Big Data:
New Opportunities for Official Statistics

MAZREHA YA'AKUB
Department of Statistics Malaysia

OUTLINES

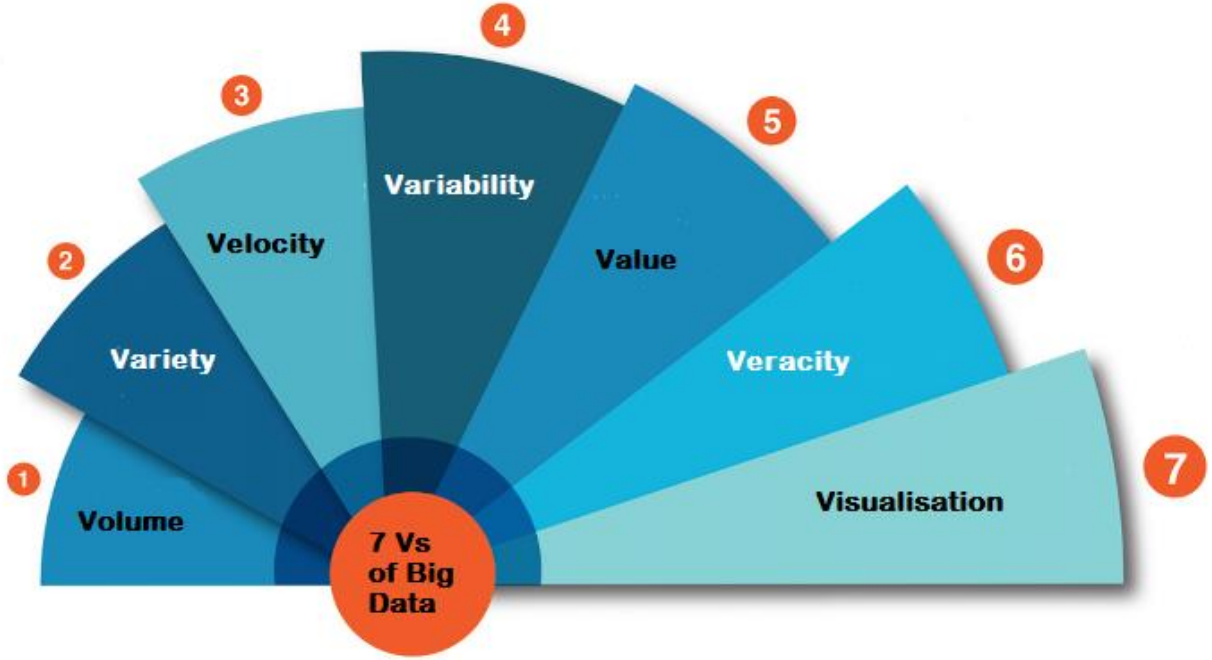
- Overview Big Data Analytics
- The Importance of Big data for Official Statistics
- The Challenges involved in using Big data for Official Statistics
- DOSM'S Big Data Analytics

