

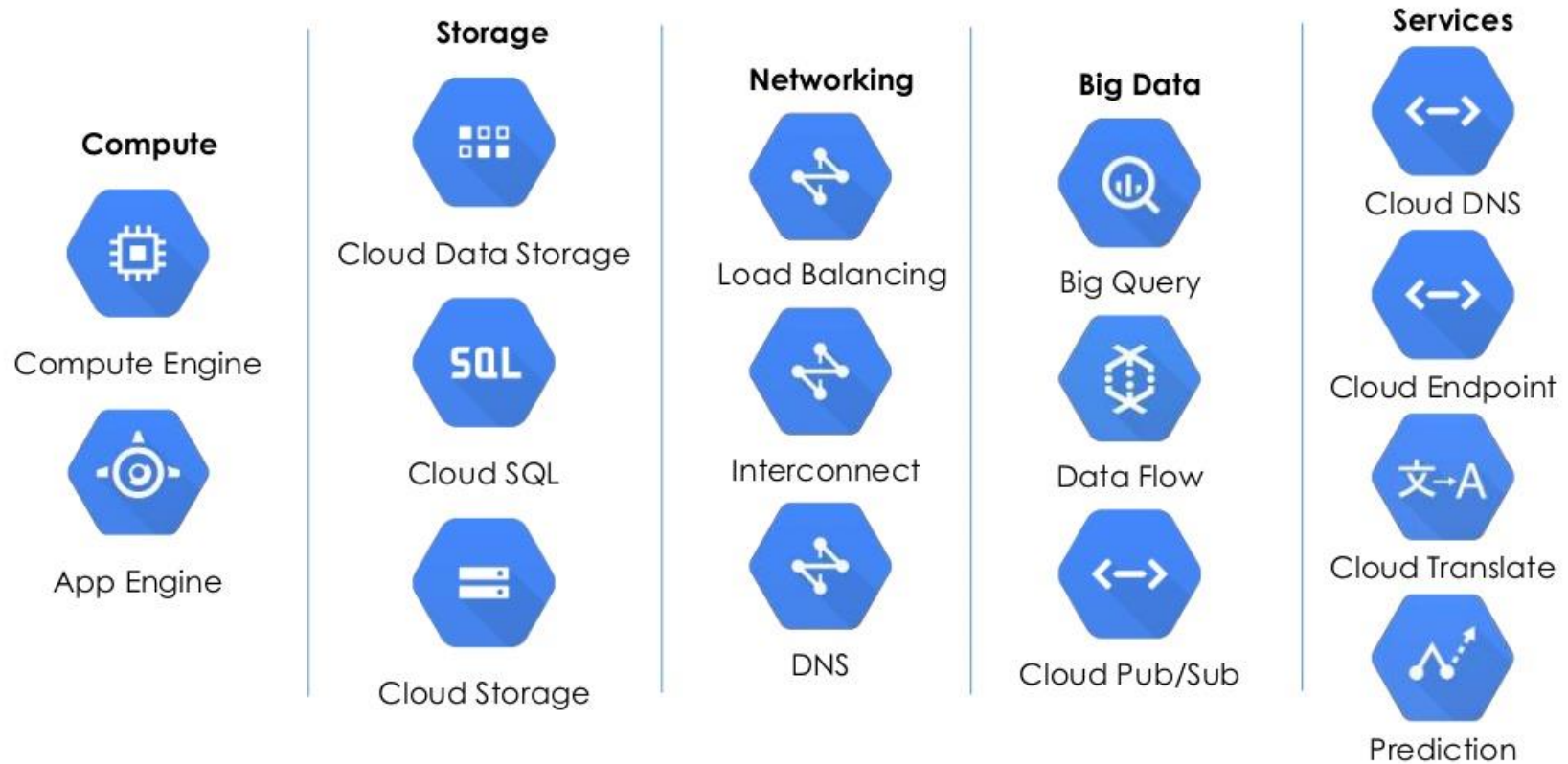
Data Warehousing and Business analytics

Lecture - 8

Dimensional modeling with Bigquery (Google Cloud Platform)

Google Cloud Platform (GCP)

