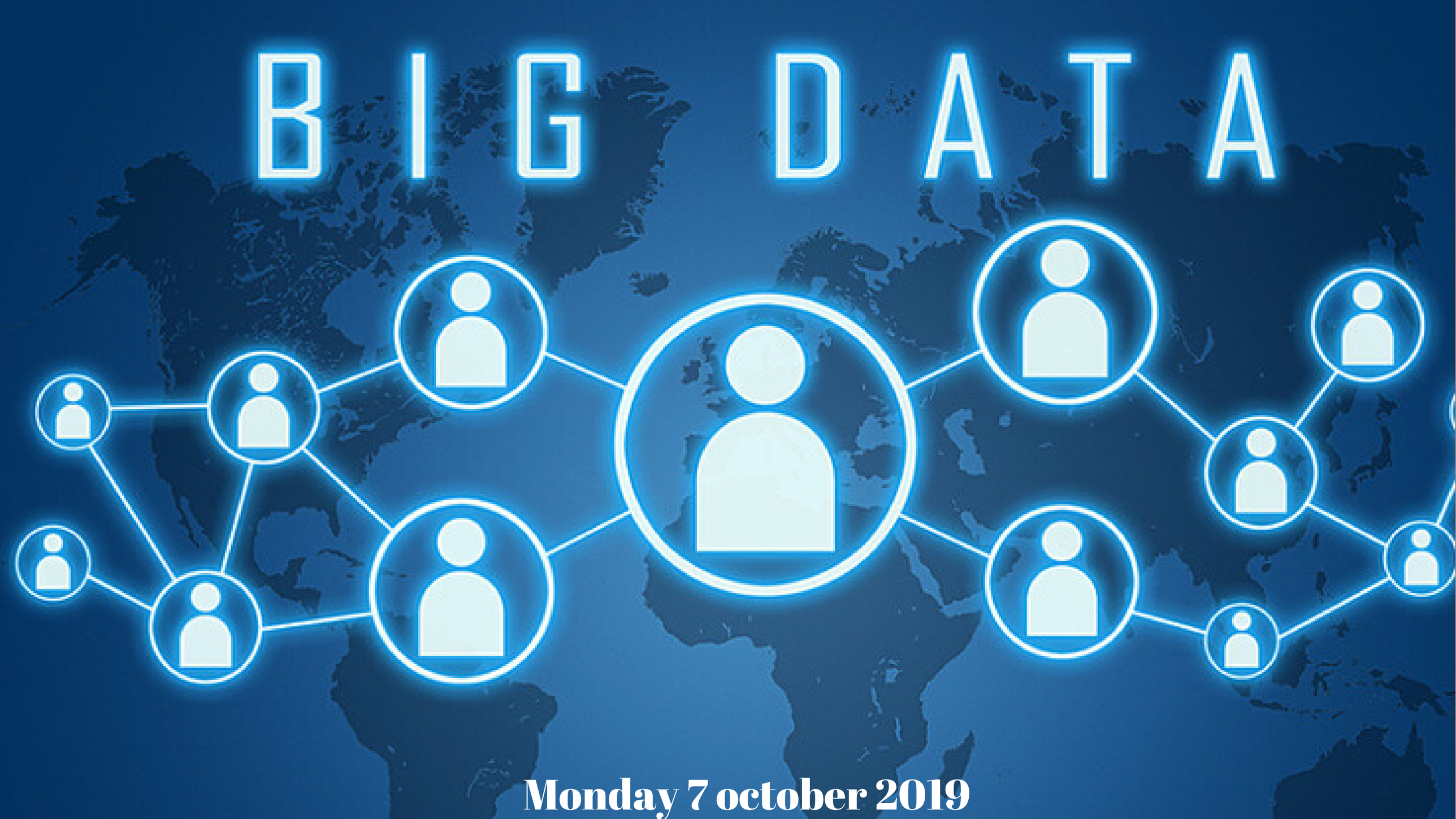


Monday 7 october 2019

The exponential increase of digital information reveals data sets that are becoming so large it is difficult (if not impossible) to process it with adequate response times via traditional tools.

This requires rethinking the very basics of data processing.

A single figure already to begin: it is admitted that every 12 to 18 months, a volume of data doubles.
This involves several strategic issues:
- maintain the performance of the production tasks that are necessary for the proper functioning of the company
- make best use of these valuable data to derive the finest information possible
- best manage the storage and backup of these data.

But if our volumetric problems affect production? Or if the volumes to be processed, even in decision-making are exponential ?

Think about the information stored for example by Google or Facebook. We are in volumes of data that can easily make you dizzy, and here we can really talk about Big Data.

This is where the NoSQL solution started to emerge. NoSQL is not necessarily to be taken as "not" SQL, but rather "not only".
It offers solutions that can complement traditional RDBMS solutions.

The NoSQL offers several forms of information storage:

01

Key / value.

in the manner of associative arrays of programming languages. A key has a value.

02

Column oriented:

a key is a set of columns, each with a value.

