



Google BigQuery

Yunus Solak
Doğa Yağmur Yılmaz
Banu Özdeveci

PLAN



1) Qu'est-ce que BigQuery ?

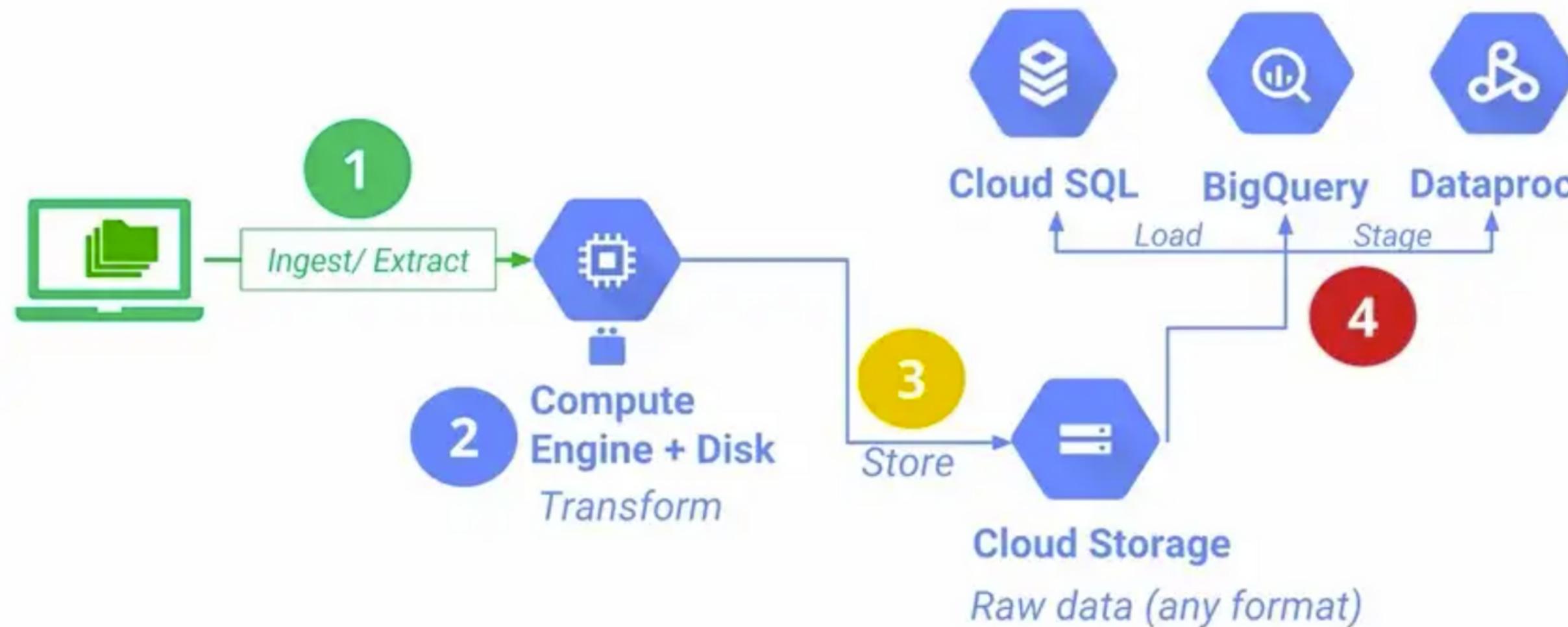
2) Comment fonctionne BigQuery ?

3) GDELT

4) Requêtes

QU'EST-CE QUE C'EST BIGQUERY ?

Data processing in the cloud



Where does BigQuery fit in on the Google Cloud Platform GCP

QU'EST-CE QUE C'EST BIGQUERY ?



QU'EST-CE QUE C'EST BIGQUERY ?



COMMENT FONCTIONNE BIGQUERY?