Overview Big Data Analytics

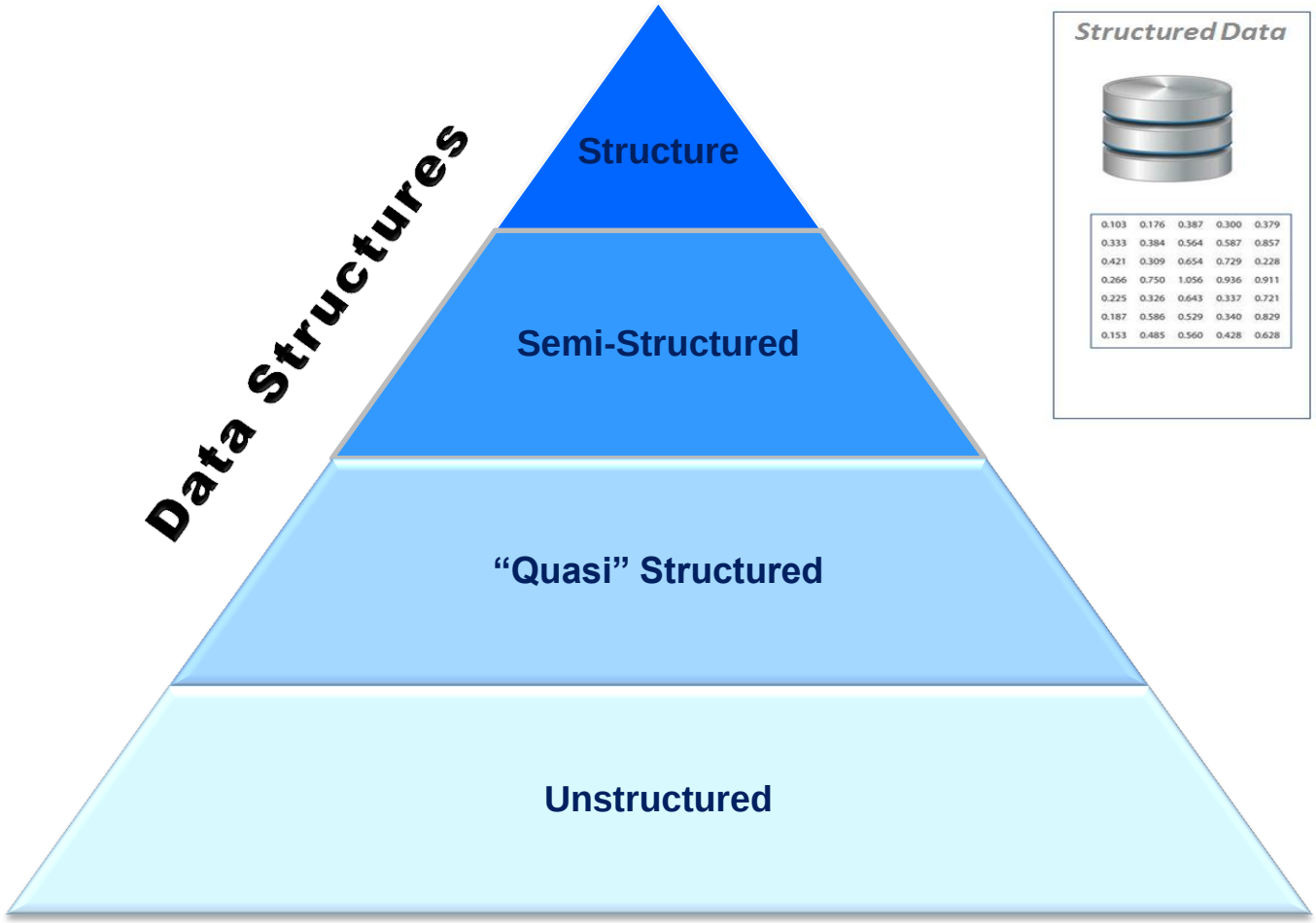
WHAT IS BIG DATA?

- Huge volume of Data
- Complexity of data types and structures
- Speed of new data creation and growth

"To Get Data is EASY To Get The Right Data is Hard"