03

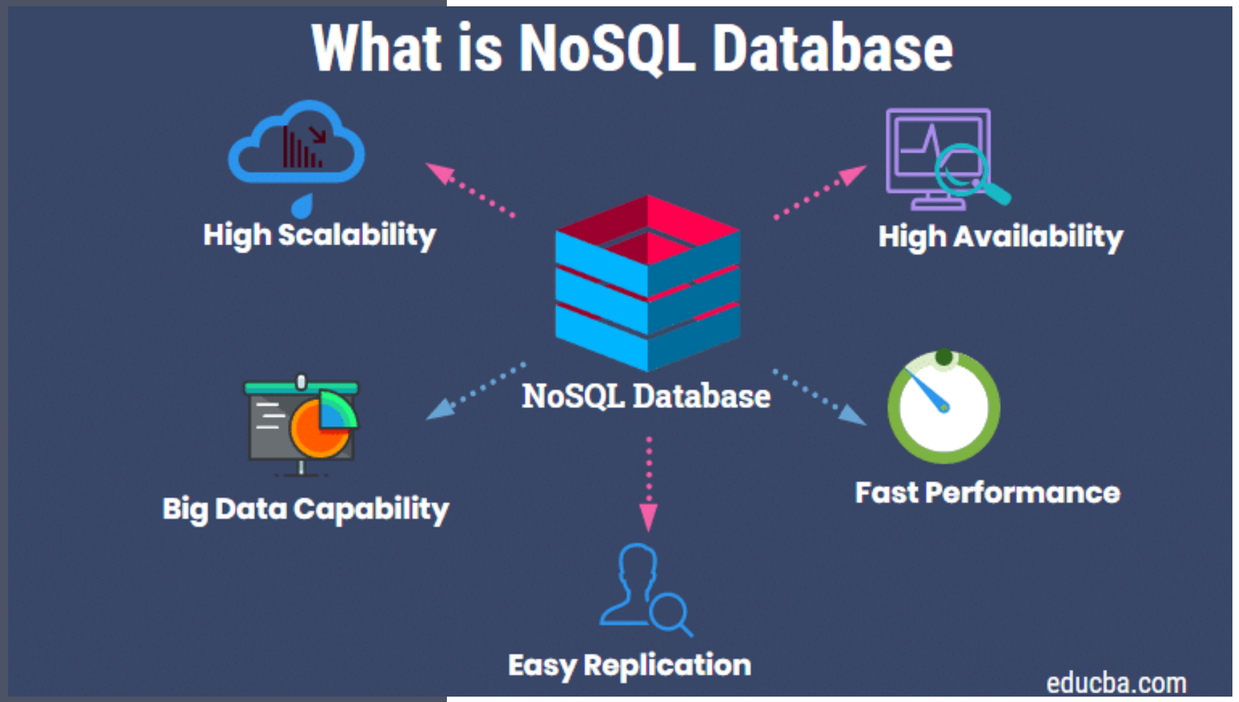
Document-oriented:

to a key corresponds sets fields / values that can be hierarchical

04

Graph oriented:

the data is modeled as nodes that have links between them.

