

HARSH JOSHI

Senior Data Engineering & BI Analytics Professional

📞 +1-(469)-922-7662

✉️ harshkjoshi888@gmail.com

LinkedIn linkedin.com/in/harshjoshi888

🌐 harshportfolio.com

Professional Overview

Senior Data Engineering & Analytics Professional with 6+ years of experience designing end-to-end data and solution architectures, building scalable ETL/ELT pipelines, and managing enterprise SQL/NoSQL data platforms. Highly skilled in Python, PySpark, SQL, and Java, with hands-on experience across Azure services, Snowflake, and distributed data systems. Proven expertise in data architecture, integration patterns, and governance frameworks, along with developing technical documentation, architecture diagrams, and design standards. Experienced in delivering business-driven insights through Power BI dashboards, optimizing data workflows, and supporting secure, scalable, cloud-based analytics solutions.

Core Competencies Toolset

- Python
- Apache Spark
- PySpark
- SQL
- ETL/ELT Pipelines
- Data Engineering
- Data Integration
- Data Extraction
- Data Modeling
- Data Warehousing
- Schema Design
- Data Quality
- Data Governance
- Metadata Management
- Azure Data Factory
- Azure Databricks
- Azure SQL
- Snowflake
- Delta Live Tables
- Unity Catalog
- AWS Glue
- AWS Lambda
- AWS S3
- EC2
- REST API Development
- API Integration
- Power BI
- Data Visualization
- CI/CD Pipelines
- GitHub
- GitHub Actions
- Azure DevOps
- Airflow
- Process Automation
- Automated Testing
- Agile Methodologies
- Big Data
- MongoDB
- Java
- Solution Architecture
- Cloud Architecture
- Security & Access Control
- Logging & Monitoring

Impactful Contributions

Enterprise ETL Architecture

Implemented ETL solutions improving processing efficiency and scalability by 40%.

Large-Scale Data Processing

Redesigned a data pipeline handling over 100 TB of data, improving performance.

Power BI Analytics

Designed and developed interactive dashboards, enhancing data visualization and decision-making processes.

Database Performance Optimization

Increased database efficiency by 50% by optimizing data management strategies.

Solution Architecture Leadership

Created architecture diagrams and collaborated with teams to deliver aligned data solutions.

Cross-Functional Collaboration

Bridged technical and business teams to deliver aligned data solutions, fostering seamless collaboration.

Academic Milestones

University of Missouri, Columbia, Missouri

Master's in Data Science and Analytics with High-Performance Computing

08/2021 – 05/2023

Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, MP, India

Bachelor's in Computer Science

07/2014 – 12/2018

Career Timeline

Cummins, United States

Senior IT Data Engineer

02/2025 – Present

- Led solution and data architecture design for enterprise data platforms by defining end-to-end ingestion, transformation, storage, and consumption patterns across Oracle ERP, SQL systems, APIs, and file-based sources, ensuring scalability, security, and alignment with business objectives.

- Owned technical documentation and architectural artifacts, including data flow diagrams, system architecture diagrams, network/security views, data dictionaries, and solution design documents to enable clear communication, governance compliance, and long-term maintainability.
- Evaluated and guided platform and technology decisions (e.g., Azure services, Snowflake, ETL frameworks) by assessing cost, scalability, security, and operational impact, and translating requirements into approved architectural standards and implementation roadmaps.
- Established data governance and integration standards covering metadata management, access controls, data quality checks, lineage, and retention policies, ensuring consistent, compliant, and trusted data usage across teams and projects.
- Acted as a technical bridge between stakeholders and engineering teams, facilitating requirement discussions, design reviews, and peer architecture sessions while ensuring solutions were pragmatic, future-ready, and aligned with enterprise architecture and security guidelines.

Children's Mercy, Kansas City, United States

05/2023 – 11/2024

Data Engineer I

- Orchestrated seamless data workflows in Azure Data Factory, designing and managing efficient ETL pipelines to optimize data integration processes.
- Developed and automated ETL solutions using Apache Spark and Azure Databricks, significantly improving data processing efficiency and scalability.
- Using Python and SQL to perform complex data analysis, ensure data accuracy, and perform data-driven testing, driving data integrity and actionable insights.
- Implemented and managed Delta Live Tables within Azure Databricks, enabling real-time streaming and batch data processing for timely analytics.
- Created comprehensive analysis reports with Power BI and Microsoft Office, transforming raw data into meaningful business insights for key stakeholders.

University of Missouri, Columbia, United States

06/2022 – 05/2023

Data Engineer

- Designed and implemented Machine Learning components to enhance predictive analytics and decision-making.
- Leveraged AWS, Hadoop, and Python for data processing and analysis, delivering actionable insights to stakeholders.
- Extracted and transformed structured and unstructured datasets, improving data usability with advanced tools like Snowflake and Databricks.
- Developed analysis reports and collaborated with multidisciplinary teams to align strategies and foster impactful outcomes.

The Housing Collective, New Haven, United States

06/2022 – 08/2022

Data Analyst Intern

- Conducted in-depth analyses of customer, competitor, and market data to identify growth opportunities, optimizing product and service strategies and enhancing advertising effectiveness.
- Designed and implemented data management strategies that improved database efficiency by 50%, including removing redundant data and optimizing filtering processes.
- Created interactive dashboards and impactful visualizations in Power BI, translating complex datasets into clear, actionable insights that guided strategic decision-making.
- Applied advanced statistical methods and data mining algorithms to uncover trends and patterns, contributing to data-driven strategies and fostering collaboration with cross-functional teams.