



Google BigQuery

Yunus Solak
Doğa Yağmur Yılmaz
Banu Özdeveci

PLAN



1) Qu'est-ce que BigQuery ?

2) Comment fonctionne BigQuery ?

3) GDELT

4) Requêtes

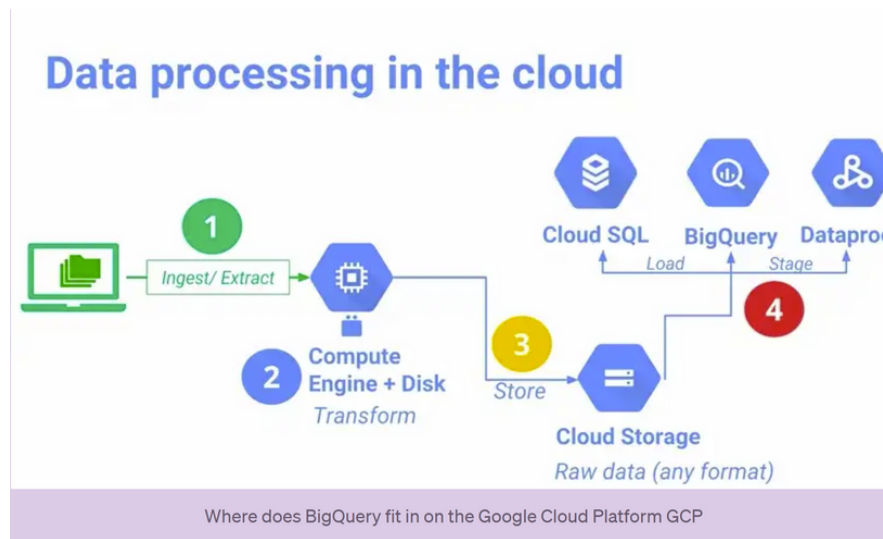
1-yunus

2-banu

3-yunus

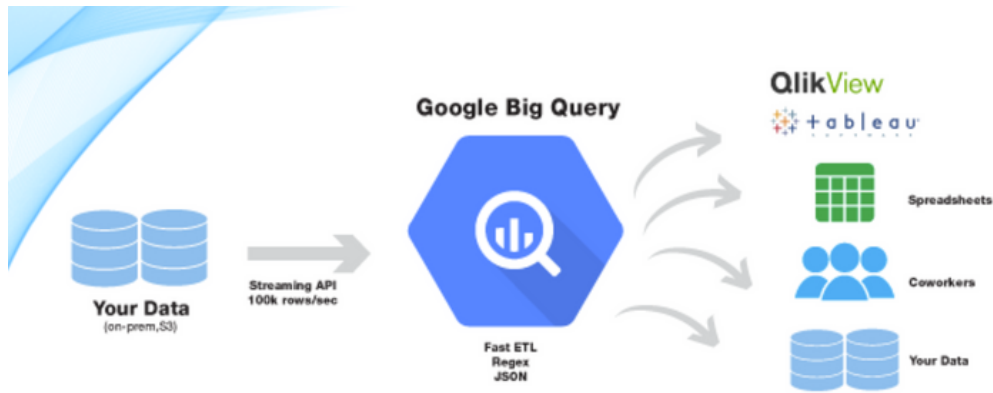
4-doğa

QU'EST-CE QUE C'EST BIGQUERY ?



BigQuery est le puissant service de base de données analytique en nuage de Google, conçu pour les plus grands ensembles de données de la planète. Il permet aux utilisateurs d'exécuter en quelques secondes des requêtes rapides, de type SQL, sur des ensembles de données de plusieurs téraoctets. Évolutif et facile à utiliser, BigQuery vous donne un aperçu en temps réel de vos données.

QU'EST-CE QUE C'EST BIGQUERY ?



BigQuery est un entrepôt de données d'entreprise entièrement géré qui vous aide à gérer et à analyser vos données grâce à des fonctionnalités intégrées telles que l'apprentissage automatique, l'analyse géospatiale et la veille stratégique.

L'architecture sans serveur de BigQuery vous permet d'utiliser des requêtes SQL pour répondre aux plus grandes questions de votre entreprise sans aucune gestion de l'infrastructure.

QU'EST-CE QUE C'EST BIGQUERY ?



Le moteur d'analyse distribué et évolutif de BigQuery vous permet d'interroger des téraoctets en quelques secondes et des pétaoctets en quelques minutes.

BigQuery maximise la flexibilité en séparant le moteur de calcul qui analyse vos données de vos choix de stockage. Vous pouvez stocker et analyser vos données dans BigQuery ou utiliser BigQuery pour évaluer vos données là où elles se trouvent. Les requêtes fédérées vous permettent de lire des données provenant de sources externes, tandis que le streaming prend en charge les mises à jour continues des données. Des outils puissants comme BigQuery ML et BI Engine vous permettent d'analyser et de comprendre ces données.

COMMENT FONCTIONNE BIGQUERY?

