



المعهد الوطني للبريد والمواصلات
ⵎⴰⵔⴰⵏⵉⵏⵓ ⵏ ⵉⵎⵓⵔⵓⵏⵉⵔ ⵏ ⵉⵎⵓⵔⵓⵏⵉⵔ
Institut National des Postes et Télécommunications

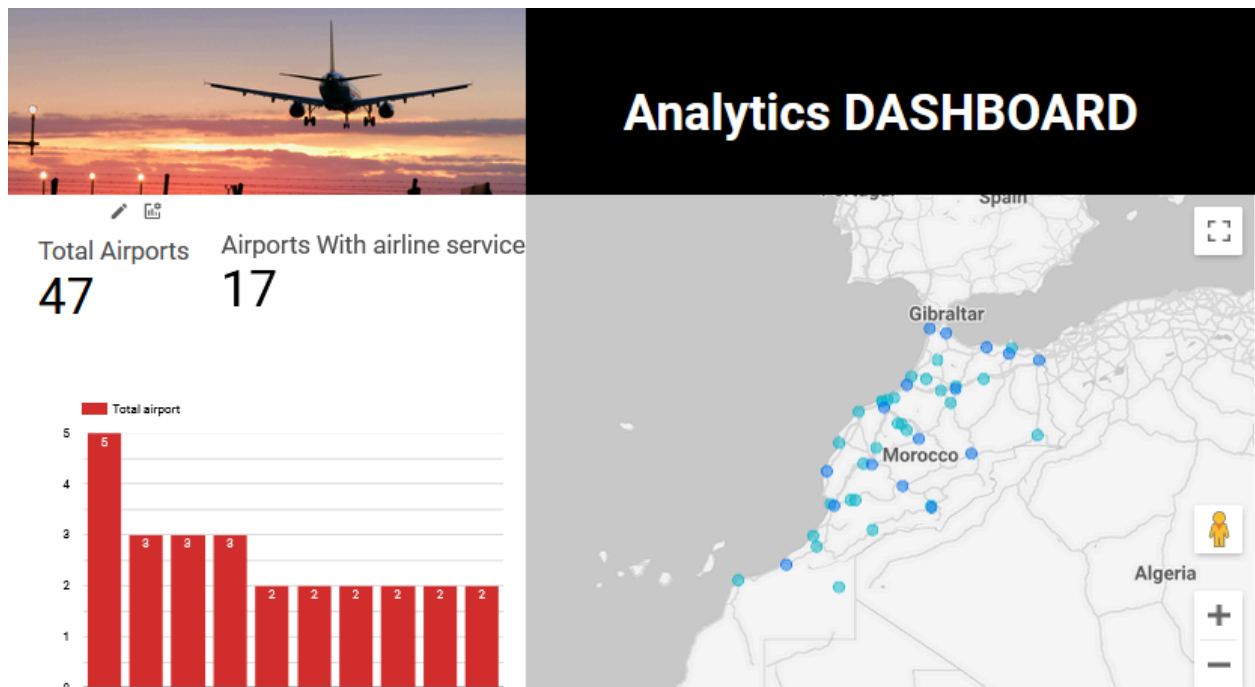


agence nationale de réglementation
des télécommunications
الوكالة الوطنية لتقنين المواصلات
ⵜⴰⵎⴰⵔⴰⵏⵉⵏⵓ ⵜⴰⵎⴰⵔⴰⵏⵉⵏⵓ ⵏ ⵉⵎⵓⵔⵓⵏⵉⵔ ⵏ ⵉⵎⵓⵔⵓⵏⵉⵔ