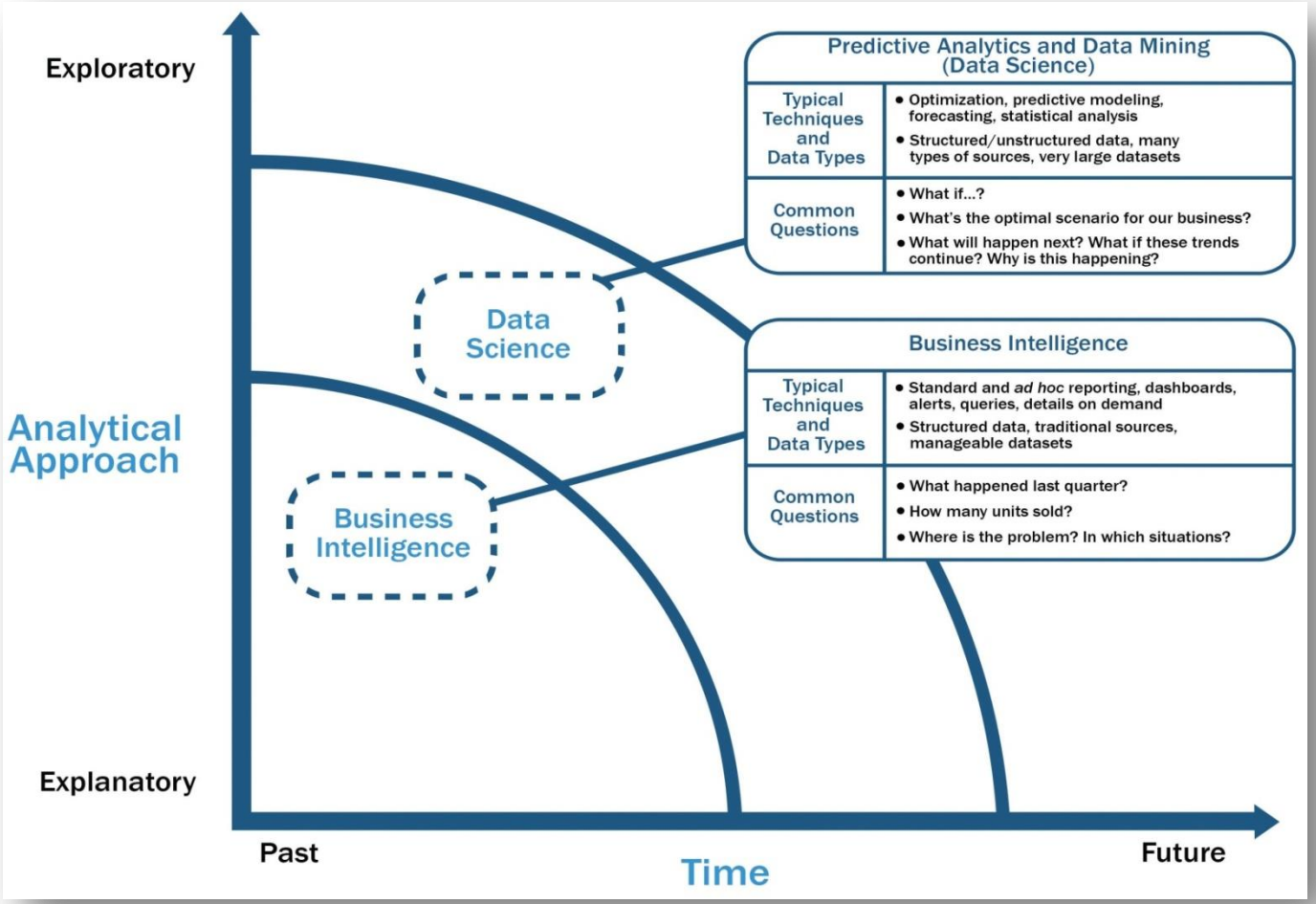


BIG DATA CHARACTERISTICS: DATA STRUCTURES



Source :EMC Education Services

BUSINESS INTELLIGENCE VS DATA SCIENCE



Source :EMC Education Services

The Importance of Big data for Official Statistics

National Statistical Office (Primary Data)

CENSUS

Administrative bodies

+

New Official Statistics

Data of 'others'

=

Compilation of official statistics for evidence-based decision–making

Big data, the future of statistics….

The Importance of Big data for Official Statistics

Reduce the response burden on enterprise and households as well as the cost for the data collections and improve the quality of the statistics.

Data growth rate are skyrocketing..

