

Google Cloud PlatForm

Mr DIATTARA Ibrahima

Sommaire

- Cloud Vs on premise
- Cloud Storage
- Big Query
- Cloud Sql
- Cloud Spanner
- Datastore
- Big Table
- IAM
- Pub/Sub
- Data Fusion/Dataproc
- Dataflow
- DataPrep
- Data Studio
- Datalab
- Cloud composer
- Tenorsflow
- Auto ML Vision
- Dialogflow

Cloud vs on premise

Les environnements locaux nécessitent:

- Un équipement matériel pour exécuter les applications et les services.
- Cet équipement comprend des serveurs physiques, une infrastructure de réseau et du stockage. L'équipement nécessite une alimentation électrique, un système de refroidissement et une maintenance périodique effectuée par un personnel qualifié.
- Un serveur a besoin d'au moins un système d'exploitation. Il peut avoir besoin de plusieurs systèmes d'exploitation si l'organisation utilise une technologie de virtualisation
- Chaque système d'exploitation installé sur le serveur peut avoir un coût de licence différent. Les licences du système d'exploitation et des logiciels sont généralement vendues par serveur ou par licence d'accès client (CAL). À mesure que les entreprises se développent, les accords de licence deviennent plus restrictifs.
- Les systèmes locaux nécessitent une maintenance du matériel, des microprogrammes, des pilotes, du BIOS, du système d'exploitation, des logiciels et de l'antivirus. Les organisations tentent de réduire le coût de cette maintenance lorsque cela est possible.

Project



Google Cloud Platform



New Project

Project name *

My Project 31317



Project ID: graceful-creek-264520. It cannot be changed later. [EDIT](#)

Location *



No organisation

[BROWSE](#)

Parent organisation or folder

[CREATE](#)

[CANCEL](#)

IAM(Identity and Access Management)

Le rôle de IAM (Identity and Access Management) dans Google Cloud est de gérer les permissions et les identités des utilisateurs et des ressources dans un environnement cloud. IAM permet de définir qui a accès à quoi

Service accounts

+ Create service account

🗑 Delete

👤 Manage access

🔄 Refresh

🎓 Learn

Service accounts for project "diattara"

A service account represents a Google Cloud service identity, such as code running on Compute Engine VMs, App Engine apps, or systems running outside Google. [Learn more about service accounts.](#)

Organization policies can be used to secure service accounts and block risky service account features, such as automatic IAM Grants, key creation/upload, or the creation of service accounts entirely. [Learn more about service account organization policies.](#)

Filter

Enter property name or value

?

☰

<input type="checkbox"/>	Email	Status	Name ↑	Description	Key ID	Key creation date	OAuth 2 Client ID ?	Actions
No rows to display								

Les type de role sont:

Owner

Editor

Reader

.....

Role IAM

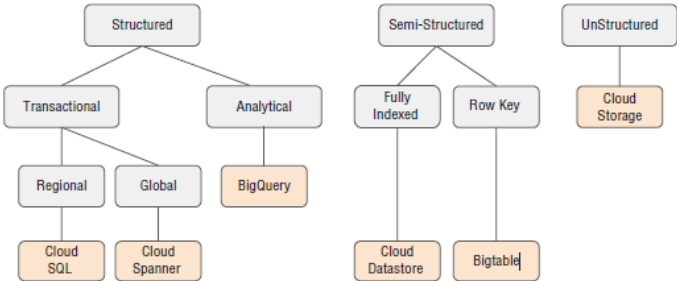
Voici les principales fonctions d'IAM dans Google Cloud :

- **Gestion des identités** : IAM permet de gérer les identités des utilisateurs, des groupes, des services et des ressources. Cela inclut la création, la modification et la suppression de comptes utilisateurs et de rôles.
- **Gestion des accès** : IAM permet de définir les permissions spécifiques pour les utilisateurs, groupes ou services afin de restreindre ou accorder l'accès à des ressources cloud comme les machines virtuelles, les bases de données, les buckets de stockage, etc.
- **Contrôle des accès au niveau des ressources** : IAM permet de spécifier des politiques de contrôle d'accès (IAM policies) pour chaque ressource au niveau de l'organisation, du projet, ou même de ressources individuelles comme des instances VM ou des bases de données.
- **Audits et sécurité** : IAM est également utilisé pour effectuer un suivi des actions des utilisateurs à travers les journaux d'audit, permettant d'assurer la traçabilité et la sécurité des accès et des modifications sur les ressources.







Google cloud Platform:Stockage

Choosing Google Database products

FIGURE 1.1 Choosing a storage technology in GCP



Choosing Google Database products

Relational		Non-relational		Object - Unstructured	Data Warehouse
					
Cloud SQL	Cloud Spanner	Cloud Datastore	Cloud Bigtable	Cloud Storage	BigQuery

Cloud storage

- supports unstructured data storage like binary
- supports data encryption at rest and in transit
- Consider using Cloud Storage, if you need to store immutable blobs larger than 10 MB, such as large images or movies.
- This storage service provides petabytes of capacity with a maximum unit size of 5 TB per object.

Storage Type

- **Choose where to store your data**

Location: us (multiple regions in United States)

Location type: Multi-region

- **Choose a default storage class for your data**

A storage class sets costs for storage, retrieval, and operations. Pick a default storage class based on how long you plan to store your data and how often it will be accessed. [Learn more](#)

☐ **Standard** 

Best for short-term storage and frequently accessed data

☒ **Nearline**

Best for backups and data accessed less than once a month

☐ **Coldline**

Best for disaster recovery and data accessed less than once a quarter

☐ **Archive**

Best for long-term digital preservation of data accessed less than once a year

guidance only. [Pricing details](#)

Storage and retrieval

Storage size

GB

\$0.010 per GB-month

Data retrieval size

GB

\$0.010 per GB-month

Object versioning: Off

If you enable versioning, note that noncurrent versions are billed at the same rate as live objects.