Projet de Data Warehousing et Dashboarding

Présenté par : **Anas CHAIRI**

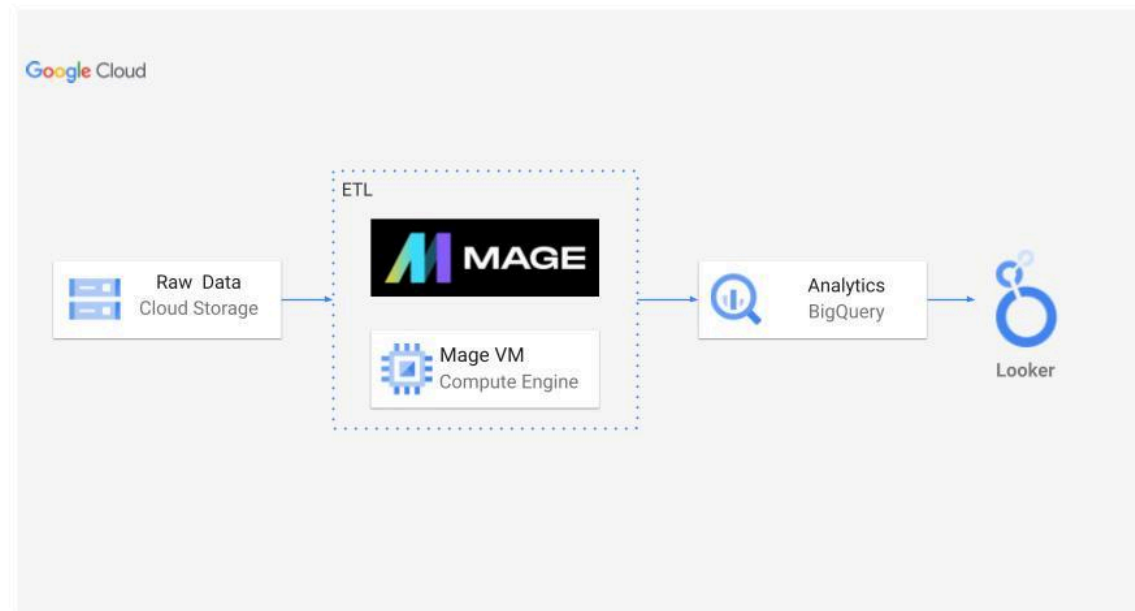
Encadré par : **Mr.Amine Baina**



Introduction

The goal of this project is to perform data analytics on Moroccan Dataset using various tools and technologies, including GCP Storage, Python, Compute Instance, Mage Data Pipeline Tool, BigQuery, and Looker Studio.

Architecture



Google Cloud Platform (GCP) offre une pléthore de services de pointe pour répondre aux besoins informatiques modernes. Avec une infrastructure évolutive et des outils de gestion puissants, GCP permet aux entreprises de déployer, de gérer et de scaler leurs applications avec une agilité remarquable. Sa polyvalence offre des solutions pour le stockage, le calcul, l'analyse de données et bien plus encore, faisant de GCP un acteur majeur dans le domaine du cloud computing.