Source : Statistical Journal of the IAOS

The Importance of Big data for Official Statistics

- Improved the timeliness of statistical products
- More detailed break-downs of statistics possible
- Improve accuracy
- Production of new statistical indicators

The Challenges involved in using Big data for Official Statistics

- Suitable statistical and IT methods
- New tools and skills are needed to handle alternative data
- Quality issues on each dataset and application

The Challenges involved in using Big data for Official Statistics

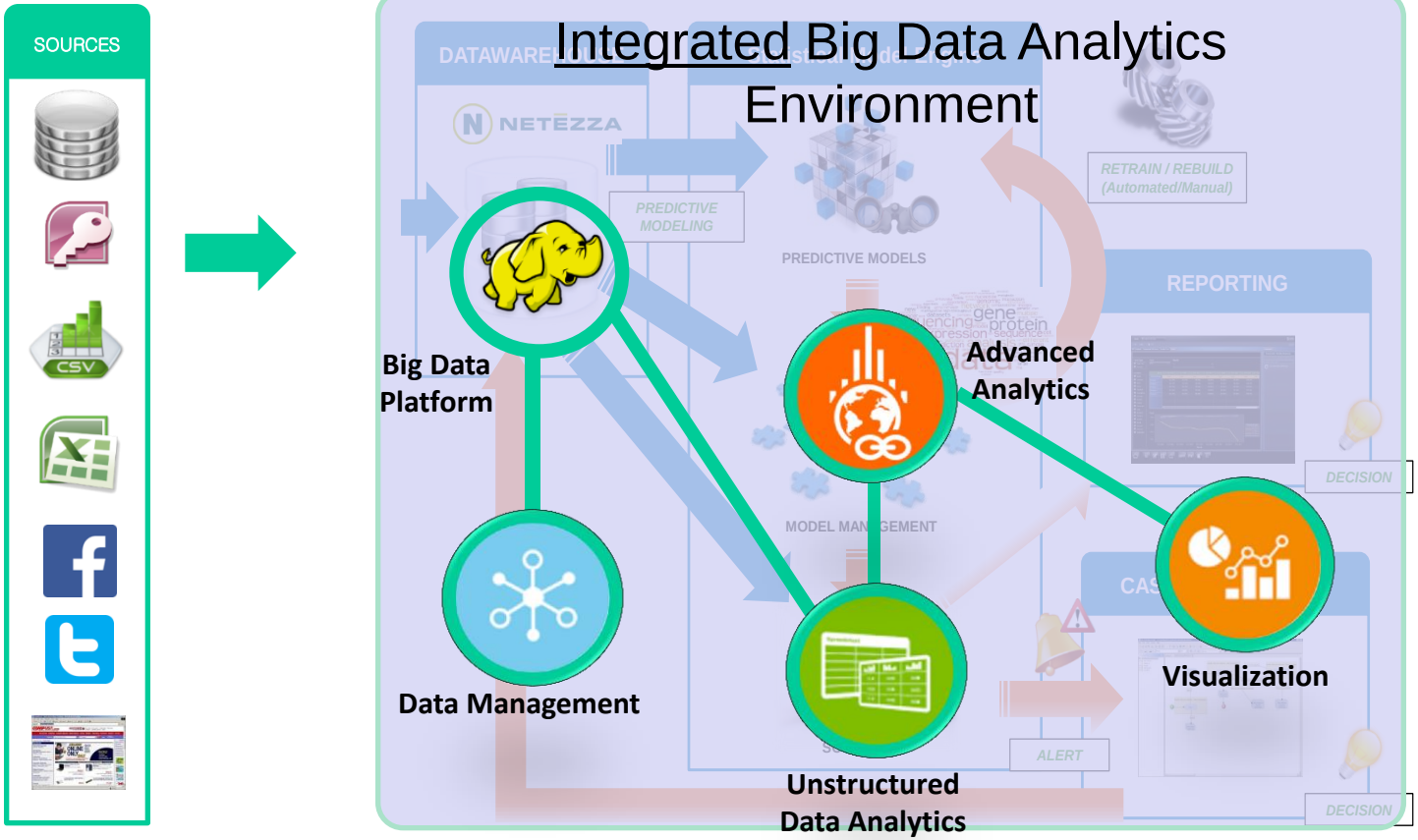
- The costs of sourcing the data must in balance with the benefits
- Legislative requirement for getting access and using the data
- Legal issues (e.g. personal data protection)

DOSM'S Big Data Analytics

- Used as supplements for existing data in production of certain statistics.
- Allow high actuality
- Reduce response burden on some respondents

