Part 2: Training Programme Details (ADE)

Section A: Course Details

1	Course Title	Applied Data Engineering
2	Type	Technical
3	Training Methodology	Classroom (physical; with virtual as option)
4	Skill Area	Data Analytics Lifecycle and Implementation Options Data Analysis Techniques Build Real-time Stream Analytics Complex Data Management Techniques with Various Technologies Build Application Visualization Techniques Analytics Presentation Skills
5	Duration	5 days / 35 hours
6	Certification	Certificate of Completion
7	Certification Body	N/A.
8	Course Overview	Most of the companies are generating/collecting more and more data by day. Efficient data management shorten the data exploration time and improves the quality of the data. These are essentially important in enabling finding insights to improve competitiveness. This course enables participant to familiarize with all the relevant components of Big Data. Participants will also be able to perform data analysis, build real-time stream analytics, execute complex data management, build apps, and create visualization on top of data.
9	Prerequisites	Diploma / Degree or Equivalent
10	Course Objective	Upon completion of this course, participants will be familiarized with all the relevant components of Big Data. Participants will also be able to perform data analysis, build real-time stream analytics, execute complex data management, build apps, and create visualization on top of data.
11	Learning Outcome	By the end of the training, participants will be familiarized with all the relevant components of Big Data. Participants will also be able to perform data analysis, build real-time stream analytics, execute complex data management, build apps, and create visualization on top of data.

12	Course Content	Day1 - Introduction to Big Data and Hadoop
		- Course Introduction and Overview of Big Data - Industry use cases - Big Data Adoption Success Factor - What is Data Analytics - Hadoop Architecture and Ecosystem - Apache Hadoop Overview - Data Ingestion, Storage, Processing, Exploration, Analysis Tools - Getting Data into Hadoop - Data Ingestions - Accessing HDFS - Import and manage with HUE Day2 - Data Processing and Analysis - Data Ingestion - import from RDBMS - Import RDBMS Data with Sqoop - Features for importing RDBMS data into Hadoop - Introduction to Hive and Impala - Query with Hive and Impala - Contrasting Hive and Impala with relational databases - Data Processing and Management - Creating databases and tables - Impala metadata caching Day3 - Processing Log Files - Data Analysis with Hive/Impala - Data partitioning - Data analysis using Hive and Impala - Visual Based Data Processing and Analysis - Data ingestion by connecting to sources - Manage Metadata - Data Cleansing & Transformation - Log File Ingestion with Flume - Flume architecture - Configuration of Flume for data ingestion Day4 - Spark and RDDs - Analyzing Log File data with Hive/Impala - Create a table with Hive - Parsing - Introduction to Spark - Spark Use cases - Hadoop MapReduce (MR) versus Spark - RDDs - Single-RDD and Multi-RDD transformations - PairRDD: MapReduce - Spark Application - Spark application configuration - Spark application web UI

		Day5 – Visualization
		  • Spark Streaming
		    ◦ Initializing spark streaming context
		    ◦ Overview of Kafka
		  • Spark SQL
		    ◦ SparkSQL architecture
		    ◦ DataFrame versus RDDs
		  • Data Visualization
		    ◦ The importance of Data Visualization
		    ◦ D3 data visualization
		  • Project

13	Learning Activities	Lecture, Practical Exercise, Case Studies, Learning Activities, Video Presentation, Training
14	Target Group	This course prepares participants to become Data Engineers and set them on the path to become Data Scientists. Participants should preferably have some knowledge in Python, SQL, HTML, or JavaScript.