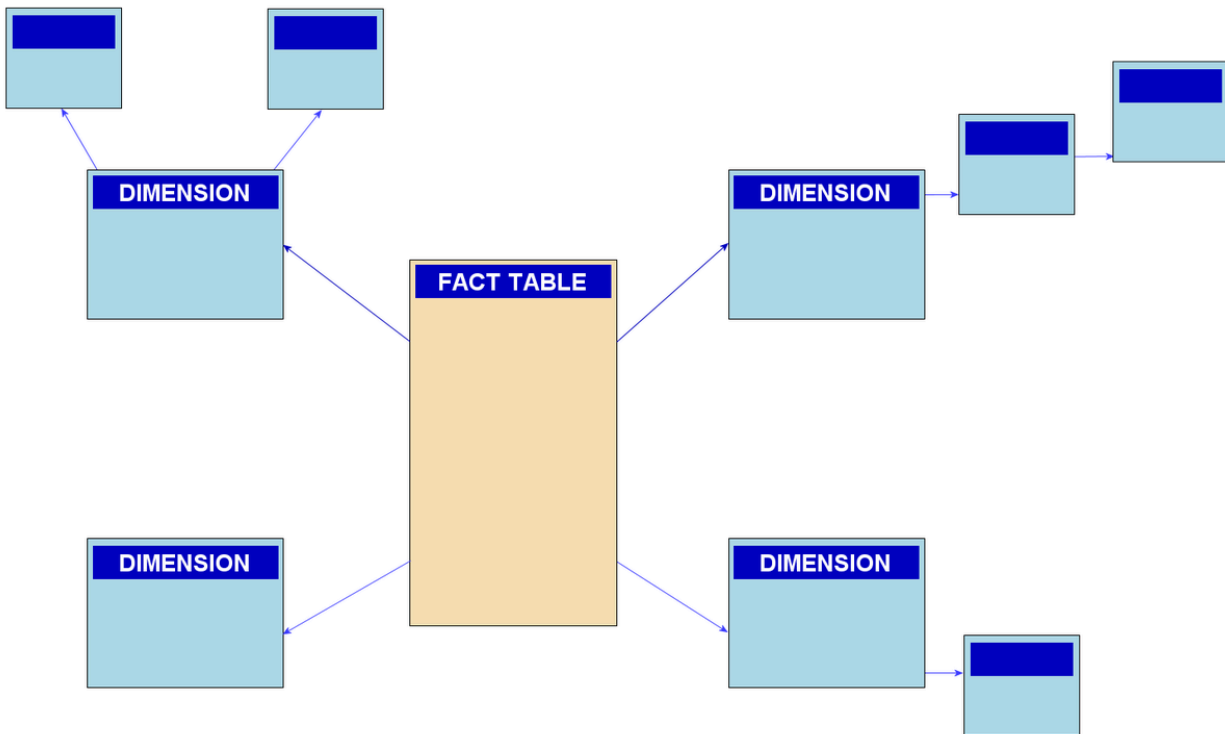
BigQuery, au sein de GCP, est une solution d'entrepôt de données entièrement gérée qui se démarque par sa rapidité et sa capacité à gérer des ensembles de données massives. Cette plateforme offre des fonctionnalités de requêtage interactif et permet une analyse de données en temps réel, offrant ainsi aux entreprises des informations précieuses pour prendre des décisions stratégiques éclairées.

Mage ETL est un outil d'extraction, de transformation et de chargement de données qui permet aux entreprises de gérer et de manipuler efficacement de grands ensembles de données. Son principal objectif est de simplifier et d'automatiser le processus de préparation des données avant leur analyse.

Enfin, **Looker Studio** est une plateforme d'analyse de données qui permet aux utilisateurs de visualiser, d'explorer et de partager des informations exploitables à partir de données variées. Intégrée avec GCP, cette solution offre des outils puissants pour créer des tableaux de bord

interactifs, des visualisations personnalisées et des rapports dynamiques, aidant les entreprises à transformer leurs données en connaissances exploitables.

Data Modeling



Le modèle de base de données Snowflake est une architecture de base de données cloud conçue pour offrir une approche novatrice et flexible pour le stockage et l'analyse de données. Cette solution se démarque par sa conception unique, offrant une séparation entre le stockage des données, le calcul et l'accès aux données, ce qui permet une évolutivité et une performance exceptionnelles.

Snowflake adopte une architecture multi-cluster, ce qui signifie qu'il peut mettre à l'échelle les ressources de calcul de manière dynamique pour répondre aux besoins de charge de travail variables. Cette approche permet une utilisation efficace des ressources, car les clusters de calcul peuvent être dimensionnés indépendamment du stockage des données, offrant ainsi une flexibilité et une agilité significatives.

Une autre caractéristique clé de Snowflake est sa capacité à prendre en charge plusieurs charges de travail concurrentes sans compromettre les performances. En permettant l'isolation des ressources pour chaque requête ou charge de travail, Snowflake garantit que les opérations ne se gênent pas mutuellement, assurant ainsi une expérience utilisateur fluide même lors de l'exécution de tâches complexes.

