

# Explorer un ensemble de données avec BigQuery

## Introduction:

Le stockage et l'interrogation d'ensembles de données volumineux peuvent prendre du temps et être coûteux sans le matériel et l'infrastructure appropriés. BigQuery est un [entrepôt de données d'entreprise](#) qui résout ce problème en permettant des requêtes SQL ultra-rapides en utilisant la puissance de traitement de l'infrastructure de Google.

Transférez simplement vos données dans BigQuery et laissez-nous nous occuper de tout le travail. Vous pouvez contrôler l'accès à la fois au projet et à vos données en fonction des besoins de votre entreprise, par exemple en donnant aux autres la possibilité d'afficher ou d'interroger vos données.

Vous pouvez accéder à BigQuery dans la console , l' interface utilisateur Web classique ou un outil de ligne de commande , ou en appelant l' API REST BigQuery à l' aide de diverses bibliothèques clientes telles que Java, .NET ou Python. Il existe également une variété d' outils tiers que vous pouvez utiliser pour interagir avec BigQuery, tels que la visualisation des données ou le chargement des données.

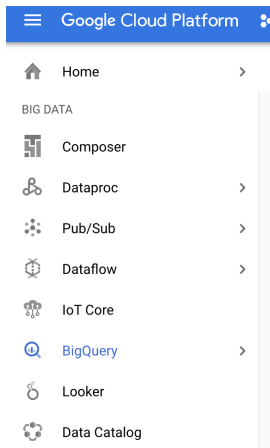
Cet atelier pratique vous montre comment utiliser l'interface utilisateur Web pour interroger des tables publiques et charger des exemples de données dans BigQuery.

## 1- Ouvrir BigQuery

La console BigQuery fournit une interface pour interroger les tables, y compris [les ensembles de données publics](#) proposés par BigQuery. La requête que vous exécuterez accède à une table à partir d'un ensemble de données public fourni par BigQuery. Il utilise un langage de requête standard pour rechercher l'ensemble de données et limite les résultats renvoyés à 10.

- **Ouvrir la console BigQuery**

Dans Google Cloud Console, sélectionnez le **menu Navigation > BigQuery** :



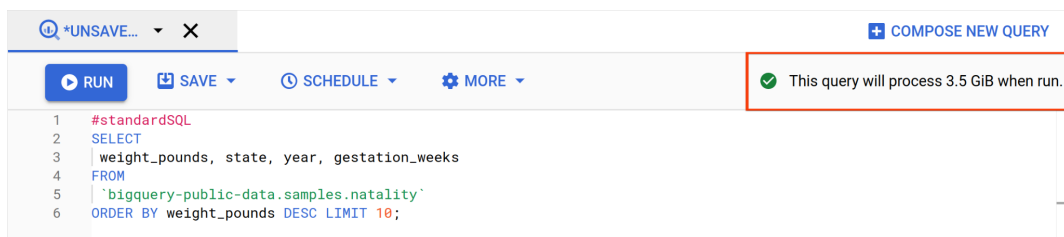
- **Interroger un ensemble de données public**

1. Copiez et collez la requête suivante dans l'éditeur de requêtes BigQuery :

```
#standardSQL
SÉLECTIONNER
  poids_livres, état, année, semaines de gestation
À PARTIR DE
  `bigquery-public-data.samples.natality`
ORDER BY weight_pounds DESC LIMIT 10;
```

Cet échantillon de données contient des informations sur la natalité américaine (taux de natalité).

Une coche verte ou rouge s'affiche selon que la requête est valide ou non. Si la requête est valide, le validateur décrit également la quantité de données à traiter après l'exécution de la requête.



Ces informations permettent de déterminer le coût d'exécution d'une requête.

2. Cliquez sur le bouton **Exécuter** .

Les résultats de votre requête doivent ressembler à ce qui suit :

## Query results

[SAVE RESULTS](#) ▼

Query complete (2.626 sec elapsed, 3.49 GB processed)

Job information [Results](#) JSON Execution details

Row	weight_pounds	state	year	gestation_weeks
1	18.0007436923	KY	2004	47
2	18.0007436923	OR	1972	40
3	18.0007436923	null	2007	39
4	18.0007436923	null	2008	null
5	18.0007436923	TX	1969	null
6	18.0007436923	null	2005	40
7	18.0007436923	null	2007	45
8	18.0007436923	null	2005	null
9	18.0007436923	GA	1979	34
10	18.0007436923	null	2007	38

*Vous pouvez parcourir le schéma d'autres ensembles de données publics dans BigQuery en cliquant sur **+ AJOUTER DES DONNÉES** > **Explorer** les ensembles de données publics , puis recherchez « données publiques bigquery » dans le champ de recherche.*

