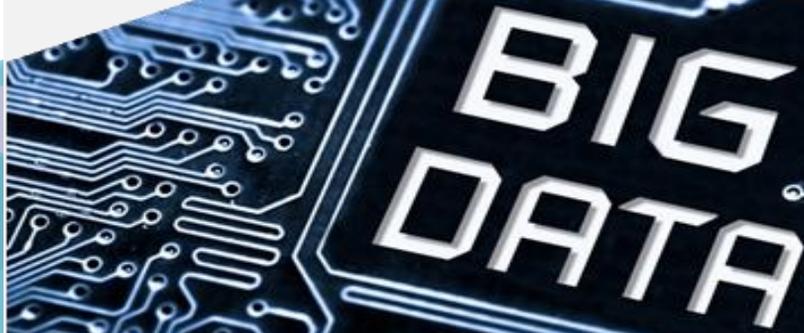


STATSBDA BULLET

Vol. 1

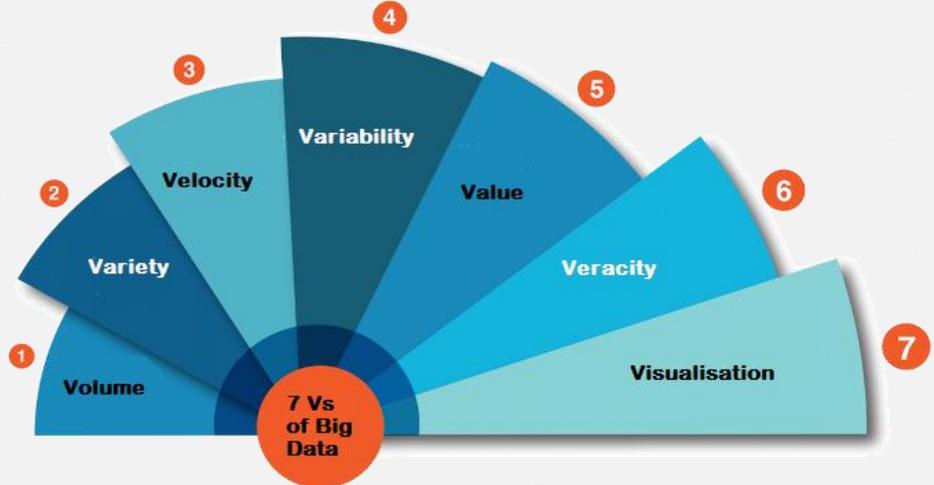


STATSBDA QUICK FACTS

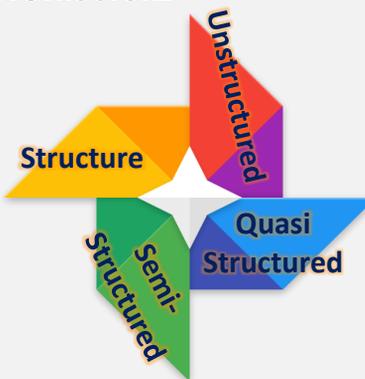
-  **STATSBDA Kick Off**
2nd Dec 16
-  **UBRS 1.0**
7-9th Dec 16
-  **UBRS 2.0**
9th – 12th Jan 17
-  **UBRS Walkthrough With Management**
15 -16th Feb 17
-  **UBRS Sign Off**
-  **BizCode@Stats PFDS 1**
27th Feb 17
-  **PFDS 1**
7-10th Mar 17
-  **PFDS 2**
12-14th Apr 17
-  **PFDS Walkthrough With Management**
23-24th May 17

BIG DATA: IS IT FANTASY?

Big data is a broad term for **gigantic digital datasets** that traditional data processing applications are inadequate.



DATA STRUCTURE



SOURCES OF BIG DATA



Streaming data



Social media data



Publicly available sources

“Data is the currency of business and its value is increasing exponentially”
by Suni Munshani, Big Data Quarterly, 2016



EDITORIAL BOARD

STATSBDA Team
c/o Methodology & Research Division

HOW CAN WE MAKE SENSE OF BIG DATA?

“ There is a lot of ‘noise’ around big data technologies and it can be challenging to know how to move forward”

World Wide Technology, Inc., 2014

Big Data is not only concerned to data big in volume but also data with big velocity, big variety and big veracity. A standard database management system (DBMS) may not be robust enough to manage the sheer enormity of the data.

Apache Hadoop is the most revolutionary technology which opened the door of infinite possibilities in Big Data. Initially Hadoop developed with two core components Hadoop Distributed File System (HDFS) and MapReduce. YARN, the NextGen MapReduce framework turns Hadoop as a general purpose data operating system. Impala raises the bar for SQL query performance on Apache Hadoop while retaining a familiar user experience. Apache is not the only organization that develops tools and projects for Big Data, many other organizations are also contributing and some provides their own Hadoop distributions.

Another area for big data analytics is the use of geographical information systems (GIS). Due to the complexity of map data, even an assessment at the municipality level would constitute a big data situation. When we are looking at the entire planet, we are analyzing big data. A typical example of GIS software would be the commercial product, ArcGIS, or the open-source product, Quantum GIS. GIS is interesting not only because it involves raw numbers, but it also involves data representation and visualization, which must then relate to a map with a clear interpretation. Google earth adds the extra complexity of zooming, decluttering, and overlaying, as well as choosing between political maps and satellite images. We now add the extra complexities of color, line, contrast, shape, and so on.

The need for low latency, another way of saying short lag time, between a request and the delivery of the results, Kafka works in combination with Apache Storm, Apache HBase and Apache Spark for real-time analysis and rendering of streaming data. Kafka can message geospatial data and whatever the industry or use case, Kafka brokers massive message streams for low-latency analysis in Enterprise Apache Hadoop.

Approached correctly, Big Data is a tool that helps you see the why behind the what.

References:

1. Pries K. H. & Dunnigan R. Big Data Analytics : A Practical Guide for Managers, 2015
2. Book Excerpt: The Trick to Making Sense of Big Data, The Globe and Mail, 26 January 2015

5 KEY ENABLERS OF THE BDA ECOSYSTEM

AWARENESS



The target is to provide opportunities for organizations to **exchange ideas** and **share knowledge**.

PEOPLE & TALENT



Malaysia 's target is to grow and nurture **16,000** data professionals, including **2,000** data scientists, by 2020

DATA GOVERNANCE & POLICY



When governments make data open, they naturally **stimulate innovation** from all members of society.

TECHNOLOGY PLATFORM



Establish more **strategic partnerships** with technology partners, universities and research institutes.

INDUSTRY-DRIVEN OPEN INNOVATION



One way to build the ecosystem is through partnerships with **subject matter experts** in BDA