Operations

	Stockage standard	Stockage Nearline	Stockage Coldline	Stockage Archive
Récupération des données	0 \$ par Go	0,01 \$ par Go	0,02 \$ par Go	0,05 \$ par Go
Durée minimale de stockage	Aucune	30 jours	90 jours	365 jours

<https://cloud.google.com/storage/docs/storage-classes#descriptions>


Exo1 GCS

1. Créer un compte de service nommé "mosef" en lui affectant le rôle de "Owner".
2. Créer un bucket.
3. Coder un script Python qui permet de mettre le fichier JSON suivant dans votre bucket, dans un répertoire appelé "datasorbonne"

Lien JSON https://github.com/idiattara/Spark_DIATTARA/blob/main/json_to_push.json


Correction


1. Create a billing account
2. Create a project
3. Create a account service (service accounts) with owner role
4. Go to manager key and genere a key in json format
5. Create a Bucket **Warning** choose your storage type (standard, near, cold, archive)
6. Python code : https://github.com/idiattara/Spark_DIATTARA/blob/main/push_to_bucket.py


 **diattar**


Location: us (multiple regions in United States) | Storage class: Standard | Public access: Not public | Protection: Soft Delete

OBJECTS | CONFIGURATION | PERMISSIONS | PROTECTION | LIFECYCLE | OBSERVABILITY **NEW** | INVENTORY REPORTS | OPERATIONS



Folder browser 





diattar 

datasorbonne/ 

Buckets > diattar > datasorbonne 

CREATE FOLDER | UPLOAD ▾ | TRANSFER DATA ▾ | OTHER SERVICES ▾

Filter by name prefix only ▾  Filter Filter objects and folders Show Live objects only ▾ 

<input type="checkbox"/>	Name	Size	Type	Created 	Storage class	Last mod
<input type="checkbox"/>	 data_20250324_153817.json	141 B	application/json	Mar 24, 2025, 4:38:17 PM	Standard	Mar 24,  

Big Query

BigQuery est un data warehouse cloud proposé par Google Cloud, conçu pour stocker et analyser de grandes quantités de données de manière rapide grâce à son architecture distribuée

- Provides fully managed, no-ops, OLAP solution
- Provides high capacity, data warehousing analytics solution
- Ideal for big data exploration and processing
- Not ideal for operational or transactional databases
- provides SQL interface
- Provides Machine learning

Table external VS manager

Sur Big query on nous avons deux type de table

Table Managée : Une table managée dans BigQuery est une table où les données sont directement stockées dans BigQuery. BigQuery gère automatiquement le stockage, la partition, et la gestion des données. Lorsque tu supprimes une table managée, les données sont également supprimée

Table Externe : Une table externe dans BigQuery est une table qui ne stocke pas directement les données dans BigQuery, mais référence des données stockées dans un autre emplacement, comme Google Cloud Storage (GCS). Les données restent dans leur emplacement d'origine et sont lues au moment de l'exécution des requêtes. Cela permet d'éviter les coûts liés au stockage des données dans BigQuery.

Avantages

1. Réduction des coûts de stockage dans BigQuery

Cela signifie que tu ne payes pas pour le stockage dans BigQuery, mais seulement pour le traitement des données lors de l'exécution de requêtes. Le stockage dans GCS est généralement moins cher que dans BigQuery.

2. Flexibilité dans la gestion des données

Avec des tables externes, tu peux garder tes données dans leur format original dans GCS, tout en utilisant BigQuery pour les analyser. Cela permet de séparer le stockage et l'analyse des données, ce qui peut simplifier la gestion des données tout en optimisant les coûts

3. Simplicité dans l'utilisation de données historiques

Si tu as des données anciennes ou moins fréquemment utilisées, tu peux les garder dans Google Cloud Storage tout en continuant à les analyser dans BigQuery via des tables externes, sans devoir les importer dans BigQuery à chaque fois

Incov

Coûts de lecture supplémentaires

Les tables externes ne sont pas stockées dans BigQuery, ce qui signifie que les données doivent être lues à partir d'un autre emplacement (comme Google Cloud Storage) chaque fois que tu exécutes une requête. Cela peut entraîner des temps de réponse plus lents par rapport aux tables managées où les données sont directement stockées dans BigQuery.

Gestion de la cohérence des données

Les données dans une table externe restent dans leur emplacement d'origine (par exemple, Google Cloud Storage). Si ces données sont mises à jour ou modifiées en dehors de BigQuery, la cohérence des données peut devenir un problème. Tu ne peux pas garantir que la version la plus récente des données est toujours utilisée lors des requêtes.