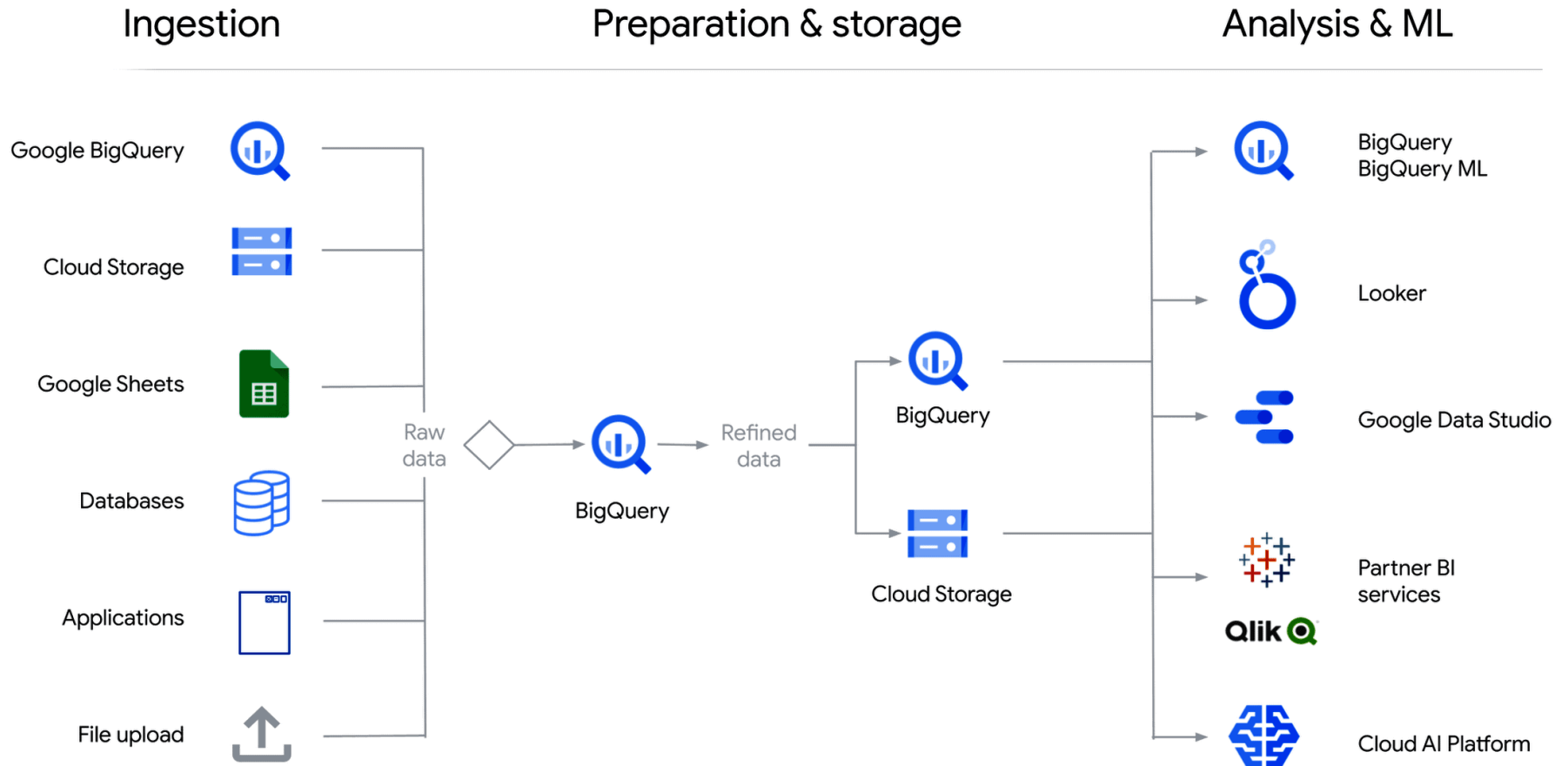
Google Cloud Platform



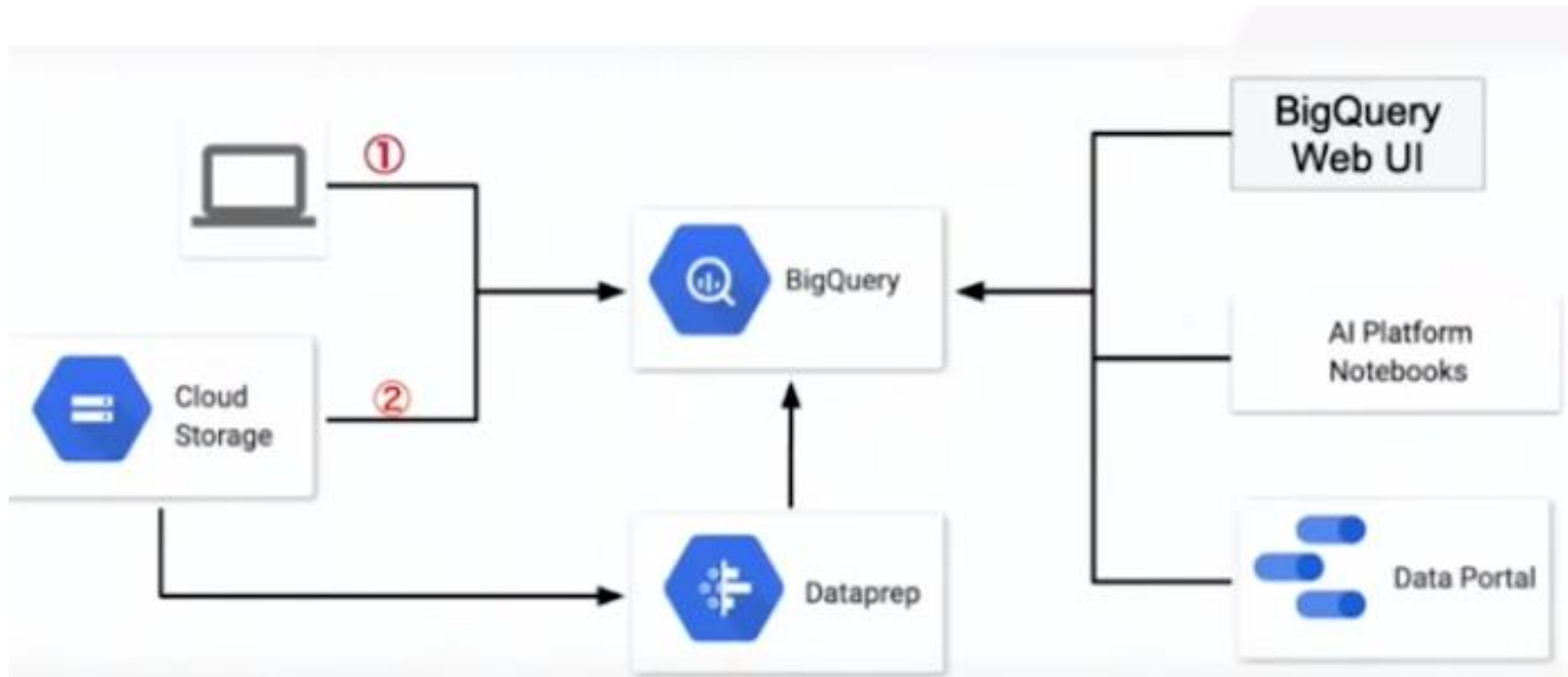
Google Cloud Platform (GCP) Services



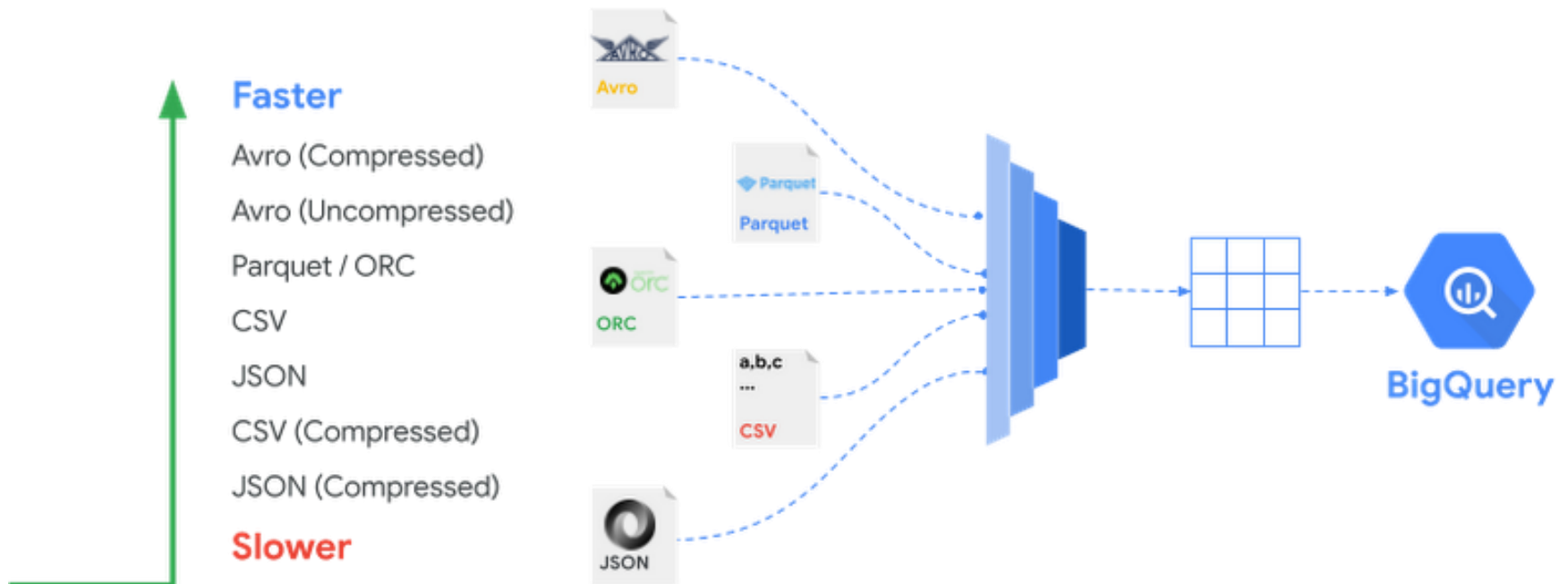
Google BigQuery: Cloud Data



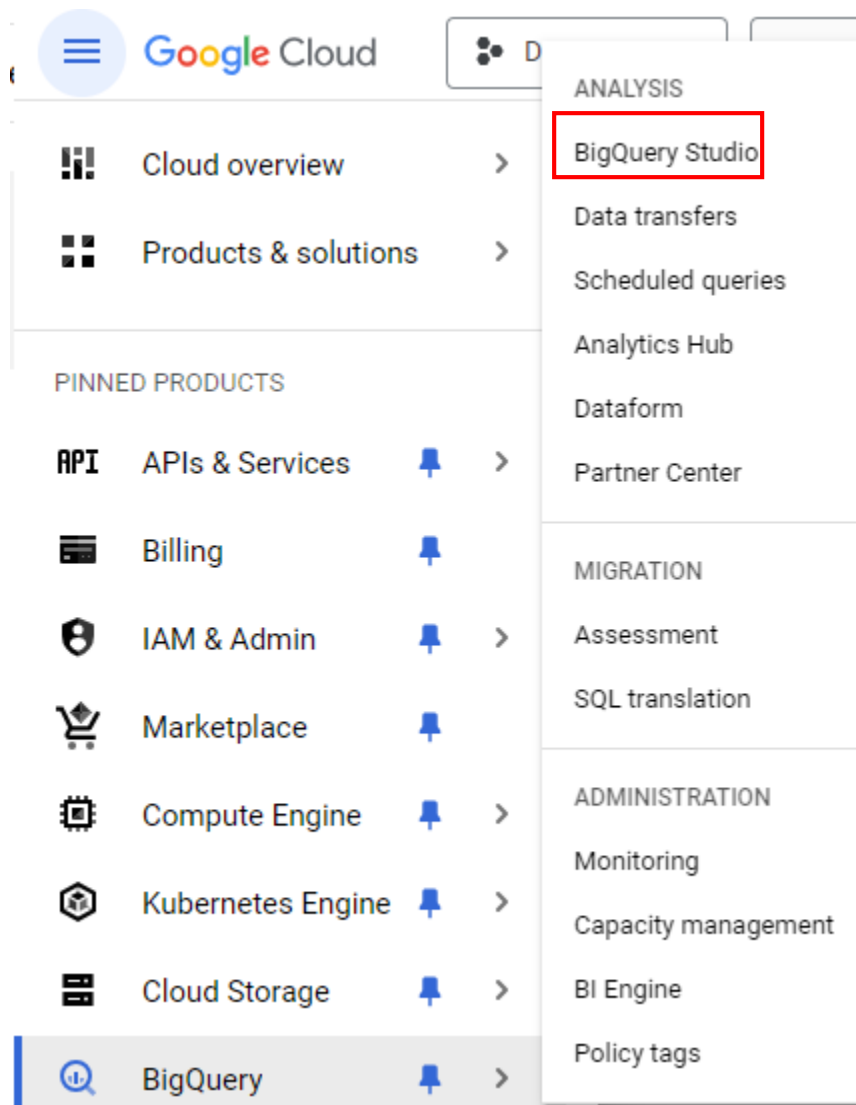
Loading Data into BigQuery



Import Data to BigQuery

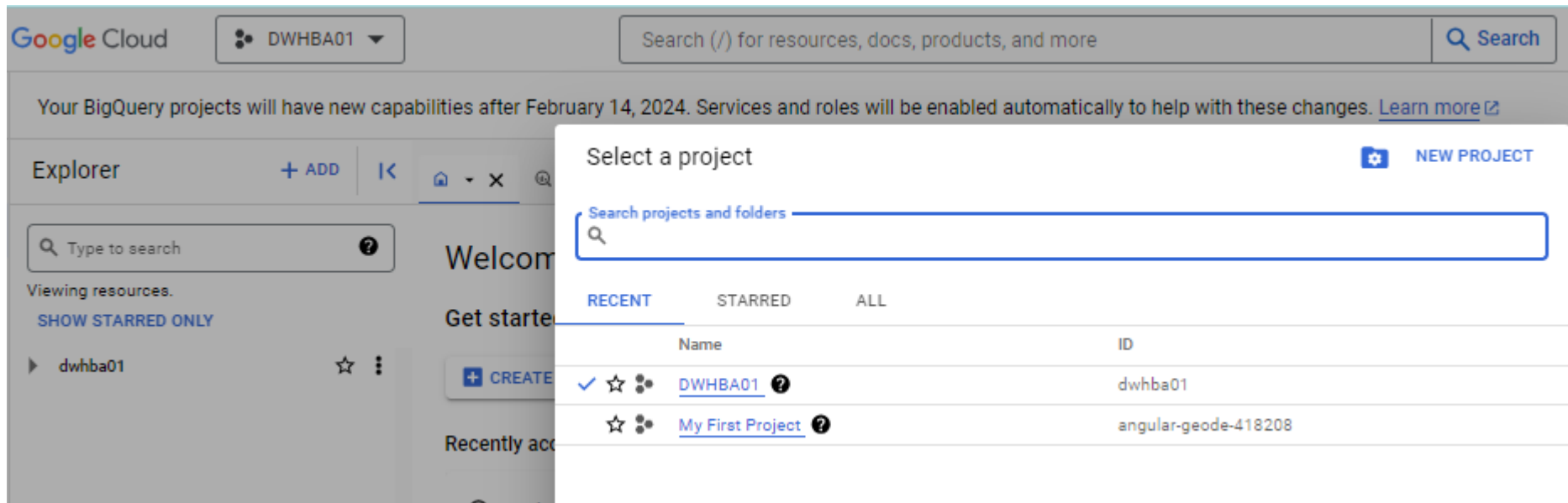


Starting with BigQuery



Starting with BigQuery

Choose the project



The screenshot shows the Google Cloud console interface. At the top, the Google Cloud logo is on the left, and a search bar is on the right. Below the logo, the current project 'DWHBA01' is displayed. A notification banner states: 'Your BigQuery projects will have new