Harsimran – Data Engineer |

PROFILE SUMMARY

Seasoned Data Engineer with 9 years of experience in backend development and data engineering, including designing, developing, and maintaining scalable data pipelines and microservices. Strong expertise in Java, Spring Boot, GCP/AWS, ETL frameworks, and Big Data tools. Adept at migrating legacy systems to the cloud, optimizing system performance, and enabling data-driven decision-making across enterprise environments. Known for a collaborative mindset, attention to detail, and delivering solutions that scale.

CORE SKILLS

- **Backend Development:** Java, Spring Boot, Hibernate, Maven, RESTful APIs
- **Cloud Platforms:** GCP (Dataflow, BigQuery, Cloud Storage, Looker), AWS (S3, RDS, Lambda, EMR, Glue, Athena, Redshift, EC2)
- **Big Data & ETL:** Spark, Hadoop, Hive, Sqoop, MapReduce, HDFS, Python, PySpark
- **Data Orchestration & Observability:** Apache Airflow, Cloud Orchestration, Pub/Sub, Grafana, Prometheus, OpenTelemetry, Cloud Logging, ELK, Splunk
- **Databases:** SQL, GBQ, DynamoDB, RDS
- **DevOps & Containers:** Docker, Kubernetes
- **Streaming & Messaging:** Kafka, Kinesis
- **Other Tools:** Dataform, Workflows, Dataform
- **Machine Learning (Basic):** Anomaly detection and SLA forecasting using scikit-learn

---

PROFESSIONAL EXPERIENCE

Tech Lead
Jan 2024 – Present

- Designed and maintained backend data pipelines for optimized data flow using Java, Spring Boot, and PySpark on GCP.
- Built and managed robust ETL frameworks using Airflow and Dataflow with monitoring support from **Prometheus and Grafana dashboards**.
- Applied basic **OpenTelemetry instrumentation** for tracing ETL pipeline latencies.
- Led cross-functional collaborations to ensure high data quality, logging insights via **Splunk** and **Cloud Logging**.
- Implemented anomaly detection scripts in Python to forecast job failures and improve SLA management.

Associate Consultant – Data Engineer
Nov 2021 – Jan 2024

- Migrated Splunk dashboards to Looker using Dataform and integrated monitoring via **Prometheus** and **Grafana**.
- Developed scalable data pipelines on **GCP (BigQuery, Dataflow, Cloud Storage)** and **AWS (Glue, Redshift, Lambda)**.
- Created Python-based anomaly detection workflows integrated into monitoring stack via **OpenTelemetry**.
- Managed data quality rules and built alerts for Looker reports using **Cloud Logging** and automated Slack integration.

Senior Associate Technology
Jan 2020 – Jan 2020

- Built and deployed efficient **Java Spring Boot microservices** to support backend APIs and scalable orchestration.
- Integrated **Kafka** and **Kinesis** pipelines for real-time message processing.
- Worked on **RESTful APIs** supporting distributed workloads with embedded logging via **ELK and OpenTelemetry traces**.

Java Spring Microservices Backend Developer
Sep 2016 – Jan 2020

- Developed scalable backend modules using **Java**, **Spring**, and **Hibernate**, supporting millions of requests.
- Supported job orchestration pipelines using early versions of **Airflow** and **custom cron-based DAGs**.
- Implemented basic Prometheus exporters to track microservice health and shared metrics to **Grafana**.

CERTIFICATIONS

- Oracle Certified Java Programmer
- AWS Certified Big Data – Specialty (Valid till Dec 2024)
- Google Cloud Certified – Professional Cloud Architect (Valid till Mar 2027)

EDUCATION

- PG Diploma, Guru Nanak Dev University (GNDU), 2020
- B.Tech/B.E., Guru Nanak Dev University (GNDU), 2016
- 12th, CBSE Board, 2011
- 10th, CISCE (ICSE), 2009

PROJECT HIGHLIGHTS

1. **GCP-Based Observability Pipeline**

    - Migrated legacy data pipeline monitoring to a modern **Grafana + Prometheus + OpenTelemetry** stack.
    - Used **BigQuery** and **Cloud Logging** to trace issues across microservices and ETL stages.
    - Built real-time job status dashboards for stakeholders using Looker and Pub/Sub.
2. **Anomaly Detection for SLA Violations**
    - Created **Python-based forecasting models** using scikit-learn to predict SLA breaches.
    - Integrated predictions with **Prometheus AlertManager** and displayed status in **Grafana**.
    - Reduced incident noise and enhanced alert precision.
3. **Cross-Cloud ETL Automation Project**
    - Developed end-to-end ETL across **AWS Glue**, **GCP Dataflow**, and **Airflow**.
    - Used **Kafka** and **Kinesis** for stream ingestion, storing results in **BigQuery** and **Redshift**.
    - Enabled consistent metric tracking across platforms via unified **Prometheus exporters** and dashboards.