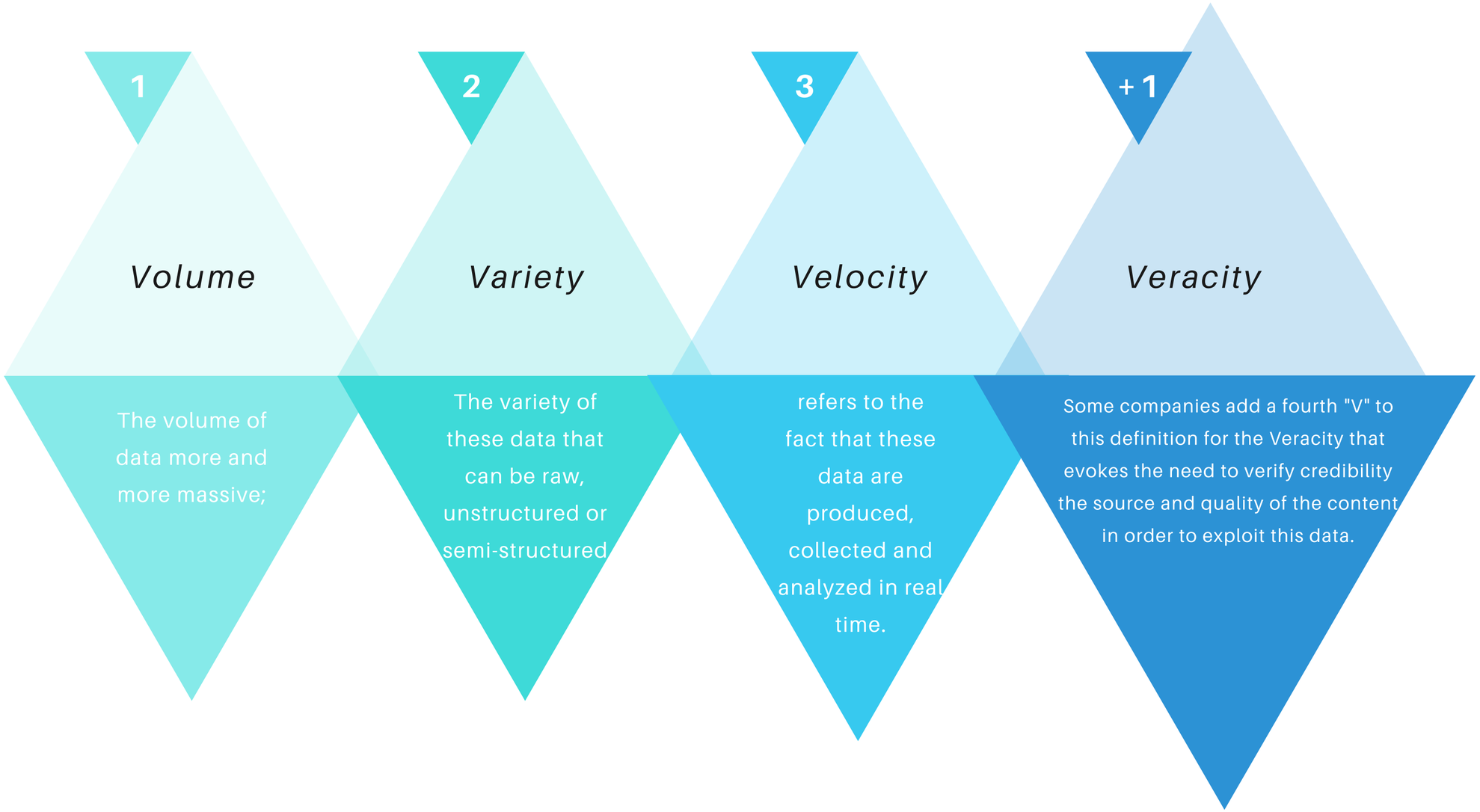


What are the current solutions?
If you want to use a NoSQL DBMS, there are many solutions available.

Some of the best known include:

- **Cassandra: http://cassandra.apache.org/**
- **MongoDB: http://www.mongodb.org/**
- **SimpleDB: http://aws.amazon.com/en/simpledb/**

The principle of 3 V (or 4):



Exemple of company that use Big Data:

NETFLIX

1. *Time*

   time that users take to choose an film/moovie

2. *High Moment*

   The highlight where users put the movie in break

3. *Title of*

   what users have seen before and after a title

4. *Start stop and Play*

   the series or movie that users stopped

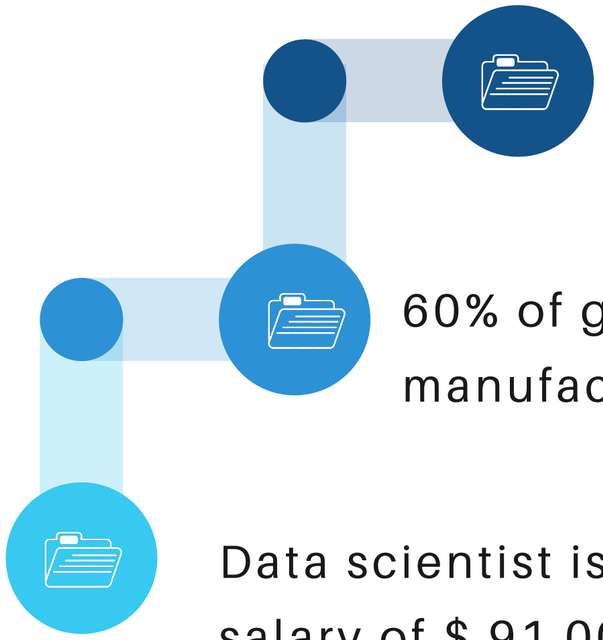
5. *Devices*

   which devices they used to...

6. *and others...*

   like geocalisation data...

Some numbers :



1.2 trillion dollars more for companies that use big data by 2020

60% of global manufacturers will use Big Data to analyze manufacturing processes and identify opportunities for optimization

Data scientist is the job number 1 ranking in 2017 with a median salary of $ 91,000

The big data upstream is a human who will create a script to be able to collect data, "clean" them, and send them back later,

Knowing that the more data volume is huge, the more accurate the script will be in order to be able to pull statistics.

Company collect, process, and process visual graphics to make decisions : according to the customers for an in-depth analysis of their client,

this allows in particular to increase the opening rate of an emailing campaign by targeting and personalizing a newsletter with the data collected upstream

Example: Predictive marketing, knowing the expectations of its target

Predictive maintenance or the renewal of industrial maintenance reduces the costs of industrial installations thanks to Big Data.

In fact, thanks to the introduction of sensors on machines, companies can now use data to plan the maintenance of machines.

What about financial services ?

Financial services companies can use Big Data to reduce risks and costs, example: analyze the data of a market to make a buisness model...