Jabatan Perangkaan
MALAYSIA

BIG DATA ANALYTICS FOR DOSM



DOSM'S Big Data Analytics

MAIN PROJECTS

1. TRADE BY ENTERPRISE CHARACTERISTICS (TEC)

2. PRICE INTELLIGENCE (PI)

3. OPINION MINING ON OFFICIAL STATISTICS

DOSM'S Big Data Analytics

SUPPORT PROJECTS

1. REAL TIME BUSINESS STATUS

2. REAL TIME NEWS ON OFFICIAL STATISTICS

3. BizCode@Stats

DOSM'S Big Data Analytics

DOSM BDA INITIATIVES 1:
TRADE BY ENTERPRISE CHARACTERISTICS (TEC)

Traditional trade statistics

Record what types of goods are trading across border between countries

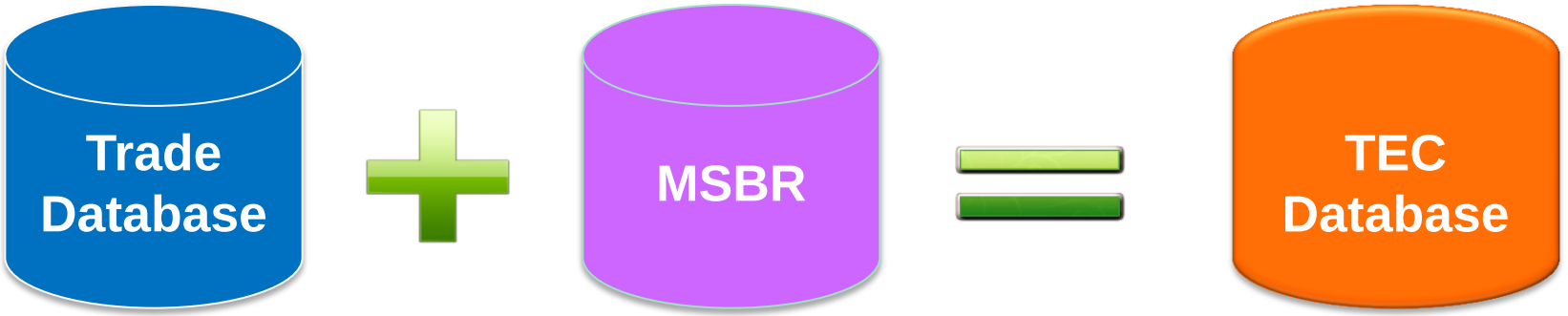
Trade by enterprises characteristics (TEC)

The linkage of trade statistics with business registers allows us to describe those who are engaged in global market and what are their characteristics.

DOSM'S Big Data Analytics

DOSM BDA INITIATIVES 1:
TRADE BY ENTERPRISE CHARACTERISTICS (TEC)

TEC DATABASE



Integrated Trade Database with MSBR

DOSM'S Big Data Analytics

DOSM BDA INITIATIVES 2: INTERNAL PORTAL FOR PRICE INTELLIGENCE (PI)

Modernization of data collection tools for improving quality of Consumer Price Index (CPI). The modernization of data collection mainly consist of the adoption of web scraping techniques to scrape price data from related website for CPI compilation.

DOSM'S Big Data Analytics

DOSM BDA INITIATIVES 2:
INTERNAL PORTAL FOR PRICE INTELLIGENCE (PI)

