



Polash ICT, Manikganj, Dhaka, Bangladesh

# Big data analytics for better life

Better lives means being able to access jobs and income, environmental quality and a good social environment in the places where people live.

Darraj Hossain

November 28, 2020, Saturday, Austin, TX



# Agenda

Warmup

About Me

Big data Analytical Model

Lab Google FORM

Live Demo Exit Ticket & 10 min Break

Road map of pilot project

Volunteer Sign Up

Workshop



**Project, Developer** of Doctors Prescription Databank Multimedia under Ministry of Science Information & Communications IT Fair, Participated in IT Fair in NSU, Notre Dame College, S. University, SAAP Conference



**IT Reporting**, Optical Fiber & ISP in Satellite TV ATN (Pic1 & Pic2)  
**IT Lecture**, Information Retrieving Process and Career Goal Setup (Village School and College-Pic3 & Pic4)  
**International IT Workshop**: Participated on Int. Nationwide Internet Access and Online Application (Pic5)



**Write-up**: Oracle & IT articles in ICON(3), Computer Bichitra(26), Financial Mirror(20), Dhaka Courier(1) & Ittef(1)  
**Content Management**: BBC, VOA, ETV, BTV, NTV, ATN, HC, CI (Pic5 & Pic6)



**Exclusive Training**, Oracle, Web Technology, HP, Telecommunication, TQM, & BPM,



## About ME



## Portfolio of Data Analyst skill set

& Fellow member



Association for  
Computing Machinery



Canada's Association of  
I.T. Professionals

1985

Research Assistant,  
 -Dhaka University and National Academy of Science, D.C., USA  
 -National Institute of Population Research and Training (NIPORT)  
 -The Department of Urban and Regional Planning at BUET  
 SW: dBase III Plus, Lotus, SPSS/PC+, Harvard Graphics

1995

Data Processor and Analyst, Data Computerization Project,  
 -National Board of Revenue- The World Bank  
 SW: Paradox RDMS, dBase III Plus, Lotus, SPSS/PC+, Harvard

Project Director, a research project as granted by  
 -The Ministry of Science Information & Communications  
 Technology, Govt. of Bangladesh, Medical SW for public health.  
 SW: MS Access, Excel, Micro Media Flash and Web Application

2000

Datacenter CCBS Manager, CDR (Called Details Record), Telco Robi  
 in 1997 as TMI (Bangladesh) with the brand name AKTEL GSM.  
 SW: Sybase, Oracle RDMS, Excel

2010

Data Analyst and IT support  
 -SYNAP IT Corporation, New York, NY  
 -Document Technologies, Inc., Philadelphia, PA  
 -ISP, Food Bank, YSCP Charter School, State College, PA  
 -Micro Datacenter and Data Warehouse, Round Rock, TX.  
 SW: Microsoft SQL Server, Oracle,

2020

Big Data and Data Warehouse Analyst and Hadoop Open Source  
 -Micro Datacenter and Data Warehouse  
 SW: Linux, Hadoop, HiveQL, Spark, Sqoop, Hbase

## Bytes Information

Processor or Virtual Storage

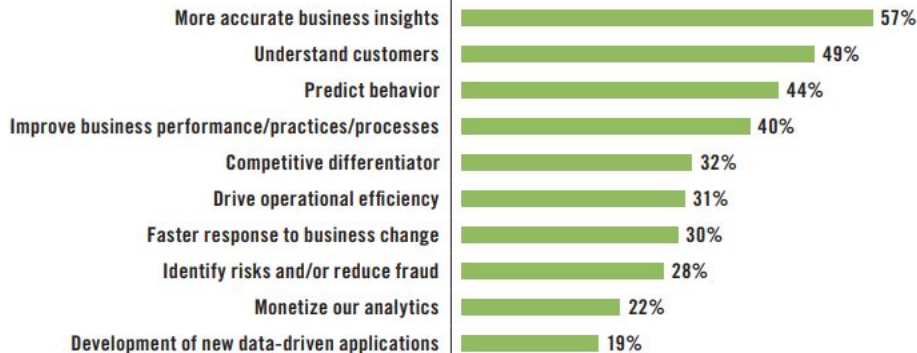
Disk Storage

• 1 bit	=Binary digit	• 1 bit	=Binary digit
• 8 bits	=1 Byte	• 8 bits	=1 Byte
• 1024 byte	=1 kb	• 1000 byte	=1 kb
• 1024 kb	=1 mb	• 1000 kb	=1 mb
• 1024 mb	=1 gb	• 1000 mb	=1 gb
• 1024 gb	=1 Terabyte	• 1000 gb	=1 Terabyte
• 1024 tb	=1 Petabyte	• 1000 tb	=1 Petabyte
• 1024 pb	=1 Exabyte	• 1000 pb	=1 Exabyte
• 1024 eb	=1 Zettabyte	• 1000 eb	=1 Zettabyte
• 1024 zb	=1 Yottabyte	• 1000 zb	=1 Yottabyte
• 1024 yb	=1 Brontobyte	• 1000 yb	=1 Brontobyte
• 1024 bb	=1 Geopbyte	• 1000 bb	=1 Geopbyte