STOCKAGE

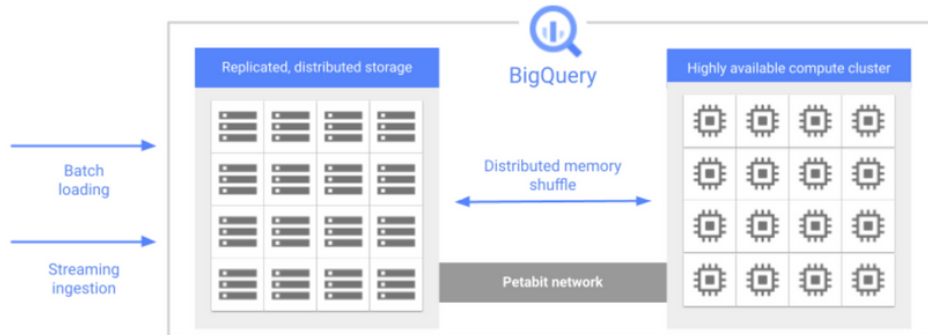
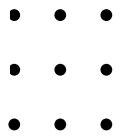


Figure 1. Architecture de BigQuery.

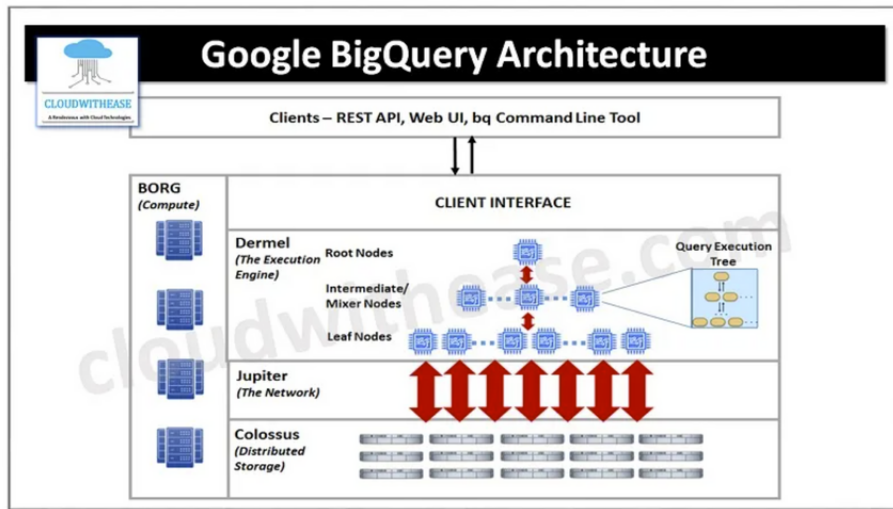


Le stockage BigQuery est optimisé pour l'exécution de requêtes sur de grands ensembles de données.

L'une des caractéristiques principales de l'architecture de BigQuery est la séparation du stockage et du calcul.

Cela permet à BigQuery de faire évoluer le stockage et le calcul de manière indépendante, en fonction de la demande.

STOCKAGE



Lorsque vous exécutez une requête, le moteur de requête distribue le travail entre plusieurs nœuds en parallèle, qui analysent les tables pertinentes dans le stockage, traitent la requête, puis montrent les résultats.

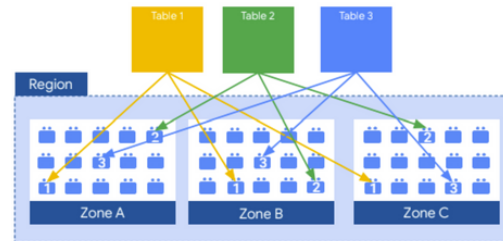
BigQuery exécute les requêtes entièrement en mémoire, à l'aide d'un réseau pétabits pour garantir que les données se déplacent extrêmement rapidement vers les nœuds.

QUELQUES FONCTIONNALITÉS CLÉS DU STOCKAGE BIGQUERY



GÉRÉE

DURABLE



Voici quelques fonctionnalités clés du stockage BigQuery:

BigQuery Storage est un service entièrement géré. Vous n'avez pas besoin de provisionner des ressources de stockage ou de réserver des unités de stockage. BigQuery alloue automatiquement du stockage pour vous lorsque vous chargez des données dans le système.

Vous ne payez que pour la quantité de stockage que vous utilisez.

*Bigquery est durable.

le stockage bigquery a une durabilité annuelle de presque cent pour cent.

BigQuery réplique vos données sur plusieurs zones de disponibilité pour vous protéger contre la perte de données .

QUELQUES FONCTIONNALITÉS CLÉS DU STOCKAGE BIGQUERY

CRYPTÉ



EFFICACE

BigQuery aussi crypte automatiquement toutes les données avant qu'elles ne soient écrites sur le disque. Vous pouvez fournir votre propre clé de cryptage ou laisser Google gérer la clé de cryptage.

Le stockage BigQuery utilise un format de codage efficace optimisé pour le travail analytiques.

DISPOSITION DE STOCKAGE

Maintenant on va parler du dispositif de stockage.

La majorité des données que vous stockez dans BigQuery sont des données de table.

Vous êtes facturé pour le stockage que vous utilisez pour ces ressources.

Une table externe a un schéma de table, tout comme une table standard, mais la définition de la table pointe vers le stockage de données externe.

Les tables externes sont un type spécial de table, dans lequel les données résident dans un stockage de données externe à BigQuery, tel que le stockage en cloud. Une table externe a un schéma de table, tout comme une table standard, mais la définition de la table pointe vers le stockage de données externe.

product	quantity	warehouse
dryer	30	warehouse #2
microwave	20	warehouse #1
top load washer	10	NULL
dishwasher	30	warehouse #3
...

Figure 2. Row-oriented format.

De nombreux systèmes de bases de données traditionnels stockent leurs données dans un format orienté ligne,

ce qui signifie que les lignes sont stockées ensemble, les champs de chaque ligne apparaissant séquentiellement sur le disque.

