

**Rapport Cassandra**

By ALGERA Pieter, BAALI Karim and INSERTYOURNAME Grégoire

Table des matières

[Dataset and queries 3](#_Toc506238823)

[Loading data from JSON to Cassandra 4](#_Toc506238824)

[Performing queries to the database 4](#_Toc506238825)

# Dataset and queries

For this project on Cassandra database, we decided to choose the dataset named “Companies2.json”. It contains different kind of information about a very large set of company object, such as a name, a website URL or the lists of customers or the number of employees. It allows us to be able to query a lot of different information based on various parameters. For example, a little part of the first object in this dataset :



We decided to create 5 queries to test our dataset on a Cassandra database :

* [EASY] Get the name of companies which have more than N products.

INSERT CQL REQUEST HERE

* [EASY] Get the id of companies with a given number of employees.

INSERT CQL REQUEST HERE

* [MEDIUM] Get the number of companies with a given category code.

INSERT CQL REQUEST HERE

* [MEDIUM] Get the name of companies which have offices in a given city.

INSERT CQL REQUEST HERE

* [HARD] Get the name of companies which were created before a given date and have more than N providers.

INSERT CQL REQUEST HERE

# Loading data from JSON to Cassandra

To be able to load our data from the json file we decided to use a Node.js application and we didn’t do it yet so I can’t tell you how it works for now. But I’m sure that in a few hours everything gonna be okay, believe me ! I can feel it !

# Performing queries to the database