## 2- Charger des données personnalisées dans une table

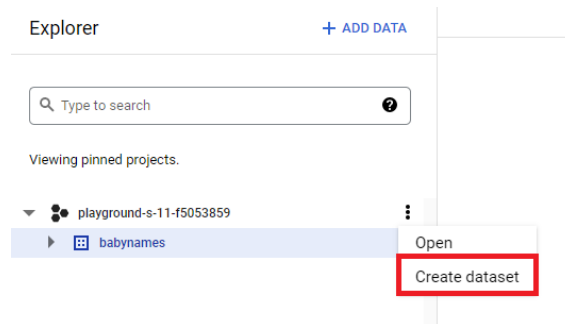
Pour charger des données personnalisées dans une table, vous effectuez les tâches suivantes :

- Créer un jeu de données
- Créer un tableau
- Ajouter des données à votre projet (à un bucket de stockage)
- Charger les données du bucket dans la table que vous avez créée

### ● Créer un jeu de données

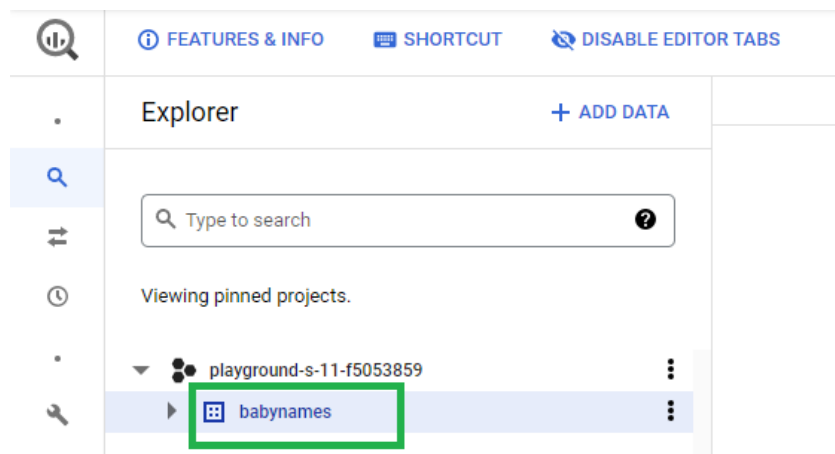
Les ensembles de données permettent de contrôler l'accès aux tables et aux vues d'un projet. Cet atelier n'utilise qu'une seule table, mais vous avez toujours besoin d'un ensemble de données pour contenir la table.

1. Dans le volet de gauche, près de l'ID de votre projet, cliquez sur **Afficher les actions**, puis sur **Créer un jeu de données** dans la section **Explorateur** .



2. Définissez l' **ID** de l'ensemble de données sur **babynames** .
3. Laissez tous les autres champs à leurs paramètres par défaut. Cliquez sur **Créer un jeu de données** .

Vous avez maintenant un ensemble de données.



- **Ajouter des données personnalisées**

Le fichier de données personnalisé que vous utiliserez contient environ 7 Mo de données sur les prénoms populaires, fournies par l'Administration de la sécurité sociale des États-Unis. Vous ajouterez le fichier zip à votre projet, puis créez un compartiment de stockage pour le fichier spécifique que vous devrez interroger.

Dans Cloud Shell, exécutez les commandes suivantes pour ajouter les fichiers de données à votre projet :

```
> gsutil cp gs://splis/gsp072/baby-names.zip .  
> unzip baby-names.zip
```

- **Créer un bucket Cloud Storage**

Créez maintenant un bucket Cloud Storage pour stocker les fichiers de données que vous avez téléchargés.

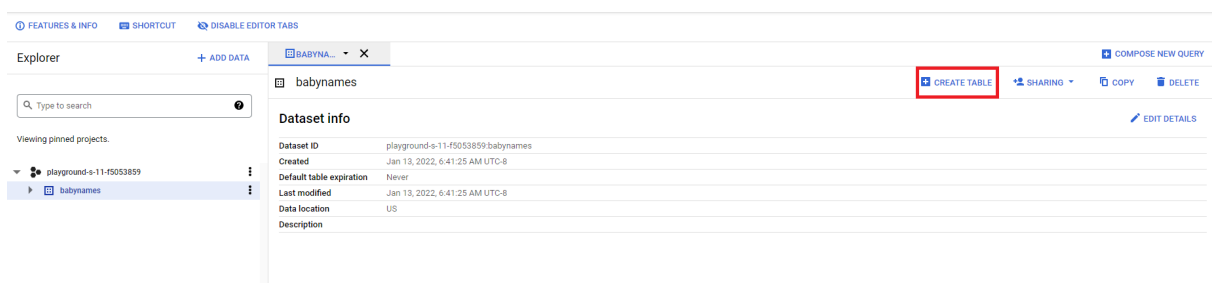
1. Dans Cloud Console, sélectionnez le **menu Navigation > Cloud Storage > Browser** , puis cliquez sur **Create bucket** .
2. Donnez à votre compartiment un **nom universellement unique** , puis cliquez sur **Créer** .
3. Dans Cloud Shell, exécutez la commande suivante pour déplacer le fichier yob2014.txt dans votre bucket. Remplacez <your\_bucket> le par le nom du bucket que vous venez de créer :  

```
> gsutil cp yob2014.txt gs://<your_bucket>
```

