Introduction to Big Data: Utilising Analytics in Investigation Process

Do you know how to leverage the industrial revolution as your business weapon?
Always remember, if you can measure, you can plan. Do you know how to capture quality data?

Introduction

This Intro to Big Data is a unique approach to help you act on data for real business gain – not what a tool can do, but what you can do with the output from the tool. Big data, as defined by Wiki, is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. In this hands-on Introduction to Big Data Course, learn to leverage big data analysis tools and techniques to foster better business decision-making – before you get into specific products like Hadoop training. Learn ways of storing data that allow for efficient processing and analysis, and gain the skills you need to store, manage, process, and analyse massive amounts of unstructured data to create an appropriate data lake.

Program Objectives

This program aims to:

- Understand the dimensions and component of big data
- Understand the challenges in compiling data and identifying the importance of each data
- Learn how to store the data the right way by the needs and legal perspective
- Understand how big data stores work and its functions
- Learn on how to use the tools to process big data
- Tools to be used in analysing big data
- How to use big data in planning and developing strategies

Learning Outcomes

After completing the training, participants should be able to:

- Store, manage and analyse unstructured data
- Select the correct big data stores for disparate data sets
- Process large data sets using Hadoop to extract value
- Query large data sets in near real-time with Pig and Hive
- Plan and implement a big data strategy for your organisation

Who should attend?

Anyone needing to implement, enhance your big data environment and looking to advance their analytics career by ensuring foundational knowledge, Typical job roles include: Project Managers and IT Managers, Database Administrators & Data Architects, Developers & SQL Developers, Data Scientists & Business Intelligence

Program Outline

Time	Day One
9.00am– 10.30am	**Introduction to Big Data** In this module, the participants would learn the definition of Big Data. The participants would learn the four dimensions of Big Data: volume, velocity, variety, veracity. The participants would be exposed to the Storage, MapReduce and Query Stack.
10.30am-11.00am	**Break and Networking**
11.00am-1.00pm	**Delivering Business Benefit from Big Data** The participants would learn how to establish the business importance of Big Data, address the challenge of extracting useful data and how to integrate Big Data with traditional data
1.00pm-2.00pm	**Lunch Break and Networking**
2.00pm-3.30pm	**Analysing Your Data Characteristics** In this module, the participants would learn the methodology to select data sources for analysis, eliminate redundant data and establish the role of NoSQL.
3.30pm-4.00pm	**Break and Networking**
4.00pm-5.00pm	**Overview of Big Data Stores and Selecting Big Datastores** The participants would learn the topics such as data models: key-value, graph, document, column–family, Hadoop Distributed File System, HBase, Hive, Cassandra, Hypertable, Amazon S3, BigTable, DynamoDB, MongoDB, Redis, Riak and Neo4J. At the same time, the participants would learn how to choose the correct data stores based on your data characteristics, move code to data, implement polyglot data store solutions, align business goals to the appropriate datastore

Time	Day Two
9.00am– 10.30am	**Processing Big Data** In this module, the participants would learn four main topics - Integrating disparate data stores, employing Hadoop MapReduce, handling streaming data and the building blocks of Hadoop MapReduce. When integrating disparate data stores, the participants would learn mapping data to the programming framework, connecting and extracting data from storage, transforming data for processing and subdividing data in preparation for Hadoop MapReduce. In employing Hadoop MapReduce, the participants would learn the method of creating the components of Hadoop MapReduce jobs, distributing data processing across server farms, executing Hadoop MapReduce jobs and monitoring the progress of job flows. In this last topic, the building blocks of Hadoop MapReduce, the participants would learn how to distinguish Hadoop daemons, investigate the Hadoop Distributed File System, select appropriate execution modes: local, pseudo-distributed and fully distributed. In the last topics, the participants would learn comparing real–time processing models, leveraging Storm to extract live events and lightning–fast processing with Spark and Shark.
10.30am-11.00am	**Break and Networking**
11.00am-1.00pm	**Abstracting Hadoop MapReduce jobs: Tools of Analysis** The participants would start communicating with Hadoop, executing commands using the Grunt Shell and streamlining high–level processing.
1.00pm-2.00pm	**Lunch Break and Networking**
2.00pm-3.30pm	**Performing Ad Hoc Big Data Querying with Hive and Creating Business Value from Extracted Data** In this module, the participants would learn the topics includes persisting data in the Hive Mega Store, performing queries with HiveQL and investigating Hive file formats. Then, the participants would learn the techniques of mining data with Mahout, visualising processed results with reporting tools and querying in real-time with Impala

3.30pm-4.00pm	**Break and Networking**
4.00pm-5.00pm	**Developing a Big Data Strategy** In this last module, the participants would learn to develop a big data strategy and enable analytic innovation. First, the participants are required in establishing your Big Data needs, meeting business goals with timely data, evaluating commercial Big Data tools and managing organisational expectations. In enabling analytic innovation, the participants would look at how-to focusing on business importance, focusing on business importance, framing the problem, selecting the correct tools and achieving timely results.