Detailed Daily Training Schedule (Course Content / Hours)

No.	Content/Activity	Objectives	Outcome	Hours
1	Day1 [9am-11am] - Overview of Big Data Analytics - Industry Use Cases	This section covers overview of Big Data Analytics, and Industry Use Cases	After this section, participants are able to understand Big Data Analytics, and aware of Industry Use Cases	2.0
2	Day1 [11.15am-12.45pm] - Hadoop Architecture and Ecosystem	This section covers Hadoop Architecture and Ecosystem	After this section, participants are able to understand Hadoop Architecture and Ecosystem	1.5
3	Day1 [1.45pm-3.45pm] - Getting Data into Hadoop - Data Ingestions	This section covers Data Ingestions into Hadoop	After this section, participants are able to perform Data Ingestions into Hadoop	2.0
4	Day1 [4pm-5.30pm] - Getting Data into Hadoop - Data Ingestions (continue)	This section covers Data Ingestions into Hadoop (continue)	After this section, participants are able to Import data and manage with HUE	1.5

5	Day2 [9am-11am] - Data Ingestion - import from RDBMS	This section covers Data Ingestion - import from RDBMS	After this section, participants are able to understand and perform Data Ingestion - import from RDBMS	2.0
6	Day2 [11.15am-12.45pm] - Introduction to Hive and Impala	This section covers Introduction to Hive and Impala	After this section, participants are able to understand Hive and Impala	1.5
7	Day2 [1.45pm-3.45pm] - Hive and Impala (continue)	This section covers Query with Hive and Impala; Contrasting Hive and Impala with relational databases	After this section, participants are able to Query with Hive and Impala; and contrasting Hive and Impala with relational databases	2.0
8	Day2 [4pm-5.30pm] - Data Processing and Management	This section covers creating databases and tables; Impala metadata caching	After this section, participants are able to create databases and tables; and perform Impala metadata caching	1.5
9	Day3 [9am-11am] - Data Analysis with Hive/Impala	This section covers Data partitioning; Data analysis using Hive and Impala	After this section, participants are able to perform Data partitioning; Data analysis using Hive and Impala	2.0
10	Day3 [11.15am-12.45pm] - Visual Based Data Processing and Analysis	This section covers Data ingestion by connecting to sources and Manage Metadata	After this section, participants are able to performance Data ingestion by connecting to sources and Manage Metadata	1.5
11	Day3 [1.45pm-3.45pm] - Visual Based Data Processing and Analysis (continue)	This section covers Data Cleansing & Transformation	After this section, participants are able to perform Data Cleansing & Transformation	2.0
12	Day3 [4pm-5.30pm] - Data Processing and Management	This section covers Flume architecture and Configuration of Flume for data ingestion	After this section, participants are able to understand Flume architecture and	1.5

			Configure Flume for data ingestion	
13	Day4 [9am-11am] - Analyzing Log File data with Hive/Impala	This section covers Create a table with Hive and Parsing	After this section, participants are able to Create a table with Hive and Parsing	2.0
14	Day4 [11.15am-12.45pm] - Introduction to Spark	This section covers Spark Use cases and Hadoop MapReduce (MR) versus Spark	After this section, participants are able to understand Spark Use cases and Hadoop MapReduce (MR) versus Spark	1.5
15	Day4 [1.45pm-3.45pm] - RDDs	This section covers Single-RDD and Multi-RDD transformations and PairRDD: MapReduce	After this section, participants are able to understand Single-RDD and Multi-RDD transformations and perform PairRDD: MapReduce	2.0
16	Day4 [4pm-5.30pm] - Spark application	This section covers Spark application configuration and Spark application web UI	After this section, participants are able to understand Spark application and configure Spark application web UI	1.5
17	Day5 [9am-11am] - Spark Streaming	This section covers Initializing spark streaming context; Overview of Kafka	After this section, participants are able to understand Initializing spark streaming context and Overview of Kafka	2.0
18	Day5 [11.15am-12.45pm] - Spark SQL	This section covers SparkSQL architecture and DataFrame versus RDDs	After this section, participants are able to understand SparkSQL architecture and DataFrame versus RDDs	1.5
19	Day5 [1.45pm-3.45pm] - Data Visualization	This section covers the importance of Data Visualization and D3 data visualization	After this section, participants are able to understand the importance of Data	2.0

			Visualization and D3 data visualization	
20	Day5 [4pm-5.30pm] - Project	This section has a project to implement what have learnt	After this section, participants are able to implement what have learnt through a project	1.5