capabilities after February 14, 2024. Services and roles will be enabled automatically to help with these changes. [Learn more](#)'. The main content area is divided into two sections: 'Explorer' on the left and 'Welcome' on the right. The 'Explorer' section has a search bar and a list of resources, including 'dwhba01'. The 'Welcome' section has a 'Get started' heading and a 'CREATE' button. A 'Select a project' dialog box is overlaid on the right side of the screen. It has a title bar, a search bar, and a 'NEW PROJECT' button. Below the search bar, there are tabs for 'RECENT', 'STARRED', and 'ALL'. The 'RECENT' tab is active, showing a list of projects. The list has two columns: 'Name' and 'ID'. The first project is 'DWHBA01' with ID 'dwhba01', marked with a checkmark. The second project is 'My First Project' with ID 'angular-geode-418208'.

Google Cloud

DWHBA01

Search (/) for resources, docs, products, and more

Search

Your BigQuery projects will have new capabilities after February 14, 2024. Services and roles will be enabled automatically to help with these changes. [Learn more](#)

Explorer

+ ADD

Type to search

Viewing resources.

[SHOW STARRED ONLY](#)

dwhba01

Get started

[CREATE](#)

Recently acc

Select a project

NEW PROJECT

Search projects and folders

RECENT

STARRED

ALL

Name	ID
<input checked="" type="checkbox"/> DWHBA01	dwhba01
<input type="checkbox"/> My First Project	angular-geode-418208

Starting with BigQuery

Choose the project

Create a new project

The screenshot shows the Google Cloud console interface. At the top, the Google Cloud logo is on the left, and a search bar is on the right. Below the logo, a dropdown menu shows 'DWHBA01' with a red box around it. A blue callout bubble points to the 'NEW PROJECT' button in the top right corner. In the center, a 'Select a project' dialog is open, showing a search bar and a table of projects. A red box highlights the first two rows of the table: 'DWHBA01' (which is selected with a checkmark) and 'My First Project'. A blue callout bubble points to the 'DWHBA01' row. The table has columns for 'Name' and 'ID'.