STOCKAGE

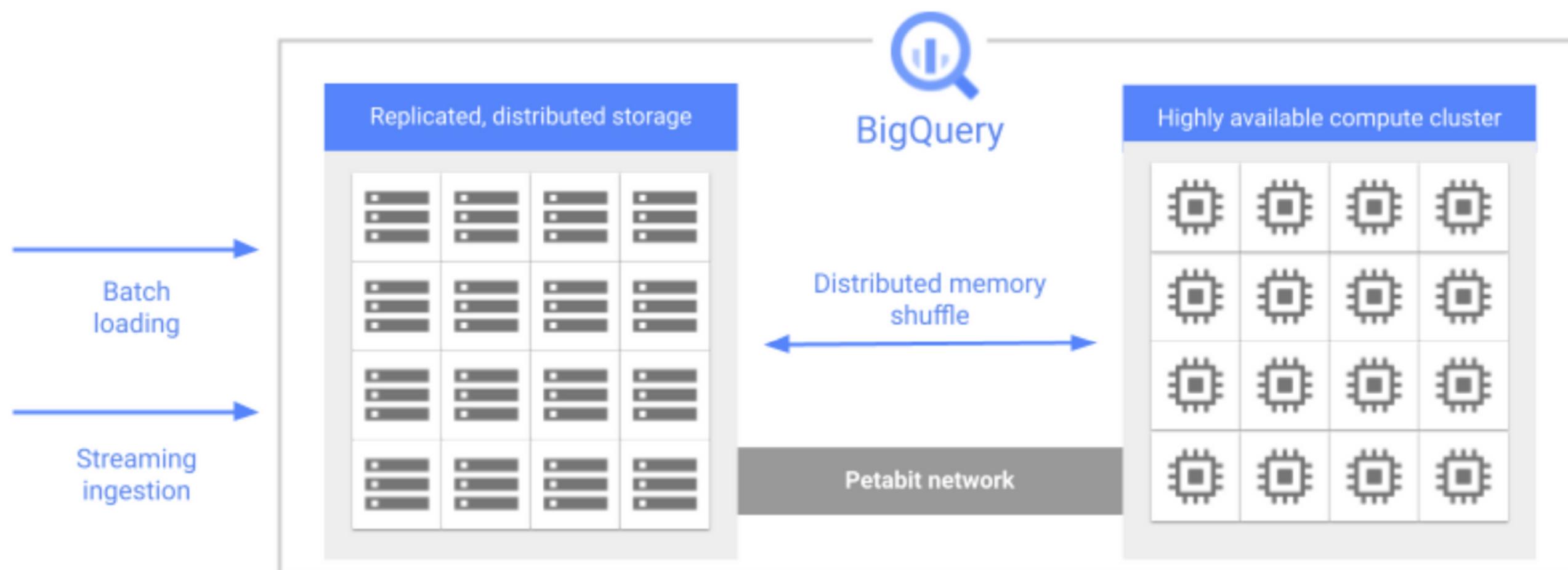
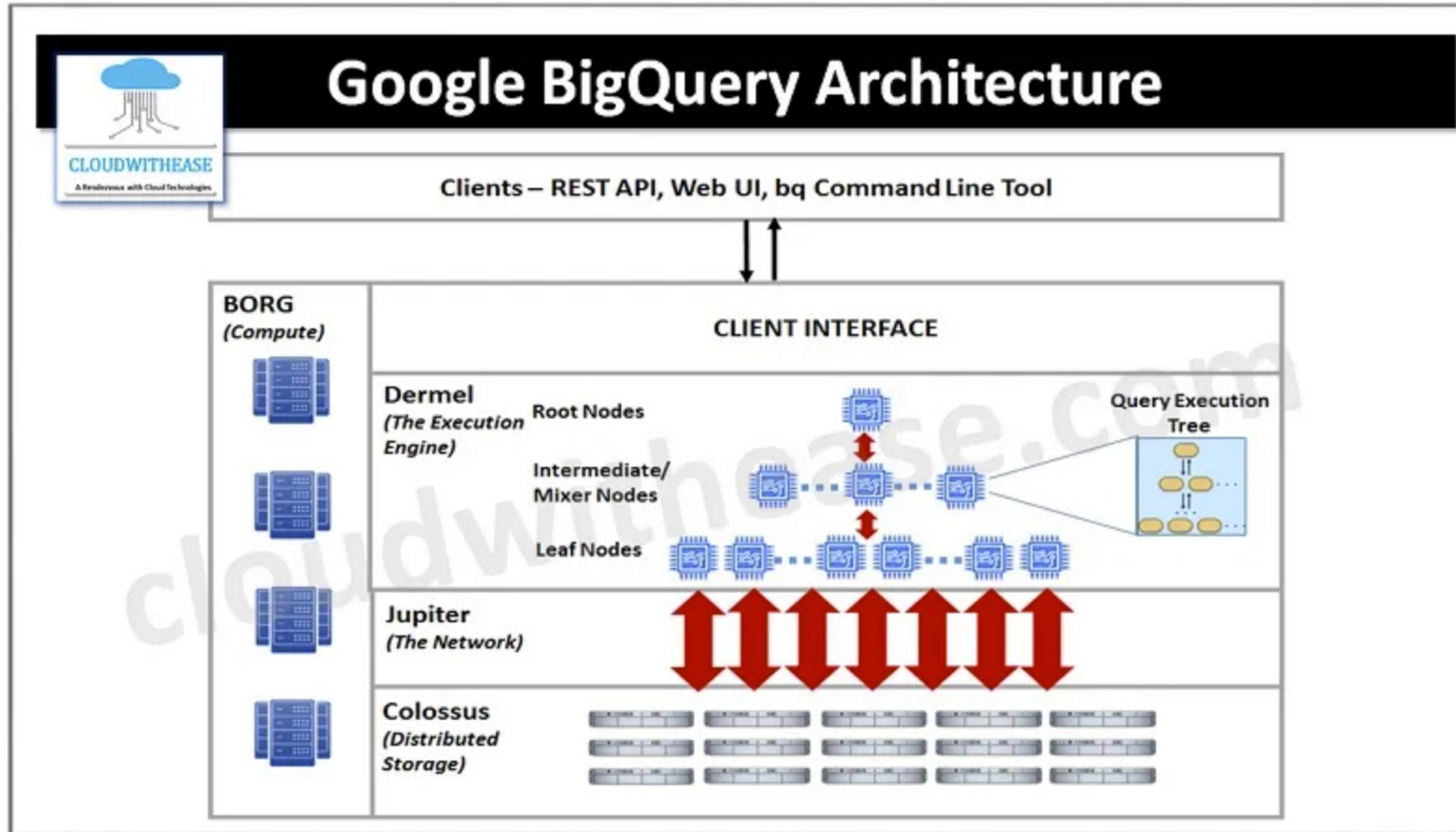