- **Charger les données dans une nouvelle table**

Ensuite, vous créez une table dans l'ensemble de données babynames, puis chargez le fichier de données de votre compartiment de stockage dans la nouvelle table.

1. Dans Cloud Console, sélectionnez le **menu Navigation > BigQuery** pour revenir à la console BigQuery.
2. Accédez à l' ensemble de données **babynames** en cliquant sur **afficher les actions** à proximité de votre ensemble de données, puis cliquez sur **Ouvrir** et sur **Créer une table**



3. Dans la boîte de dialogue Créer une table, définissez les champs suivants, laissez tous les autres sur la valeur par défaut :

maine	Valeur
Créer un tableau à partir de :	<b>Stockage Google Cloud</b>
Sélectionnez le fichier dans le bucket GCS :	<bucket_name> /yob2014 . txt, remplacez <bucket_name> par le nom du bucket que vous avez créé précédemment.
Format de fichier:	<b>CSV</b>
Nom de la table	names_2014
<b>Schéma</b> > Modifier en tant que texte	Faites glisser, puis ajoutez ce qui suit dans la zone de texte : <i>name:string,gender:string,count:integer</i>

Create table

**Source**

Create table from Google Cloud Storage

Select file from GCS bucket \*

ahmedhosni-gta13012022/yob2014.txt BROWSE

File format CSV

☐ Source Data Partitioning

**Destination**

Project \* playground-q-11-f5053859 BROWSE

Dataset \* babynames

Table \* names\_2014

Unicode letters, marks, numbers, connectors, dashes or spaces allowed.

Table type Native table

**Schema**

☐ Auto detect

☒ Edit as text

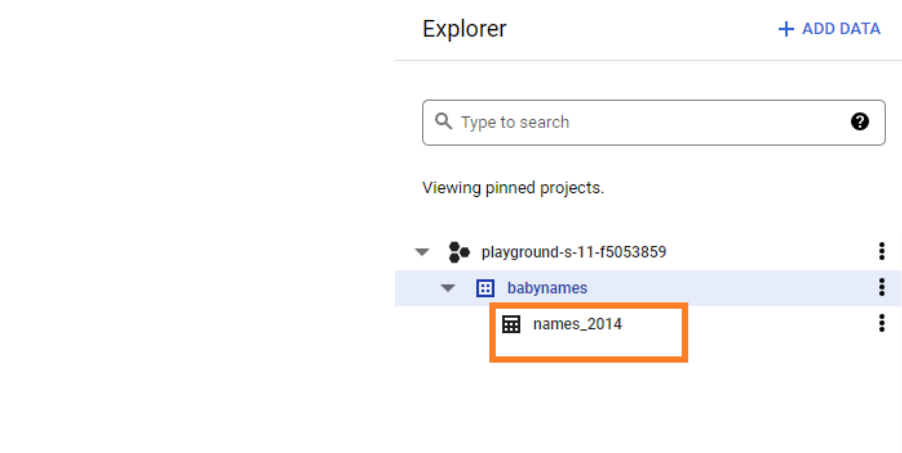
Press Alt+F1 for Accessibility Options.

name:string,gender:string,count:integer

CREATE TABLE CANCEL

4. Cliquez sur le bouton **Créer un tableau** .

Lorsque BigQuery a terminé de créer la table et de charger les données, la table **names\_2014** s'affiche sous l' ensemble de données **babynames**.



- **Aperçu du tableau**

Vérifiez votre tableau ! Affichez les premières lignes des données. Cliquez sur le tableau `names_2014` dans le menu de gauche, puis sur Aperçu .

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	10215
10	Charlotte	F	10111
11	Harper	F	9606
12	Sofia	F	9591
13	Avery	F	9563
14	Elizabeth	F	9557
15	Amelia	F	8789
16	Evelyn	F	8739
17	Ella	F	8547
18	Chloe	F	8533

Votre table est prête pour les requêtes.