Les bases de données orientées lignes sont efficaces pour rechercher des enregistrements individuels. Cependant, ils peuvent être moins efficaces pour exécuter des fonctions analytiques sur de nombreux enregistrements, car le système doit lire chaque champ lors de l'accès à un enregistrement.

product	quantity	warehouse
dryer	30	warehouse #2
microwave	20	warehouse #1
top load washer	10	NULL
dishwasher	30	warehouse #3
...

Figure 3. Column-oriented format.

Donc BigQuery stocke les données de la table au format colonne, ce qui signifie qu'il stocke chaque colonne séparément. Les bases de données orientées colonnes sont particulièrement efficaces pour analyser des colonnes individuelles sur un ensemble de données entier.

Les bases de données orientées colonnes sont optimisées pour les charges de travail analytiques qui regroupent les données sur un très grand nombre d'enregistrements. Souvent, une requête analytique n'a besoin de lire que quelques colonnes d'une table. Par exemple, si vous souhaitez calculer la somme d'une colonne sur des millions de lignes, BigQuery peut lire les données de cette colonne sans lire chaque champ de chaque ligne.



L'ADMINISTRATION DE BIGQUERY



GÉRER DES RESSOURCES, TELLES QUE DES PROJETS, DES ENSEMBLES DE DONNÉES ET DES TABLES.



SÉCURISER DES RESSOURCES, AFIN QUE L'ACCÈS SOIT LIMITÉ AUX COMPTES PRINCIPAUX QUI EN ONT BESOIN.

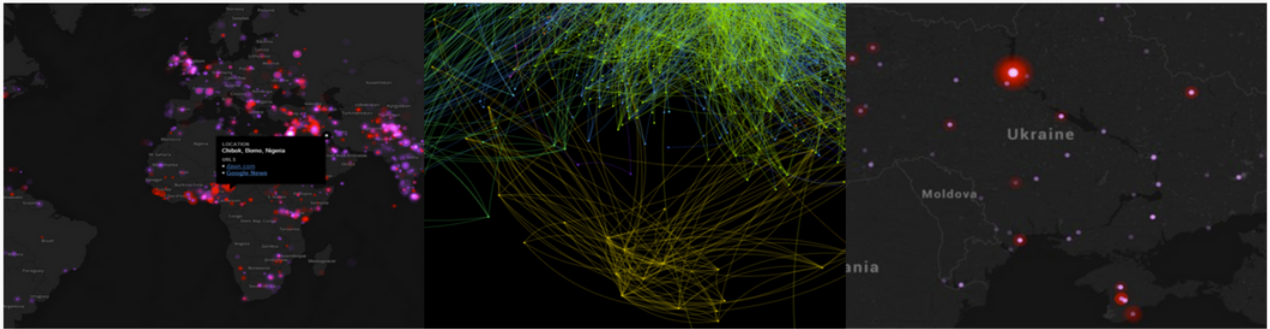


SURVEILLER LES RESSOURCES, Y COMPRIS LES QUOTAS, LES TÂCHES ET L'UTILISATION.

Les administrateurs de BigQuery effectuent généralement les types de tâches suivants



GDELT: UNE BASE DE DONNÉES MONDIALE DE LA SOCIÉTÉ

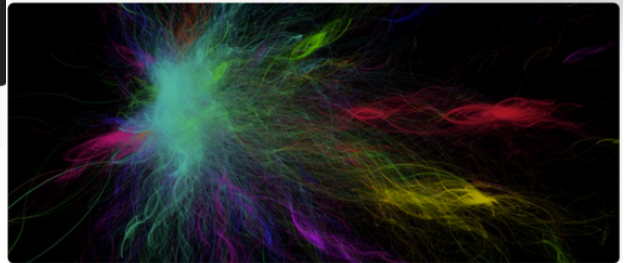


Soutenu par Google Jigsaw, le projet GDELT surveille l'actualité mondiale diffusée par les médias audiovisuels, la presse écrite et le web, dans presque tous les coins de tous les pays et dans plus de 100 langues. Il identifie les personnes, les lieux, les organisations, les thèmes, les sources, les émotions, les chiffres, les citations, les images et les événements qui animent notre société mondiale à chaque seconde de chaque jour, créant ainsi une plateforme ouverte et gratuite de calcul sur le monde entier.

