

Synthèse du laboratoire: BigQuery: Qwik Start - Console

Présentation :

BigQuery est un entrepôt de données entièrement géré qui permet des requêtes SQL ultra-rapides en utilisant l'infrastructure de Google. Ce laboratoire pratique guide les utilisateurs pour interroger des ensembles de données publics, créer des ensembles de données personnalisés, et charger des données dans des tables BigQuery.

Étapes principales :

1. Configuration et Connexion

- **Création de l'environnement temporaire** : Accès fourni via des identifiants étudiants temporaires pour éviter tout conflit avec un compte Google Cloud personnel.
- **Activation de Cloud Shell** : Configuration du projet et vérification des paramètres avec des commandes gcloud.

The screenshot shows the Google Cloud console interface. The top navigation bar includes the Google Cloud logo, the project ID 'qwiklabs-gcp-02-848081ec44a6', a search bar, and various utility icons. Below the navigation bar, there are tabs for 'Navigation menu', 'ACTIVITY', and 'RECOMMENDATIONS'. The main content area is divided into several panels: 'Project info' (showing project name, number, and ID), 'APIs' (showing a graph of requests per second), 'Google Cloud Platform status' (indicating all services are normal), and 'Monitoring'. A 'CLOUD SHELL' terminal window is open at the bottom, displaying the following commands and output:

```
Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to quwiklabs-gcp-02-848081ec44a6.
Use "gcloud config set project [PROJECT_ID]" to change to a different project.
student_03_1278ba845365@cloudshell:~ (qwiklabs-gcp-02-848081ec44a6) $ gcloud auth list
Credentialed Accounts

ACTIVE: *
ACCOUNT: student-03-1278ba845365@quwiklabs.net

To set the active account, run:
$ gcloud config set account 'ACCOUNT'

student_03_1278ba845365@cloudshell:~ (qwiklabs-gcp-02-848081ec44a6) $ gcloud config list project
[core]
project = quwiklabs-gcp-02-848081ec44a6
```

2. Interrogation d'un ensemble de données public

- Requête exécutée :

The screenshot shows the Google Cloud BigQuery console interface. The top navigation bar includes the Google Cloud logo, the project ID 'qwiklabs-gcp-02-848081ec44a6', a search bar, and various utility icons. Below the navigation bar, there are tabs for 'Navigation menu', 'ACTIVITY', and 'RECOMMENDATIONS'. The main content area is divided into several panels: 'Analysis' (showing 'BigQuery Studio'), 'Data transfers', 'Scheduled queries', 'Analytics Hub', 'Dataform', 'Partner Center', 'Orchestration', 'Migration', 'Assessment', 'Settings', and 'Release Notes'. The 'Explorer' panel shows a search bar and a list of resources. The 'Untitled query' panel shows a SQL query:

```
1 #standardSQL
2 SELECT
3   weight_pounds, state, year, gestation_weeks
4 FROM
5   `bigquery-public-data.samples.natality`
6 ORDER BY weight_pounds DESC LIMIT 10;
```

The 'Query results' panel shows the following table:

Row	weight_pounds	state	year	gestation_weeks
1	18.0007436923	null	2005	39
2	18.0007436923	KY	2004	43
3	18.0007436923	IL	2002	39
4	18.0007436923	PA	1970	null
5	18.0007436923	KY	2004	39
6	18.0007436923	UT	1982	33
7	18.0007436923	null	2007	null

The bottom of the console shows 'Results per page: 50' and '1 - 10 of 10'.

- **Objectif** : Récupérer des informations sur la natalité aux États-Unis. La validation a confirmé une exécution correcte de la requête.

3. Création d'un nouvel ensemble de données

The screenshot shows the Google Cloud BigQuery console. On the left is a navigation menu with options like Analysis, Migration, and Settings. The main area is divided into three panes. The 'Explorer' pane on the left shows a search bar and a list of resources, including the 'babynames' dataset. The 'DETAILS' pane on the right shows the 'Dataset info' for 'babynames', including its ID, creation time, expiration, and location. The 'SUMMARY' pane at the bottom shows a table with columns for 'Data' and 'US'.

