**Five project topics for a Data Engineering professional**

**1. Real-Time Data Pipeline with Kafka & Spark Streaming**

* **Why?** Streaming is critical for modern apps (fraud detection, IoT, live analytics).
* **Tech Stack:** Apache Kafka, Spark Streaming/Flink, Python/Java, Docker.
* **Outcome:** Process & analyze live data (e.g., clickstream, sensor logs).

**2. Cloud-Based Data Lake with ETL/ELT Automation**

* **Why?** Data lakes are foundational for analytics and AI.
* **Tech Stack:** AWS S3/Azure Data Lake, Delta Lake/Iceberg, Airflow/dbt, PySpark.
* **Outcome:** Scalable storage + automated pipelines for structured/unstructured data.

**3. Modern Data Warehouse Migration (e.g., to Snowflake/BigQuery)**

* **Why?** Companies are shifting from on-prem to cloud warehouses for scalability.
* **Tech Stack:** Snowflake/BigQuery/Redshift, SQL, Terraform (IaC), CDC tools.
* **Outcome:** Faster queries, cost optimization, and seamless data integration.

**4. Data Quality & Observability Platform**

* **Why?** Poor data quality costs businesses millions annually.
* **Tech Stack:** Great Expectations/SodaCL, Python, Grafana/Prometheus.
* **Outcome:** Alerts for anomalies, lineage tracking, and reliability metrics.

**5. Serverless ETL Pipeline (AWS Lambda/GCP Cloud Functions)**

* **Why?** Cost-efficient, scalable, and low-maintenance processing.
* **Tech Stack:** AWS Lambda/Azure Functions, Step Functions, Parquet/JSON.
* **Outcome:** Event-driven pipelines (e.g., processing user uploads or logs).