Figure 1. Architecture de BigQuery.

STOCKAGE

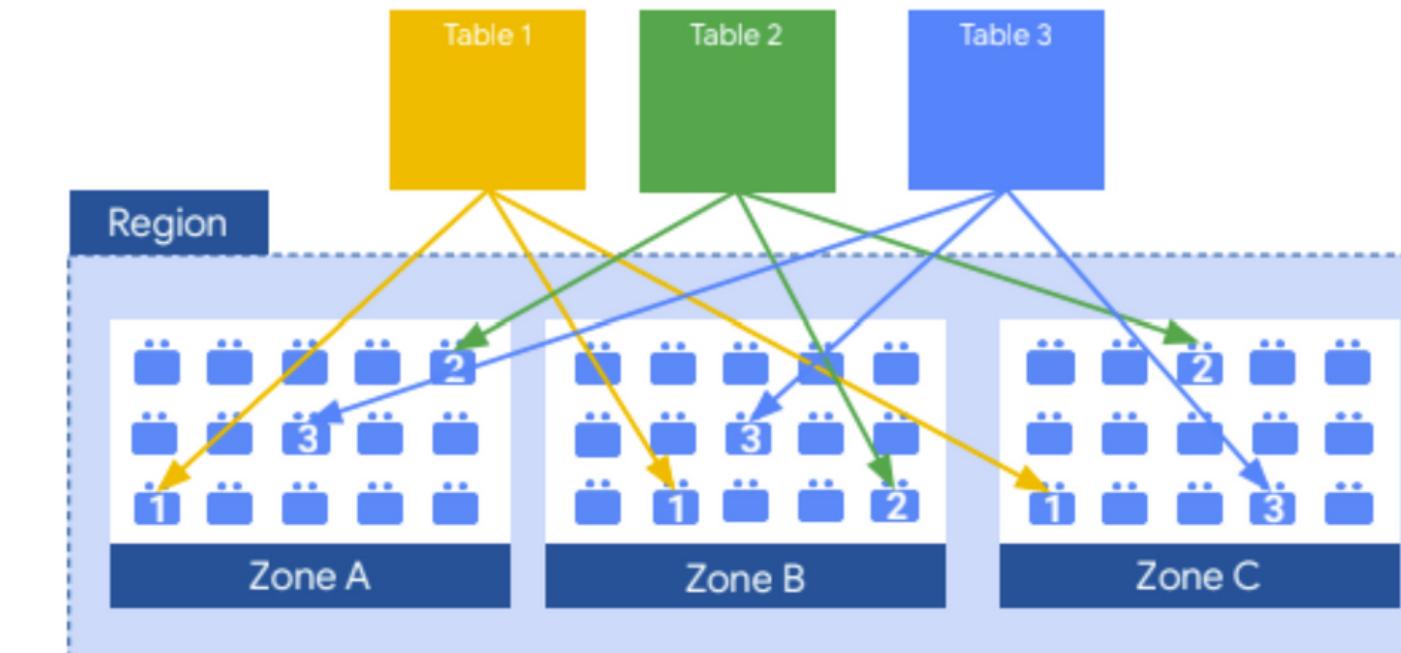


QUELQUES FONCTIONNALITÉS CLÉS DU STOCKAGE BIGQUERY



DURABLE

GÉRÉE



QUELQUES FONCTIONNALITÉS CLÉS DU STOCKAGE BIGQUERY

CRYPTÉ



EFFICACE

DISPOSITION DE STOCKAGE

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.

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.

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.

GDELT: UNE BASE DE DONNÉES MONDIALE DE LA SOCIÉTÉ



GDELT



GDELT Event Database



GDELT Global Knowledge Graph

REQUÊTES

Query results

 SAVE RESULTS •

Press Alt+F1 for Accessibility Options

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW	
N	I	V2Themes	Locations	V2Locations	Persons	V2Persons	Organization
1		null	3#Miami, Florida, United States#US#USFL#25.7743#-80.1937#295004;3#Boston, Massachusetts, United States#US#USMA#MA025#4	3#Boston, Massachusetts, United States#US#USMA#MA025#4 2.3584#-71.0598#617565#2 829;3#Boston,	kathryn bigelow;donald trump;chris rock;john krasinski;usain bolt;david ortiz;micahel bloomberg;missy	Kathryn Bigelow,1349;Donald Trump,101;Chris Rock,2416;John Krasinski,2985;Usain Bolt,3408;David Bloomberg,113;Michael Ortiz,100;Missy	google;facebook
2		;TAX_FNCACT_SPEAKER,41;TAX_RELIGION_ISLAMIC,268;TAX_ETHNICITY_ALGERIAN,16;TAX_WORLDLANGUAGES_ALGERIAN,166;	4#Ouagadougou, Kadiogo, Burkina Faso#UV#UV53#12.3703#-1.52 472#-1721728;1#Algeria#AG#AG#28#3#AG	1#Algerian#AG#AG##28#3#AG #166;4#Ouagadougou, Kadiogo, Burkina Faso#UV#UV53#154478#12.37 03#-1.52472#-1721728#11	slimane chenine	Slimane Chenine,94	speaker of the assembly;algerian parliament;council of the islamic committee
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-benz	Nicholas Pongratz,66;Mercedes-Benz,113;Daimler,113	null

