



# HOW BIG DATA DRIVE GREATER MSBR

SITI HASLINDA MOHD DIN, NUR AZIHA MANSOR

2016 WIESBADEN GROUP MEETING IN TOKYO: SESSION 2  
DEPARTMENT OF STATISTICS MALAYSIA

# OUTLINES

**MODERNIZE  
MSBR IN BIG  
DATA  
PLATFORM**

**2**

**3**

**MSBR QUALITY  
IMPROVEMENT  
WITH BIG DATA**

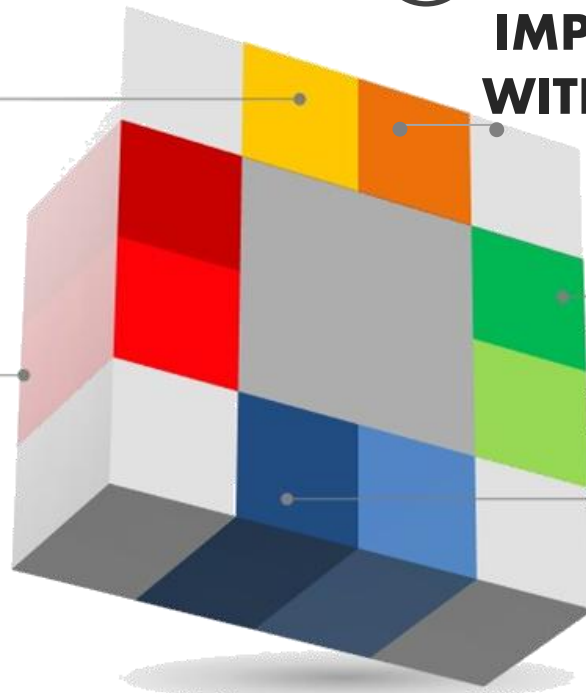
**4**

**OUTCOME:  
BUSINESS  
INTELLIGENCE  
& ANALYTICS**

**1**

**5**

**CONCLUSION**



**INTRODUCTION**

# INTRODUCTION

**1** SBR provides a comprehensive businesses' and companies' frame with a set of stratification variables

**2** MSBR is an important DOSM's asset, thus the data quality is a highly significant consideration

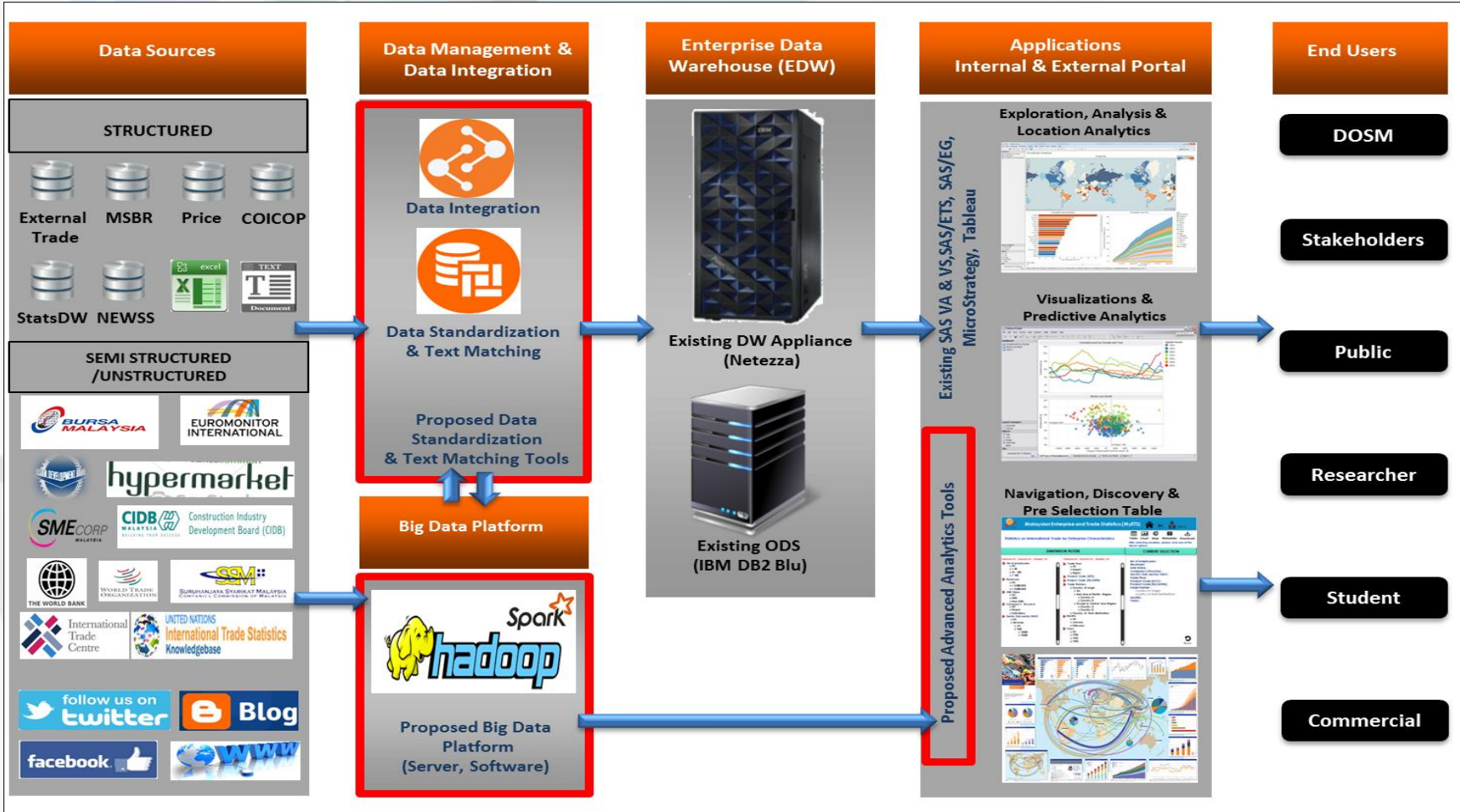
**3** With emerging trends of big data, data quality management offered by big data helps DOSM to enhance the quality of MSBR