GDELT



GDELT Event Database



GDELT Global Knowledge Graph

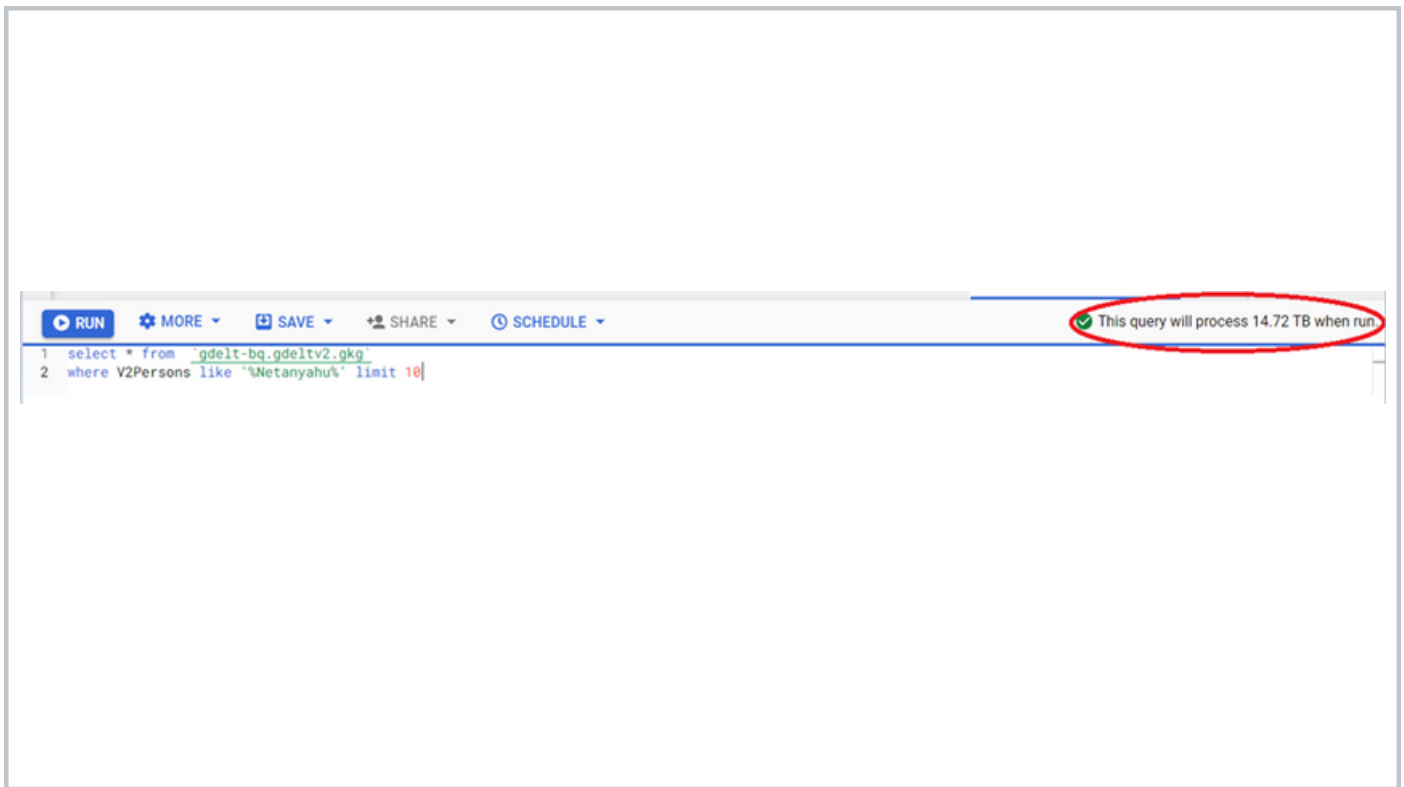
La surveillance de la quasi-totalité des médias d'information du monde n'est qu'un début - même la plus grande équipe d'humains ne pourrait pas commencer à lire et à analyser les milliards et milliards de mots et d'images publiés chaque jour. GDELT utilise certains des algorithmes informatiques les plus sophistiqués au monde, conçus sur mesure pour les médias d'information mondiaux, fonctionnant sur "l'un des réseaux de serveurs les plus puissants de l'univers connu", ainsi que certains des algorithmes d'apprentissage profond les plus puissants au monde, pour créer un enregistrement informatique en temps réel de la société mondiale qui peut être visualisé, analysé, modélisé, examiné et même prévu.

REQUÊTES

*Editor		*Unsaved query 2			
RUN		SAVE		SHARE	
		SCHEDULE		MORE	
1 select * from `gdelv-bq.gdelv2.gkg` limit 10				This query will process 14.71 TB when run.	
Query results		SAVE RESULTS		EXPLORE DATA	
JOB INFORMATION		RESULTS		EXECUTION DETAILS	
		JSON		EXECUTION GRAPH	
				PREVIEW	
V2Themes		Locations		V2Locations	
1 null		3#Miami, Florida, United States#US#USFL#25.7743#-80.1937#295004;3#Boston, Massachusetts, United States#US#USMA#42.3584#-		3#Boston, Massachusetts, United States#US#USMA#MA025#4 2.3584#-71.0598#617565#2 829;3#Boston,	
				Persons	
				Kathryn Bigelow,donald trump;chris rock;john krasinski;usain bolt;david ortiz;michael bloomberg;missy	
				V2Persons	
				Kathryn Bigelow,1349;Donald Trump,101;Chris Rock,2416;John Krasinski,2985;Usain Bolt,3408;David	
				Organization	
				google;facel	
2 ;TAX_FNCACT_SPEAKER,41;TAX_RELIGION_ISLAMIC,268;TAX_ETHNICITY_ALGERIAN,166;TAX_WORLDLANGUAGES_ALGERIAN,166;		4#Ouagadougou, Kadiogo, Burkina Faso#UV#UV53#12.3703#-1.52472#-1721728;1#Algeria#AG#AG28#3#AG		1#Algerian#AG#AG#28#3#AG #166;4#Ouagadougou, Kadiogo, Burkina Faso#UV#UV53#154478#12.3703#-1.52472#-1721728#11	
				slimane chenine	
				Slimane Chenine,94	
				speaker of t assembly;al parliament;c islamic coop	
3 EPU_ECONOMY,299;EPU_ECONOMY_HISTORIC,299;TAX_ETHNICITY_GERMAN,200;TA		1#Germany#GM#GM#51.5#10.5#GM;4#Budapest, Budapest, Hungary#HU#HU05#47.5#19.08		4#Budapest, Budapest, Hungary#HU#HU05#17514# 47.5#19.0833#-850553#272;	
				nicholas pongratz;mercedes-be...	
				Nicholas Pongratz,66;Mercede...	
				null	

“gdelv-bq.gdelv2.gkg” est une table de l'ensemble de données GDEL (Global Database of Events, Language, and Tone), qui est disponible sur Google BigQuery.

