🛠️ Stage 1: Foundations (Beginner)

Goal: Build strong programming, database, and data fundamentals.

- **Programming:** Python (pandas, NumPy), SQL (joins, indexing, optimization).
- **Databases:** Relational (PostgreSQL, MySQL) + NoSQL (MongoDB, Cassandra).
- **Linux & Shell:** Basic commands, scripting, cron jobs.
- **Version Control:** Git/GitHub for collaboration.
- **Projects:**

- Build a small ETL pipeline (extract CSV → transform in Python → load into SQL).
- Create dashboards with Power BI or Tableau.

🚀 Stage 2: Core Data Engineering (Intermediate)

Goal: Learn how to move, store, and process large-scale data.

- **Big Data Tools:** Apache Spark, Hadoop basics.
- **Data Warehousing:** Snowflake, Redshift, BigQuery.
- **Workflow Orchestration:** Apache Airflow, Prefect.
- **Cloud Basics:** AWS (S3, RDS, Lambda), Azure (Blob Storage, Synapse).
- **Data Modeling:** Star schema, normalization, denormalization.
- **Projects:**

- Build a data warehouse for a mock retail company.
- Automate daily ETL jobs with Airflow.

🌐 Stage 3: Advanced Engineering (Expert Track)

Goal: Handle streaming, scalability, and advanced architectures.

- **Streaming Data:** Kafka, AWS Kinesis, Apache Flink.
- **Data Lakes:** Delta Lake, Lakehouse architecture.
- **Advanced Cloud:** AWS Glue, Azure Data Factory, GCP Dataflow.
- **DevOps & CI/CD:** Docker, Kubernetes, Terraform for infrastructure automation.
- **Security & Governance:** Data privacy (POPIA compliance in South Africa), IAM roles.
- **Projects:**

- Real-time analytics pipeline (Kafka → Spark → Dashboard).
- Deploy scalable pipelines on AWS/Azure.

🎓 Certifications (Boost Employability & Salary)

- **AWS Certified Data Engineer – Associate** (or AWS Solutions Architect).
- **Microsoft Azure Data Engineer Associate (DP-203)**.

- **Google Professional Data Engineer**.
- Optional: **Databricks Certified Data Engineer Associate**.

💼 Career Path

- **Entry-level:** Junior Data Engineer / BI Developer.
- **Mid-level:** Data Engineer / Cloud Data Engineer.
- **Senior-level:** Big Data Architect / Solutions Architect.
- **Global Remote Roles:** Cloud Data Engineer (USD salaries).

📈 Timeline (Approximate)

- **Months 1–3:** Foundations (Python, SQL, databases).
- **Months 4–6:** Intermediate (Spark, Airflow, cloud basics).
- **Months 7–12:** Advanced (streaming, DevOps, certifications).
- **Year 2+:** Specialize in cloud + big data, aim for senior roles.