**4** The function of MSBR is further expanded by integrating it with trade database by using big data platform.

**4** The paper is to highlight how big data drives greater MSBR in the production of Trade by Enterprise Characteristics (TEC) statistics.

# GREATER MSBR IN BIG DATA PLATFORM

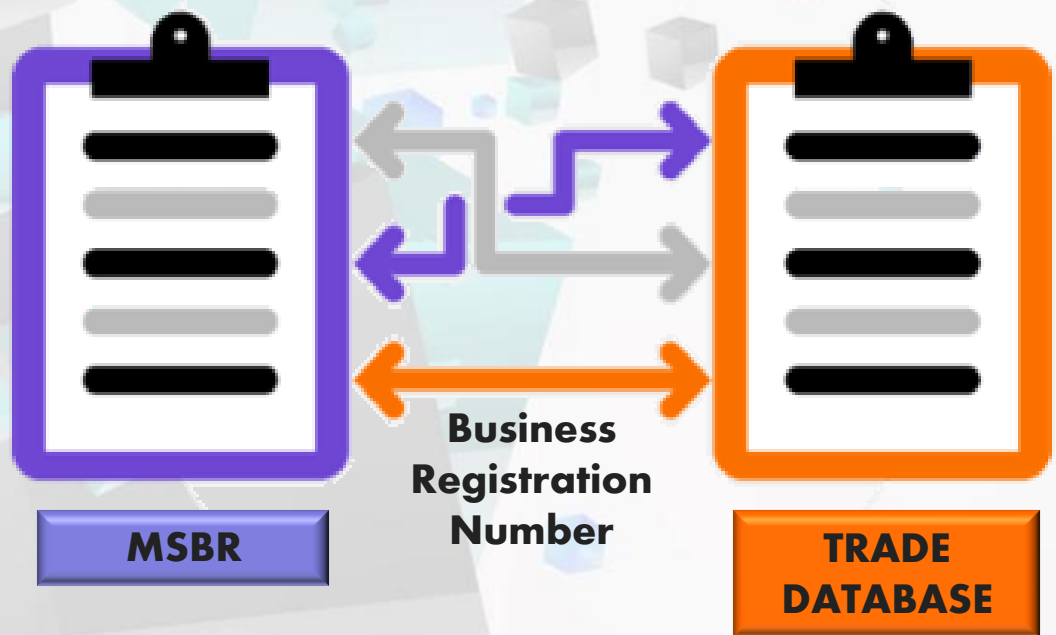
TEC project has become one of the DOSM Big Data Analytics (StatsBDA) initiatives.



StatsBDA Proposed Architecture

# MODERNIZE MSBR IN BIG DATA PLATFORM

## Conventional Approach



# MODERNIZE MSBR IN BIG DATA PLATFORM

## String Matching Algorithm Approach



# MODERNIZE MSBR IN BIG DATA PLATFORM

## Data Profiling 01

The data is examined and analysed to come out with a brief and informative summary

## Data Standardization 02

Bringing data into a common format to allow for string matching algorithm approach.

