# Perin Shah

Bothell, Washington | +1-469-943-3090 | shahperin72@gmail.com | linkedin.com/in/shahperin

## PROFESSIONAL SUMMARY

Accomplished Data Engineer with 5 years of experience in developing and optimizing data infrastructure solutions. Proficient in big data technologies and cloud platforms, with a strong track record of enhancing data pipeline efficiency, reducing latency, and lowering costs. Dedicated to leveraging data engineering expertise to drive business success and innovation.

## SKILLS

**Languages**: Python, SQL, Java, Scala

**Big Data:** Apache Spark, Hadoop, Kafka, Hive

**Database Management:** PostgreSQL, MongoDB, Cassandra

**Cloud Platforms**: AWS (EMR, Redshift, S3), Google Cloud Platform (BigQuery, Dataflow)

**Data Warehousing**: Snowflake, Amazon Redshift

**ETL/ELT**: Apache Airflow, DBT

**Containerization**: Docker, Kubernetes

**Version Control**: Git

**Certifications:** Microsoft Azure Fundamentals, Google Data Analytics Professional, IBM Data Engineering Professional

## WORK EXPERIENCE

**Alaska Air (Contract)**                                                                                          **Seattle, WA, USA**
*Sr. Data Engineer*                                                                                                   *May 2022 – Current*

**Project: Real-time Analytics Platform**
- Engineered a real-time data processing system using Kafka and Spark Streaming, decreasing data latency from hours to minutes and enabling timely decision-making.
- Pioneered data quality checks that led to a 40% decrease in error rates.

**Project: Cloud Migration Initiative**
- Directed the migration of an on-premises data warehouse to Snowflake, resulting in a 65% improvement in reporting speed through optimized data models and queries.
- Achieved a 35% reduction in infrastructure costs while boosting scalability and performance.

**Deltek (Onvia)**                                                                                                    **Seattle, WA, USA**
*Data Engineer*                                                                                                       *July 2020 – May 2022*

**Project: Customer 360 Data Platform**
- Developed a centralized customer data platform by integrating data from over 10 sources, enhancing data accessibility and analytics capabilities.
- Instituted data governance policies that ensured compliance with CCPA and GDPR regulations.
- Automated data pipelines using Apache Airflow, effectively processing over 1TB of data daily.

**Project: Predictive Maintenance ML Pipeline**
- Partnered with data scientists to establish an ML pipeline for predictive maintenance, cutting model deployment time from weeks to hours with MLflow and Kubernetes.
- Devised an A/B testing framework to assess model performance, improving predictive accuracy.

**Johnson & Johnson (Contract)**                                                                               **Pune, India**
*Jr. Data Engineer*                                                                                                   *May 2019 – Jul 2020*

**Project: Sales Analytics Dashboard**
- Crafted ETL processes for the integration of sales data from diverse CRM systems, enabling comprehensive sales analysis.
- Designed interactive dashboards using Tableau, enhancing sales team efficiency by 25%.
- Optimized slow-running queries, resulting in a 50% improvement in dashboard load times.

## EDUCATION

**University Of Texas At Arlington**                                                                          **Arlington, Texas, USA**
*Masters in Information and Computer Sciences*                                                       *August 2021 - May 2023*

**Navrachana University**                                                                                         **Vadodara, Gujarat, India**
*Bachelor of Technology - Computer Science and Engineering*                                 *July 2016 - June 2020*