- **Interroger un ensemble de données personnalisé**

L'exécution d'une requête sur des données personnalisées est identique à l'[interrogation d'un ensemble de données public](#) que vous avez fait précédemment, sauf que maintenant vous interrogez votre propre table au lieu d'une table publique.

Dans BigQuery, cliquez sur le bouton Composer une nouvelle requête dans l'angle supérieur droit pour effacer votre requête précédente.

Collez ou tapez la requête suivante dans l' éditeur de requêtes .

```
#standardSQL

SÉLECTIONNER

nom, compte

À PARTIR DE

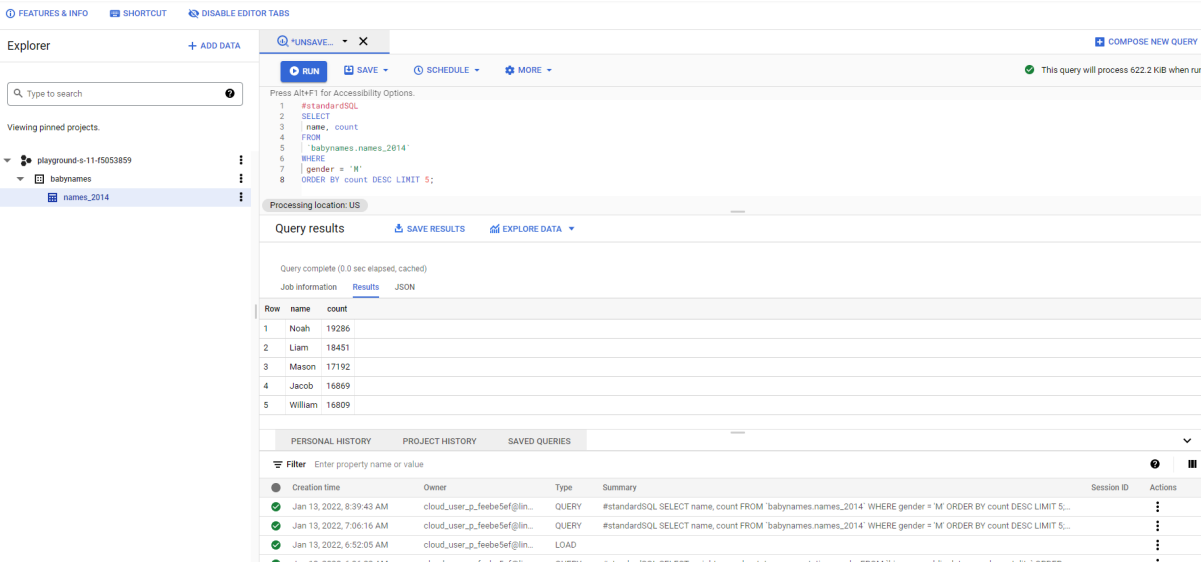
`babynames.names_2014`

OÙ

genre = 'M'

ORDER BY count DESC LIMIT 5;
```

Cliquez sur le bouton Exécuter . La requête affiche les 5 meilleurs noms de garçons pour l'année des données (2014) que vous avez chargées dans le tableau.



The screenshot shows the BigQuery web interface. On the left is the Explorer panel with a project named 'playground-s-11-f5053859' and a table 'babynames.names\_2014'. The main editor shows a SQL query in standard SQL dialect. The query results are displayed in a table with 5 rows, showing the top names by count for males in 2014. Below the results is a 'PERSONAL HISTORY' tab showing a list of recent queries and their execution details.

Query results

Row	name	count
1	Noah	19286
2	Liam	18451
3	Mason	17192
4	Jacob	16869
5	William	16809

PERSONAL HISTORY

Creation time	Owner	Type	Summary	Session ID	Actions
Jan 13, 2022, 8:39:43 AM	cloud_user_p_feebeSef@lin...	QUERY	#standardSQL SELECT name, count FROM `babynames.names_2014` WHERE gender = 'M' ORDER BY count DESC LIMIT 5...		⋮
Jan 13, 2022, 7:06:16 AM	cloud_user_p_feebeSef@lin...	QUERY	#standardSQL SELECT name, count FROM `babynames.names_2014` WHERE gender = 'M' ORDER BY count DESC LIMIT 5...		⋮
Jan 13, 2022, 6:52:05 AM	cloud_user_p_feebeSef@lin...	LOAD			⋮
Jan 13, 2022, 6:36:22 AM	cloud_user_p_feebeSef@lin...	QUERY	#standardSQL SELECT weight_pounds, state, year, gestation_weeks FROM `bigquery-public-data.samples.natality` ORDER ...		⋮