RUN

MORE ▾

SAVE ▾

SHARE ▾

SCHEDULE ▾

This query will process 14.72 TB when run.

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

Editor Unsaved query 2 +
RUN SAVE SHARE SCHEDULE MORE
Query completed.

```

1 select * from `gdel-t-bq.gdeltv2.gkg`
2 where V2Persons like '%Netanyahu%' limit 10
  
```

Press Alt+F1 for Accessibility Options.

Query results

SAVE RESULTS EXPLORE DATA

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS	EXECUTION GRAPH	PREVIEW		
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;sergel	Jews,1359;Benjamin Netanyahu,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 sawafta;vladimir putin;jon alterman;mike pompeo;stephen farrell;mahmoud abbas;benny	Zomlot,1750;Angus Macswan,5482;Benjamin Netanyahu,1703;Dlab Al- Louh,953;Steven Holland,5426	u s embassy;wh jordan valley;ara league;reuters;u		
3	CRISISLEX_T11_UPDATESSY MPATHY,1709;CRISISLEX_T1 1_UPDATESSYMPATHY,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 Netanyahu,977;Benny Gantz	supreme court;i		

RUN**SAVE** ▾**SHARE** ▾**SCHEDULE** ▾**MORE** ▾

This query will process 1.34 TB when run.

```
1 SELECT V2Themes from `gdelt-bq.gdeltv2.gkg` where DATE>20150302000000 and DATE < 20150304000000 and V2Persons like '%Netanyahu%' limit 10
```

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	V2Themes
1	LEADER,410;LEADER,1563;LEADER,2153;LEADER,2553;TAX,_FNCACT_PRESIDENT,410;TA_X_FNCACT_PRESIDENT,1563;TAX_FNCACT_PRESIDENT,21
2	ECON_WORLDCURRENCIES_DOLLAR,2916;TAX_RELIGION,_SHAKERS,2215;TAX_ETHNI,CITY_ASIAN,2725;TAX_FNCA,CT_ECONOMISTS,2440;MAN
3	MEDIA_MSM,2086;TAX_POLITICAL_PARTY_REPUBLICANS,,1546;TAX_RELIGION_ISLAM,

RUN**SAVE** ▾**SHARE** ▾**SCHEDULE** ▾**MORE** ▾

This query will process 1.34 TB when run.

