|  |
| --- |
| **Java Fundamentals** |
| Basic java concepts |
| Multi-threading |
| File I/O –Java. IO |
| Collections –Java.Util.\*, Java.Math, Java.Lang |
| Java Generics |
| Java Serialization |
| Java Database Connectivity –JDBC |
| Java Common Design Patterns |
| Java Open Source Frameworks (Spring, Apache Maven, Logging, etc...) |
| Java Apache Hadoop Frameworks (Hadoop Common, Map Reduce etc.) |
| Understand Web Servers & Application Servers - JBoss Application server, Apache Tomcat server |
| Java Unit testing Frameworks (Junit / TestNG) |
| Eclipse IDE – Java Development. |
| Version Control – GIT, SVN, etc. |
| Java Continuous Integration frameworks – Husdson, Jenkins, etc. |
| Handling XML and XSD using Java frameworks |
| Java XML Parsers frameworks – DOM and SAX |
| Java Web services concepts – SOA, SOAP, XML, JAXB, |
| SOAP Web services |
| REST web services |
|  |
| **Hadoop Fundamentals** |
| What is Big Data? Why Big Data? |
| Hadoop Architecture & Components |
| Hadoop Storage & File Formats (ASCII, Avro, Parquet, RC4, JSON, EBCDIC etc.) |
| Hadoop Processing – Map Reduce, Spark Frameworks |
|  |
| **Map Reduce** |
| What Is MapReduce? |
| Basic MapReduce Concepts |
| Concepts of Mappers, Reducers, Combiners and Paritioning |
| Inputs and Output formats to MR Program |
| Error Handling and creating UDFs for MR |
|  |
| **Spark** |
| What Is Spark? |
| Basic Spark Concepts |
| How Spark differs from Map Reduce? |
| Working with RDD’s |
| Parallel Programming with Spark |
| Spark Streaming |
|  |
| **Hive** |
| What is Hive, why we need it and its importance in DWH? |
| How Hive is different from Traditional RDBMS |
| Modeling in Hive, creating Hive structures and data load process. |
| Concepts of Partitioning, Bucketing, Blocks, Hashing, External Tables etc. |
| Concepts of serialization, deserialization |
| Different Hive data storage formats including ORC, RC, and Parquet. |
| Introduction ton HiveQL and examples. |
| Hive as an ELT tool and difference between Pig and Hive |
| Performance tuning opportunities in Hive, learnings and Best Practices. |
| Writing and mastering Hive UDFs |
| Error Handling and scope of creating Hive UDFs. |
|  |
| **Pig and Latin** |
| Basics of Pig and Why Pig? |
| Grunt |
| Pig’s Data Model |
| Writing Evaluation |
| Filter |
| Load & Store Functions |
| Benefits of Pig over SQL language |
| Input and Output formats to MR program. |
| Error Handling and scope of creating UDFs for Pig. |
|  |
| **HBase** |
| HBase – Introduction |
| When to use HBase |
| HBase Data Model |
| HBase Families & Components |
| Data Storage and Distribution |
| HBase Master |
|  |
| **MongoDB** |
| Introduction to In-Memory Computing |
| When to use MongoDB |
| MongoDB API |
| Indexing and Data Modeling |
| Drivers / Replication / Sharding |
|  |
| **ETL / ELT Solutions Build Workshop** |
| Java |
| MapReduce |
| Pig |
| Hive |
| HBase |
| Cassandra |
| Talend Open Studio |
| Cloudera Morphlines - Kite SDK |
| Impala |
| Mongo DB |