Limitations de certains types de données

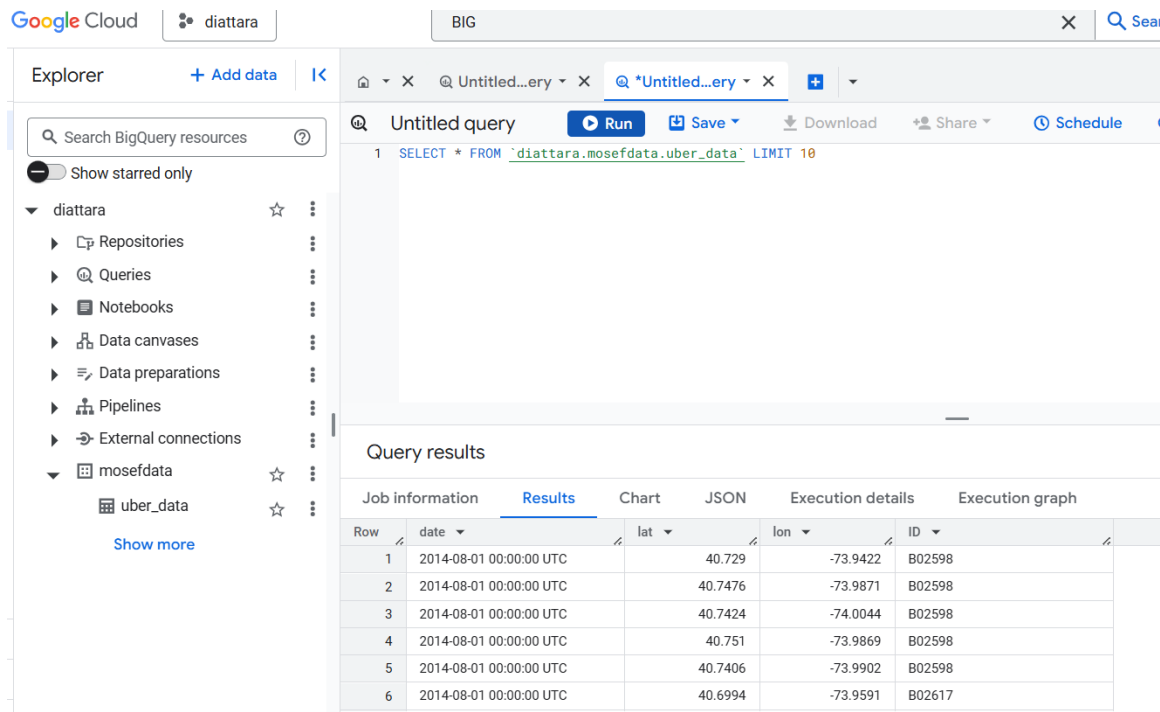
Les tables externes ne supportent pas tous les types de données et formats de fichiers. Par exemple, les fichiers très volumineux ou très complexes peuvent poser des problèmes de performance ou de compatibilité.

Optimisation automatique dans BigQuery pour les tables managées

Exo1 Bigquery

Créer un table uber_data dans un dataset mosefdata en upload le fichier suivant:

https://github.com/idiattara/Spark_DIATTARA/blob/main/uber.csv



The screenshot displays the Google Cloud BigQuery interface. On the left, the 'Explorer' panel shows the project 'diattara' with various resources. The 'mosefdata' dataset is expanded, revealing the 'uber_data' table. The main panel shows an 'Untitled query' with the following SQL statement:

```
1 SELECT * FROM `diattara.mosefdata.uber_data` LIMIT 10
```

The query has been executed, and the results are displayed in a table format. The table has 6 rows and 6 columns: Row, date, lat, lon, ID, and an additional column. The data represents Uber ride records from August 1, 2014.

Row	date	lat	lon	ID	
1	2014-08-01 00:00:00 UTC	40.729	-73.9422	B02598	
2	2014-08-01 00:00:00 UTC	40.7476	-73.9871	B02598	
3	2014-08-01 00:00:00 UTC	40.7424	-74.0044	B02598	
4	2014-08-01 00:00:00 UTC	40.751	-73.9869	B02598	
5	2014-08-01 00:00:00 UTC	40.7406	-73.9902	B02598	
6	2014-08-01 00:00:00 UTC	40.6994	-73.9591	B02617	

Exo2 Bigquery

Créer un modele de 4 cluster en utilisant l'algo Kmeans en se basant sur les variables lon et lat

Correction: https://github.com/idiattara/Spark_DIATTARA/blob/main/kmeans.sql

Untitled query

Run

Save

Download

Share

Schedule

Open in

More

Query completed

```
1 SELECT * FROM ML.PREDICT(MODEL `diattara.mosefdata.kmeansmodele`,(SELECT * FROM `diattara.mosefdata.uber_data` ))
```

Press Alt+F1 for Accessibility Options.

Query results

Save results

Open in

ob information

Results

Chart

JSON

Execution details

Execution graph

w	CENTROID_ID	CENTROID_ID	N...DISTANCE	date	lat	lon	ID
1	1	1	0.280338508095...	2014-08-01 00:02:00 UTC	40.7407	-74.0077	B02598
		3	1.255352629175...				
		2	1.582643819351...				
		4	4.137851234597...				
2	1	1	0.073297227082...	2014-08-01 01:22:00 UTC	40.7348	-73.9927	B02682
		3	1.219362949448...				

Activer Windows

Exo3

Créer un table external qui va pointer sur la fichiers de GCS de l'exo1

https://github.com/idiattara/Spark_DIATTARA/blob/main/external.sql

The screenshot shows the Google BigQuery web interface. On the left is the 'Explorer' sidebar with a search bar and a tree view of resources. The main area is divided into a query editor at the top and a 'Query results' section at the bottom. The query editor contains a SQL script to create an external table and query it. The 'Query results' section shows a table with 2 rows and 4 columns: Row, location, price, and typeProduit.

Explorer + Add data | <

Search BigQuery resources ?

