UNIVERSITY DATA ANALYTICS: A CASE OF THE TECHNICAL UNIVERSITY OF KENYA

AUTHORS: OBONYO ISHMAEL, OPIYO ELISHA AND ODONGO WILLIAM

PRESENTER: ISHMAEL OBONYO

ORGANIZATION: UNIVERSITY OF NAIROBI

VENUE: INTEL COLLEGE, NAIROBI - KENYA
OVERVIEW

- Definitions of Major Terms
- Background Information
- Problem Statement
- Methodology
- Results
- Conclusion
- References
**DEFINITIONS OF MAJOR TERMS**

- **Big Data**: described by 3V's - Volume, Velocity and Variety (Chen et al., 2012 & Kwon et al., 2014)

- In addition to the 3 V's, there are other V's: Veracity, Variability, and Value (Gandomi & Haider, 2015)

- It is also a term describing large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information (Das & Kalita, 2015).

---

**Social media** (Facebook, Twitter, LinkedIn, Youtube)

**ERP Data**

**Machine sensor data** (RFID)
DEFINITIONS OF MAJOR TERMS

- **Business Intelligence (BI):** set of methodologies, processes, architectures, and technologies used in transforming raw data into meaningful and useful information that is applied in business for competitive advantage (Foley & Guillemette, 2012).

- **Data Analytics:** science that incorporates various disciplines including, but not limited to, data engineering, mathematics, statistics, computing, and domain-specific expertise geared towards transforming data into useful information (Rodgers, Talbut, & Baranovic, 2015, February). It is argued that Business intelligence and analytics (BI&A) has emerged as an important area of study for both practitioners and researchers, reflecting the magnitude and impact of data-related problems to be solved in contemporary business organizations; Data analytics therefore refers to the BI&A technologies that are grounded mostly in data mining and statistical analysis. (Chen, Chiang & Storey, 2012)

- **Hadoop:** open source advanced Data Analytics platform for Big Data adopted in this study for data integration, analytics, and data visualization

- **University Data Analytics:** a process of transforming data pertaining to students, staff, research and other stakeholders into actionable information that can aid a University management during decision making

- **R:** Open source statistical software adopted in this study
Good Management Decisions

Results to

Effective Management of Organization

Requires

Timely access to right information pertaining to the organization processes in the right format by the right person

Challenge is access to this information

What brings about the challenge?

Data from multiple sources, in various formats, different definitions/schemas

Only machine knows!

Decision maker – human being

No way to know what information the machine has captured

(Al-Dabbagh, 1992; Malik, 2005)

(Erion & Prusak, 1998)
BACKGROUND INFORMATION

THE MAJOR CHALLENGE! DATA => INFORMATION => KNOWLEDGE. THE MORE THE DATA THE GREATER THE CHALLENGE!

WHAT HAS BEEN DONE TO ADDRESS THE CHALLENGE

- Decision Support Systems
  - Data models, interface

- Management Information System
  - Databases, algorithms of processing

- Expert Systems
  - Reasoning Bases, Knowledge Bases

- Executive Information Systems

- Business Intelligence
  - Data Mining, OLAP, Warehouse

(Olszak & Ziemba, 2007)
BACKGROUND INFORMATION

EMERGENCE OF DATA ANALYTICS

1970s 1990s 2010

Decision Support Systems → Business Intelligence → Analytics (Kerren et al., 2014)

Evolution

- B&A 1.0: DBMS-Based, Structured Content
- B&A 2.0: Web-Based, Unstructured Content
- B&A 3.0: Mobile and Sensor-Based Content

Applications

- E-Commerce and Market Intelligence
- E-Government and Politics 2.0
- Science & Technology
- Smart Health and Wellbeing
- Security and Public Safety

Emerging Research

- (Big) Data Analytics
- Text Analytics
- Web Analytics
- Network Analytics
- Mobile Analytics

(Chen, Chiang & Storey, 2012)

OBJECTIVES

The main objective: to perform data analytics pertaining to university operations using open source tools like Hadoop framework and R, taking a case of the Technical University of Kenya
PROBLEM STATEMENT

THE TECHNICAL UNIVERSITY OF KENYA

TOP MANAGEMENT
(VICE CHANCELLOR, DEPUTY VICE CHANCELORS)