Name	ID
✓ ☆ DWHBA01 ?	dwhba01
☆ My First Project ?	angular-geode-418208

Choose a
available
project

Create a Project

Project name *

My Project 46349



Project ID: model-creek-418217. It cannot be changed later. [EDIT](#)

Location *

 No organization

[BROWSE](#)

Parent organization or folder

CREATE

CANCEL

BigQuery workspace

Projects, datasets, tables

Create a new dataset

The screenshot displays the BigQuery Studio interface. On the left, the 'Explorer' panel shows a search bar and a list of resources for the project 'dwhba01'. The resources include 'Queries' (with sub-items 'Shared queries' and 'selectbycity'), 'Notebooks', 'External connections', and the 'Adventure_works' dataset. Under 'Adventure_works', the tables 'Adventureworks' and 'used_cars' are listed. A red box highlights the 'dwhba01' project name, and another red box highlights the 'Adventure_works' dataset. A third red box highlights the 'used_cars' table. A blue callout points to the 'dwhba01' project with the text 'Create a new dataset'. Another blue callout points to the 'Adventure_works' dataset with the text 'Create a new table'. The main workspace on the right shows a 'Welcome to BigQuery Studio!' message, 'Get started' buttons for 'CREATE SQL QUERY' and 'CREATE PYTHON NOTEBOOK', and a 'Recently accessed' section with cards for 'usedcarshig...', 'used_cars', 'selectbycity', and 'Adventureworks'. At the bottom, there is a 'Toggle preview panel' button and an upload icon.

Explorer

+ ADD

Untitled 2

Welcome to BigQuery Studio!

Get started

+ CREATE SQL QUERY

+ CREATE PYTHON NOTEBOOK

Recently accessed

usedcarshig... US : dwhba01 OPEN

used_cars US : dwhba01 OPEN

selectbycity us-central1 : dwhba01 OPEN

Adventureworks

used_cars

usedcarshigher20

Add your own data

Toggle preview panel

Create a dataset

Create dataset

Project ID

dwhba01

[CHANGE](#)

Dataset ID *

Letters, numbers, and underscores allowed

Location type ?

☐ Region

Specify a region to colocate your datasets with other Google Cloud services.

☒ Multi-region

Allow BigQuery to select a region within a group to achieve higher quota limits.

Multi-region *

US (multiple regions in United States) ▼

Default table expiration

☐ Enable table expiration ?

Default maximum table age

Days

Advanced options



CREATE DATASET

CANCEL

Create a table in dataset

Create table

Source

Create table from

Empty table

Google Cloud Storage

Upload

Drive

Google Bigtable

Amazon S3

Azure Blob Storage

Maximum name size is 1,024 UTF-8 bytes. Unicode letters, marks, numbers, connectors, dashes, and spaces are allowed.

Table type

Native table

Schema



Edit as text

CREATE TABLE

CANCEL

Create a table in dataset

Create table

Source

Create table from
Upload

Select file *

BROWSE



File format

CSV

JSONL (Newline delimited JSON)

Avro

Parquet

ORC

Table *

Maximum name size is 1,024 UTF-8 bytes. Unicode letters, marks, numbers, connectors, dashes, and spaces are allowed.

Table type

CREATE TABLE

CANCEL

Create a table in dataset

Create table

dwhba01

BROWSE

Dataset *

staging

Table *

categories

Maximum name size is 1,024 UTF-8 bytes. Unicode letters, marks, numbers, connectors, dashes, and spaces are allowed.

Table type

Native table



Schema



Auto detect



Schema will be automatically generated.

Partition and cluster settings

Partitioning

No partitioning



CREATE TABLE

CANCEL

Working with tables

Explorer

+ ADD

<

Type to search

?

Viewing resources.

SHOW STARRED ONLY

Adventure_works

datawarehouse

staging

categories

customers

employee-territories

employees

order-details

categories

+

+

categories

QUERY

SHARE

COPY

SNAPSHOT

DELETE

EXPORT

REFRESH

<

SCHEMA

DETAILS

PREVIEW

LINEAGE

DATA PROFILE

DATA QUALITY

>

Working with tables

The screenshot shows a database management interface for a table named 'categories'. The interface includes a top navigation bar with a tab for 'categories', a search icon, and a dropdown menu. Below the navigation bar, there are several action buttons: 'QUERY' (highlighted with a red box in the context menu), 'SHARE', 'COPY', 'SNAPSHOT', 'DELETE', 'EXPORT', and 'REFRESH'. The 'QUERY' button has a dropdown menu with three options: 'In new tab' (highlighted with a red box), 'In split tab', and 'In Python notebook'. The 'In Python notebook' option has a 'PREVIEW' button next to it. Below the action buttons, there are tabs for 'SCHEMA', 'DATA PROFILE', and 'DATA QUALITY'. The 'SCHEMA' tab is selected, showing a table with columns: 'Field name', 'Key', 'Collation', 'Default Value', 'Policy Tags', and 'Description'. The table contains four rows of data: 'categoryID' (INTEGER, NULLABLE), 'categoryName' (STRING, NULLABLE), 'description' (STRING, NULLABLE), and 'picture' (FLOAT, NULLABLE). At the bottom of the interface, there are two buttons: 'EDIT SCHEMA' and 'VIEW ROW ACCESS POLICIES'.