Show starred only

- diattara
 - Repositories
 - Queries
 - Notebooks
 - Data canvases
 - Data preparations
 - Pipelines
 - External connections
- mosefdata
 - Models (1)
 - kmeansmodele
- external_table
- uber_data

Untitled query Run Save Download

```
1 CREATE OR REPLACE EXTERNAL TABLE `diattara.mosefdata.external_table`  
2 OPTIONS (  
3   format = 'NEWLINE_DELIMITED_JSON',  
4   uris = ['gs://diattar/*.json']  
5 );  
6  
7 select * from `diattara.mosefdata.external_table`
```

Query results

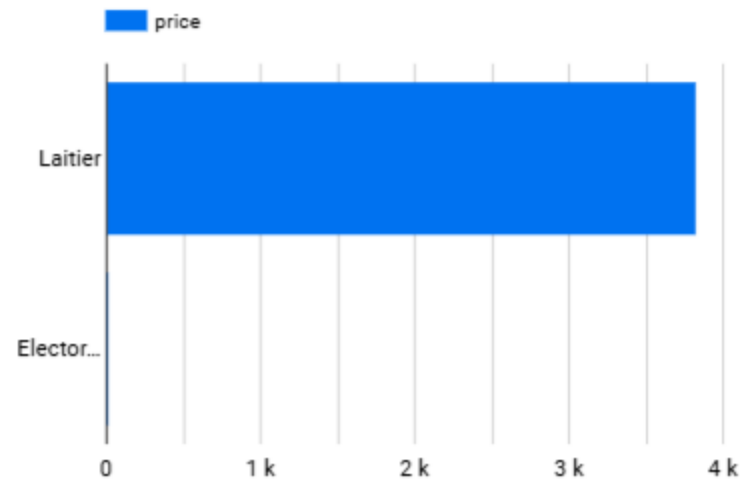
Job information Results Chart JSON Execution details

Metadata caching is disabled. You can accelerate queries over external tables

Row	location	price	typeProduit
1	3.7038, 8.4168	3826.4	Laitier
2	7.4, 6	12.0	Electornic

How to Viz data in bigquery using Looker Studio

<https://datastudio.google.com/>



DataStorage

Cloud Datastore non-relational NoSQL database(document)

fully managed with no-ops and no planned downtime and no need to provision database instances

provides SQL-like query language

supports strong and eventual consistency – ensures that entity lookups and ancestor queries always receive strongly consistent data. All other queries are eventually consistent.

supports data encryption at rest and in transit

provides terabytes of capacity with a maximum unit size of 1 MB per entity (vs Bigtable)

Consider using Cloud Datastore if you need to store semi-structured objects, or if require support for transactions

Embedded Array

Concept	Relational Database	Datastore
Category of object	Table	Kind
Single Object	Row	Entity
Individual data for an object	Column	Property
Unique ID for an object	Primary key	Key

```
{  
  "properties": {  
    "ville": {  
      "stringValue": "paris"  
    },  
    "region": {  
      "stringValue": "idf"  
    }  
  }  
}
```

DataStorage

	features, with offline support and real-time synchronization. SELECT NATIVE MODE	behavior on top of Cloud Firestore's powerful storage layer. SELECT DATASTORE MODE
API	Firestore	Datastore
Scalability	Automatically scales to millions of concurrent clients	Automatically scales to millions of writes per second
App engine support	Not supported in the App Engine standard Python 2.7 and PHP 5.5 runtimes	All runtimes
Max writes per second	10,000	No limit
Real-time updates	✓	✗
Mobile/web client libraries with offline data persistence	✓	✗

Big Table

Empld	Nom	Prénom	Salaire
1	Durant	Jacques	40000
2	Dupont	Marie	50000
3	Martin	Jeanne	44000

- Une table SGBDR sérialise toutes les valeurs d'une ligne ensemble, puis les valeurs de la ligne suivante, etc.
1,Durant, Jacques, 40000; 2,Dupont, 50000; 3,Martin,Jeanne,44000;

Une base de données orientée colonne sérialise les valeurs d'une colonne, puis celles de la suivante, etc.
1,2,3; Durant,Dupont,Martin; Jacques,Marie,Jeanne; 40000,50000,44000

Requêtage: Cbt ou hbase

Big Table

- Bigtable is a non-relational NoSQL (Column)
- supports large quantities (>1 TB) of semi-structured or structured data (vs Datastore)
- Does not support transactions or strong relational semantics (vs Datastore)
- Does not support SQL queries (vs BigQuery and Datastore)
- provides petabytes of capacity with a maximum unit size of 10 MB per cell and 100 MB per row
- ideal for time-series or natural semantic ordering data
- Consider using Cloud Bigtable, if you need to high performance datastore to perform analytics on a large of structured objects.

Big Table

Le modèle se base sur six concepts, qui sont :

Table						
RowKey	CF1			CF2		
	Colonne1	Colonne2	Colonne3	Colonne4	Colonne5	Colonne6
			Timestamp3: Valeur3			
			Timestamp3: Valeur3			
			Timestamp2: Valeur2			
			Timestamp1: Valeur1			

Figure 1. Vue synthétique du modèle de données dans HBase.

1. **Table**
2. **RowKey** : Une lignes d'une table est identifiée par une clé unique (RowKey)
3. **Column Family** : Les données au sein d'une ligne sont regroupées par column family. Seule les colonnes column family sont définit à la création de la table
4. **Column qualifier ou column** : L'accès aux données au sein d'une column family se fait via le column qualifier
5. **Cell** = Les données stockées dans une cellule sont appelée les valeurs

Version : Les valeurs au sein d'une cellule sont versionnées. Les versions sont identifiés par leur timestamp , le nombre de versions est configuré via la Column Family. Par défaut, ce nombre est égale à trois.