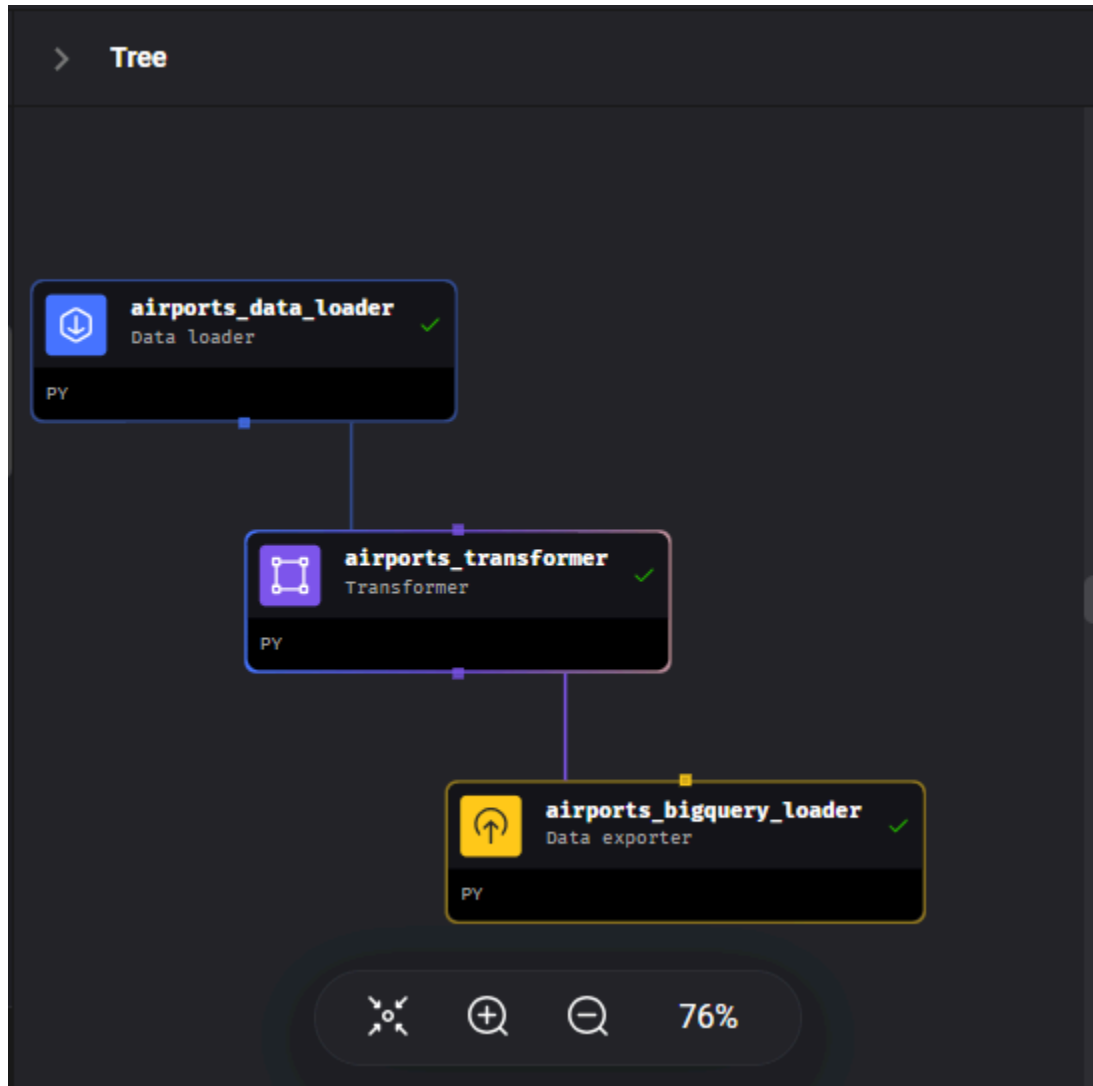
De plus, Snowflake offre des fonctionnalités intégrées de gestion de données, y compris la gestion des versions, la sécurité avancée, ainsi que des fonctionnalités de partage et de collaboration. Son modèle de données semi-structurées permet de stocker et d'analyser des données semi-structurées et non structurées aux côtés des données relationnelles, offrant ainsi une approche unifiée pour l'analyse de données variées.