categories

QUERY

SHARE COPY SNAPSHOT DELETE EXPORT REFRESH

SCHEMA DATA PROFILE DATA QUALITY

Filter Enter properties

Field name	Key	Collation	Default Value	Policy Tags	Description
categoryID	INTEGER	NULLABLE	-	-	-
categoryName	STRING	NULLABLE	-	-	-
description	STRING	NULLABLE	-	-	-
picture	FLOAT	NULLABLE	-	-	-

EDIT SCHEMA VIEW ROW ACCESS POLICIES

Working with tables

Table tab

Query tab

The screenshot shows a web-based SQL editor interface. At the top, there are two tabs: 'order-details' and '*Untitled'. Below the tabs, there is a toolbar with buttons for 'RUN', 'SAVE', 'DOWNLOAD', 'SHARE', 'SCHEDULE', and 'MORE'. The main area displays a SQL query: `1 SELECT | FROM 'dwhba01.staging.order-details' LIMIT 1000`. A red error message is visible at the bottom right: 'Syntax error: SELECT list must not be empty at [1:5]'. A blue arrow points from this error message to the 'SELECT' keyword in the query.

Untitled

RUN SAVE DOWNLOAD SHARE SCHEDULE MORE

1 SELECT | FROM 'dwhba01.staging.order-details' LIMIT 1000

Syntax error: SELECT list must not be empty at [1:5]

Error notification

Working with tables

customers

QUERY SHARE COPY SNAPSHOT DELETE EXPORT

SCHEMA DETAILS DATA PROFILE DATA QUALITY

Filter Enter properties

In new tab

In split tab

In Python notebook PREVIEW

<input type="checkbox"/>	Field name	Data Type	Nullable	Primary Key	Collation	Default Value	Policy Tags	Description
<input type="checkbox"/>	string_field_0	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	string_field_1	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	string_field_2	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	string_field_3	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	string_field_4	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	string_field_5	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	string_field_6	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	string_field_7	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	string_field_8	STRING	NULLABLE	-	-	-	-	-
<input type="checkbox"/>	string field 9	STRING	NULLABLE	-	-	-	-	-

Working with tables

Writing the query in split tab and save the results
(after running the query)

The screenshot displays the Google Cloud BigQuery interface. At the top, there's a search bar and a notification about new capabilities. The left sidebar shows the 'Explorer' with a tree view of resources including 'datawarehouse', 'staging', 'categories', 'customers', 'employee-territories', 'employees', 'order-details', and 'orders'. The main area is split into two panes. The left pane shows the 'SCHEMA' of a table named 'customers', with columns 'Row' and 'string_field_1'. The right pane shows the 'Query results' table, which has columns 'Row', 'postalCode', and 'country'. A dropdown menu is open over the 'Query results' table, showing options to save the results: 'CSV (Google Drive)', 'CSV (local file)', 'JSON (local file)', 'JSONL (newline delimited)', 'BigQuery table', 'Google Sheets', and 'Copy to Clipboard'. The 'BigQuery table' option is highlighted.

Search (/) for resources, docs, products, and more

Your BigQuery projects will have new capabilities after February 14, 2024. Services and roles will be enabled.

Explorer + ADD <

Type to search ?

Viewing resources.

SHOW STARRED ONLY

- datawarehouse ☆ :
- staging ☆ :
 - categories ☆ :
 - customers ☆ :
 - employee-territories ☆ :
 - employees ☆ :
 - order-details ☆ :
 - orders ☆ :

SUMMARY

Nothing currently selected

customers

SCHEMA

Row	string_field_1
1	(171) 555-67
2	NULL
3	(171) 555-91
4	(171) 555-33
5	NULL
6	(171) 555-25
7	(171) 555-56
8	NULL
9	(503) 555-23
10	(509) 555-62

Untitled

RUN

```
1 SELECT
2 string_field_0 as custo
3 string_field_1 as compa
4 string_field_2 as cont
5 string_field_3 as cont
6 string_field_4 as addr
7 string_field_5 as city
8 string_field_6 as regi
9 string_field_7 as post
10 string_field_8 as coun
11 string_field_9 as phone
12 string field 10 as fav
```

Query results

SAVE RESULTS EXPLORE DATA

Row	postalCode	country
31	05022	Mexico

Create an account on GCP



google bigquery



All

Images

Videos

News

Shopping

More

Tools

About 20,700,000 results (0.29 seconds)

Sponsored



Google Cloud

<https://cloud.google.com/bigquery>

Google Cloud BigQuery

Data warehouse solution — Serverless, highly scalable, cost-effective data warehouse designed for business agility.

