

Contact

www.linkedin.com/in/atharvakurle
(LinkedIn)
github.com/atharva-kurle (Portfolio)

Top Skills

Azure Databricks
Microsoft Azure
Fabric

Certifications

AWS Certified Solutions Architect – Associate
Microsoft Certified: Azure Data Engineer Associate
Big Data Hadoop and Spark Developer

Atharva Kurle

Data Engineer @ Makutu | Azure | AWS | Databricks | Spark |
Building efficient data pipelines to store and analyse big data
Sheffield, England, United Kingdom

Summary

Big Data Engineer and Solutions Architect with an industrial exposure of 3.5+ years in the field of data, programming and distributed systems. Successful history of building optimised, robust data pipelines and data warehousing solutions on Azure, Databricks and AWS. Passionate about learning new technologies and data driven tools.

- Python, Java, SQL
- Hadoop, Hive, Spark, Sqoop
- Docker, Airflow
- Data modelling (Kimball) and ER Diagram (OLTP), Lakehouse (Delta Lake)
- AWS (S3, EMR, ECS, Kinesis, Lambda, Aurora, Redshift, Glue, QuickSight)
- Azure (Synapse, Azure Databricks, Data Lake Gen 2, Azure SQL, Stream Analytics, Azure Data Factory, PowerBI)
- GitHub, Jira, Jenkins (CI/CD), Great Expectations, Azure Devops
- Shell scripting

I love to contribute to the community by writing blogs in a newsletter.

Experience

Makutu

Data Engineer

September 2024 - Present (1 year)

Derby, England, United Kingdom

- Designed a Delta Lakehouse architecture with Azure Synapse Analytics and Azure Data Lake Storage (ADLS), storing 600 GB of data using the Medallion architecture
- Optimised Spark job by applying salting to eliminate data skew, reducing processing time from 3 hours to 30 minutes

- Cut ADLS storage costs by 5x by resolving the small file issue using Delta Lake's Z-Order OPTIMIZE feature
- Used Azure Devops for CICD which improved overall development speed by 25%
- Worked on CRUD-based APIs and GraphQL services for third-party integrations
- Built event-driven pipelines using Azure Function Apps and Fabric SQL Database

The University of Sheffield
 Research Assistant - Data
 May 2024 - September 2024 (5 months)
 Sheffield, England, United Kingdom

- Designed and managed secure data infrastructure for the Computer Science Department using AWS VPC to enhance network isolation and security
- Extracted news related data from 100+ U.S websites using Pyspark and stored in parquet format.
- Implemented transformation logic to process and transform 200+ GBs of data to store in a data lake
- Designed data pipeline architecture to process and load the data efficiently
- Optimised code by using Pyspark partitioning to enhance the processing speed by 20%
- Used GitHub for collaboration and version control

AST LLC
 2 years 6 months
 Consultant
 June 2023 - July 2023 (2 months)
 Pune, Maharashtra, India

- Developed Python programs to automate data pipelines, significantly improving operational efficiency
- Implemented AWS integrations using Python, aligning technology solutions with business goals to achieve a scalable and flexible infrastructure.
- Proficient in writing precise SQL queries, extracting relevant data to provide valuable insights and data transformations.

- Designed robust API systems using Postman, ensuring seamless communication between applications. Deployed APIs across various frameworks to facilitate efficient data exchange.

Associate Consultant

February 2021 - June 2023 (2 years 5 months)

Pune, Maharashtra, India

Education

The University of Sheffield

Master's degree, Data Analytics · (September 2023 - August 2024)

Vishwakarma Institute Of Technology

Bachelor of Technology - BTech, Computer Science · (2017 - 2021)