


# Harsh Joshi

## Senior Data Engineer

 <https://harshportfolio.com/>

 469-922-7662

 [harshkjoshi888@gmail.com](mailto:harshkjoshi888@gmail.com)

Results-driven Senior Data Engineer with 6+ years' experience in Python for ETL, SQL/No-SQL databases, and ETL pipeline automation. Proficient in Azure (Data Bricks, DevOps) and AWS (S3, Ec2), with expertise in CI/CD (GitHub, JIRA). Skilled in Spark (PySpark, streaming) and data analysis (Statistical, PowerBI). Excels in data management (governance, integration) and eager to leverage my skills and experience to drive innovative solutions and contribute to the success of a forward-thinking team.

## Skills

**Programming** - Python, PySpark, Java

**Data Base** - Database Design, SQL, No-SQL, PostgreSQL

**ETL Pipeline** - Real Time Processing, Data Warehousing

**Azure** - Data Bricks, Delta Live Tables, Data Factory

**AWS** - S3, EC2, Glue, Lambda, RedShift

**Continuous Integration** - CI/CD, GitHub, JIRA, DevOps

**Data Analysis** - Data Analysis, Power BI, Data Detective

**Data** - Data Governance, Data Integration, Data Privacy

## Education

**Master's of Science: Data Science and Analytics with High Performance Computing**

Aug 2021 – May 2023

University of Missouri - Columbia, MO

## Work History

**Data Engineer I - Children's Mercy Kansas City**

May 2023 – Present

- Leveraged Azure Data Factory, Apache Spark, and Azure Databricks to orchestrate seamless data workflows and optimize data pipelines.
- Championed Extract, Transform, Load (ETL) processes using Apache Spark and Azure Databricks, optimizing data pipelines for enhanced efficiency. Led the project, creating and refining Power BI dashboards for comprehensive data analysis.
- Demonstrated strong time management skills, ensuring the timely execution of tasks and projects to meet organizational goals.

**Graduate Research Assistant - University of Missouri**

June 2022 – May 2023

- Extensive expertise in entity extraction from structured and unstructured data, leveraging advanced techniques to extract valuable insights with 60% efficiency.
- Skilled in working with AWS, Hadoop, and Python for efficient data processing and analysis, including experience in data crawling using these technologies.
- Responsible for creating comprehensive analysis reports and maintaining effective communication with lab researchers, while also presenting findings to clients, stakeholders, directors, and presidents.

**Data Analyst Intern - The Housing Collective, Bridgeport, CT**

June 2022 – Aug 2022

- Generated detailed studies on potential third-party data handling solutions, verifying compliance with internal needs and stakeholder requirements.
- Examined data from web traffic, database and from other sources to provide visualization and insights to business team.
- Analyzed complex data and identified anomalies, trends, and risks to provide useful insights to improve internal controls

**Data Engineer Intern - Edgevanta LLC, Nashville, TN**

May 2022 – July 2022

- Developed, implemented, and maintained data analytics protocols, standards, and documentation.
- Developed and implemented data quality metrics and automated data quality checks, resulting in a 50% reduction in data errors and improved data accuracy.
- Developed an automated AWS pipeline solution that processed over 5TB of data per month while reducing operating costs by 30%.

**Data Engineer - Affimintus Technologies, Indore, India**

Jan 2020 – July 2021

- Established a high-availability data infrastructure, boosting data availability by 90% and ensuring uninterrupted access of data.
- Mentored junior team members, conducted code operations and design reviews, and enforced coding standards and best practices.