



millionlights[®]

Nayi Shiksha, Nayi Soch

BIG DATA
COURSE CONTENT



[I] Get Started with Big Data

Microsoft Professional Orientation: Big Data

Duration: 12 hrs

Course Content:

❖ **Introduction**

- Course Introduction

❖ **Data Fundamentals**

- Introduction to Data and Data Files
- Lab

❖ **Introduction to Relational Databases**

- Database Basics
- Lab

❖ **Introduction to NoSQL Databases**

- Getting Started with NoSQL Databases
- Lab

❖ **Introduction to Big Data Processing**

- Getting Started with Big Data Processing
- Lab

❖ **Next Steps**

- Further Learning





[II] Analyze and Visualize Data

2 TRAINING OPTIONS AVAILABLE

➤ **Option 1: Analyzing and Visualizing Data with Power BI**

Duration:24 hrs

Course content:

❖ **Start Here!**

- Welcome to the Course!
- Set up the Lab Environment
- Data Set / Examples Usage
- Pre-course survey

❖ **Introduction**

- What is Power BI?

❖ **Power BI Desktop Data Transformations**

- Transforming Data using Power BI Desktop
- Lab 1

❖ **Power BI Desktop Modelling**

- Data Modelling in Power BI Desktop
- Lab 2

❖ **Power BI Desktop Visualization**

- Visualizing Your Data
- Working with Multiple Visualizations
- Lab 3

❖ **Power BI Service**

- Working with Power BI Service
- Viewing Power BI Dashboard
- Lab 4

❖ **Working with Excel**

- Connecting and Collaborating with Excel
- Lab 5

❖ **Organization Packs, Security and Groups**

- Working with Others
- Lab 6





❖ **Direct Connectivity**

- Direct Connectivity
- Lab 7

❖ **Developer API**

- Developer API
- Lab 8

❖ **Mobile App**

- Mobile App

❖ **Course Wrap-up**

➤ **Option 2: Analyzing and Visualizing Data with Excel**

Duration:24 hrs

Course content:

❖ **Start Here!**

- Welcome to the Course!
- Set up the Lab Environment
- DataSet / Examples Usage
- Pre-course Survey

❖ **Module 1**

- Data Analysis in Excel
- Lab: Explore and Extend a Classic Excel Dashboard

❖ **Module 2**

- The Excel Data Model
- Basic DAX
- Lab: Explore an Excel Data Model

❖ **Module 3**

- Importing Data from a CSV File
- Lab: Importing Data from a CSV File

❖ **Module 4**

- Importing Data from Databases
- Importing Data from Multiple Files
- Create a Date Table in Excel Data Model
- Lab: Creating Mash-ups of Data from Multiple Sources





❖ **Module 5**

- Creating and Formatting Measures
- Using Advanced DAX Functions
- Lab: Creating Measures using Advanced DAX Functions

❖ **Module 6**

- Importing Data from a Formatted Excel Report
- Lab: Advanced Text Query

❖ **Module 7**

- Visualizing Data in Excel
- Lab: Data Visualizations in Excel

❖ **Module 8**

- Using Excel with Power BI
- Power BI Mobile App

❖ **Course Wrap-up**





[III] Work with NoSQL Data

Introduction to NoSQL Data Solutions

Duration:9 hrs.

Course Content:

❖ Introducing NoSQL Solutions in Azure

- Welcome to the course!
- NoSQL
- NoSQL Solutions in Azure

❖ Azure Storage Tables

- Overview
- Data
- Query and Programming
- Management
- Lab

❖ SQL API in Azure Cosmos DB

- Overview
- Query and Programming
- Data
- Management
- Lab

❖ MongoDB

- Overview
- Query and Programming
- Data
- Management
- Lab

❖ More NoSQL Database Solutions

- Cassandra
- Lucene & Solr
- Azure Search
- HBase
- Redis
- Lab

❖ Final Evaluation

- Exam





[IV] Query Relational Data

Querying Data with Transact-SQL

Duration:30Hrs

❖ **Course content:**

- Before You Start
- Introduction

❖ **Section 1: Modules 1-2**

- Module 1: Introduction to Transact-SQL
- Lab 1: Introduction to Transact-SQL
- Module 2: Querying Tables with SELECT
- Lab 2: Querying Tables with SELECT

❖ **Section 2: Modules 3-5**

- Module 3: Querying Multiple Tables with Joins
- Lab 3: Querying Multiple Tables with Joins
- Module 4: Using Set Operators
- Lab 4: Using Set Operators
- Module 5: Using Functions and Aggregating Data
- Lab 5: Using Functions and Aggregating Data