Field	Value
Dataset ID	qwiklabs-gcp-02-848081ec44a6.babynames
Created	Dec 29, 2024, 12:52:45 AM UTC+1
Default table expiration	60 days
Last modified	Dec 29, 2024, 12:52:45 AM UTC+1
Data location	US
Description	
Default collation	
Default rounding mode	ROUNDING_MODE_UNSPECIFIED
Time travel window	7 days
Case insensitive	false
Labels	
Tags	

- **Nom de l'ensemble de données** : babynames.
- **Fonctionnalité** : Permet de structurer et gérer des tables personnalisées sous le projet.

4. Chargement des données dans une table

- **Table créée** : names_2014 dans l'ensemble de données babynames.
- **Source des données** : Fichier CSV provenant d'un bucket Google Cloud Storage : spls/gsp072/baby-names/yob2014.txt.
- **Schéma défini** :

The screenshot shows the 'Create table' dialog in the Google Cloud BigQuery console. The 'Source' section is set to 'Google Cloud Storage' with the file path 'spls/gsp072/baby-names/yob2014.txt' selected. The 'File format' is set to 'CSV'. The 'Destination' section shows the 'Project' as 'qwiklabs-gcp-02-848081ec44a6', the 'Dataset' as 'babynames', and the 'Table' as 'names_2014'. The 'Table type' is set to 'Table'. The 'CREATE TABLE' button is highlighted.

Source

Create table from: Google Cloud Storage

Select file from GCS bucket or use a URI pattern: ☒ spls/gsp072/baby-names/yob2014.txt **BROWSE**

File format: CSV

☐ Source Data Partitioning

Destination

Project: qwiklabs-gcp-02-848081ec44a6 **BROWSE**

Dataset: babynames

Table: names_2014

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

Table type: Table

CREATE TABLE CANCEL

5. Prévisualisation des données

- **Action** : Affichage des premières lignes des données de la table `names_2014` pour validation.

The screenshot shows the Google Cloud BigQuery interface. On the left is the navigation menu with options like Analysis, Migration, and Settings. The main area is divided into an Explorer pane on the left and a table preview on the right. The Explorer pane shows a project named 'qwiklabs-gcp-02-848081ec44a6' with a dataset 'babynames' containing a table 'names_2014'. The table preview on the right shows the first 14 rows of data.

Row	name	gender	count
1	Emma	F	20924
2	Olivia	F	19791
3	Sophia	F	18598
4	Isabella	F	17068
5	Ava	F	15688
6	Mia	F	13506
7	Emily	F	12642
8	Abigail	F	12076
9	Madison	F	10315
10	Charlotte	F	10111
11	Harper	F	9606
12	Sofia	F	9591
13	Avery	F	9563
14	Elizabeth	F	9557

6. Interrogation d'un ensemble de données personnalisé

- **Requête exécutée** :

The screenshot shows the Google Cloud BigQuery interface with a custom SQL query executed. The query is displayed in the main editor, and the results are shown in a table below it.

```
1 #standardSQL
2 SELECT
3   name, count
4 FROM
5   `babynames.names_2014`
6 WHERE
7   gender = 'M'
8 ORDER BY count DESC LIMIT 5;
```

Row	name	count
1	Noah	19286
2	Liam	18451
3	Mason	17192
4	Jacob	16869

- **Objectif** : Identifier les 5 prénoms masculins les plus populaires en 2014.

7. Validation et Questions

- **Question posée :**

BigQuery est-il un entrepôt de données entièrement géré permettant des requêtes SQL ultra-rapides ?

- **Réponse : Vrai**

Résultats obtenus et compétences acquises :

- Requête sur des ensembles de données publics dans BigQuery.
 - Création et gestion d'ensembles de données et de tables personnalisées.
 - Chargement de données à partir de Google Cloud Storage.
 - Interrogation de tables personnalisées en SQL.
-

Recommandations :

Ce laboratoire est un excellent point de départ pour explorer les capacités d'AppSheet. Il est conseillé d'explorer d'autres fonctionnalités de BigQuery ou continuer avec des laboratoires de la série **Qwik Starts** pour approfondir nos connaissances sur Google Cloud.