

# DATA SCIENCE OVERVIEW | TECHNOLOGIES, TOOLS, AND ROLES IN THE DATA-DRIVEN ENTERPRISE

Course Code: 8779

Gain a baseline understanding of the core concepts, tools, and roles in Data Science.

This foundation-level level course introduces the multi-disciplinary Data Science team to the many evolving and related terms. It includes a focus on Big Data, Data Science, Predictive Analytics, Artificial Intelligence, Data Mining, and Data Warehousing. You'll also explore the current state of the art and science, the major components of a modern data science infrastructure, team roles and responsibilities, and level-setting of possible outcomes for your investment.

This course provides a high-level view of current data science related technologies, concepts, strategies, skillsets, initiatives and supporting tools in common business enterprise practices. This goal of this course is to provide you with a baseline understanding of core concepts.

**Learn more about this topic. View the recorded webinar [AI + Coronavirus + DI: Using Technology to Restart Your Business Safely](#)**

## What You'll Learn

Join an engaging learning environment, where you'll explore:

- Foundations: Grids & Virtualization; SOA, ESB/EMB and the Cloud
- The Hadoop Ecosystem: HDFS, Resource Navigators, MapReduce, Spark, and Distributions
- Big Data, NOSQL, and ETL
- ETL: Exchange, Transform, Load
- Handling Data and 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? Roles and Functions in Data Science
- Growing your Infrastructure

This is a seminar-style course that combines engaging expert lectures, pertinent skills, tool demonstrations, and group discussions.

## Who Needs to Attend

Business Analysts, Data Analysts, Data Architects, Database Administrators, Network Administrators (Grid), Developers, Technical Manager, or anyone else in the data science realm who needs to have a baseline understanding of the core areas of modern Data Science technologies, practices, and tools.

## Prerequisites

Attendees should have:

- Exposure to Enterprise Information Technology
- Familiarity with Relational Databases

# DATA SCIENCE OVERVIEW | TECHNOLOGIES, TOOLS, AND ROLES IN THE DATA-DRIVEN ENTERPRISE

Course Code: 8779

VIRTUAL CLASSROOM LIVE

\$995 USD

1 Day

## Virtual Classroom Live 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

### **Developing in Hadoop Ecosystem**

- Languages: R, Python, Java, Scala, Pig, and BPMN
- Libraries and Frameworks
- Development, Testing, and Deployment

### **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 Team**

- Agile Data Science
- NOSQL Data Architects and Administrators
- Developers
- Grid Administrators
- Business and Data Analysts
- Management
- Evolving your Team
- Growing your Infrastructure

Sep 12 - 12, 2025 | 10:00 AM - 6:00 PM EST

Oct 24 - 24, 2025 | 10:00 AM - 6:00 PM EST

Dec 5 - 5, 2025 | 10:00 AM - 6:00 PM EST



# DATA SCIENCE OVERVIEW | TECHNOLOGIES, TOOLS, AND ROLES IN THE DATA-DRIVEN ENTERPRISE

Course Code: 8779

PRIVATE GROUP TRAINING

1 Day

Visit us at [www.globalknowledge.com](http://www.globalknowledge.com) or call us at 1-866-716-6688.

Date created: 8/30/2025 3:12:02 PM

Copyright © 2025 Global Knowledge Training LLC. All Rights Reserved.