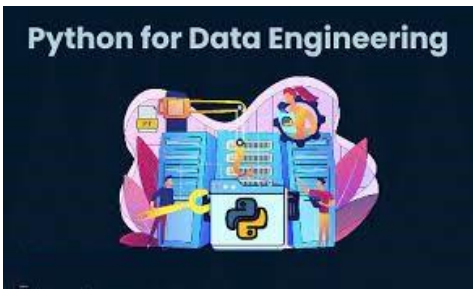

Data Engineering – Python

Duration: 5 Months(~120 Hours)

Price: 39999/-

Data Engineering is the process of designing, building, and maintaining the systems and infrastructure that enable organizations to collect, store, process, and analyze large volumes of data. Python is a popular programming language that can be used for data engineering tasks, such as data ingestion, data integration, data transformation, and data storage. Here is a detailed syllabus for learning Data Engineering through Python:

Introduction to Data Engineering:

Understanding the role of data engineering in the data ecosystem

Key concepts in data engineering, such as data ingestion, data integration, data transformation, and data storage

Overview of Python for data engineering

Data Ingestion:

Understanding different types of data sources

Using Python libraries for data ingestion, such as pandas and NumPy

Techniques for handling data in different formats, such as CSV, JSON, and XML

Data Integration:

Combining data from multiple sources

Techniques for data matching, cleansing, and deduplication

Using Python libraries for data integration, such as pandas and Dask

Data Transformation:

Techniques for data cleaning and preparation

Using Python libraries for data transformation, such as pandas and PySpark

Understanding the basics of data pipelines and workflow management

Data Storage:

Overview of different types of databases, such as relational and NoSQL databases

Using Python libraries for database access, such as SQLAlchemy and pymongo

Understanding the basics of distributed storage and data lakes

Big Data Processing:

Introduction to distributed computing frameworks, such as Apache Hadoop and Apache Spark

Understanding the basics of data parallelism and distributed processing

Using Python libraries for big data processing, such as PySpark and Dask

Cloud Computing for Data Engineering:

Understanding cloud computing and its benefits for data engineering

Overview of cloud computing platforms, such as AWS, Google Cloud, and Azure

Using Python libraries and tools for cloud computing, such as Boto3 and Terraform

Real-world Data Engineering Projects:

Applying the concepts and techniques learned in the course to real-world data engineering projects

Working with large datasets and real-world use cases

Understanding best practices for data engineering and working with data engineering teams

This syllabus is designed to provide a comprehensive introduction to Data Engineering through Python, covering key concepts, techniques, and tools used in the field. It is recommended to have a basic understanding of Python programming and data analysis before starting this course.

Data Analytics – Python (Tentative)

Introduction to Data Analytics:

- Understanding the role of data analytics in business decision-making
- Key concepts in data analytics, such as data wrangling, data exploration, and data visualization
- Overview of Python for data analytics

Data Wrangling:

- Understanding data types and data structures in Python
- Techniques for data cleaning and preparation
- Using Python libraries for data wrangling, such as pandas and NumPy
- Techniques for handling missing data, outliers, and inconsistent data

Data Exploration:

- Understanding data distributions and summary statistics
- Techniques for data visualization and exploratory data analysis
- Using Python libraries for data exploration, such as Matplotlib, Seaborn, and Plotly
- Techniques for identifying patterns, trends, and relationships in data

Statistical Analysis:

- Understanding statistical concepts and hypothesis testing
- Techniques for statistical analysis and inference
- Using Python libraries for statistical analysis, such as scipy and statsmodels
- Techniques for building predictive models and evaluating model performance

Machine Learning:

- Introduction to machine learning and its applications in data analytics
- Understanding different types of machine learning algorithms, such as supervised learning, unsupervised learning, and reinforcement learning
- Using Python libraries for machine learning, such as scikit-learn and TensorFlow
- Techniques for model selection, feature engineering, and hyperparameter tuning

Big Data Analytics:

- Introduction to distributed computing frameworks, such as Apache Hadoop and Apache Spark

- Understanding the basics of data parallelism and distributed processing
- Using Python libraries for big data analytics, such as PySpark and Dask
- Techniques for data processing in distributed systems, such as MapReduce and Spark RDDs

Cloud Computing for Data Analytics:

- Understanding cloud computing and its benefits for data analytics
- Overview of cloud computing platforms, such as AWS, Google Cloud, and Azure
- Using Python libraries and tools for cloud computing, such as Boto3 and Terraform
- Techniques for building and deploying data analytics workflows on cloud platforms

Real-world Data Analytics Projects:

- Applying the concepts and techniques learned in the course to real-world data analytics projects
- Working with large datasets and real-world use cases
- Understanding best practices for data analytics and working with data analytics teams

This syllabus is designed to provide a comprehensive introduction to Data Analytics through Python, covering key concepts, techniques, and tools used in the field. It is recommended to have a basic understanding of Python programming and statistics before starting this course. By the end of this course, you should be able to analyze data, build predictive models, and make data-driven decisions using Python and related tools.

Website: www.vidyaxcel.com

Address: BC-74 3rd Floor Calcutta Greens Commercial Complex, Kolkata -700075

Contact No: +91-9123672473