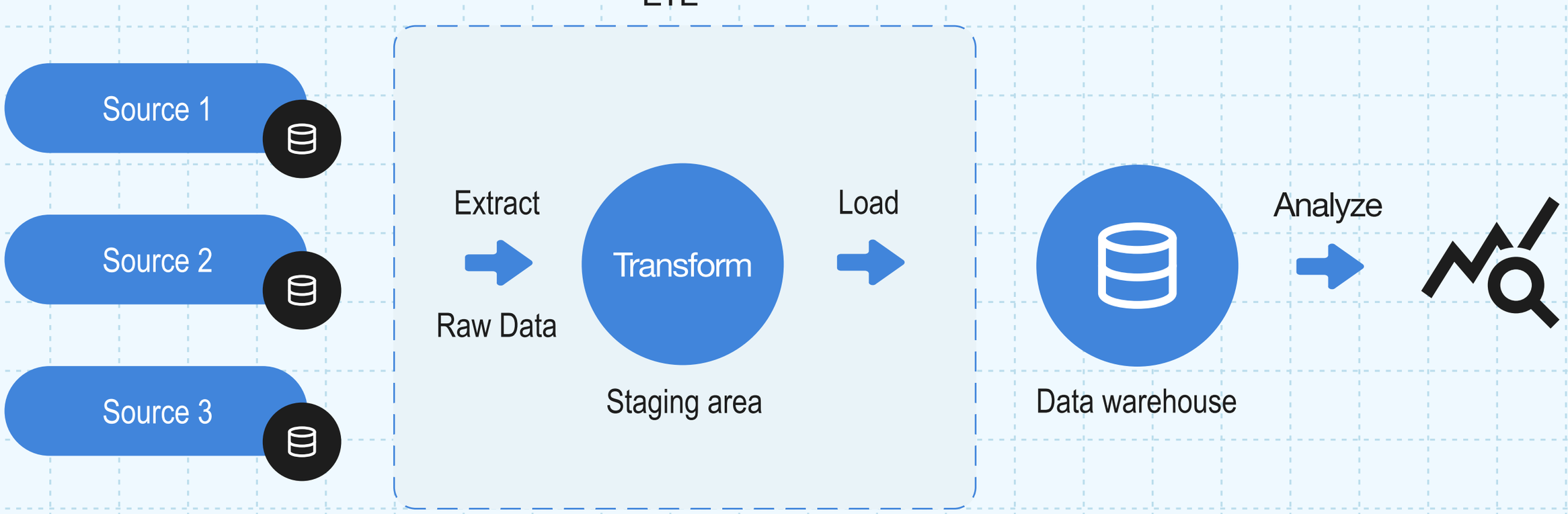
BOSSCODER ACADEMY

UNDERSTANDING ETL PIPELINE FOR DATA ENGINEERS



Concepts & Practice Real-World Problems

Disclaimer

EVERYONE LEARNS UNIQUELY.

This roadmap is a guide to help you navigate the journey of becoming a Data Engineer.

Treat this as direction, not a fixed rulebook. Real growth comes from consistent practice and solving real-world problems.

TABLE OF CONTENTS

01 **INTRODUCTION** ---- 4-9

What Is ETL And Why It Matters

Importance For Analytics, Machine Learning, And Operations.

ETL Vs ELT Explained Simply.

02 **ETL BASICS** ---- 10-16

Extract → Gather Data From Sources

Transform → Clean, Standardize, Enrich

Load → Store Into Warehouse/Lake

Batch Vs Real-Time Pipelines + Challenges

03 **ETL ARCHITECTURE** ---- 17-21

Layers: Sources → Staging → Transformation → Warehouse → Consumption

Shows The Structured Flow From Raw Data To Insights

04 **TOOLS & TECHNOLOGIES** ---- 22-31

Ingestion: Kafka, Fivetran, Glue

Transformation: Spark, Dbt, Pandas

Storage: Snowflake, BigQuery, S3

Orchestration: Airflow, Prefect

Monitoring: Great Expectations, Monte Carlo

05 **BEST PRACTICES** ---- 32-33

Modular Design, ELT In Modern Warehouses

Automated Data Quality Checks

Schema Evolution, Monitoring, Scalability

Shows Real Business Value Of ETL Pipelines

06 **USE CASES** ---- 34

Daily Sales Reporting → Walmart

Customer Personalization → Amazon

Inventory Optimization → Target

Shows Real Business Value Of ETL Pipelines