(rowkey, column family, column,, timestamp) ---> value

Consistance : Toutes les modifications sont atomiques et les lectures se font toujours sur la dernière valeur validée (commit)

Big Table Hbase commande

```
create 'sda_sorb', 'prof_sorb'
```

```
put 'sda_sorb', 'key-row1', 'prof_sorb:cloud', 'helmi'
```

```
put 'sda_sorb', 'key-row1', 'prof_sorb:cloud', 'diattara'
```

```
put 'sda_sorb', 'key-row1', 'prof_sorb:math', 'smitch'
```

```
scan 'sda_sorb'
```

```
scan 'sda_sorb', {VERSIONS => 3}
```

```
get 'sda_sorb', 'key-row1', {COLUMN => 'prof_sorb:cloud', VERSIONS => 3}
```

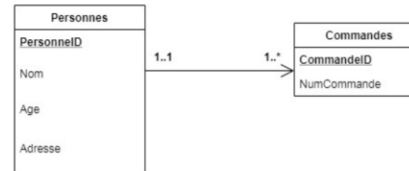
<https://cloud.google.com/bigtable/docs/quickstart-hbase?hl=fr>

Commande

<https://www.guru99.com/hbase-shell-general-commands.html>

<https://stackoverflow.com/questions/7681933/hbase-getting-all-timestamped-values-for-a-cell>

Bases de données relationnelles OLTP



```
CREATE TABLE Personnes ( PersonneID int AUTO_INCREMENT PRIMARY KEY,  
                           Nom VARCHAR(20) NOT NULL,  
                           Age int, Adresse VARCHAR(100) );
```

```
CREATE TABLE Commandes ( CommandeID int AUTO_INCREMENT PRIMARY KEY,  
                           NumCommande int NOT NULL,  
                           PersonneID int,  
                           FOREIGN KEY (PersonneID) REFERENCES Personnes(PersonneID) );
```

insert into Personnes (Nom, Age) values("diattara", 30);

```
mysql> insert into Commandes (NumCommande, PersonneID) values(123, 30);  
ERROR 1452 (23000): Cannot add or update a child row: a foreign key constraint fails (`test`.`Commandes`, CONSTRAINT `Commandes_ibfk_1` FOREIGN KEY (`PersonneID`) REFERENCES `Personnes` (`PersonneID`))  
mysql> █
```

insert into Commandes (NumCommande, PersonneID) values(123, 10);

Cloud SQL

- Cloud SQL provides managed, relational SQL databases
- Offers MySQL and PostgreSQL databases as a service
- single region only – although it now supports cross region read replicas (vs Cloud Spanner)
- supports data encryption at rest and in transit
- provides up to 10,230 GB, depending on machine type (vs Cloud Spanner)
- Consider using Cloud SQL for full relational SQL support for OTLP and lift and shift of MySQL, PostgreSQL databases

Cloud Spanner

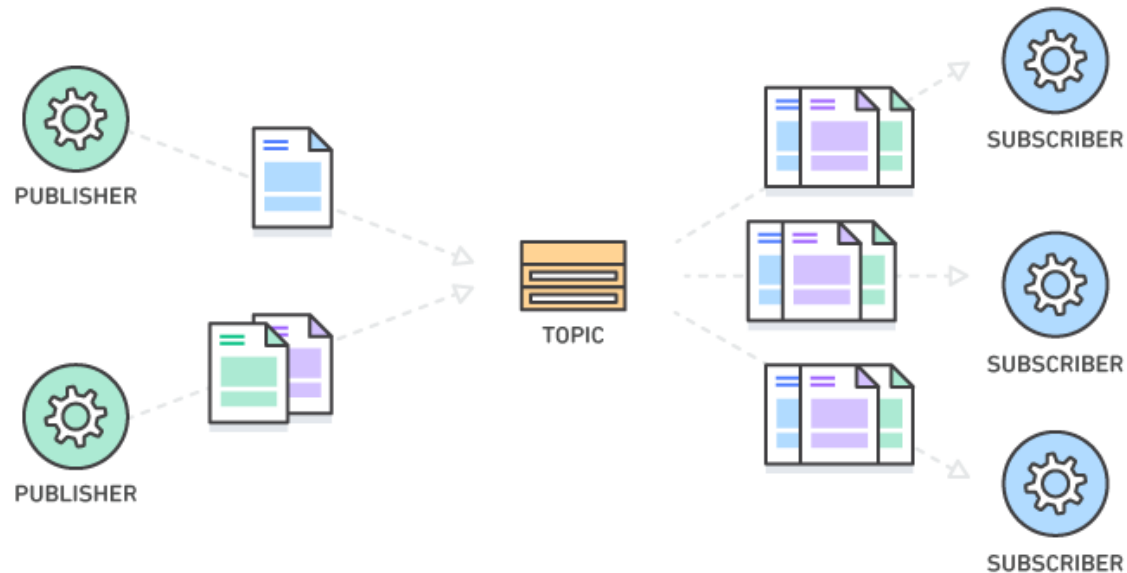
- Cloud Spanner provides fully managed, relational SQL databases with joins and secondary indexes
- Provides cross-region,
- supports database sizes exceeding ~2 TB (vs Cloud SQL)
- does not provide direct lift and shift for relational databases (vs Cloud SQL) expensive as compared to Cloud SQL
- Consider using Cloud SQL for full relational SQL support, with horizontal scalability spanning petabytes for OTLP

Google cloud Platform:Stockage

	Cloud Storage	Cloud SQL	Datastore	Bigtable	BigQuery
Capacity	Petabytes +	Gigabytes	Terabytes	Petabytes	Petabytes