## Data Clustering 03

The data is partitions into groups based on similarity.

## Data Cleansing 04

Amend/ remove data in a database that is incorrect, incomplete, improperly formatted/ duplicated within the cluster.

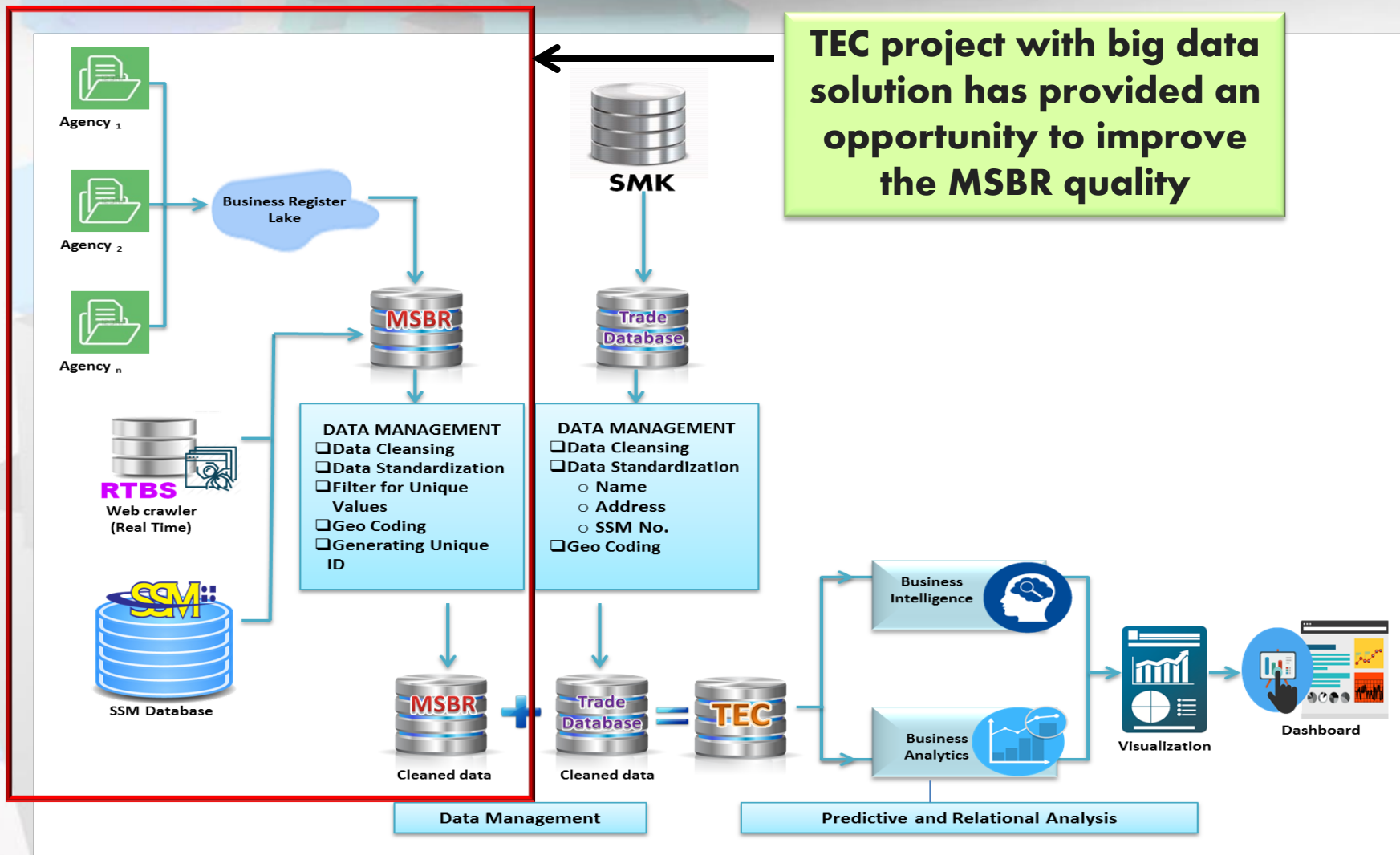
## Text Matching 05

Finding strings in traders' name that exactly or approximately match with business registration names in MSBR.