Notre **Fact table**

	name	ident	scheduled_service	score	type_id	city_id	coordinates_id	datetime_id
0	Menara Airport	GMMX	1	750	1	1	1	1
1	Al Massira Airport	GMAD	1	750	1	6	3	3
2	Tangier Ibn Battuta Airport	GMTT	1	500	1	9	4	4
3	Rabat-Salé Airport	GMME	1	750	1	10	5	5
4	Fes Saïss International Airport	GMFF	1	750	1	11	6	6
5	Oujda Angads Airport	GMFO	1	750	1	13	7	7
6	Ouarzazate Airport	GMMZ	1	750	1	14	8	8
7	Essaouira-Mogador Airport	GMMI	1	500	1	15	9	9
8	Nador Al Aaroui International Airport	GMMW	1	750	1	16	10	10
9	Cherif Al Idrissi Airport	GMTA	1	750	1	18	11	11
10	Moulay Ali Cherif Airport	GMFK	1	750	1	19	12	12
11	Tan Tan Airport	GMAT	1	750	1	20	15	15
12	Sania Ramel Airport	GMTN	1	500	1	21	16	16
13	Kenitra Air Base	GMMY	0	250	1	23	17	17

ETL Pipeline

Ma pipeline sur Mage ETL (hébergé sur GCP Virtual Machine)

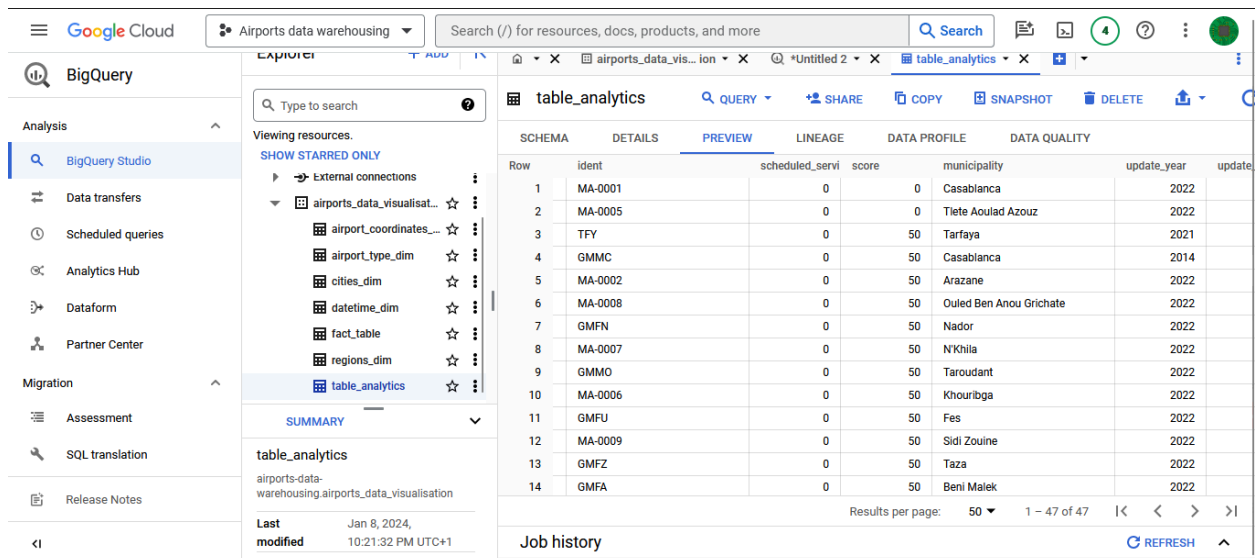


Un pipeline ETL (Extraction, Transformation, Loading) est un processus crucial dans le domaine de l'analyse de données, où les données sont extraites de différentes sources, transformées pour répondre aux besoins spécifiques, puis chargées dans un entrepôt de données ou une base de données pour l'analyse ultérieure.

1. **Extraction (Extraction)** : Ce processus implique la collecte de données à partir de sources multiples telles que des bases de données, des fichiers plats, des API, des flux de données en temps réel, etc. Cette étape vise à rassembler toutes les informations pertinentes pour l'analyse à venir.

