

# Rendu n°1 : CASSANDRA

Groupe : Adèle MONTTOYA, Eleonor KIOULOU, Paul RUNAVOT  
Classe : ESILV S8, DIA 4

## Presentation of the dataset

We worked on the Companies (2) dataset.

This database contains:

- basic information about companies like the name, the location, the url of the homepage, the category of company
- specific data like the amount about investments, acquisitions, type, currency...

## Insertion of the dataset

We encountered some problems about the insertion of our dataset in cassandra.

Firstly, the limit of the data insertion limit. Our dataset has a lot of data which exceeds the limit of data supported by the cassandra “by default”, we had to find a way to unlock this limitation.

```
> docker cp cqlshrc 84fa0c09a52f:/root/.cassandra/
```

After this, another problem arose. after several research, we found that the “delimiter” was the problem and we change it by “|” but even with that delimiter, some rows do not insert into the table. So we didn’t have the choice to continue like this.

## Simple queries (6)

1) Sélectionner les entreprises ayant pour catégorie “Software” :

```
cqlsh:comp> SELECT * FROM companies WHERE category_code='software' ALLOW FILTERING;
```

id	alias_list	blog_feed_url	category_code	created_at	crunchbase_url	deadpooled_day	deadpooled_month	deadpooled_url	deadpooled_year	email_address	founded_day	founded_month	founded_year	homepage_url	ipo	number_of_employees	permalink	phone_number	total_money_raised	twitter_username	updated_at
1	Constant Wave	http://www.constantwave.com	software	Sat Jul 11 17:17:27 UTC 2009	http://www.crunchbase.com/company/constant-wave								2008	http://www.constantwave.com							
2	Constant Wave	http://www.constantwave.com	software	Sat Jul 11 17:17:27 UTC 2009	http://www.crunchbase.com/company/constant-wave								2008	http://www.constantwave.com							
3	GlobeRanger	http://www.globeranger.com	software	Mon Mar 16 17:49:32 UTC 2009	http://www.crunchbase.com/company/globeranger								1999	http://www.globeranger.com			972-744-9977	\$8.3M	GlobeRanger	2012-10-29 02:02:05+00:00	
4	QualiLife	http://www.qualilife.com	software	Sat Oct 04 04:10:36 UTC 2008	http://www.crunchbase.com/company/qualilife								2008	http://www.qualilife.com			41-(0)-91-980-09-51	\$692K		2013-10-04 06:56:34+00:00	
5	Madgex	http://www.madgex.com	software	Tue Apr 01 00:28:49 UTC 2008	http://www.crunchbase.com/company/madgex								2005	http://www.madgex.com			+44 (0) 1273 775 100	\$0	madgex	2012-01-05 06:53:54+00:00	
6	Advent Software	http://www.advent.com	software	Wed May 27 03:33:13 UTC 2009	http://www.crunchbase.com/company/advent-software								1993	http://www.advent.com				\$0	adventsoftware	2013-09-07 18:37:57+00:00	
7	Adlucent	http://www.adlucent.com	software	Tue Apr 08 21:29:23 UTC 2008	http://www.crunchbase.com/company/adlucent								2005	http://www.adlucent.com				\$0	Adlucent	2013-05-02 04:01:18+00:00	
8	Universal Robotics	http://www.universalrobotics.com	software	Wed May 27 18:52:38 UTC 2009	http://www.crunchbase.com/company/universal-robotics								2008	http://www.universalrobotics.com			(615) 366-7281	\$1.04M		2009-06-17 23:48:01+00:00	
9	Tivix	http://www.tivix.com	software	Tue Nov 04 17:52:56 UTC 2008	http://www.crunchbase.com/company/tivix								2008	http://www.tivix.com				\$0	tivix	2013-08-21 06:41:37+00:00	
10	MarketCetera	http://www.marketcetera.com	software	Wed May 20 05:11:33 UTC 2009	http://www.crunchbase.com/company/marketcetera								2006	http://www.marketcetera.com				\$4M	MarketCetera	2013-11-13 02:19:36+00:00	
11	WorldTVRadio	http://www.worldtvradio.com	software	Tue Jul 29 19:51:23 UTC 2008	http://www.crunchbase.com/company/worldtvradio								2003	http://www.worldtvradio.com			(905) 296-4264	\$0		2009-04-24 07:42:24+00:00	
12	DigitalOfficePro	http://www.digitalofficepro.com	software	Fri Jan 16 05:59:12 UTC 2009	http://www.crunchbase.com/company/digitalofficepro								2009	http://www.digitalofficepro.com						2013-08-21 06:41:37+00:00	

2) Compter les entreprises ayant pour catégorie “Software” :

```
cqlsh:comp> SELECT COUNT(*) FROM companies WHERE category_code='software' ALLOW FILTERING;
```

count
1593

(1 rows)