1 SELECT SPLIT(V2Themes, ';') from `gdelt-bq.gdeltv2.gkg` where DATE>20150302000000 and DATE < 20150304000000 and V2Persons like '%Netanyahu%' limit 10

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	
	f0_
	WB_2462_POLITICAL_VIOLENCE_AND_WAR,185
	WB_2473_DIPLOMACY_AND_NEGOTIATIONS,2331
	WB_2471_PEACEKEEPING,2331
	TAX_POLITICAL_PARTY_DEMOCRAT,1934
	TAX_FNCACT_AMBASSADOR,1388
	TAX_FNCACT_AMBASSADOR,2166
	MIL_WEAPONs_PROLIFERATION,540
	GENERAL_GOVERNMENT,1946
	ACT_MAKESTATEMENT,1641
	WB_2470_PEACE_OPERATIONS_AND_CONFLICT_MANAGEMENT,708
	WB_2503_WEAPONs_PROLIFERATION_AND_ARMS_CONTROL,708
	TAX_FNCACT_SENATOR,874

- > ^ 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

|< < > >|

Editor × Editor 2 × *Unsaved query 3 × +

RUN SAVE SHARE SCHEDULE MORE Query completed.

```
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
```

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	GENERAL_GOVERNMENT	925			
2	ECON_DEVELOPMENTORGS_I...	125			
3	LEADER	64			
4	PRIVATIZATION	43			
5	ELECTION	43			
6	SOVEREIGNTY	42			
7	TAX_ETHNICITY_GERMAN	41			
8	TAX_FAIRNESS_INEQUALITY	37			

Results per page: 50 ▾ 1 – 50 of 98 |< < > >|

RUN**SAVE** ▾**SHARE** ▾**SCHEDULE** ▾**MORE** ▾

```
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 > 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

PREV

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

RUN**SAVE** ▾**SHARE** ▾**SCHEDULE** ▾**MORE** ▾**Query completed.**

```
1 SELECT theme, COUNT(*) as count
2 FROM (
3   SELECT REGEXP_REPLACE(ARRAY_TO_STRING(SPLIT(V2Themes, ','), ','), r',.*', '') as theme
4   FROM `gdelb-bq.gdeltv2.gkg`
5   WHERE DATE > 20220202000000 AND DATE < 20220504000000 AND V2Persons LIKE '%Zelenski%'
6 )
7 group by theme
8 ORDER BY 2 DESC
9 LIMIT 100
```

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	ARMEDCONFLICT		4194			
2	GENERAL_GOVERNMENT		3046			
3	TAX_FNCACT_DEPUTY		2782			
4	LEADER		2744			
5	EPU_CATS_NATIONAL_SECURI...		1516			
6	TAX_FNCACT_CHIEF		1389			
7	BAN		933			
8	MILITARY		896			
9	TAX_WORLDSL LANGUAGES_RUS...		809			
10	KILL		791			
11	WB_2433_CONFLICT_AND_VIO...		703			
12	CRISISLEX_CRISISLEXREC		703			

"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 RÉSOLUTION GÉOGRAPHIQUE DE L'EMPLACEMENT.

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

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

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

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 AMÉRICAİN).

RUN

SAVE

SHARE

SCHEDULE

MORE

Query completed