La table GKG de l'ensemble de données GDEL contient des enregistrements de la couverture médiatique mondiale, avec un enregistrement pour chaque article ou rapport unique. Chaque enregistrement comprend des informations sur l'article, telles que la date de publication, la langue dans laquelle il a été rédigé, le ton de l'article et le lieu mentionné dans l'article. Il comprend également une liste d'entités nommées mentionnées dans l'article, telles que des personnes, des organisations et des lieux. La table GKG est mise à jour quotidiennement avec les dernières nouvelles du monde entier.



Ici, on a vu que on a chance de travailler avec le big data. au niveau de terebaytes

cette requete dure au peupres 30 seconds. on pense que cette plateforme est tres performante.

dans cette requete on veut filtrer v2persons avec le prime minister de Israel "netenyahu" avec une limite

The screenshot shows the KQL Playground interface in a web browser. At the top, there's a toolbar with icons for home, search, editor, unsaved queries, and a plus sign. Below the toolbar are buttons for RUN, SAVE, SHARE, SCHEDULE, and MORE. A status bar indicates "Query completed." with a green checkmark icon.

```

1 select * from `gdelt-bg.gdeltv2.gkg`
2 where V2Persons like '%Netanyahu%' limit 10

```

Below the query editor, the "Query results" section is active. It features tabs for JOB INFORMATION, RESULTS, JSON, EXECUTION DETAILS, EXECUTION GRAPH, and PREVIEW. The RESULTS tab displays a table with the following columns: Row, V2Themes, Locations, V2Locations, Persons, V2Persons, and Organizations.

Row	V2Themes	Locations	V2Locations	Persons	V2Persons	Organizations
1	ELECTION,1255;TAX_FNCACT_L EADERS,32;TAX_FNCACT_LEAD ERS,643;LEADER,192;LEADER,48 2;	4#Moscow, Moskva, Russia#RS#RS48#55.7522#3 7.6156#-2960561;4#Jerusale m, Israel (General), Israel#IS#IS00#31.7667#35.	1#Russia#RS#RS##60#100# RS#202;1#Russia#RS#RS# 60#100#RS#2606;1#Russia# RS#RS##60#100#RS#2982;1 #Russia#RS#RS##60#100#R	benny gantz,moshe kantor,donald trump,andrzej duda,europe jews,benjamin netanyahu,adolf hitler,mike pence,bashar al-assad,sergei	Jews,1359,Benjamin <u>Netanyahu</u> ,629,Adolf Hitler,3273,Mike Pence,252,Mike Pence,2081,Bashar Al-	european jewish
2	ARMEDCONFLICT,832;EPU_C ATS_NATIONAL_SECURITY,83 2;TAX_FNCACT_PRIME_MINI STER,1684;TAX_FNCACT_LE ADERS,489;TAX_FNCACT_LE	4#Jerusalem, Israel (General), Israel#IS#IS00#31.7667#35. 2333#-797092;1#United States#US#US#39.828175#-	1#Britain#UK#UK##54#-4#U K#1795;4#Gaza, Israel (General), Israel#IS#IS00#18315#31.41 67#34.3333#-797156#44.4#	mark heinrich,ali sawaf,vladimir putinjon alterman,mike pompeo,stephen farrell,mahmoud abbas,benny	Zomlot,1750,Angus Macswan,5482,Benjamin <u>Netanyahu</u> ,1703,Diab Al- Louh,953,Steven Holland,5426	u s embassy,wh jordan valley,ara league,reuters,u
3	CRISISLEX_T11_UPDATESY MPATHY,1709;CRISISLEX_T1 1_UPDATESYMPATHY,1864;	1#Israel#IS#IS#31.5#34.75#IS	1#Israel#IS#IS#31.5#34.75 #IS#6;1#Israel#IS#IS#31.5# 34.75#IS#526;1#Israel#IS#IS	naftali bennett,avigdor lieberman,benjamin netanyahu;benny gantz kahol	Naftali Bennett,1035,Avigdor Lieberman,1005,Benjamin <u>Netanyahu</u> ,977,Benny Gantz	supreme court;i

Au resultat, on a vu qu'il existe une chaos a niveau de enregistrement.

les noms des personnes et le nombre de fois qu'elles apparaissent dans le texte sont adjacents.

on voit cette probleme dans le v2themes aussi.

<div> <div> <div>RUN</div> <div>SAVE</div> <div>SHARE</div> <div>SCHEDULE</div> <div>MORE</div> </div> <div> <div> <div>✓</div> <div>This query will process 1.34 TB when run.</div> </div> </div> </div>	
<div> <div>1</div> <div>SELECT V2Themes from `gdelt-bq.gdeltv2.gkg` where DATE>20150302000000 and DATE < 20150304000000 and V2Persons like '%Netanyahu%' limit 10</div> </div>	
<div> <div>Processing location: US</div> <div>Press Alt+F1 for Accessibility Options</div> </div>	
<div> <div>Query results</div> <div> <div>SAVE RESULTS</div> <div>EXPLORE DATA</div> </div> </div>	
<div> <div>JOB INFORMATION</div> <div>RESULTS</div> <div>JSON</div> <div>EXECUTION DETAILS</div> <div>EXECUTION GRAPH</div> <div>PREVIEW</div> </div>	
Row	V2Themes
1	LEADER,410;LEADER,1563;LEADER,2153;LEADER,2553;TAX_FNCACT_PRESIDENT,410;TAX_FNCACT_PRESIDENT,1563;TAX_FNCACT_PRESIDENT,21
2	ECON_WORLD CURRENCIES_DOLLAR,2916;TAX_RELIGION_SHAKERS,2215;TAX_ETHNICITY_ASIAN,2725;TAX_FNCACT_ECONOMISTS,2440;MAN
3	MEDIA_MSM,2086;TAX_POLITICAL_PARTY_REPUBLICANS,1546;TAX_RELIGION_ISLAM,

Par exemple, imaginons créer un histogramme des principaux thèmes associés au Premier ministre israélien Benjamin Netanyahu lors de sa visite au Congrès américain le 3 mars 2015. Demandons une liste des thèmes qui apparaissent dans chaque article dont le nom est mentionné.