3) Sélectionner et compter les entreprises ayant plus de 5000 employés :

```
cqlsh:comp> SELECT * FROM companies WHERE number_of_employees > 5000 ALLOW FILTERING;
```

id	crunchbase_url	alias_list	blog_feed_url	deadpooled_day	deadpooled_month	deadpooled_url	deadpooled_year	blog_url	category_code	created_at	
page_url	number_of_employees	permalink	ipo	phone_number	total_money_raised	twitter_username	updated_at	founded_day	founded_month	founded_year	home
											name
['\$id': '5a5c533e942d09e481c16943']									hardware	Mon Jun 23 15:43	
:20 UTC 2008		http://www.crunchbase.com/company/xerox								1906	
		http://www.xerox.com									
	57400			1-800-275-9376	\$1.1B	XeroxCorp	2013-06-07 22:44:33+00:00				
['\$id': '5a5c533e942d09e481c18250']									other	Sun Dec 14 11:38	
:32 UTC 2008		http://www.crunchbase.com/company/kalimatapla								1992	
		http://www.kalimatapla.com									
	25617			080-28361456	\$0		2009-04-13 16:01:13+00:00				
['\$id': '5a5c533e942d09e481c17b87']											
:27 UTC 2008		http://www.crunchbase.com/company/guitar-center								Mon Oct 27 06:24	
		http://www.guitarcenter.com								1959	
	7810			818-735-8800	\$0		2013-05-31 22:50:45+00:00				
['\$id': '5a5c533e942d09e481c17621']									hardware	Sat Sep 06 16:43	
:30 UTC 2008		http://www.crunchbase.com/company/mitsubishielectric								1921	
		http://global.mitsubishielectric.com									
	1,07e+05			81 3 3216 2111	\$0		2013-07-04 21:53:15+00:00				
['\$id': '5a5c533e942d09e481c165cd']									other	Thu May 22 12:03	
:30 UTC 2008		http://www.crunchbase.com/company/dentsau								1901	
		http://www.dentsau.com									
	6000				\$0	DentsauAustralia	2014-01-01 10:24:12+00:00				
['\$id': '5a5c533e942d09e481c193bc']										Mon May 18 23:31	
:30 UTC 2009		http://www.crunchbase.com/company/digital-equipment-corporation								1927	
		http://www.digit-eq.co.uk									
	1,4e+05				\$0		2013-09-30 07:49:05+00:00				
['\$id': '5a5c533e942d09e481c14422']									other	Wed May 07 03:58	
:53 UTC 2008		http://www.crunchbase.com/company/daily-mail-and-general-trust									
		http://www.dailymail.co.uk									
	22000				\$0		2010-04-28 22:09:22+00:00				
['\$id': '5a5c533e942d09e481c190f1']									semiconductor	Thu Apr 09 18:49	
:11 UTC 2009		http://www.crunchbase.com/company/microchip-technologies								1989	
		http://www.microchip.com/steilient/idsig7IdcService=SS_GET_PAGE&nodeId=64									
	5200			(480) 792-7200	\$0		2013-06-05 06:34:16+00:00				
['\$id': '5a5c533e942d09e481c175e7']									other	Sun Aug 31 23:32	
:40 UTC 2008		http://www.crunchbase.com/company/hilti									
		http://www.hilti.com									
	19000				\$0		2009-04-26 08:12:20+00:00				
['\$id': '5a5c533e942d09e481c1585d']									software	Fri Jan 04 07:41	
:24 UTC 2008		http://www.crunchbase.com/company/netapp								1994	
		http://blogs.netapp.com/dave/									

```
cqlsh:comp> SELECT COUNT(*) FROM companies WHERE number_of_employees > 5000 ALLOW FILTERING;
```

count
41

(1 rows)

4) Compter le nombre d'entreprises dans chaque pays en utilisant le code pays “country\_code”

```
cqlsh:comp> SELECT name,founded_year FROM companies WHERE founded_year<1950 ALLOW FILTERING ;
```

name	founded_year
See's Candies	1921
Athenix	1863
Thomas Publishing	1898
Newsweek	1933
DuBois	1920
GCHQ	1919
ewa-marine	1947
Xerox	1906
ACNielsen	1923
Leucadia National	1854
Mitsubishi Electric	1921
Dentsu	1901
EMI Group	1931
Altrasoft	1800
Short's Travel Management	1946
Random House	1925
Publicis Groupe	1926
Trammell Crow Company	1948
EFJohnson Technologies	1923
Cornell University	1865
Corning	1851
IBM Japan	1937
Shell	1907
University of Washington	1861
Harris	1890
Consumers Union	1936
Sumitomo Corporation	1919
Louis Vuitton	1854
The Washington Post Company	1947
Fortune Magazine	1930
Fluor Corporation	1912
Associated Press	1846
Solutia	1901
Wall Street Journal	1889
Millage, Inc	1867
Safeway	1915
Financial Times	1888
Landor Associates	1941
The Walt Disney Company	1923
Pearson	1844
Telenor	1885
NRC	1899
ADP	1949
Adidas	1920
Edison Electric Institute	1933
Cision	1892
Raytheon	1922
Bachmann Industries	1833
DuPont	1802
Acme Scale Systems	1909
Beloit College	1846
Sanistal	1923
Baker Commodities	1937
Deloitte	1900
Scotts Miracle-Gro	1868

5)

```
cqlsh:comp> SELECT COUNT(*) FROM companies WHERE founded_year<1950 ALLOW FILTERING;
```

count
73

6) (1 rows)

7) Sélectionner les entreprise avec plus de cinq acquisitions:

```
SELECT name, acquisitions FROM companies WHERE acquisitions > 5;
```

8) Sélectionnez les entreprises dont les coordonnées GPS sont définies

9) Sélectionnez toutes les entreprises qui ont au moins un partenaire

## Complex queries (2)

1) Groupement des entreprises en fonction de leur catégorie et calcul du nombre d'entreprises pour chaque catégorie

```
SELECT category_code, count(id) FROM companies GROUP BY category_code
```

- 2) Sélectionner les entreprises avec plus de cinq acquisitions

## Hard queries (1)

- 1) Tri des entreprises en fonction de la distance entre leur siège social (lat et lon dans coordonnées de la colonne "offices") et Paris.