❖ **Section 3: Modules 6-8**

- Module 6: Using Subqueries and APPLY
- Lab 6: Using Subqueries and APPLY
- Module 7: Using Table Expressions
- Lab 7: Using Table Expressions
- Module 8: Grouping Sets and Pivoting Data
- Lab 8: Grouping Sets and Pivoting Data

❖ **Section 4: Modules 9-11**

- Module 9: Modifying Data
- Lab 9: Modifying Data
- Module 10: Programming with Transact-SQL
- Lab 10: Programming with Transact-SQL
- Module 11: Error Handling and Transactions
- Lab 11: Error Handling and Transactions

❖ **Final Assessment**

- Final Assessment





[V] Create a Data Warehouse

Delivering a Data Warehouse in the Cloud

Duration: 18 hrs

Course content:

❖ **Course Introduction**

- Before You Begin.

❖ **Module 1: Introducing SQL Data Warehouse**

- Introducing SQL Data Warehouse
- Provisioning and Configuring SQL DW
- Reading
- Lab 1
- Module Review and Assessment

❖ **Module 2: Designing and Querying Data Warehouses**

- Table Design and Implementation
- Partitioning Tables
- Indexes and Statistics
- Monitoring Queries
- Reading
- Lab 2
- Module Review and Assessment

❖ **Module 3: Integrating and Ingesting Data**

- Loading Data
- Integrating with other Azure and third-party tools
- Migrating to SQL DW
- Other data movement and load options
- Reading
- Lab 3
- Module Review and Assessment.

❖ **Module 4: Managing Data Warehouses**

- SQL DW Performance
- SQL DW Security
- Managing Compute
- Monitoring SQL DW Workloads with DMVs
- SQL DW Backups





- Reading
- Lab 4
- Module Review and Assessment.

❖ **Final Exam and Survey**

- Final Exam.
- Post-Course Survey
- Quiz





[VI] Process Big Data at Rest

2 TRAINING OPTIONS AVAILABLE

➤ **Option 1: Processing Big Data with Azure Data Lake Analytics**

Duration: 16 hrs

Course content:

❖ **Introduction**

- Course Introduction
- Pre-Course Survey

❖ **Getting Started with Azure Data Lake Analytics**

- Introduction to Azure Data Lake Analytics
- U-SQL Fundamentals
- Lab: Getting Started with Azure Data Lake
- Module Review

❖ **Using a U-SQL Catalog**

- Introduction to U-SQL Catalogs
- U-SQL Database Objects
- Lab: Using a U-SQL Catalog
- Module Review

❖ **Using C# Functions in U-SQL**

- Using C# in U-SQL
- Lab: Using C# in U-SQL
- Module Review

❖ **Monitoring and Optimizing U-SQL Jobs**

- Monitoring U-SQL Jobs
- Optimizing U-SQL Jobs
- Lab 4: Monitoring U-SQL Execution
- Module Review

❖ **Final Challenge**

- Post-Course Survey





➤ **Option 2: Processing Big Data with Azure HDInsight**

Duration:25 hrs.

Course content:

❖ **Course Introduction**

- Introduction
- Pre-Course Survey.

❖ **Module 1: Getting Started with HDInsight**

- Lesson 1: Big Data, Hadoop, and HDInsight
- Lesson 2: Working with HDInsight
- Lab: Getting Started with HDInsight
- Module 1 Review

❖ **Module 2: Processing Big Data with Hive**

- Lesson 1: Working with Hive Tables
- Lesson 2: Developing Hive Applications
- Lab: Processing Big Data with Hive
- Module 2 Review

❖ **Module 3: Going Beyond Hive with Pig and Python**

- Lesson 1: Processing Data with Pig
- Lesson 2: Extending Pig and Hive with UDFs
- Lab: Beyond Hive - Pig and Python
- Module 3 Review

❖ **Module 4: Building a Big Data Workflow**

- Lesson 1: Implementing Workflows with Oozie
- Lesson 2: Transferring Data with Sqoop
- Lab: Orchestrating Big Data Workflows
- Module 4 Review.