	Cloud Datastore	Bigtable	Cloud Storage	Cloud SQL	Cloud Spanner	BigQuery
Type	NoSQL document	NoSQL Wide Column	Blob Storage	Relational SQL - OLTP	Relational SQL - OLTP	Relational SQL - OLAP
Transactions	Yes	Single-Row	No	Yes	Yes	No
Complex Queries	No	No	No	Yes	Yes	Yes
Capacity	Terabytes	Petabytes	Petabytes	Upto ~10TB	Petabytes	Petabytes
Unit Size	1MB/entity	~10 MB/cell ~100 MB/row	5TB per object	Depends on DB Engine	10,240 MiB/row	19 MB/row

Pub Sub



How to publish and consum data from pub sub

- 1 Creat a topic
- 2 Créer a subscription
- 3 Grant role(Owner, Editor pub Sub ,)

Kafka vs Pub Sub

Pub sub Scalabilité automatique

Utilisé pour des notification

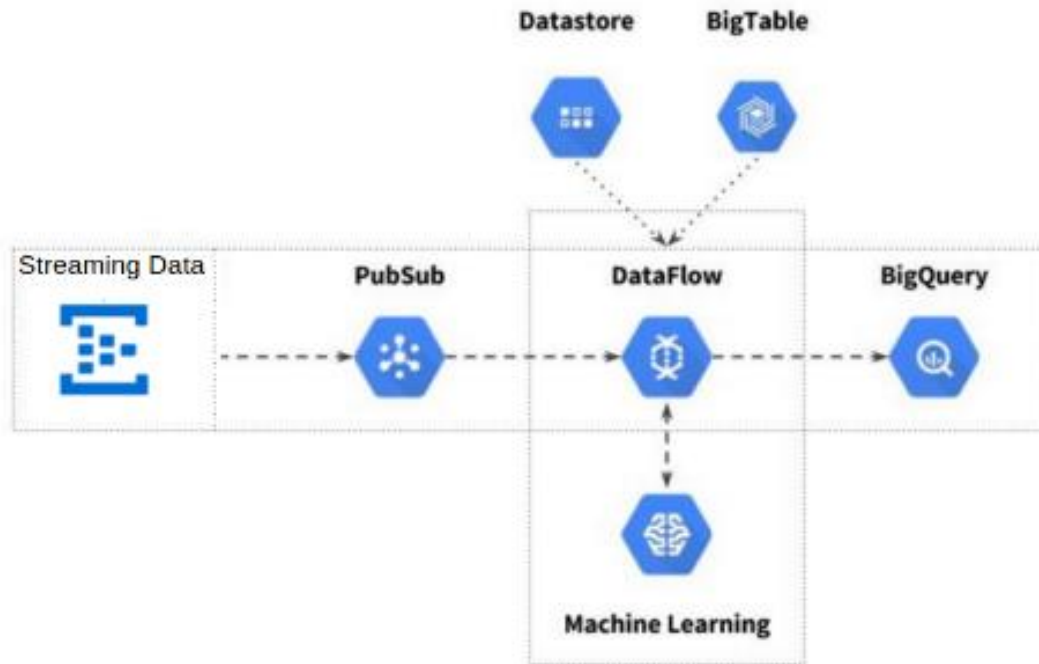
Kafka est plus rapide

Kafka pour des volum élevés

DataFlow

Cloud Dataflow est un service de traitement de données e, compatible avec l'exécution de pipelines par flux (stream) et par lots (batches).

Cet outil sont également à l'origine du modèle de programmation Apache Beam



DataLab ou Colab

Write a script that push data to kafka with Colab

```
!pip install kafka-python
from kafka import KafkaProducer
producer = KafkaProducer(bootstrap_servers='20.107.76.88:9092')
producer.send('datascientest', b'some_message_bytes from cloab')
producer.flush()
```

DataPrep

Dataprep by Trifacta est un service intelligent qui permet d'explorer visuellement, de nettoyer et de préparer des données structurées et non structurées à des fins d'analyse, de création de rapports et de machine learning

La transformation idéale suivante est suggérée et prédite à chaque entrée d'UI afin que vous n'ayez pas à écrire de code.



DataPrep

Nous avons un fichier json avec le format:

```
{"nom": "diattara", "prix": 100, "ville": "paris"}
```

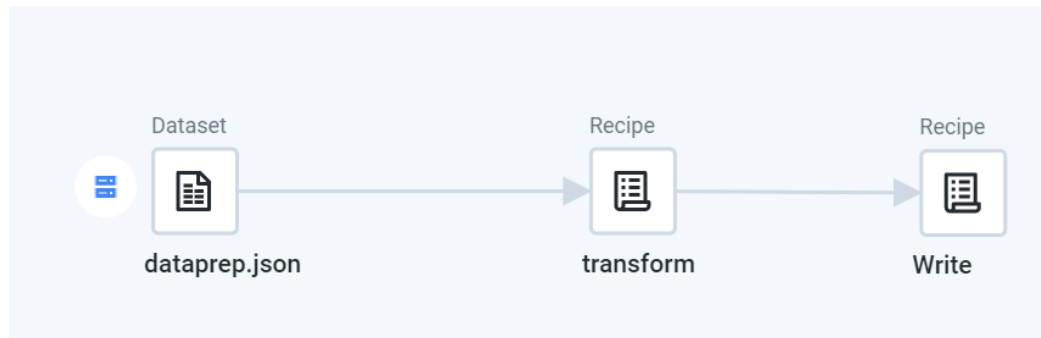
```
{"nom": "joe", "ville": "marseille"}
```

```
{"nom": "smitch", "prix": 50, "ville": "lyon"}
```

```
{ "prix": 50, "ville": "lyon"}
```

