Vamsi Krishna

Hyderabad, Telangana, India vamsi.krishna26101989@gmail.com +91-9347719165 in/vamsi-krishna-bigdata/

SUMMARY

Data Engineering Architect with over 13 years of experience leads a team in developing and refining data pipelines using **AWS EMR, Apache Spark and Azure Databricks**. This role emphasizes scalable solutions, enhanced performance, and cost efficiency within the **Bigdata** ecosystem, focusing on solving complex business problems and creating value across domains.

EXPERIENCE

Data Engineering Architect

Paytm R & D **September 2022 - Present, Hyderabad**

- Built **Unified Ingestion Platform** with both **Streaming** and **Batch** accommodating 200+ pipelines ingesting 20k tables processing 5B events reconciling 70TB daily.
- Designed and implemented a scalable, fault-tolerant **Azure** data system handling petabytes, boosting retrieval speeds by 50% and enhancing reliability at peak loads.
- Built and managed scalable **Azure big data solutions**, growing from hundreds of terabytes to petabytes and tripling annual data volume capacity.
- Developed scalable ETL pipelines on Databricks **on Azure**, processing multi-terabyte datasets daily into **ADLS GEN2**, resulting in a 40% faster data processing time and enabling real-time analytics.
- Demonstrated expertise in Azure Big Data services like **Azure Databricks**, **Azure key Vault (AKV)**, **ADLS GEN2** optimizing workflows to handle over 15 petabytes annually, increasing efficiency and reducing costs by 25%.

Data Engineering Architect

Honeywell R & D **April 2021 - August 2022, Bangalore**

- Spearheaded the design and development of the **Data Enrichment Engine** utilizing the **Medallion** architecture pattern, enhancing data integration and analytics capabilities across business units at Honeywell, processing up to 5 terabytes of data daily.
- Developed a Real-Time Streaming Analytics framework that processes over 2 million IoT sensor data points per hour, enhancing operational insights by 30% through real-time analytics like filtering and transforming.
- Incorporated **Spark Structured Streaming** on **Azure Data Bricks** to extract data from **Kafka/EventHub** topics, pre-process, and store in a **Delta Lake**, increasing data throughput by 50% and reducing latency in data availability.
- Deployed **Prometheus** and **Grafana** for robust monitoring and alerting, improving system uptime by 99.9% and reducing incident response time by 40%.
- Defined a high-capacity data processing framework for IoT devices, handling millions of data points daily and storing in Time Series Databases and **ADLS Gen2**, enhancing retrieval speed and storage efficiency by 25%.

Lead Data Engineer

Ericsson R & D **August 2016 - April 2021, Bangalore**

- Engineered expansive **big data infrastructures on AWS, accommodating** data expansion from hundred terabytes to over 2 petabytes, which doubled the system's capacity to manage an annual data growth of 200%.
- Directed high-efficiency ETL pipelines with **AWS EMR** and **PySpark**, processing terabytes of data daily into **AWS S3**, cutting transformation time by 45% and facilitating real-time analytics.
- Combined **Kafka** and **Spark Streaming** to develop a real-time data processing system. This integration handled streaming data, boosting throughput by more than 50% and cutting down latency in analytics reporting by 30%.

Senior Data Engineer

Altran **October 2014 - April 2016, Bangalore**

- Defined Hive tables using **HiveQL**, handling and analysing more than 5 TB of data across these tables to facilitate data driven decision making in organisational projects.
- Demonstrated solid experience with Big Data services including **Spark, HDFS** and **Hive**, optimizing data processing workflows that handled over 25 petabytes of data annually, which increased operational efficiency and reduced costs by 15%.

BigData Developer

Wipro **June 2011 - September 2014, Bangalore**

- Pioneered excellence in managing a Hive data warehouse by creating and maintaining over 50 tables, **optimizing** data distribution through partitioning and bucketing techniques, and enhancing query performance by 30% through HiveQL optimizations.
- Migrated 20+ MapReduce programs to **Spark** transformations using **Scala**, resulting in a 40% reduction in processing time and improving job performance and scalability within the data processing workflows.

EDUCATION

Computer Science & Engineering (CSE), B.Tech

Sri Venkateswara University College Of Engineering (SVUCE) · Tirupati, Andhra Pradesh · 8.0
· Awarded with full scholarship and gold medal for 4 years for being class topper

CERTIFICATIONS

AWS Certified Cloud Practitioner

https://www.credly.com/badges/afadcdf2-d39b-4313-b9c0-a55d85d055df/linked_in_profile · 2024

Astronomer Certification for Airflow

https://www.credly.com/badges/6bb3b9b0-7860-41b6-8275-f30952d45836/linked_in_profile · 2024

Databricks Certified Associate Developer for Apache Spark 3.0 - (Credential ID - 60654282)

https://credentials.databricks.com/6268a072-d351-4e11-b37f-fcd0c08853f5#gs.9zb7e3 · 2024

SKILLS

AWS Cloud: AWS EMR, AWS Glue, AWS Athena, AWS S3, AWS IAM, AWS KMS, AWS Redshift
Azure Cloud: Azure Databricks (ADB), Azure Data Factory (ADF), Azure Data Lake Gen2 (ADLS Gen2), Azure Key Vault (AKV), Azure Event Hub
Data Engineering Tools: Spark, Scala, Python, SQL, Pyspark, Spark Structured Streaming, Hive, Kafka
Devops Tools: BitBucket, Jenkins, Bamboo, GitHub
SDLC: Agile, Scrum, Project management, Sprint planning
IDE: PyCharm, IntelliJ, Eclipse
Domains: Telecom, E-Commerce, Banking, Electrical & Electronics