

**Ravi Sharma**  
**Data Engineer**

667-354-9019 | ravijayprakashsharma@gmail.com | Long Beach, CA | [LinkedIn](#)

## SUMMARY

- 3 years of experience as a Data Engineer with a proven track record of designing and building scalable data solutions. Skilled in leveraging Cloud Data Engineering tools (Azure) to deliver Data Warehouses, Hadoop ecosystems, Big Data analytics pipelines, and Data Visualization and Reporting solutions.
- Expert in implementing Airflow to manage a complex data pipeline with dependent tasks, ensuring reliable data flow and reducing operational costs through automation.
- Experience with Azure Blob Storage and Azure SQL Database for additional data storage and management needs.
- Proficient in developing Spark notebooks and jobs within Databricks to perform complex data transformations and machine learning algorithms.
- Ability to integrate data pipelines (SSIS, ADF) within Azure DevOps pipelines for continuous integration and continuous delivery (CI/CD) practices.

## SKILLS

- **Programming Language:** Scala, Python, R, SQL
- **IDE's:** PyCharm, Jupyter Notebook
- **Big Data Ecosystem:** Hadoop, Hive, HDFS, Sqoop, Apache Airflow, Apache Kafka, Apache Spark, Apache Flink
- **ETL and Cloud Technologies:** SSIS, Azure (Azure DevOps, Azure Data Lake, Azure Data Factory, Azure Databricks)
- **Visualizations:** Tableau, Power BI, Excel
- **Packages & Data Processing:** NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, PySpark, Data Pipelines
- **Version Control & Database:** GitHub, Gitlab, SQL Server, PostgreSQL, MongoDB, MySQL, Snowflake
- **Operating Systems:** Windows, MacOS

## EDUCATION

<b>Masters of Science, Information Systems</b> University of Maryland, Baltimore County, Baltimore, Maryland	<b>May 2024</b>
<b>Bachelor of Engineering in Computer Engineering</b> University of Mumbai, India	<b>May 2022</b>

## EXPERIENCE

<b>McKesson, MD</b> <b>Data Engineer</b>	<i>Jan 2024 – May 2024</i>
---	----------------------------

- Developed and optimized Apache Spark jobs for large-scale data processing tasks, including filtering, aggregation, and transformations using Spark SQL, DataFrames, and RDDs.
- Leveraged Apache Flink to integrate with various data sources like databases, message queues, and data lakes for real-time data processing.
- Created interactive dashboards and reports to communicate complex data insights to stakeholders using Power BI.
- Use Python for data analysis tasks, including data cleaning, manipulation, and feature engineering, employed libraries like NumPy, Pandas, and Scikit-learn to extract insights from data
- Optimized SQL queries using Snowflake features, achieving a 25% improvement in average query execution time.
- Design and manage Azure Data Lake Storage (ADLS) for scalable storage of structured, semi-structured, and unstructured data.

<b>Zensar Technologies, India</b> <b>Data Engineer</b>	<i>Jan 2020 – July 2022</i>
---	-----------------------------

- Utilized Airflow for data processing workflows, resulting in a 20% decrease in pipeline execution time due to streamlined scheduling and dependency management.
- Created and managed Hive tables with appropriate partitioning and bucketing for efficient querying of structured data in HDFS using HiveQL.
- Leveraged PySpark for large-scale data processing and analytics tasks within the Apache Spark ecosystem and utilized DataFrames and SQL functionalities for efficient data manipulation.
- Calculated and implemented scalable data warehouses on Snowflake, a 30% reduction in query execution time compared to the previous data platform
- Built real-time data pipelines using Apache Kafka to ingest streaming data feeds from various sources into HDFS and other storage systems.
- Facilitated cross-functional collaboration by integrating Azure DevOps with collaboration tools, a 30% improvement in communication and information sharing across development teams.
- Successfully integrated Azure Data Lake Storage and Data Lake Analytics to streamline data processing workflows, resulting in a 15% reduction in data processing times.
- Designed and developed SSIS packages for efficient extraction, transformation, and loading (ETL) workflows, utilized Data Flow Tasks, Script Tasks, and Control Flow logic to automate data movement and manipulation.