2. **Transformation (Transformation)** : Une fois que les données sont extraites, elles peuvent nécessiter des modifications pour être utilisables. Cela peut inclure le nettoyage des données, la normalisation des formats, la fusion de différentes sources, la suppression des doublons, l'enrichissement avec des métadonnées, etc. Cette phase est souvent la plus complexe et nécessite des opérations de traitement de données avancées.
3. **Loading (Chargement)** : Une fois les données transformées, elles sont chargées dans une base de données ou un entrepôt de données pour l'analyse ultérieure. Ce processus peut impliquer le chargement de grandes quantités de données de manière incrémentielle ou en lots réguliers.

Data Analyze



The screenshot displays the Google Cloud BigQuery console. On the left, a sidebar lists navigation options like 'Analysis', 'Data transfers', 'Scheduled queries', 'Analytics Hub', 'Dataform', 'Partner Center', 'Migration', 'Assessment', 'SQL translation', and 'Release Notes'. The main area is divided into a 'Viewing resources' pane on the left and a 'table_analytics' preview pane on the right. The 'Viewing resources' pane shows a tree of external connections and tables, with 'table_analytics' selected. The 'table_analytics' preview pane shows a table with columns: 'ident', 'scheduled_servi', 'score', 'municipality', 'update_year', and 'update'. The table contains 14 rows of data. Below the table, there is a 'Job history' section with a 'REFRESH' button.

Row	ident	scheduled_servi	score	municipality	update_year	update
1	MA-0001	0	0	Casablanca	2022	
2	MA-0005	0	0	Tlete Aoulad Azouz	2022	
3	TFY	0	50	Tarfaya	2021	
4	GMMC	0	50	Casablanca	2014	
5	MA-0002	0	50	Arazane	2022	
6	MA-0008	0	50	Ouled Ben Anou Grichate	2022	
7	GMFN	0	50	Nador	2022	
8	MA-0007	0	50	NKhila	2022	
9	GMMO	0	50	Taroudant	2022	
10	MA-0006	0	50	Khouribga	2022	
11	GMFU	0	50	Fes	2022	
12	MA-0009	0	50	Sidi Zouine	2022	
13	GMFZ	0	50	Taza	2022	
14	GMFA	0	50	Beni Malek	2022	

L'analyse de données sur BigQuery offre une plateforme puissante pour explorer, interroger et analyser des ensembles de données massifs de manière rapide et évolutive. Voici quelques points clés à considérer :