❖ **Module 5: Course Exam**

- Course Wrap-Up
- Exam





[VIII] Process Big Data in Motion

2 TRAINING OPTIONS AVAILABLE

➤ Option:1 Processing Real-Time Data Streams in Azure

Duration:16Hrs

Course Content:

❖ Introduction

- Course Introduction
- Pre-Course Survey

❖ Ingesting Real-Time Data with Azure Event Hubs

- Getting Started with Event Hubs
- Lab: Using Azure Event Hubs
- Module Review

❖ Ingesting Real-time Data with Azure IoT Hubs

- Getting Started with IoT Hubs
- Lab: Using IoT Hubs
- Module Review

❖ Getting Started with Azure Stream Analytics

- Introduction to Azure Stream Analytics
- Lab: Getting Started with Azure Stream Analytics
- Module Review

❖ Working with Temporal Windows

- Aggregating Data in Temporal Windows
- Lab: Aggregating Streaming Data
- Module Review

❖ Final Challenge and Post-Course Survey

- Final Challenge
- Post-Course Survey





➤ **Option -2 Processing Real-Time Data with Azure HDInsight**

Duration: 12 hrs

Course Content:

❖ **Course Introduction**

- Introduction
- Pre-Course Survey.

❖ **Using HBase for NoSQL Data**

- Introduction to HBase
- Accessing Data in HBase
- Additional Resources
- Lab
- Review.

❖ **Using Storm for Streaming Data**

- Introduction to Storm
- Implementing Storm Topologies
- Additional Resources
- Lab
- Review.

❖ **Using Spark for Interactive Analysis**

- Introduction to Spark
- Exploring Data with Spark
- Additional Resources
- Lab
- Review.

❖ **Introducing Kafka**

- Introduction to Kafka

❖ **Final Exam and Post-Course Survey**





[IX] Orchestrate Big Data Solutions

Orchestrating Big Data with Azure Data Factory

Duration: 16 hrs

Course content:

❖ **Introduction**

- Course Introduction
- Pre-Course Survey.

❖ **Introduction to Azure Data Factory**

- Getting Started with Azure Data Factory
- Lab: Getting Started with Azure Data Factory
- Module Review Pipelines

❖ **Pipelines**

- Introduction to Pipelines
- Lab: Creating a Pipeline
- Module Review.

❖ **Scheduling Pipelines**

- Overview of Pipeline Scheduling
- Lab: Scheduling a Pipeline
- Module Review.

❖ **Transformations**

- Introduction to Transformations
- Lab: Transforming Data
- Module Review.

❖ **Final Assessment and Post-Course Survey**

- Final Assessment.





[X] Build Big Data Analysis Solutions

3 TRAINING OPTIONS AVAILABLE

➤ **Option 1: Developing Big Data Solutions with Azure Machine Learning**

Duration: 16 hrs

Course content:

❖ **Introduction**

- Before You Start

❖ **Introduction to Azure Machine Learning**

- Getting Started with Azure Machine Learning Studio
- Working with Big Data Sources
- Lab: Getting Started with Azure Machine Learning
- Module Review

❖ **Building Predictive Models with Azure Machine Learning**

- Introduction to Machine Learning
- Clustering and Recommenders
- Lab: Building Predictive Models
- Module Review

❖ **Operationalizing Machine Learning Models**

- Predictive Experiments and Web Services
- Working with Web Services
- Lab: Publishing Predictive Web Services
- Module Review

❖ **Using Azure Machine Learning in Big Data Solutions**

- Using Azure Machine Learning in Batch Processes
- Using Azure Machine Learning in Streaming Processes
- Lab: Building Predictive Big Data Solutions
- Module Review

❖ **Final Exam**

- Final Exam





➤ **Option 2: Analyzing Big Data with Microsoft R**

Duration:16Hrs

Course content:

❖ **Getting Started**

- Getting Started

❖ **Introduction**

- Introduction

❖ **Reading and Preparing Data**

- Reading the Data
- Preparing the Data
- LAB

❖ **Examining and Visualizing Data**

- Examining the Data
- Visualizing the Data
- LAB

❖ **Clustering and Modeling**

- Clustering
- Predictive Modelling.
- LAB.

❖ **Deploying and Scaling**

- Deploying and Scaling

❖ **Final Exam and Wrap-up**

- Final Exam

➤ **Option 3: Implementing Predictive Analytics with Spark in Azure HDInsight**

Duration:24 Hrs

Course Content:

❖ **Course Introduction**

- Introduction
- Pre-Course Survey

❖ **Introduction to Data Science with Spark**

- Getting Started with Spark





- Exploring Data with Spark
 - Further Reading
 - Lab
 - Review
- ❖ **Getting Started with Machine Learning**
- Introduction to Machine Learning in Spark
 - Pipelines and Text Analysis
 - Further Reading
 - Lab
 - Review
- ❖ **Evaluating and Optimizing Machine Learning Models**
- Evaluating Machine Learning Models
 - Optimizing Model Parameters
 - Further Reading
 - Lab
 - Review
- ❖ **Recommenders and Unsupervised Models**
- Recommenders
 - Clustering
 - Further Reading
 - Lab
 - Review
- ❖ **Final Exam and Post-Course Survey**
- Final Exam