Data Warehouse

Analysis

Data Mining

Data Visualization

Report & Dashboards

Alert

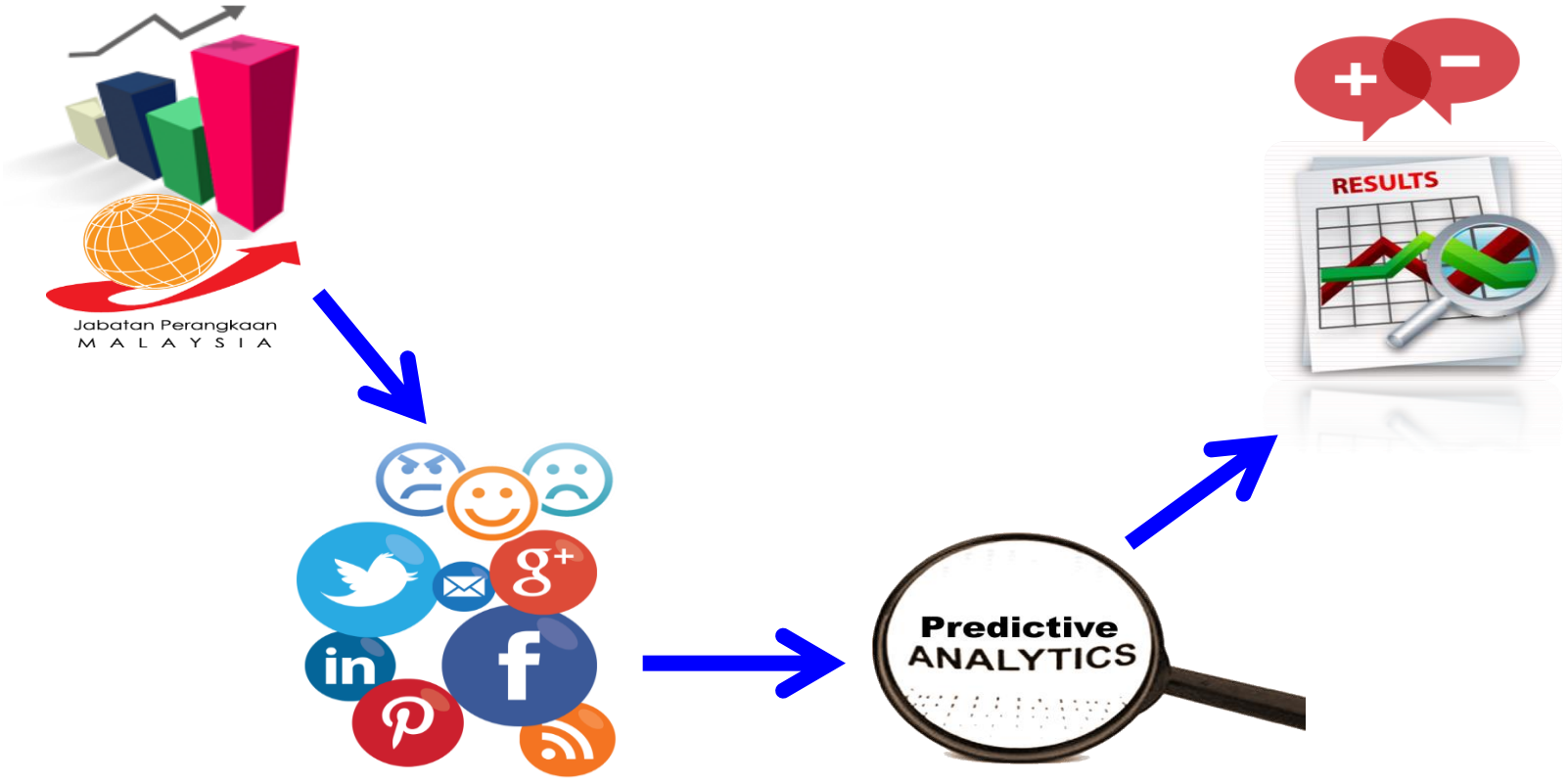
DOSM'S Big Data Analytics

DOSM BDA INITIATIVES 3:
Opinion Mining on Official Statistics

The analysis and assessment of the degree of "happiness" of Malaysia community with regards to official statistics published by DOSM. The data is obtained from online social media.

DOSM'S Big Data Analytics

DOSM BDA INITIATIVES 3:
Opinion Mining on Official Statistics



DOSM utilizes Big Data as supplements for existing data to produce comprehensive and quality statistical products, services and new indicators.

Source : Statistical Journal of the IAOS

Thank you!

Mazreha Binti Ya'akub

Mazreha@stats.gov.my
Methodology & Research Division
Department of Statistics Malaysia