## Data Integration 06

Combination of MSBR and trade data to create a meaningful and valuable information

# MSBR QUALITY IMPROVEMENT WITH BIG DATA



**Proposed TEC Architecture**



# OUTCOME: BUSINESS INTELLIGENCE & ANALYTICS

## Business Intelligence

**Transforms MSBR and trade data into more meaningful and useful information for analysis purposes**

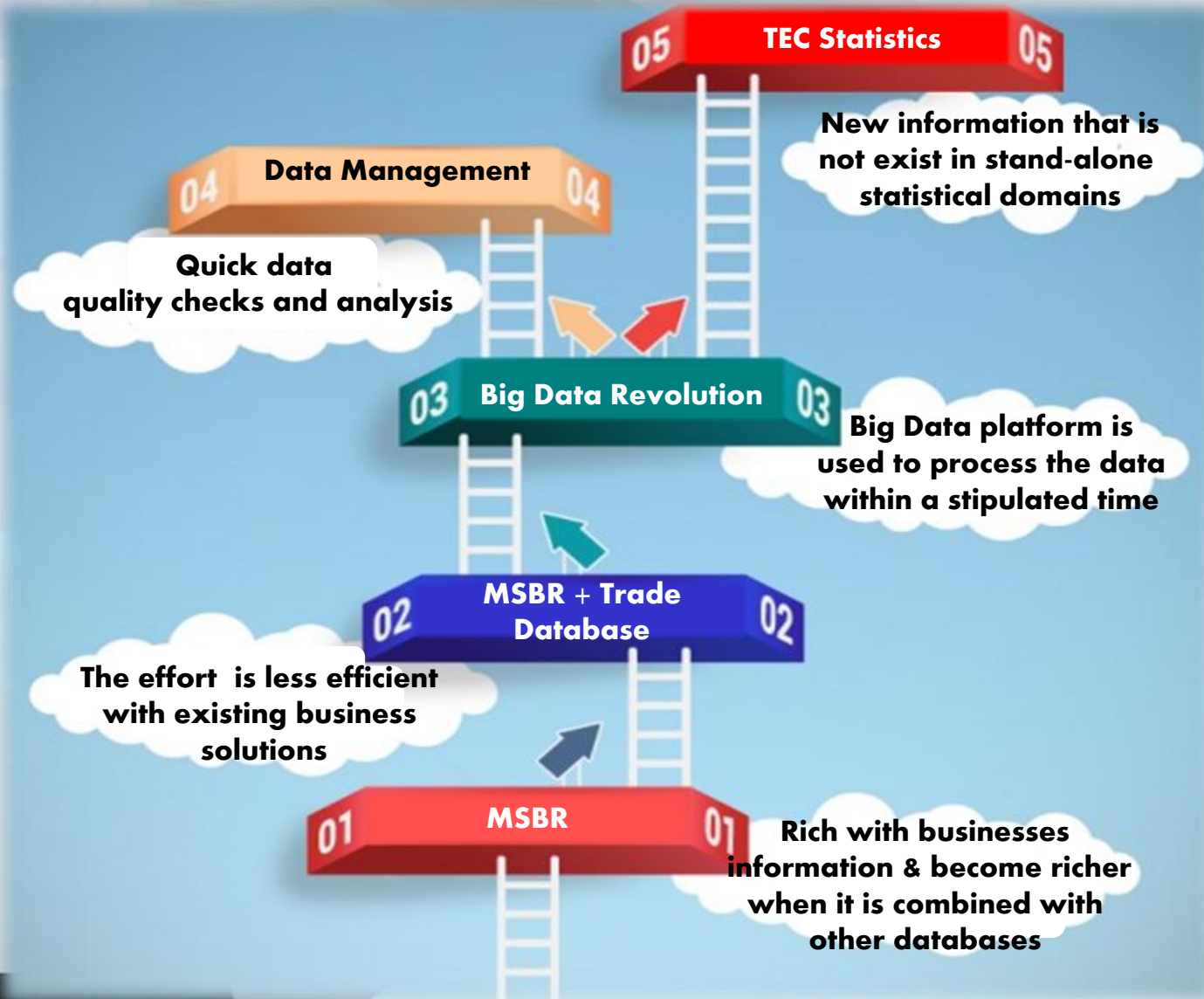
## Business Analytics

**Determine the patterns and predict future trade outcomes and trends**

## Visualisation

**Manipulate complex pools of data to visually display the data's patterns, trends, and correlations**

# CONCLUSION





**THANK YOU**

**Nur Aziha Mansor**

**[nuraziha@stats.gov.my](mailto:nuraziha@stats.gov.my)**

**Methodology & Research Division  
Department of Statistics Malaysia**