[Sign Up for Free](#)

Kick start your development. Get started with a free trial.

Create free account

Learn & build with our Free Tier & \$300 free credit!

Pricing calculator

Enter what you need to run in the Cloud & we'll calculate the cost.

Case Studies

Discover how innovative companies are using Google Cloud.

Create an account on GCP

Solve real business challenges on Google Cloud

Get started for free

Contact sales

Run workloads for free

20+ free products for all customers

All customers get [free hands-on experience](#) with popular products, including Compute Engine and Cloud Storage, [up to monthly limits](#).


\$300 in free credits for new customers

New customers get [\\$300 in free credits](#) to fully explore and conduct an assessment of Google Cloud. You won't be charged until you upgrade.

Start deploying pre-built solutions free

New customers get [\\$300 in free credits](#) on signup to use on deploying a dynamic website, launching a VM, building a three tier web app, and more [pre-built solutions templates](#).

Create an account on GCP

 Try Google Cloud for free

Step 1 of 2 Account Information



Ngân Trần Thị
ngantt@vnuis.edu.vn

[SWITCH ACCOUNT](#)



Good news! You're eligible for an additional \$100.00 in Free Trial credits for a total of \$400.00. You'll receive these credits within 24 hours of completing signup.

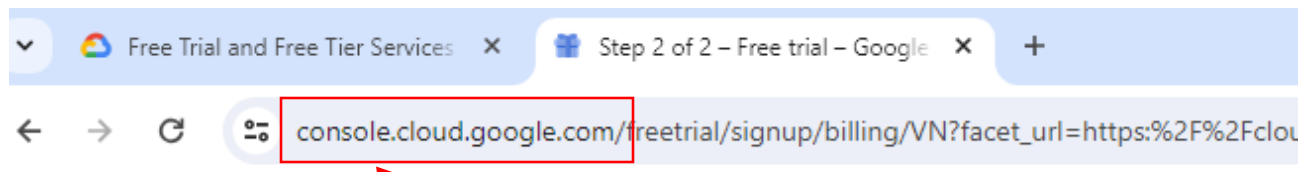
Country

Vietnam

By using this application, you agree to the [Google Cloud Platform](#), [Supplemental Free Trial](#), and [any applicable services and APIs](#) Terms of Service.

AGREE & CONTINUE

Create an account on GCP




Keep this, delete others for your instant using



Step 2 of 2 Payment Information Verification

Your payment information helps us reduce fraud and abuse. If using a credit or debit card, you won't be charged until you manually activate your full account.



Account type 

Business

Only Business accounts can have multiple users. You cannot change the account type after signing up. In some countries, this selection affects your tax options. If you choose Individual as your account type, you agree that use of your account is for your trade, business, craft, or profession. [Learn more](#)



Business name

|

Business name is a required field

Exercise 1

- Create a new project on BigQuery
- Create a dataset
- Import all tables in **Northwind** dataset to BigQuery
(<https://github.com/neo4j-contrib/northwind-neo4j/tree/northwind/data>)
- Transform these tables and create fact tables and dimensional tables (stored in another dataset after transforming).

<https://www.linkedin.com/pulse/oltp-operational-database-olap-transactional-data-singleton>

Dimension tables

- Contains descriptive textual or categorical data that provides context to the fact table.
- Holds attributes support grouping, filtering, and categorizing the data in the fact table.
- Used to provide business context and details about specific aspects of data in the fact table.
- Typically have fewer records compared to the fact table.
- Enables querying and analyzing data from different perspectives, enhancing data insights.
- Used to define hierarchies that allow drilling down into data.
- Can contain redundant data due to data de-normalization, improving query performance.
- Typically not directly connected but linked to the fact table via foreign keys.

Fact, fact tables, factless fact tables

- **Fact** A business performance measurement, typically numeric and additive, that is stored in a fact table
- **Fact table** In a dimensional model, the central table with numeric performance measurements characterized by a composite key, each of whose elements is a foreign key drawn from a dimension table.
- **Factless fact table** A fact table that has no facts but captures the many-to-many relationships between the dimension keys. Most often used to represent events or coverage information that does not appear in other fact tables.

Questions?