SHI TTDS6000 Data Science and Big Data Overview User Guide

Contents

SHI TTDS6000 Data Science and Big Data Overview

About this course

The Data Science & Big Data Overview | Tools, Tech & Modern Roles in the Data-Driven Enterprise is an introductory level course that introduces the entire multi-disciplinary Data Science team to the many evolving and related terms, with a focus on Big Data, Data Science, Predictive Analytics, Artificial Intelligence, Data Mining, Data Warehousing.

The overview explores the current state of the art and science, the major components of a modern data science infrastructure, team roles and responsibilities, and level-setting realistic possible outcomes for your investment. The goal of this course is to provide students with a baseline understanding of core concepts and technologies to a conversant level.

This course provides a high-level view of a variety of core, current data science-related technologies, strategies, skillsets, initiatives and supporting tools in common business enterprise practices. This list covers a general range of topics current to the time of course distribution. We will collaborate with your team to refine the level of depth of coverage, understand areas of greater importance to your team, where you would like to add demos, etc.

Audience profile

This introductory-level / primer course is an overview intended for Business Analysts, Data Analysts, Data Architects, DBAs, Network (Grid) Administrators, Developers or anyone else in the data science realm who needs to have a baseline understanding of some of the core areas of modern Data Science technologies, practices and available tools. Attendees should have prior exposure to Enterprise Information Technology. As well as familiarity with Relational Databases.

At course completion

After completing this course, students will be able to:

- Foundations: Grids & Virtualization; SOA, ESB / EMB, The Cloud
- The Hadoop Ecosystem: HDFS; Resource Navigators, MapReduce, Spark, Distributions
- Big Data, NOSQL, and ETL
- ETL: Exchange, Transform, Load
- Handling Data & a Survey of Useful tools
- Enterprise Integration Patterns and Message Busses
- Developing in Hadoop Ecosystem: R, Python, Java, Scala, Pig, and BPMN
- Artificial Intelligence and Business Systems
- Who’s on the Team? Evolving Roles and Functions in Data Science
- Growing your Infrastructure

Course Outline Foundations

- Grids and Virtualization
- Service-Oriented Architecture
- Enterprise Service Bus
- Enterprise Message Bus
- The Cloud
- The Hadoop Ecosystem
- HDFS: Hadoop Distributed File System

Resource Negotiators

- YARN, Mesos, and Spark; ZooKeeper
- Hadoop Map/Reduce
- Spark
- Hadoop Ecosystem Distributions: Cloudera, Hortonworks, OpenSource
- Big Data, NOSQL, and ETL
- Big Data vs. RDBMS
- NOSQL: Not Only SQL
- Relational Databases: Oracle, MariaDB, DB/2, SQL Server, PostgreSQL
- Key/Value Databases: JBoss Infinispan, Terracotta, Dynamo, Voldemort
- Columnar Databases: Cassandra, HBase, BigTable
- Document Databases: MongoDB, CouchDB/CouchBase
- Graph Databases: Giraph, Neo4J, GraphX
- Apache Hive
- Common Data Formats
- Leveraging SQL and SQL variants
- ETL: Exchange, Transform, Load
- Data Ingestion, Transformation, and Loading
- Exporting Data
- Sqoop, Flume, Informatica, and other tools
- Enterprise Integration Patterns and Message Busses
- Enterprise Integration Patterns: Apache Camel and Spring Integration
- Enterprise Message Busses: Apache Kafka, ActiveMQ, and other tools
- An Overview of Developing in the Hadoop Ecosystem
- Languages: R, Python, Java, Scala, Pig, and BPMN
- Libraries and Frameworks
- Development, Testing, and Deployment

Exploring Artificial Intelligence and Business Systems

- Artificial Intelligence: Myths, Legends, and Reality
- The Math
- Statistics
- Probability
- Clustering Algorithms, Mahout, MLLib, SciKit, and Madlib
- Business Rule Systems: Drools, JRules, Pegasus

The Modern Data Team

- Agile Data Science
- NOSQL Data Architects and Administrators
- Developers

- Grid Administrators
- Business and Data Analysts
- Management
- Evolving your TeamGrowing your Infrastructure

Documents / Resources

	SHI TTDS6000 Data Science and Big Data Overview [pdf] User Guide TTDS6000 Data Science and Big Data Overview, TTDS6000, Data Science and Big Data Over view, Science and Big Data Overview, Big Data Overview, Data Overview, Overview

References

- User Manual