C'est facile dans BigQuery.

voila on voit le chaos precisement.

The screenshot shows the Google BigQuery web interface. At the top, there are buttons for RUN, SAVE, SHARE, SCHEDULE, and MORE. A status message indicates the query will process 1.34 TB. The query text is: `1 SELECT SPLIT(V2Themes, ';') from `gdelt-bq.gdeltv2.gkg` where DATE>20150302000000 and DATE < 20150304000000 and V2Persons like '%Netanyahu%' limit 10`. Below the query, it says 'Processing location: US'. The 'Query results' section is active, showing a table with 13 rows. The table has two columns: 'f0_' and a count. The data rows are as follows:

Row	f0_	Count
1	WB_2462_POLITICAL_VIOLENCE_AND_WAR	185
2	WB_2473_DIPLOMACY_AND_NEGOTIATIONS	2331
3	WB_2471_PEAKEKEEPING	2331
4	TAX_POLITICAL_PARTY_DEMOCRAT	1934
5	TAX_FNCACT_AMBASSADOR	1388
6	TAX_FNCACT_AMBASSADOR	2166
7	MIL_WEAPONS_PROLIFERATION	540
8	GENERAL_GOVERNMENT	1946
9	ACT_MAKESTATEMENT	1641
10	WB_2470_PEAKE_OPERATIONS_AND_CONFLICT_MANAGEMENT	708
11	WB_2503_WEAPONS_PROLIFERATION_AND_ARMES_CONTROL	708
12	TAX_FNCACT_SENATOR	874

Le problème est que la colonne V2Themes utilise une délimitation imbriquée - chaque citation d'un thème reconnu dans un article est séparée par un point-virgule, et pour chaque citation, le thème et son décalage de caractères dans l'article sont séparés par une virgule.

Tout d'abord, nous utilisons la fonction SPLIT() pour demander à BigQuery de prendre le champ V2Themes, de le séparer par un point-virgule et de le renvoyer sous forme d'enregistrements multiples, un par citation.
mais il y a encore un problème que themes et le counts sont adjacents.

nous avons toujours le décalage de caractères indiqué à la fin de chaque citation de thème dont nous devons séparer. Pour y arriver, Nous disposons de la fonction utile REGEXP_REPLACE().

-> ^ CORRESPOND AU DÉBUT D'UNE CHAÎNE DE CARACTÈRES.

->. CORRESPOND À N'IMPORTE QUEL CARACTÈRE UNIQUE (SAUF UNE NOUVELLE LIGNE).

-> * CORRESPOND À ZÉRO OU PLUSIEURS OCCURRENCES DU CARACTÈRE OU DU MOTIF PRÉCÉDENT.

-> ? CORRESPOND À ZÉRO OU UNE OCCURRENCE DU CARACTÈRE OU DU MOTIF PRÉCÉDENT.

^ correspond au début d'une chaîne de caractères.

. correspond à n'importe quel caractère unique (sauf une nouvelle ligne).

* correspond à zéro ou plusieurs occurrences du caractère ou du motif précédent.

? correspond à zéro ou une occurrence du caractère ou du motif précédent.

[RUN](#) [SAVE](#) [SHARE](#) [SCHEDULE](#) [MORE](#)

This query will process 1.34 TB when run.

```

1 SELECT theme, COUNT(*) as count
2 FROM (
3   SELECT REGEXP_REPLACE(ARRAY_TO_STRING(SPLIT(V2Themes, ','), ', '), r'^.*$', '') as theme
4   FROM `gdelt-bq.gdeltv2.gkg`
5   WHERE DATE > 20150302000000 AND DATE < 20150304000000 AND V2Persons LIKE '%Netanyahu%'
6 )
7 group by theme
8 ORDER BY 2 DESC
9 LIMIT 300

```

Processing location: US

Press Alt+F1 for Accessibility Options.

Query results

[SAVE RESULTS](#) [EXPLORE DATA](#)

JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	theme		count		
1	TAX_POLITICAL_PARTY_REPUBLICANS		2088		
2	LEADER		1217		
3	MEDIA_MSM		989		
4	TERROR		785		
5	TAX_FNCACT_SENATORS		467		
6	GENERAL_GOVERNMENT		452		
7	TAX_FNCACT_DEPUTY		276		

Results per page: 50 1 - 50 of 300

^ correspond au début d'une chaîne de caractères.

. correspond à n'importe quel caractère unique (sauf une nouvelle ligne).

* correspond à zéro ou plusieurs occurrences du caractère ou du motif précédent.

? correspond à zéro ou une occurrence du caractère ou du motif précédent.