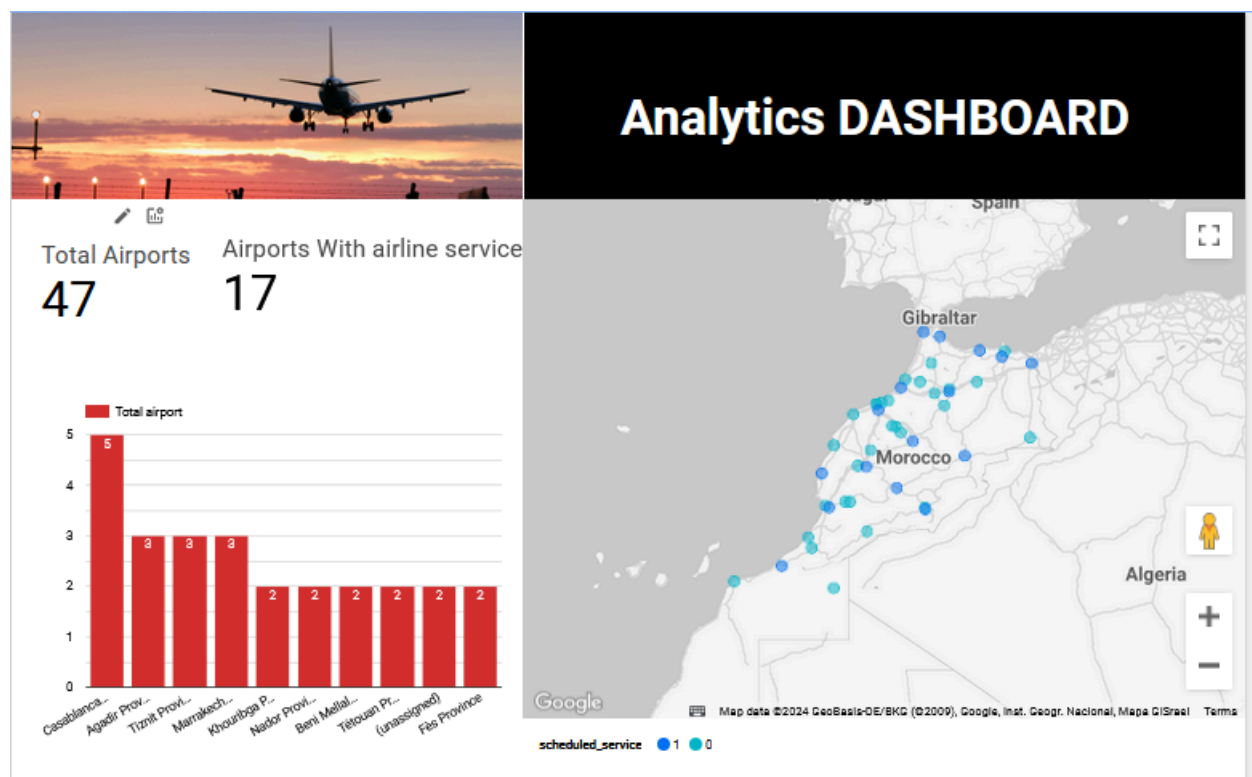
1. **Traitement rapide des requêtes** : BigQuery est conçu pour permettre des requêtes SQL rapides, même sur de grands ensembles de données. Il utilise une architecture distribuée qui répartit automatiquement les requêtes sur plusieurs nœuds de traitement, ce qui permet des temps de réponse très rapides pour des analyses complexes.
2. **Extensibilité et évolutivité** : Cette plateforme est hautement évolutive, capable de gérer des pétaoctets de données, ce qui signifie qu'elle peut s'adapter aux besoins changeants en matière de volume de données sans sacrifier les performances.
3. **Intégration aisée avec d'autres outils** : BigQuery s'intègre facilement à d'autres outils d'analyse et de visualisation de données tels que Data Studio, Looker, Tableau, etc.,

permettant ainsi aux utilisateurs de créer des visualisations et des rapports à partir des résultats de leurs requêtes.

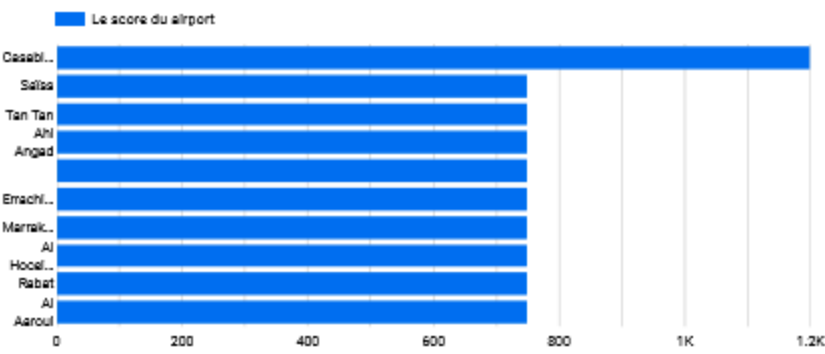
4. **Fonctionnalités avancées d'analyse** : BigQuery propose des fonctionnalités avancées telles que les fenêtres temporelles, les agrégations complexes, les fonctions analytiques avancées et la prise en charge d'algorithmes d'apprentissage automatique via BigQuery ML, permettant aux utilisateurs d'effectuer des analyses approfondies directement dans la plateforme.
5. **Sécurité et contrôle d'accès** : BigQuery offre des fonctionnalités robustes de sécurité et de contrôle d'accès, permettant de gérer les autorisations d'accès aux données de manière fine, assurant ainsi la confidentialité et la protection des données sensibles.
6. **Traitement de données semi-structurées** : BigQuery prend en charge les données semi-structurées telles que JSON, Parquet, Avro, ce qui permet d'analyser des données dans des formats variés sans nécessiter de transformation préalable.

DASHBOARDING

Et Voila



The Highest scored Airports



Airport Quality Evolution