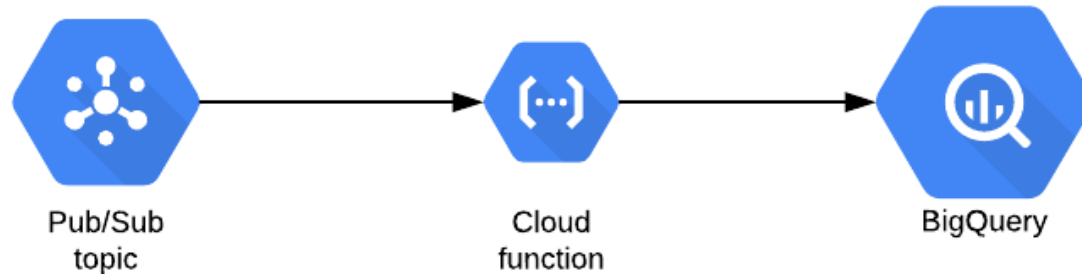
L'objectif est de mettre en majuscule tous les noms et de remplacer les valeurs nulles par le résultat sous BigQuery

Prix nul avec la médiane sans écrire de code

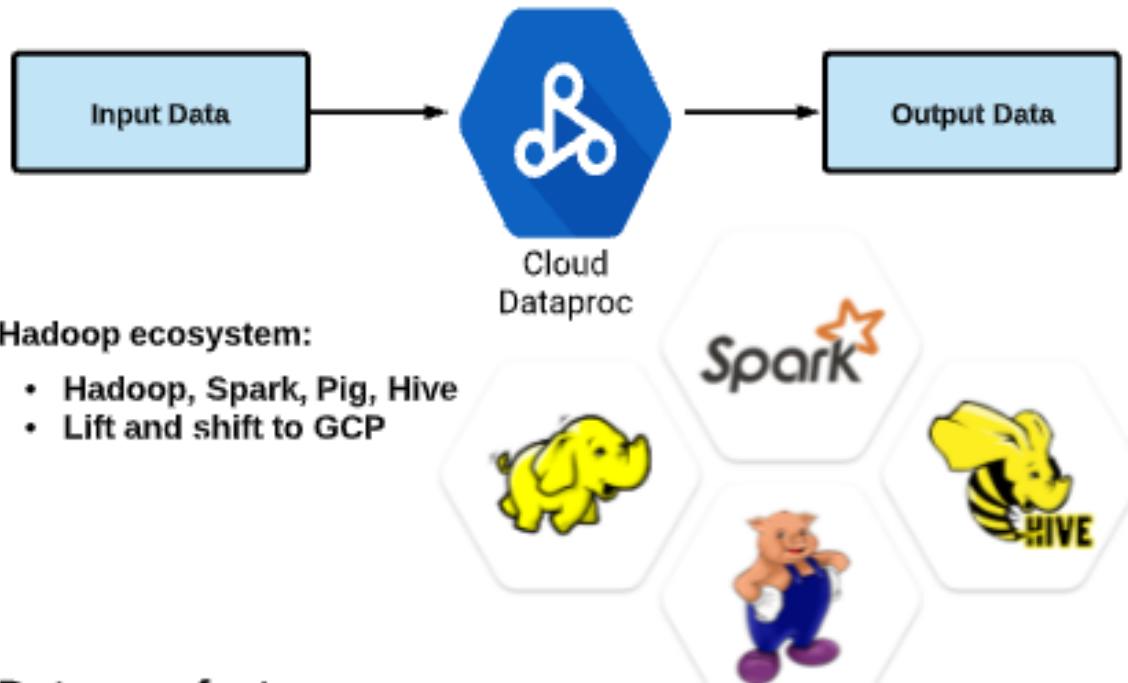


Cloud function

Cloud Functions est une solution de calcul légère permettant aux développeurs de créer des fonctions autonomes à usage unique qui répondent aux événements Cloud sans avoir à gérer de serveur ni d'environnement d'exécution.



DataProc/Datafusion



How to Create Data fusion flow

1 Create instance and enable account service data fusion to access to data proc

Warning! the instance creation can take 30 min

2 Copie the Account service and grand it the right access

1048786481523-compute@developer.gserviceaccount.com

3 Go to the UI : <https://sda-sda132510-dot-usw1.datafusion.googleusercontent.com/cdap/ns/default>

How to Create Data fusion flow

Create a flow with datafusion that pull data from GCS and push in Bigquery Every 06h 00 AM

DatesetName: datasetdatascientest

Tablename: tabledatascientest



```
[
  {
    "name": "etlSchemaBody",
    "schema": {
      "type": "record",
      "name": "fileRecord",
      "fields": [
        {
          "name": "age",
          "type": "int"
        },
        {
          "name": "nom",
          "type": "string"
        }
      ]
    }
  }
]
```