The screenshot displays the Google Cloud BigQuery web interface. At the top, there's a toolbar with buttons for RUN, SAVE, SHARE, SCHEDULE, and MORE. A status bar indicates 'Query completed.' Below the toolbar is a SQL editor with the following query:

```
1 SELECT theme, COUNT(*) as count
2 FROM (
3 SELECT REGEXP_REPLACE(ARRAY_TO_STRING(SPLIT(V2Themes, ','), ','), r',.*', '') as theme
4 FROM `gdelt-bq-gdeltv2.gkg`
5 WHERE DATE > 20150302000000 AND DATE < 20150304000000 AND V2Persons LIKE '%Tsipras%'
6 )
7 group by theme
8 ORDER BY 2 DESC
9 LIMIT 300
```

Below the query editor, the 'Query results' section is active, showing a table with two columns: 'theme' and 'count'. The table lists various themes and their corresponding counts, ordered by count in descending order.

Row	theme	count
1	GENERAL_GOVERNMENT	925
2	ECON_DEVELOPMENTORGS_L...	125
3	LEADER	64
4	PRIVATIZATION	43
5	ELECTION	43
6	SOVEREIGNTY	42
7	TAX_ETHNICITY_GERMAN	41
8	TAX_ETHNICITY_GERMAN	41

At the bottom right of the results table, there's a pagination control showing 'Results per page: 50' and '1 - 50 of 98'.

Comparons ces résultats à ceux du Premier ministre grec Alexis Tsipras durant la même période.

Comme prévu, nous observons un ensemble de thèmes très différents, qui reflètent fortement le discours sur l'économie et la crise de la Grèce.


```
1 SELECT theme, COUNT(*) as count
2 FROM (
3 SELECT REGEXP_REPLACE(ARRAY_TO_STRING(SPLIT(V2Themes, ','), ','), r',.*', '') as theme
4 FROM `gdelit-bq.gdelitv2.gkg`
5 WHERE DATE > 20200302000000 AND DATE < 20200304000000 AND V2Persons LIKE '%Erdogan%'
6 )
7 group by theme
8 ORDER BY 2 DESC
9 LIMIT 100
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW
Row	theme	count				
1	GENERAL_GOVERNMENT	859				
2	LEADER	336				
3	ARMEDCONFLICT	307				
4	TERROR	211				
5	TAX_ETHNICITY_TURKISH	186				
6	KILL	123				
7	EPU_POLICY_GOVERNMENT	111				

-Un exemple que nous pouvons comprendre la Turquie et les politiques gouvernementales en 2020. Dans cette requête, nous voyons les nombres "Erdogan" qui sont mentionnés dans différents thèmes.

türkçe söyle

"4#BERLİN, BERLİN, GERMANY#GM#GM16#16538#52.5167#13.4#-1746443#1340"

CELA INDIQUE QUE LE DOCUMENT MENTIONNE BERLİN, ALLEMAGNE

TYPE D'EMPLACEMENT : IL S'AGIT D'UN NOMBRE ENTIER QUI SPÉCIFIÉ LA RÉOLUTION GÉOGRAPHIQUE DE L'EMPLACEMENT.

1 -> INDIQUE QUE L'EMPLACEMENT EST UN PAYS,

2 ->INDIQUE QU'IL S'AGIT D'UN ÉTAT AMÉRICAIN,

3-> INDIQUE QU'IL S'AGIT D'UNE VILLE OU D'UN POINT DE REPÈRE AMÉRICAIN,

4-> INDIQUE QU'IL S'AGIT D'UNE VILLE OU D'UN POINT DE REPÈRE HORS DES ÉTATS-UNIS,

5-> INDIQUE QU'IL S'AGIT D'UNE DIVISION ADMINISTRATIVE HORS DES ÉTATS-UNIS (À PEU PRÈS L'ÉQUIVALENT D'UN ÉTAT AMÉRICAIN).

une entrée typique pourrait ressembler a ecriture au debut du page

la propriete que nous interressons est " type d'emplacement"

brevement:

1 indique le pays et les autres nombres sont pour indiquer plus precisement la location

on va utiliser cette propriete pour un extra filtration


```

1 SELECT location, COUNT(*)
2 FROM(
3 SELECT REGEXP_EXTRACT(ARRAY_TO_STRING(SPLIT(V2Locations,','), ','),r'^[2-5]#(.*?)#') as location
4 FROM `gdelit-bq.gdelitv2.gkg`
5 WHERE DATE > 20150302000000 AND DATE < 20150304000000 AND V2Persons LIKE '%Tsipras%'
6 )
7 WHERE location is not null
8 group by location
9 ORDER BY 2 DESC
10 LIMIT 100

```

Row	location	f0_
1	Madrid, Madrid, Spain	414
2	Guindo, Castilla Y LeóP, Spain	83
3	Brussels, Bruxelles-Capitale, Belgium	50
4	Samara, Samarskaya Oblast', Russia	28
5	Dublin, Dublin, Ireland	21
6	Texas, United States	20
7	Athens, AttikiR, Greece	19
8	Bruxelles, Bruxelles-Capitale, Belgium	15
9	Mediterranean Sea, Oceans (General), Oceans	14
10	Toronto, Ontario, Canada	14
11	Lemnos, Perifereia Voreiou Aiqaiou, Greece	13

Comme on peut s'y attendre, un grand nombre des premiers résultats sont des localisations au niveau national, comme "Grèce" et "Espagne", qui sont probablement moins intéressantes pour de nombreuses requêtes. Au lieu de cela, la requête suivante modifie légèrement le REGEXP_EXTRACT pour ajouter "[2-5]" au début de la requête au lieu du ".* ?" original. (tout ce qui a une résolution plus élevée qu'une correspondance au niveau du pays).