```

1 SELECT location, COUNT(*)
2 FROM(
3   SELECT REGEXP_EXTRACT(ARRAY_TO_STRING(SPLIT(V2Locations, ','), ','), r'^.*?#(.*)#') as location
4   FROM `gdelt-bq.gdeltv2.gkg`
5   WHERE DATE > 20150302000000 AND DATE < 20150304000000 AND V2Persons LIKE '%Tsipras%'
6 )
7 group by location
8 ORDER BY 2 DESC
9 LIMIT 100

```

Press Alt+F1 for Accessibility Option

Query results

SAVE RESULTS

EXPLORE DATA



JOB INFORMATION

RESULTS

JSON

EXECUTION DETAILS

EXECUTION GRAPH

PREVIEW

Row	location	f0_	Elapsed time	Slot time consumed	Bytes shuffled	Bytes spilled to disk
1	Madrid, Madrid, Spain	414	30 sec	13 hr 17 min	54.45 KB	0 B
2	Greece	276				
3	Spanish	140				
4	Spain	103				
5	Guindo, Castilla Y LeóP, Spain	83				
6	null	75				
7	Greek	55				
8	Brussels, Bruxelles-Capitale, Be...	50				
9	Russia	48				
10	Greeks	41				

SHOW AVERAGE TIME SHOW MAXIMUM TIME

Stages	Working timing	Rows
▶ S00: Input	Wait: 12 sec Read: 86 ms Compute: 9 ms Write: 3 ms	Records read: 1023815521 Records written: 1704
▶ S01: Sort+	Wait: 165 ms Read: 0 ms Compute: 204 ms Write: 1 ms	Records read: 1704 Records written: 102

```

1 SELECT location, COUNT(*)
2 FROM(
3   SELECT REGEXP_EXTRACT(ARRAY_TO_STRING(SPLIT(V2Locations, ','), ','), r'^[2-5]#(.*)#') as location
4   FROM `gdelt-bq.gdeltv2.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 Aigaiou, Greece	13

```

1 SELECT location, COUNT(*)
2 FROM(
3   SELECT REGEXP_EXTRACT(ARRAY_TO_STRING(SPLIT(V2Locations, ','), ','), r'^[2-5]#(.*)#') as location
4   FROM `gdelt-bq.gdeltv2.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, BingöU, 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		

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

```
1
2 SELECT location, COUNT(*)
3 FROM(
4 SELECT REGEXP_EXTRACT(ARRAY_TO_STRING(SPLIT(V2Locations, ','), ", "), r'^[2-5]#(.*)#(?:GR|GM|SP)#' ) as location
5 FROM `ndelt-hn_ndeltv2_nkn`
```

Query results

[SA](#)[JOB INFORMATION](#)[RESULTS](#)[JSON](#)[EXECUTION DETAILS](#)[EXECUTION GRAPH](#)[PREVIEW](#)

Row	location	f0_
1	Madrid, Madrid, Spain	414
2	Guindo, Castilla Y LeóP, Spain	83
3	Athens, AttikiR, Greece	19
4	Lemnos, Perifereia Voreiou Aigaiou, Greece	13
5	Brussels, Bruxelles-Capitale, Belgium#BE#BE11#5850#50.8333#4.33333#-1955538#13,1#Greek	6

RUN**SAVE** ▾**SHARE** ▾**SCHEDULE** ▾**MORE** ▾

Query completed.

```
1
2 SELECT person1, person2, COUNT(*) as count
3
4 FROM(
5 SELECT GKGRECORDID, REGEXP_REPLACE(ARRAY_TO_STRING(SPLIT(V2Persons, ','), ',', r'.,.*', '')) person1
6 FROM `gdelt-bq.gdeltv2.gkg`
7 where DATE>20150302000000 and DATE < 20150304000000 and V2Persons like '%Tchipras%'
8 )a
9
10 INNER JOIN
11 (
12 SELECT GKGRECORDID, REGEXP_REPLACE(ARRAY_TO_STRING(SPLIT(V2Persons, ','), ',', r'.,.*', '')) person2
13 FROM `gdelt-bq.gdeltv2.gkg`
14 where DATE>20150302000000 and DATE < 20150304000000 and V2Persons like '%Tchipras%'
15 )b
16
17 ON a.GKGRECORDID = b.GKGRECORDID
18 WHERE a.person1<b.person2
19 GROUP BY 1,2
20 ORDER BY 3 DESC
21 LIMIT 250
22
23
```

Source	Target	count	Type	Weight
Alexis	Mariano	496	Undirected	0.056998
Tsipras	Rajoy			
Alexis	Jeroen	334	Undirected	0.038382
Tsipras	Dijsselbloem			
Alexis	Simone	229	Undirected	0.026316
Tsipras	Boitelie			
Jeroen	Simone	229	Undirected	0.026316
Dijsselbloem	Boitelie			
Alexis	Wolfgang	221	Undirected	0.025396
Tsipras	Schaeuble			



MERCI BEAUCOUP!