DataPre Vs DataFusion

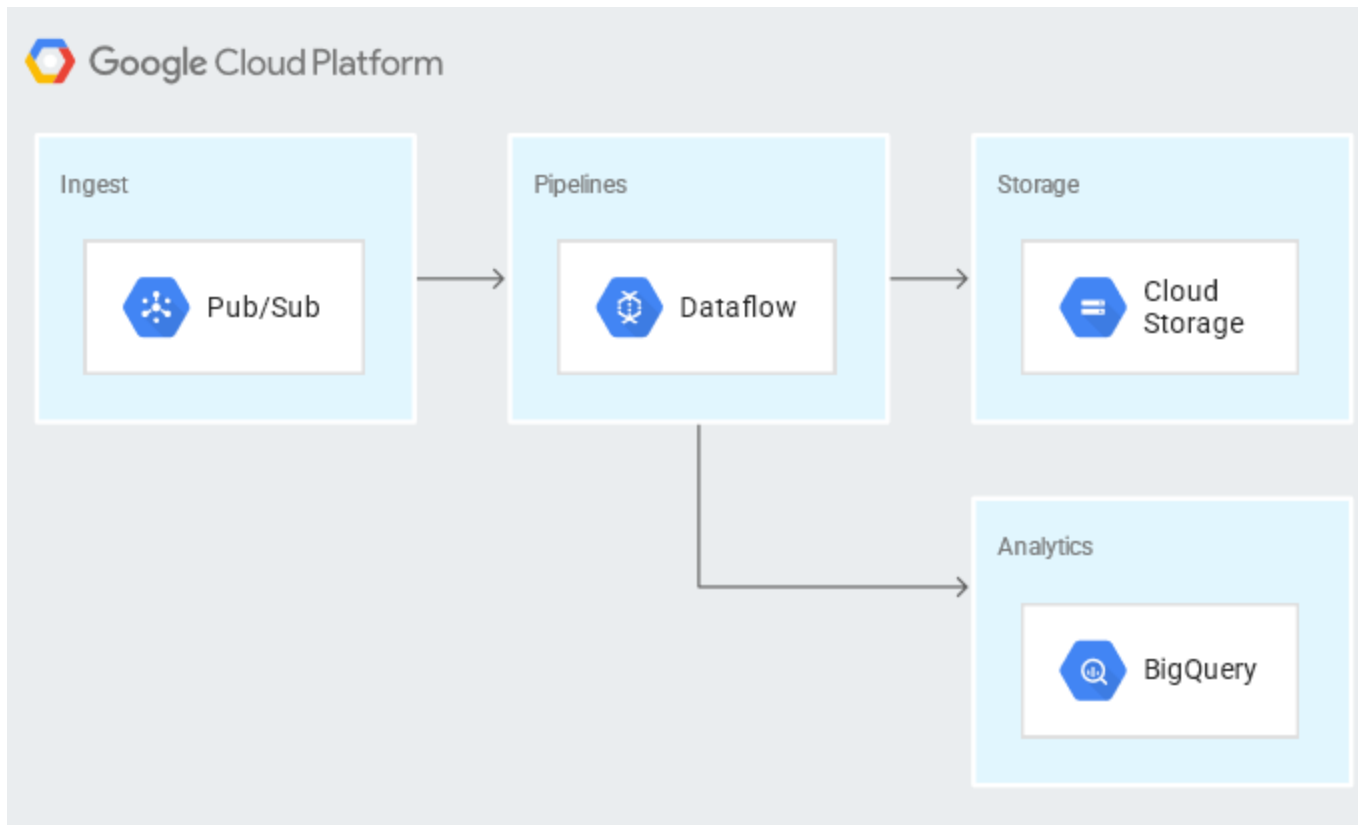
Datafusion and Dataprep can perform the same things. However their execution are different.

- Datafusion create a Spark pipeline and run it on Dataproc cluster
- Dataprep create a Beam pipeline and run it on Dataflow

IMO, Datafusion is more designed for data ingestion from one source to another one, with few transformation. Dataprep is more designed for data preparation (as its name means), data cleaning, new column creation, splitting column. Dataprep also provide insight of the data for helping you in your recipes.

In addition, Beam is a part of [Tensorflow extended](#) and your Data engineer pipeline will be more consistent if you use a tool compliant with Beam

Pipeline



Autre

- **Cloud Composer** pour l'orchestration des pipeline
- **IAM**: Gestion des compte
- **Cloud scheduler _job** Planification des taches
- **TenorsFlow**: Service pour machine Learning
- **Dialog flow** est une IA conversationnelle réaliste associée à des agents virtuels de pointe
- **Speech-to-Text** permet d'intégrer facilement les technologies de reconnaissance vocale Google aux applications en cours de développement

Exemple Envoyez des fichiers audio et recevez-en une transcription

- **Datalab** pour explorer, visualiser, analyser et transformer vos données en toute facilité et de manière interactive à l'aide de langages connus (comme Python et SQL). Les notebooks de présentation, d'exemple et de tutoriels Jupyter préinstallés montrent
- **AutoML** Vision vous permet d'entraîner des modèles de machine learning afin de classifier vos images en fonction de vos propres libellés
- **Compute Engine** Ce service de calcul sécurisé et personnalisable permet de créer et d'exécuter des machines virtuelles sur l'infrastructure de Google
- **IA Natural Language** utiliser pour l'analyser les mots => sentiment, ...

Exo

1. Mettre en place un producteur de data qui publie dans **pub/sub**
2. Lire les data de pub et les inscrire dans **Bigquery** et **Cloud Storage**
3. Faire La Viz(Chiffre d'affaire par ville , Chiffre d'affaire par type de produit et un map pour chaque ville du nombre de commande) avec **Data Studio** en utilisant Bigquery comme datasource

Exemple de message:

```
{"nomclient":"diattara","prix":15,"ville":"Marseille","typeproduit":"boisson"}
```

Dans bigquery on ne doit pas stocker l'attribut nom client

Dans Cloud Storage on doit avoir la donnée brute avec des fichiers horodatés comme suit
data_\${yyyyMMdd_hhmmss}.json et des records mergés par paquet de 10000 mais pour tester la solution utiliser par paquet de 2