the data exchange in your organization's **Forecasting** project.
- Manage other *Analytics Hub administrators*.
- Manage *publishers* by granting the Analytics Hub Publisher role to your organization's employees. If you want some employees to only be able to update, delete, and share listings but not create them, then you can grant them the Analytics Hub Listing Admin role.

- Manage *subscribers* by granting the Analytics Hub Subscriber role to a Google group consisting of all vendors. If you want some vendors to only have view access to the available exchanges and listings then you can grant them the Analytics Hub Viewer role. These vendors won't be able to subscribe to listings.

For more information, see [Manage data exchanges](/bigquery/docs/analytics-hub-manage-exchanges) (/bigquery/docs/analytics-hub-manage-exchanges).

Analytics Hub publishers

Publishers create the following listings for their datasets in the **Forecasting** project or in a different project:

- Listing A: Demand Forecast Dataset 1
- Listing B: Demand Forecast Dataset 2
- Listing C: Demand Forecast Dataset 3

For more information, see [Manage listings](/bigquery/docs/analytics-hub-manage-listings) (/bigquery/docs/analytics-hub-manage-listings).

Analytics Hub subscribers

Subscribers can browse through listings that they have access to in data exchanges. They can also subscribe to these listings and add these datasets to their projects by creating a linked dataset. Vendors can then run queries on these linked datasets and retrieve results in real time.

For more information, see [View and subscribe to listings](/bigquery/docs/analytics-hub-view-subscribe-listings) (/bigquery/docs/analytics-hub-view-subscribe-listings).

Pricing

There is no additional cost for managing data exchanges or listings. Analytics Hub publishers are charged for data storage, whereas subscribers pay for queries that run against the shared data based on either on-demand or flat-rate pricing model. For information about pricing, see [BigQuery pricing](/bigquery/pricing) (/bigquery/pricing).

Quotas

For information about Analytics Hub quotas, see [Quotas and limits](#) (/bigquery/quotas#analytics-hub).

What's next

- Learn how to [view and subscribe to listings](#) (/bigquery/docs/analytics-hub-view-subscribe-listings).
- Learn how to [grant roles to Analytics Hub users](#) (/bigquery/docs/analytics-hub-grant-roles).

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](#) (<https://creativecommons.org/licenses/by/4.0/>), and code samples are licensed under the [Apache 2.0 License](#) (<https://www.apache.org/licenses/LICENSE-2.0>). For details, see the [Google Developers Site Policies](#) (<https://developers.google.com/site-policies>). Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2022-10-03 UTC.