

Perin Shah

Jersey City, New Jersey | +1-469-943-3090 | shahperin72@gmail.com | linkedin.com/in/shahperin

PROFESSIONAL SUMMARY

Data Engineer with 6+ years of experience designing scalable ETL pipelines and lakehouse architectures for diverse data types including BIM, geospatial, imagery and IOT across analytics and operational use cases. Strong background in Python, SQL, Apache Spark, Databricks, Airflow, Docker, GitHub, and cloud environments including Azure and AWS. Experienced working with structured, semi-structured, and unstructured data, collaborating with data scientists and cross-functional teams, and mentoring junior engineers to improve data quality, performance, and reliability across production systems.

SKILLS

Languages: Python, SQL, Java, Scala

Big Data: Apache Spark, Hadoop, Kafka, SkySpark, Hive, Databricks, Power BI

Database Management: PostgreSQL, MongoDB, Cassandra, Neo4J

BIM & Geospatial: Autodesk Revit, ArcGIS, IFC, IMAGINiT Clarity

Cloud Platforms: AWS (EMR, Redshift, S3), Google Cloud Platform (BigQuery, Dataflow), Azure (Data Factory, Cosmos DB)

Data Warehousing: Snowflake, Amazon Redshift

ETL/ELT: Apache Airflow, dbt

Containerization & CI/CD: Docker, Kubernetes, Jenkins, GitHub Actions

Version Control: Git, GitHub, Terraform

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

EDUCATION

University Of Texas At Arlington

Masters in Information and Computer Sciences

Navrachana University

Bachelor of Technology - Computer Science and Engineering

Arlington, Texas, USA

August 2021 - May 2023

Vadodara, Gujarat, India

July 2016 - June 2020

WORK EXPERIENCE

DLR GROUP

Data Engineer II (Hybrid)

New York City, NY, USA

Oct 2024 – Current

- Design and maintain scalable ETL pipelines in Databricks and Apache Airflow to ingest BIM, geospatial, and imagery data into a cloud lakehouse, processing geometric building data across 2000+ projects and reducing analyst data prep time by 40%.
- Develop data models and schemas for structured and unstructured architectural data, building space-embedding and automated labeling pipelines across 50K+ spaces that improved labeling accuracy by 25% and powered recommendation workflows used by 30+ designers.
- Build Neo4J graph-based data applications on IFC/BIM geometry to power space intelligence and furniture placement analysis across 100+ active building models, cutting spatial-lookup query times by 50%.
- Partner with data scientists, software engineers and design stakeholders on audit and recommendation workflows for space programs, mentoring junior engineers on pipeline development and shortening program review cycles from 2 weeks to 3 days.

IQuest Solutions Corp (Client: Alaska Airlines)

Sr. Data Engineer (Hybrid)

Prosper, TX, USA

July 2024 – Sep 2024

- Supported ongoing data quality and monitoring work on existing Kafka and Spark Streaming pipelines, helping triage upstream issues and improve pipeline observability.
- Tuned Spark job configurations and partitioning strategies on high-volume batch workloads, contributing to improved runtime and reduced cluster resource consumption.
- Partnered with analytics stakeholders to refine Snowflake reporting models and prioritize follow-on pipeline enhancements identified during the post-migration stabilization period.

iOasiz Inc. (Client: Alaska Airlines)

Piscataway, NJ, USA

Sr. Data Engineer (Hybrid)

Aug 2023 – July 2024

- Led the migration of an on-premises data warehouse to Snowflake, modernizing the data platform and enabling a scalable cloud architecture.
- Engineered a real-time data processing system using Kafka and Spark Streaming, decreasing data latency from 6 hours to 15 minutes and accelerating decision-making processes by 80%.
- Optimized data models and query performance, improving reporting speed by 65% and delivering faster access to business insights.
- Reduced infrastructure costs by 35% and scaled the platform to handle a 50% increase in data volume without performance degradation.

Shah Enterprise (Client: Deltek)

Vadodara, India

Data Engineer (Remote)

Aug 2022 – July 2023

- Developed a centralized customer data platform integrating data from over 10 sources, improving data accessibility by 70% and enhancing analytics capabilities.
- Established data governance policies, ensuring compliance with CCPA and GDPR, reducing data handling errors by 30%.
- Automated data pipelines with Apache Airflow, processing over 1TB of data daily and reducing manual intervention by 60%, leading to a 50% increase in processing efficiency.
- Collaborated with data scientists to create an ML pipeline for predictive maintenance, cutting model deployment time from weeks to hours and increasing deployment frequency by 70%.

vCreaTek (Client: Johnson & Johnson)

Pune, India

Data Engineer (Hybrid)

Jun 2020 – Jul 2021

- Developed ETL processes to integrate sales data from diverse CRM systems, increasing the efficiency of sales analysis by 30%.
- Created interactive dashboards with Tableau, improving sales productivity by 25% through faster and accurate data insights.
- Optimized slow-running queries, resulting in a 50% reduction in dashboard load times and enhancing user experience.

Shah Enterprise

Vadodara, India

Data Engineer

Jun 2018 – Jun 2020

- Assisted in building data ingestion and ETL scripts in Python and SQL to consolidate data from internal operational systems into a SQL Server database, supporting day-to-day reporting needs.
- Added basic data validation and monitoring checks to ingestion jobs to catch load failures and improve reliability of downstream reports.
- Worked with business stakeholders to gather reporting requirements and build SQL views supporting ad-hoc analysis.

PUBLICATIONS

- **Paper: Mind Machine Interface Using Internet Of Things**
Published by IJERSM in Volume-2, Issue-12, December-2019
- **Book Chapter: A Building is a Graph: A New Perspective on Building Information Modeling**
Forthcoming, Wiley (in publication)