Data Engineering

Duration – 40 Hours (2 Months)

Phase 1: Foundations (6 hours)

- **SQL** (most important skill)
- Basics → Intermediate → Advanced
- Joins, subqueries, CTEs, Window functions, Aggregations
- Query optimization, Indexes, Execution plans
- Practice: PostgreSQL / MySQL
- **Python for Data Engineering**
- Core Python (functions, OOP basics, error handling)
- File handling (CSV, JSON, Parquet)
- Pandas / Polars for data manipulation
- Working with APIs, Requests library
- Writing clean, reusable scripts

Phase 2: Databases, Data Modeling & Warehousing (12 hours)

- Relational Databases (PostgreSQL deep dive)
- Data Modeling
- Normalization vs Denormalization
- Star Schema & Snowflake Schema
- Fact & Dimension tables
- Slowly Changing Dimensions (SCD Type 1, 2)
- Data Warehouses vs Data Lakes vs Lakehouse
- Modern warehouses: Snowflake / BigQuery / Redshift (pick one)
- Basic NoSQL concepts (MongoDB)

Phase 3: ETL/ELT Pipelines & Orchestration (8 hours)

- ETL vs ELT concepts
- **dbt** (Analytics Engineering) — highly recommended
  - Models, tests, documentation, macros
- **Apache Airflow** (or Prefect / Dagster)
  - DAGs, Operators, Scheduling, Retries, Monitoring
- Data Quality & Testing

AA/19, DESBANDHU NAGAR, KOLKATA, North 24 Parganas, West Bengal, 700059
Phone: 8282884601
Email id: partner@cronusconsultants.co.in

- Docker basics (containerization)

Phase 4: Big Data & Distributed Processing (6 hours)

Apache Spark (PySpark)
DataFrames, Transformations, Actions
Spark SQL
Performance tuning basics
Introduction to Hadoop ecosystem (optional but useful)
File formats: Parquet, ORC, Avro
Batch processing patterns
Basic cloud services (AWS)

Phase 5: Hands on Project (8 Hours)

AA/19, DESBANDHU NAGAR, KOLKATA, North 24 Parganas, West Bengal, 700059
Phone: 8282884601
Email id: partner@cronusconsultants.co.in