

DTI-DSA-001

# **Fundamentals of US Healthcare and Data Analytics**

# Program Information



**Nature of the Course**  
Theory + Practical



**Total Hours per Day**  
2 hours



**Course Duration**  
2 Months (80 Hours)

## Overview of US Health Care

### Part 1 : Key Data Components

#### Eligibility

Objective: Provide overview of Eligibility file, processing and its analysis steps

- Introduction to Eligibility
- Analysis of Eligibility Data
- Processing Raw File

#### Claims

Objective: Provide overview of Claims file, processing and its analysis steps

- Introduction to Claims
- Analysis of Claims Data
- Processing Raw File

#### Pharmacy

Objective: Provide overview of Pharmacy file, processing and its analysis steps

- Introduction to Pharmacy
- Analysis of Pharmacy Data
- Processing Raw File

#### Provider

Objective: Provide overview of Provider file, processing and its analysis steps

- Introduction to Provider
  - Analysis of Provider Data
  - Processing Raw File
-

## Part 2 : MRF

### Introduction to MRF(Machine Readable Files)

Objective Provide overview of the role and significance of MRF in the US healthcare system

- Introduction to MRF in HealthCare
- MRF standards and formats
- Use Cases and Applications

### Components of a MRF

Objective: Provide knowledge of various components and its significance

- Billing Code
  - Service Code Type
  - Modifier Code
  - Bundle Code
  - Negotiation Arrangements
  - Negotiation Type
  - Billing Class
- 

## Part 3: SQL

### Introduction to SQL

Objective: Introduce Students to the basics of SQL

- Introduction to SQL
- Install and Setup Database

### Basic SQL Queries

- Data Types, Primary-Foreign Keys & Constraints
- DDL and DML Statements

### Create Table In SQL & Create Database

- INSERT UPDATE, DELETE & ALTER Table
  - SELECT Statement & WHERE Clause
  - Aggregate Functions
  - Group By and Having clause
-

## **JOINS in SQL**

- Different Types of JOINS
- SELF JOIN, UNION & UNION ALL
- Use cases of JOIN operation

## **Managing Database Objects**

- Working with Indexes, Views, Synonyms, and Sequences
  - Partitioning and Materialized Views
  - Introduction of PL SQL, Stored Procedure, Functions, Trigger
- 

# **Part 4: Mongo**

## **Introduction to MongoDB**

Objective: Introduce students to the fundamentals of MongoDB

- Overview of NoSQL Databases
- Introduction to MongoDB
- Installing and Setting Up MongoDB

## **Data Modeling in MongoDB**

Objective: Understand the principles of data modeling in MongoDB

- Document-Oriented Data Model
- Collections and Documents
- Schema Design Best Practices

## **MongoDB CRUD Operations**

Objective: Learn the basic CRUD (Create, Read, Update, Delete) operations in MongoDB

- Introduction to Indexing
- Index Strategies
- Performance Optimization Techniques

## **Aggregation Framework**

Objective: Introduce the MongoDB Aggregation Framework for complex data transformations

- Overview of Aggregation Framework
  - Aggregation Operators
  - Real-world Examples
-

# Part 5: PySpark

## Introduction to Apache Spark and PySpark

Objective: Introduce students to the basics of Apache Spark and PySpark

- Introduction to Big Data and Apache Spark
- Introduction to PySpark
- Setting Up a PySpark Environment

## RDDs (Resilient Distributed Datasets)

Objective: Understand the fundamentals building blocks of PySpark – RDD

- Introduction to RDDs
- Creating and Transforming RDDs
- Actions on RDDs

## PySpark DataFrames

Objective: Learn about PySpark DataFrames and their advantages over RDDs

- Introduction to DataFrames
- Creating and Manipulating RDDs
- SQL Operations with DataFrames

## PySpark Use Cases in US Healthcare

Objective: Practice use case scenario with healthcare dataset.

---

# Part 6: ETL using Pyspark and SQL

## Introduction to ETL

Objective: Introduce student to the concept of ETL

- Overview of ETL

## PySpark DataFrames and SQL Queries

Objective: Introduce PySpark DataFrames and SQL queries for data manipulation

- Working with PySpark Dataframe
- Executing SQL Queries with PySpark

## Extracting Data with PySpark

Objective: Cover techniques for extracting data from various sources

- Reading data from Files
  - Connecting to Databases
  - Web Scraping with PySpark
-

## **Transforming Data with PySpark**

Objective: Explore various data transformations techniques in PySpark

- Data Cleaning and Preprocessing
- Feature Engineering
- Aggregations and Grouping

## **Loading Data with PySpark**

Objective: Learn method for loading transformed data into different destinations

- Writing data to Files
- Loading data to Databases
- Publishing data to Cloud Storage

## **Advance ETL Techniques**

Objective: Explore advanced ETL topics based on the interests and needs of the students

---

# **Part 7: Data Quality**

## **Introduction to Data Quality**

- Importance of Data Quality in Healthcare Analytics
- Understanding Data Quality Metrics and Parameters
- Data Correctness

## **Definition and Significance**

- Methods for Assessing Data Correctness
- Accuracy of Data Entry
- Data Validation Techniques
- Error Identification and Rectification Processes

## **Data Completeness**

- Definition and Importance in Healthcare Data
  - Measures to Assess Data Completeness
  - Missing Data Handling Techniques
  - Estimation and Imputation Methods
  - Evaluating the Impact of Incomplete Data on Analysis
-

## **Data Quality Control**

- Overview of Quality Control in Data Management
  - Establishing Data Quality Standards and Procedures
  - Implementation of Quality Control Processes
  - Quality Control Reporting Process
- 

# **Part 8: AWS**

## **Introduction to AWS for Data Analyst**

Objective: Provide an overview of AWS services and tools relevant to data analysis

- Introduction to Cloud computing
- Overview of AWS
- AWS pricing model

## **Data Storage in AWS**

Objective: Explore AWS services for Data Storage

- Amazon S3 (Simple Storage Service)
- Amazon RDS (Relational Database Service)
- AWS Glue for Data Catalog and ETL

## **Data Processing with AWS**

Objective: Introduce Amazon services for data processing and analysis

- AWS Athena for Querying S3 Data (With practical)
- Amazon EMR (Elastic MapReduce) (theory)
- AWS Lambda for Serverless computing (theory)

Sifal, Kathmandu, Nepal

Phone: +977 - 01 - 5913021 | 4567153

Mobile: +977 - 9765355167 | 9860422021

Email: [training@deerwalkcompware.com](mailto:training@deerwalkcompware.com)

Website: [deerwalktrainingcenter.com](http://deerwalktrainingcenter.com)