EXECUTIVE DEANS, ACADEMIC REGISTRAR, DIRECTORS

OPERATIONAL MANAGERS (HEADS OF DEPARTMENTS, REGISTRARS, ETC)

TECHNICAL STAFF

INFORMATION/KNOWLEDGE

INFORMATION REQUIRED:
- Accurate number of students & staff
- Student admission trend
- Staff composition
- Publication impact
- Programme success factor
- Webometric ranking trend
- Etc

CHALLENGES/ISSUES:
- Too much time required for report generation
- Staff busy with report generation instead of working
- Incomplete data for important decision making
- Rigid and stale reports
- Periodical reports

DATA (FACTS) CONCERNING APPLICANTS, STUDENTS, STAFF, SUPPLIERS, ETC

LOG FILES, EXTERNAL DOCUMENTS, MEMOS, EMAILS, ETC
METHODOLOGY

Surveys
Background Literature
Interviews
Observations
Experiences

Design Tool Selection
(Open source tools)
System Design
Setup and Configuration
Programming/ Coding

Data Extraction
Data Exploration
Filtering/ Cleaning
Transformation
Data Visualization
(Dashboards)

Evaluation & User Feedback
METHODOLOGY

ANALYTICAL SYSTEM ARCHITECTURE

DECISION MAKERS

BI web portal (Powered by HUE) – single source of truth

STRATEGY

KNOWLEDGE

INFORMATION

INTEGRATED DATA

DISPERSED DATA

University Data Warehouse
(HIVE, IMPALA)

ETL (SQOOP, HIVE MetaStore, Flume)

External data sources

Student MIS

Students’ Portal

E-Learning Portal

Data Mining (i)

 Dashboards
( Cloudera Search & SOLR)

Online Analysis
(Data Visualization)

Standard & Ad hoc Reports
(HIVE Queries)

Online Analysis
(Data Visualization)

Dashboards
( Cloudera Search & SOLR)

External data sources

Student MIS

Students’ Portal

E-Learning Portal

ETL (SQOOP, HIVE MetaStore, Flume)
METHODOLOGY

SYSTEM FEATURES

BI PORTAL LOGIN

ININAUGURAL RESEARCH CONFERENCE BY UNIVERSITY OF SUNDERLAND, UK AND INTEL COLLEGE, NAIROBI ON 6 NOVEMBER 2015
METHODOLOGY

SYSTEM FEATURES

SAMPLE DASHBOARD

Student Admission Pattern
RESULTS

- THERE WAS CONSISTENT AND RELIABLE INFORMATION (USERS MORE CONFIDENT ON THE INFORMATION GENERATED)

- THERE WAS FASTER ACCESS TO REQUIRED INFORMATION (EXAMPLE STAFF COMPOSITION REPORT THAT WAS REQUIRED BY THE AUDITOR GENERAL COULD BE GENERATED WITHIN HOURS INSTEAD OF WEEKS)

- THERE WAS A STUDENT PROFILE REPORT THAT WAS USED TO DETERMINE THE NEEDINESS LEVELS OF STUDENTS IN ALLOCATION OF ROOMS FOR ACCOMMODATION

- STAFF ETHNIC BALANCE WAS USED IN INFLUENCING THE HIRING OF NEW STAFF MEMBERS

- USERS WERE ABLE TO IDENTIFY NEW INFORMATION THAT WAS NOT PREVIOUSLY AVAILABLE: FOR EXAMPLE STAFF DISTRIBUTION BASED ON JOB GROUPS EXPOSED A JOB GROUP THAT WAS CROWDED
CONCLUSION

Although not all data sources were included, there were tangible benefits in the Technical University. Hence, replication can be done in other universities.

CHALLENGES FACED

- Data cleaning and transformation difficult
- Competely unretrivable data
- Management still not confident on evidence-based decision making
- Lack of data analytical skills in university community
- Data from external sources like emails, social media still do not have a lot of relevant information required by management
- Modeling for predictive analytics difficult due to missing/less data

Security

Public service (Public perception on government)

Health sector (Disease diagnosis, possibility of outbreaks, fake drugs)

Banking sector (Fraud detection, possibility of loan defaulting)

Insurance (Fraudulent claims)
REFERENCES


