WHAT IS BIG DATA ?

A. Definition

The term "Big Data" refers to the evolution and use oftechnologies that provide the right user at the right time with the right information from a mass of data that has been growing exponentially for a long time in our society. The challenge is not only to deal with rapidly increasing volumes of data but also the difficulty of managing increasingly heterogeneous formats as well as increasingly complex and interconnected data.

Being a complex polymorphic object, its definition varies according to the communities that are interested in it as a user or provider of services. Invented by the giants of the web, the Big Data presents itself as a solution designed to provide everyone a real-time access to giant databases.

Big Data is a very difficult concept to define precisely, since the very notion of big in terms of volume of data varies from one area to another. It is not defined by a set of technologies, on the contrary, it defines a category of techniques and technologies. This is an emerging field, and as we seek to learn how to implement this new paradigm and harness the value, the definition is changing.

1) *Characteristics of Big Data*

The term Big Data refers to gigantic larger datasets (volume); more diversified, including structured, semi-structured, and unstructured (variety) data, and arriving faster (velocity) thanbefore. These are the 3V.

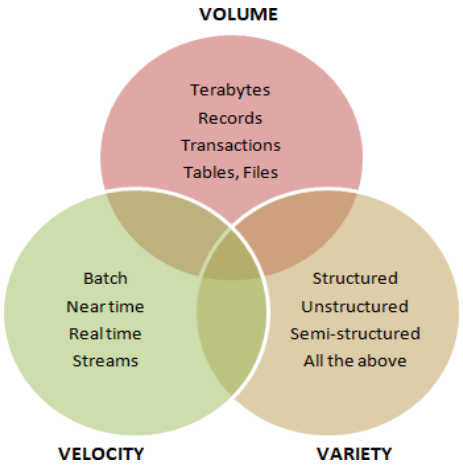

Fig. 1. *3V Concept*

-Volume: represents the amount of data generated, stored and operated within the system. The increase in volume is explained by the increase in the amount of data generated and stored, but also by the need to exploit it.

-Variety: represents the multiplication of the types of data managed by an information system. This multiplication leads to a complexity of links and link types between these data. The variety also relates to the possible uses associated with araw data.

-Velocity: represents the frequency at which data is generated, captured, and shared. The data arrive by stream and must be analyzed in real time.

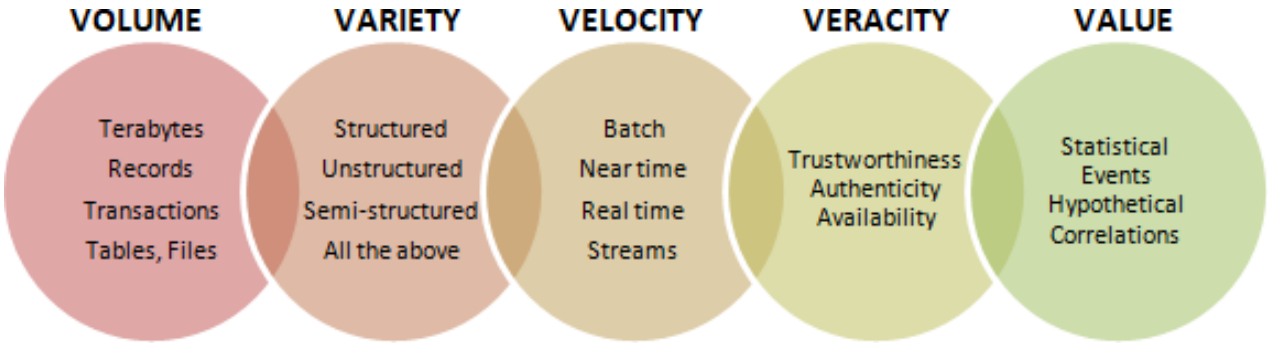

Fig. 2. *5V Concept*

To this classical characterization, two other "V"s areimportant:
-Veracity: level of quality, accuracy and uncertainty of dataand data sources.
-Value: the value and potential derived from data.