```

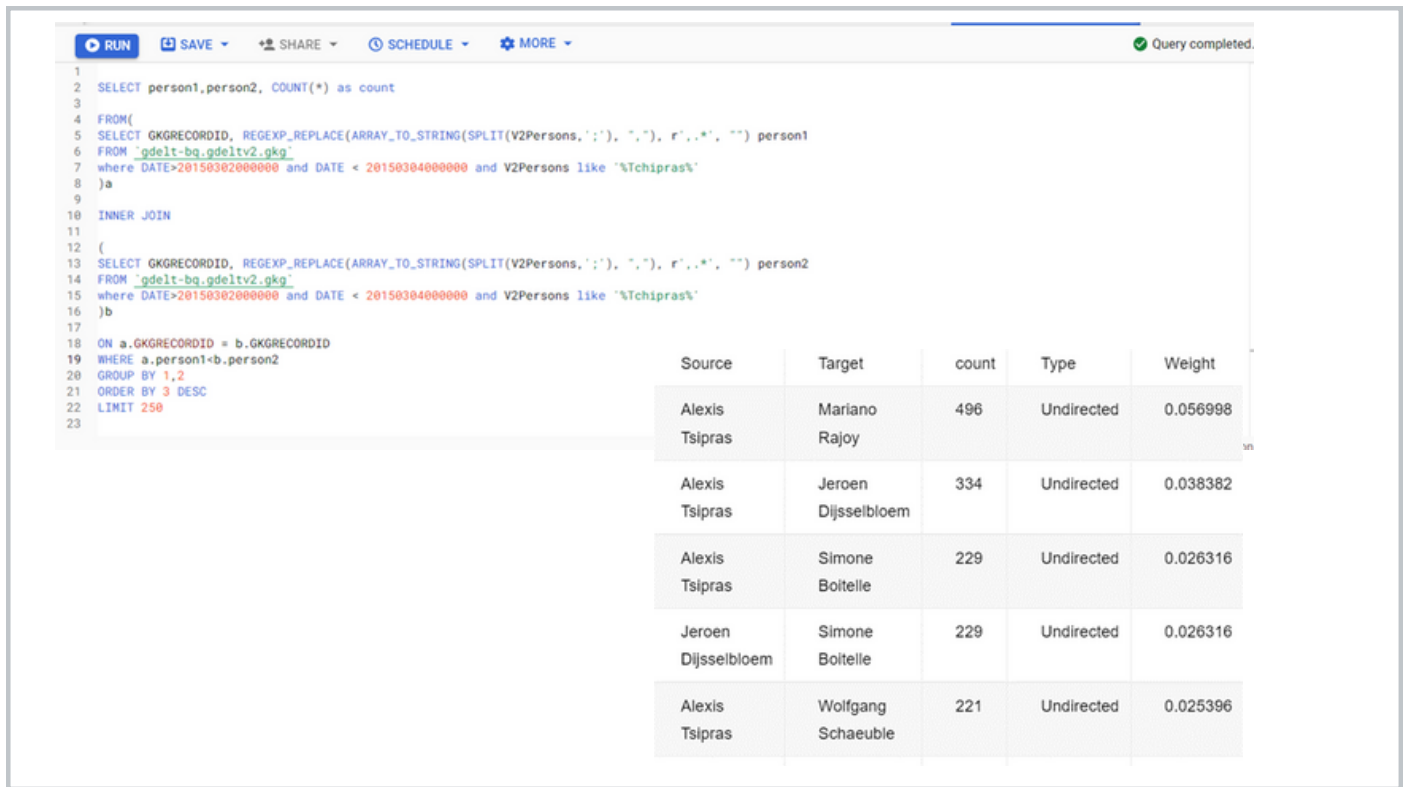
1 SELECT location, COUNT(*)
2 FROM(
3 SELECT REGEXP_EXTRACT(ARRAY_TO_STRING(SPLIT(V2Locations, ','), ','), r'^[2-5]#(.*?)#') as location
4 FROM `gdelit-bq.gdelitv2.gkg`
5 WHERE DATE > 20150302000000 AND DATE < 20220304000000 AND V2Persons LIKE '%Ataturk%'
6 )
7 WHERE location is not null
8 group by location
9 ORDER BY 2 DESC
10 LIMIT 100

```

JOB INFORMATION		RESULTS	JSON	E
Row	location	f0_		
1	Dublin, Dublin, Ireland	166		
2	Izmir, Izmir, Turkey	41		
3	Istanbul, Istanbul, Turkey	27		
4	Edremit, Van, Turkey	12		
5	Bingol, Bingöl, Turkey	4		
6	Balgownie, New South Wales, ...	4		
7	Trabzon, Trabzon, Turkey	3		
8	Silivri, Istanbul, Turkey	3		
9	Sivas, Sivas, Turkey	3		
10	Baghdad, Baghdad, Iraq	2		
11	Kalymnos, Perifereia Notiou Ai...	2		
12	Gursel, Kahramanmaras, Turkey	2		

atatürk örneği :

Intéressant que la Turquie ne soit pas au premier rang.



Comprenons le réseau de connexions qui entoure le Premier ministre grec Alexis Tsipras, en créant un diagramme de réseau montrant toutes les personnes avec lesquelles il est le plus étroitement lié dans la presse pendant une période donnée. La requête suivante crée un diagramme de réseau de cooccurrence des personnes apparaissant dans la couverture du Premier ministre Tsipras pendant la période du 2 mars 2015 au 4 mars 2015 .

MERCI BEAUCOUP!