Sharun Talapagala
Senior Data Engineer

sharuntalapagala@gmail.com
https://www.linkedin.com/in/sharun-t
303-4197-505
Denver, Colorado, 80210, United States

Professional in Databricks lake house, Snowflake and SQL databases to Azure Data Lake, Azure data lake Analytics, Azure SQL Database, DataBricks, and Azure SQL Data Warehouse, analytical, innovative leadership. Proficient in the AWS platform and its features including IAM, EC2 , EBS, VPC, RDS, Cloud Watch, Cloud Trail, Cloud Formation, AWS Configuration, Autoscaling, Cloud Front, S3 , SQS, SNS, Lambda, and Route53 Within Databricks, gained extensive experience in utilizing various Hadoop ecosystem tools, including HDFS, MapReduce, Yarn, Spark, Kafka, Hive, Impala, HBase, Sqoop, Pig, Airflow, Oozie, Zookeeper, Ambari, Flume, and Nifi. Skilled in working with EC2 instances ECS, EBeanstalk, Lambda, Glue, RDS, DynamoDB, CloudFront, CloudFormation, S3 , Athena, SNS, SQS, X-ray, Elastic load balancing (ELB) , creating auto-scaling groups, advanced. Hands-on experience in Azure Development, worked on Azure web application, App services, Azure storage, Azure SQL Database, Virtual machines, Fabric controller, Azure AD, Azure search, speech and notification hub.

Work Experience

Senior Data Engineer
Soimex

Jan 2024 - Present

- Developed, tested, and maintained critical data pipelines and architectures across various technical areas to support business objectives.
- Secured Azure Databricks ETL pipelines and protected sensitive data, ensuring robust controls for enterprise data protection for a 90%.
- Developed and optimized Synapse Pipelines, Data Factory, Snowflake, Apache Kafka, and Spark Streaming data integration and ETL pipelines for real-time data processing and Workflows, work independently.
- Designed and deployed Medallion Architecture in Databricks Data Lakehouse and Snowflake systems, enhancing data sharing, performance, scalability, and collaborative analytics, reliability, data analytics.
- Updated logical and physical data 100% in models based on new use cases to support evolving business requirements.
- Promoted a team culture of diversity, equity, inclusion, and respect, data management.

Data Engineer
Tata Consultancy Services

Mar 2022 - Jan 2024

- Converted 100% MapReduce to Spark transformations using Python and Spark DataFrames, improving data processing efficiency.
- Developed data solutions on AWS PaaS for visualization, assessing their impact on business operations to save 47% of the time.
- Created PySpark and Spark-SQL applications to analyze client usage trends by extracting, transforming, and aggregating data.
- Reduced AWS RDS expenditures by 30% through query optimization, index enhancements, and migration to Azure SQL Server Managed Instance and implement best practices, QA, retail and strong problem-solving skills.
- Integrated data from Salesforce, Marketo, and third-party APIs into Snowflake, enhancing data accessibility and analytics capabilities.
- Managed AWS resources including EC2, S3, RDS, Lambda, Glue, and Redshift, ensuring stable and scalable data infrastructure.
- Implemented automated data processes and task orchestration using AWS Step Functions and Lambda triggers.

Data Engineer
GE Healthcare

Apr 2020 - Oct 2021

- Extracted and loaded data from Azure Data Lake and Blob storage into on-premises SQL servers using Azure Databricks notebooks.

- Developed Azure Databricks ETL pipelines using Python to meet specific business needs, mentorship.
- Utilized Python modules (Pandas, NumPy) for data cleaning and preprocessing, facilitating ML model training.
- Created ETL pipelines using Python and Airflow to load data from SQL Server and PostgreSQL into data lakes.
- Managed csat Snowflake database advanced objects (tables, materialized views, schemas) for mentorship efficient data storage retrieval.
- Designed a data-centric and data requirements architecture for handling large datasets, improving decision-making processes.
- Conducted data profiling, cleaning, and validation in Snowflake to ensure data accuracy and integrity.

Core Skills

C++, Java, JavaScript, Python, SQL, PHP, ASP.NET, PLSQL, **Web Technologies :** HTML, CSS, Angular, ReactJs, Vue.js, remote, warehouses, tableau, big data, **Frameworks & Libraries :** Spring Framework, Hibernate, Servlets, JSP, Spring Boot, React, Angular, Flask, Django **Version Control & Testing :** Git, Bitbucket, JUnit, Automated Testing, Confluence, Verilog, Unit Test, Product Design, **Cloud Platforms & DevOps :** AWS, Azure, CI/CD, Docker, Kubernetes,
Infrastructure as Code : Azure, AWS CloudFormation **Other Technologies :** DDI (DNS, DHCP,
IPAM) Firewalls Load Balancers NICs (Network Interface Controllers) CI/CD, Power bi, data marts, e-commerce., AI,
Computer Engineering, Customer Requirements, Data Modeling, Data Models, Computer Engineering,
Customer Requirements, Identifying Training, MySQL, Predictive Models, Proof Of Concepts, Understanding Of Data,
Proof Of Concepts, Predictive Models, MySQL, Collaborative Skills, Great Communication, Time Management

Education

University of Colorado

Master of Computer Science