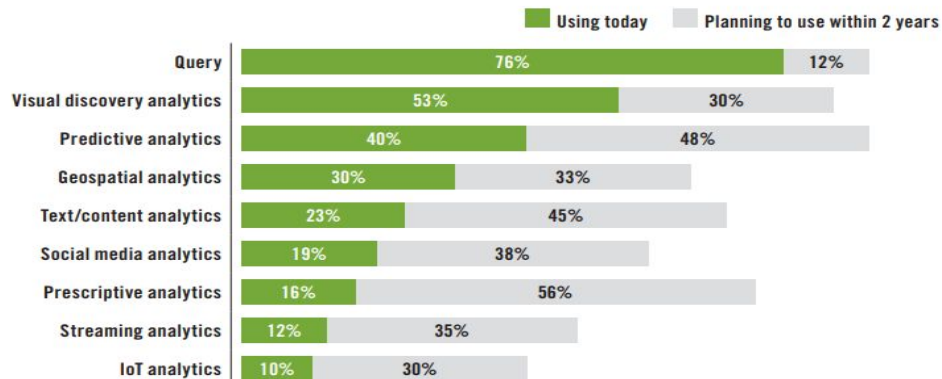
62

# Big Data Volume, Variety & Velocity

Percentage of respondents ranking the following drivers for big data deployments "extremely important" in their organization or company.



What are your plans for the big data analytics listed below?



# Big data Analytical Model

## Infrastructure V0.07



Cisco Small Business Router



Wifi Cisco Small Business Router



Cisco Small Business Smart Switch



WD Sentinel Server - Xeon E3



Seagate Personal Cloud NAS Server (Backup - Set 1)



Seagate Backup Plus External Hard Drive (Backup - Set 2)

cloudera

amazon web services

ubuntu Linux



hadoop

The Data Warehouse

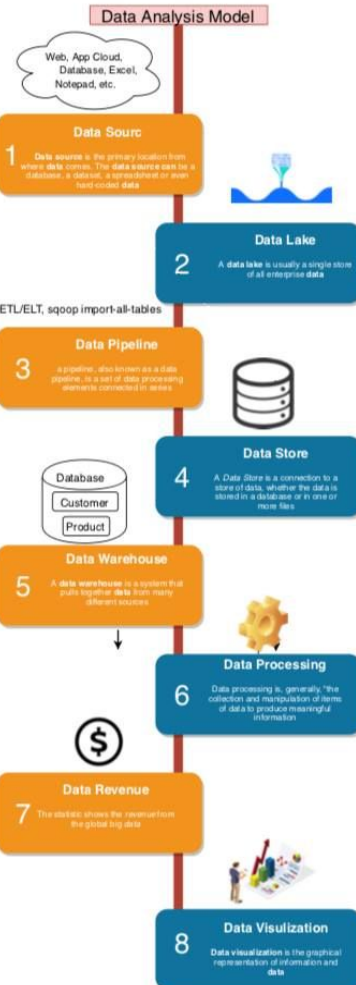


Excel



1M Data

Data Set:  
01. Healthcare (Record-46935)  
02. Weather-history (Record-4575)  
03. World Demography  
04. Census Tracts 2010 (Record-21)  
05. Animal\_Services\_Intake\_Data  
06. Average\_Daily\_Traffic\_Counts  
07. Accidental\_Drug\_Retained\_Death  
08. Popular\_Baby\_Names (Record  
09. Sales\_Tax\_Rates (Record-1911)  
10. Restaurants (Record-1309)  
11. Accidental\_Drug\_Related\_Death  
12. Census Tracts 2010 (Record-216)  
13. Employees\_Salary (Record-824)  
14. Customer\_transactional\_spending  
15. Customer\_Order (Record-1000)  
16. Employees\_Salary (Record-824)



Power by Open Source Platform



Excel/CSV/Text File Format:  
John,30,New York  
Robert,25,DC

JSON, Web Application File Format : (  
"employees":{  
"name":"John", "age":30, "city":"New York"}  
}

Text File/Sequence File  
RC File/ORC File  
AVRO File  
Parquet/Binary File

Linux O/S,  
Hadoop Big Data,  
Hadoop File System(HDFS)  
3 replication Datanode

SELECT \* FROM Customers;  
SELECT Country FROM Customers;  
SELECT \* FROM Customers  
WHERE Country='Mexico';





# Lab - Data Analytical Platform



GitHub is a Git repository hosting service, but it adds many of its own features. While Git is a command line tool, GitHub provides a Web-based graphical interface. It also provides access control and several collaboration features, such as a wikis and basic task management tools for every project.



Google Colab is a free cloud service and now it supports free GPU! You can; improve your Python programming language coding skills. develop deep learning applications using popular libraries



Apache Hive is a data warehouse system for data summarization and analysis and for querying of large data systems in the open-source Hadoop platform. It converts SQL-like queries into MapReduce jobs for easy execution and processing of extremely large volumes Dataset



Power BI is a business analytics service by Microsoft. It aims to provide interactive visualizations and business intelligence



Google Forms is a tool that allows collecting information from users through a personalized survey or quiz. The information is then collected and automatically connected to a spreadsheet. The spreadsheet is populated with the survey and quiz response



Microsoft Forms is a simple, lightweight app that lets you easily create surveys, quizzes, and polls. In educational institutions, it can be used to create quizzes, collect feedback from teachers and parents, or plan class and staff activities

Cloud H/W & S/W platform



# Exit Ticket



Darrajit@gmail.com

