

JAYA SURYA MAHADEVAN

D: [+1 8623727937](tel:+18623727937) | E: mahadevan.javasurva@gmail.com | L: [NJ](#) | Ln: <https://www.linkedin.com/in/java-surva-6799371a0/>

SUMMARY:

- Data Engineer with over 4 years of experience designing, implementing, and optimizing end-to-end data solutions.
- Proficient in **Python, Scala, and SQL**, with expertise in building robust ETL processes and data pipelines.
- Skilled in relational (**MySQL, PostgreSQL**) and non-relational (**MongoDB, Cassandra**) databases, ensuring data accuracy.
- Expert in data modeling, warehouse design, and cloud platforms (**AWS, Azure**) for scalable, reliable, and cost-effective solutions.
- Proven AWS expertise with **EMR, Glue, Redshift**, and **EC2** for data processing.
- Skilled in **Agile methodologies, Spark Streaming**, and adept at utilizing **Git/Bitbucket** for version control.
- Practical use of Snowflake tools (**SnowSQL, SnowPipe**), along with Python for data modeling techniques.
- Well-versed in big data tools using Hadoop technologies, including **Map Reduce, Hive, HDFS, and Pig**
- Worked on **Airflow**, to automate, to monitor and scheduled ETL pipelines.
- Utilized **ADF** for reliable data pipelines and demonstrated proficiency in **Apache Spark, PySpark, and Spark Structured API** for large data processing.
- Solid expertise in creating data visualization dashboards using **Power BI** and **Tableau**.

SKILLS:

Version Control Tools	Git, GitHub, Bit Bucket
CI/CD Tools	Apache Airflow, Jenkins, CircleCI, Travis CI, GitLab CI/CD, Azure DevOps
Data Engineering Tools	MS SQL Server, T-SQL, SQL Server Integration Services (SSIS), Informatica Power Center, SQL Server Reporting Services (SSRS), Apache NiFi, Apache Spark, Apache Kafka, Talend
Programming Languages	Python (NumPy, Pandas, Matplotlib, Seaborn) Scala, Java, SQL, Shell Scripting
Monitoring Tools	Datadog, Grafana
Cloud services	AWS(Amazon S3,AWS Lambda,AWS Redshift), Microsoft Azure(Azure Apache Beam, Azure Data Factory,Azure Synapse Analytics)
Database	MSSQL, MySQL, PL/SQL, MongoDB (NoSQL), Apache Hive, AWS Redshift, Amazon S3, AWS Lambda, snowflake
Operating Systems	Linux/Unix, Windows
Data Visualization Tools	Tableau, Power BI, MS Excel
Others	JIRA, Agile, Scrum, Kanban, Confluence, Terraform, Jenkins

CERTIFICATION:

- [Microsoft Azure \[DP-203\] Data Engineer Associate](#)
- [IBM Python for Data Science](#)
- [Uniq Technologies Python - ML certification](#)

PROFESSIONAL EXPERIENCE:

MetLife, USA	<u>Mar 2023 – Dec 2023</u>
Role: Data Engineer	

Responsibilities:

- Applied advanced **SQL queries** and logic for efficient data extraction from a diverse range of structured and unstructured data sources.
- Delivered end-to-end data processing pipelines using Hadoop ecosystem components (**HDFS, MapReduce, Hive**) to efficiently handle large datasets.

- Collaborated with Azure Functions in Python to load data into **Azure Synapse Analytics**, managing on-arrival CSV files in the **Azure Storage** account.
- Proficient in using **Azure Data Factory** for orchestrating and automating complex data workflows, ensuring efficient data processing and ETL operations in an Azure environment.
- Utilized Azure cloud platforms, working with **ADFs, Log Analytics, and Databricks APIs** to establish pipeline configurations.
- Implemented robust and scalable data pipelines using **Azure Apache Beam, Azure Data Factory, and Azure Synapse Analytics**, optimizing data processing efficiency and enabling real-time analytics in an Azure environment.
- Worked on real-time data processing solutions using **Spark Streaming**, enabling timely analysis of live data streams. Also, Optimized Spark jobs for performance gains, reducing processing time by 25% and lowering infrastructure costs by implementing resource-efficient configurations.
- Designed and conceptualized architecture, performed analysis, ETL operations, and visualization of large datasets extracted, **RESTful API** web service. Worked hands-on with **SQL SSIS, Python, and SSMS**.
- Executed automated testing frameworks for data pipelines, utilizing tools such as **Apache Airflow's** testing features and Testes, ensuring robustness and reliability.
- Hands-on experience with Snowflake utilities, **Snow SQL, SnowPipe, Big Data** model techniques using Python.
- Showcased expertise in creating visually appealing dashboards using **Power BI** for impact business presentations.

Bullfinch Software Private Limited, Bangalore, India
Role: Data Engineer

Jan 2020 - Aug 2022

Responsibilities:

- Collaborated with cross-functional teams in an Agile environment, experience in building and architecting multiple Data pipelines, end-to-end ETL, and ELT processes for Data ingestion and transformation in AWS.
- Demonstrated proficiency in AWS components, **EC2 and S3** showcasing a strong understanding of cloud infrastructure.
- Deployed and maintained data solutions on AWS, leveraging services like **Redshift, S3, and Lambda**.
- Built an ETL framework for Data Migration from on premise data sources such as Hadoop, Oracle to AWS using **Apache Airflow, Apache Sqoop and Apache Spark (PySpark)**.
- Utilized **Apache Spark and Apache Kafka** to process and analyze large-scale datasets, enhancing data processing speed and efficiency.
- Created on-demand tables on **S3** files using **Lambda Functions and AWS Glue** using **Python and PySpark**, coordinated intricate data workflows utilizing Apache Airflow, enhancing data processing efficiency through optimized task scheduling.
- Constructed sophisticated ETL workflows with **Spark Dataframe**, ensuring streamlined data processing.
- Devised simple and complex SQL scripts to check and validate Data flow in various applications, also scripts in Python and Shell for automating routine tasks and enhancing operational efficiency.
- Spearheaded the integration of data from various sources, utilizing **Power Query and Apache NiFi** to streamline ETL processes and optimize data workflows.
- Managed version control with **Git and GitHub**, promoting team collaboration and optimizing codebase management.
- Implemented robust monitoring solutions with **Datadog and Grafana**, providing real-time insights into system performance and data quality.
- Performed diverse data-related tasks including Data Analysis, Data Migration, Data Cleansing, Transformation, Integration, Data Import, and Data Export through Python.
- Designed and created insightful visualization dashboards in **Tableau**, translating complex data into actionable insights for informed decision-making.

EDUCATION DETAILS:

- **Master's in Computer Science**, NJIT, Newark, New Jersey **DEC 2023**
- **Bachelor's in Computer Science Engineering**, Velammal Engineering College, Chennai, India **APR 2020**