LOKESH SESHAM

• 607-206-1329 • lokeshsesham114@gmail.com • http://www.linkedin.com/in/seshamlokesh

SUMMARY

Data Engineer with 4+ years of experience in designing, building, and maintaining **high-performance** data pipelines on **cloud platforms** (**AWS**, **Azure**, **GCP**). Proven expertise in **Spark**, **Hadoop**, **Kafka**, and other big data technologies for data analysis and transformation. Successfully delivered projects that **improved data processing efficiency**, **reduced costs**, and **enhanced data quality**. Strong analytical and **problem-solving skills** with a focus on **delivering impactful data solutions** for business needs.

PROFESSIONAL EXPERIENCE

Data Engineer

June 2023 - December 2024

Prime Therapeutics | Irving, Texas

- **Improved data processing efficiency** by implementing advanced techniques such as partitioning, caching, compression, shuffling, and joining in Scala, Python, and Java, achieving a **25% increase in processing speed** for large-scale datasets.
- **Enhanced operational workflows** by integrating Spark with Hadoop, Hive, Kafka, and AWS, resulting in a **40% improvement in efficiency** for medium-scale data processing while ensuring compliance with **healthcare regulations like HIPAA.**
- **Proposed and implemented a data storage optimization strategy** using Google Cloud Storage (GCS), reducing costs by **30%** and enhancing **data security**. This approach has been adopted as a standard practice within the organization..
- **Published real-time predictive services** through REST API integrations, enabling **dynamic data clustering** and enhancing the usability of client applications.
- **Led the orchestration of data clusters** on Google Dataproc, leveraging automation and scalability to achieve a **50% improvement in data analysis efficiency** and optimize resource allocation.
- **Developed and implemented data quality frameworks** to ensure **accuracy and consistency** throughout the data pipeline, collaborating with cross-functional teams to meet stringent **data governance and compliance standards**.

Data Engineer

January 2021 - July 2022

SM Infinite Technologies | Bengaluru, India

- **Spearheaded** the **optimization** of ETL processes by **20%** through the **strategic utilization** of **Spark DataFrame APIs** for efficient data analysis.
- **Successfully implemented Kubernetes** as the container orchestration platform, enabling seamless application deployment across **17 cloud providers**. This streamlined deployment significantly reduced time-to-market by **50%**, achieving a **15-minute deployment cycle** while maintaining **high availability**.
- **Proactively identified** and resolved software defects through bug tracking and root cause analysis within the **Jira** ecosystem.
- **Significantly enhanced** operational efficiency by **100%** through the orchestration of ETL processes using **Spark** and **AWS EMR**, ensuring **robust high availability**, **fault tolerance**, and adherence to **stringent data quality standards**.
- **Leveraged** the power of **AWS cloud services** to architect scalable and cost-effective data solutions.

TECHNICAL SKILLS

Programming languages	Python, Scala, SQL, R, Java, C, C++, C#, .NET, JavaScript, Golang
Database Technologies	Database Management Systems, SQL, NoSQL databases (MongoDB, Cassandra), Data warehousing concepts, Data modeling techniques
Frameworks and Libraries	Spring Boot, Angular, HTML, CSS, NodeJS, Probability and statistics concepts
Big Data Technologies	Apache Spark, Hadoop, Kafka, Databricks, Terraform, HDFS, MapReduce, Hive, Sqoop, NIFI
Cloud Platforms	AWS (EC2, S3, EMR, EBS, Lambda functions, Glue, Athena, Kinesis, Redshift), GCP (BigQuery, Dataproc), Azure Synapse, Iaas, Docker, Kubernetes

Data Engineering and Analytics Tools	ETL tools, Data visualization tools (Tableau, Power BI), Data pipelines, Data Management, Data Architectures, Data quality frameworks, Data mining, Data migration, Data splitting technologies, Data analysis, Airflow, DAG, Jira, BitBucket, Jenkins, Snowflake, MS SQL Server, Maven
Cyber Security Skills	Firewalls, Intrusion Detection/Prevention Systems, Encryption algorithms (AES, RSA), Hashing (SHA-256, MD5), TCP/IP, DNS, External/Insider threats, Attacker tactics, techniques, Detection & Response Engineering, IP-based wireless networks, IAM(Identity and Access Management)
Salesforce Technologies	Salesforce Lightning, Apex, Visualforce, SOQL, Salesforce DX, MuleSoft, Einstein Analytics, Tableau CRM
IDE & OS	Visual Studio IDE, UTM, Android Studio, Linux, Unix, Windows, IntelliJ, Putty, Mobaxterm, Eclipse, PyCharm, Xcode
Containerization, Orchestration & CI/CD	Docker, Kubernetes, GIT, Version Control, CI/CD Tools, Agile methodologies

EDUCATION

SUNY at Binghamton | Binghamton, New York
Masters in Computer Science
SRM University | Andhra Pradesh, India
Bachelor of Technology in Computer Science

CERTIFICATIONS

- **AWS Certified:** Solutions Architect
- **AWS Certified:** Data Engineer
- **Microsoft Certified:** Azure Data Engineer Associate
- **IBM Certified:** Big Data and Data Analyst
- **Google Cloud Platform Certified Expert**
- **Salesforce Certified:** Administrator, Platform Developer

PROJECTS

- **Cloud-Native ETL Solution: Airflow-Managed Data Pipeline on AWS**
  Developed and deployed a scalable data pipeline using **Apache Airflow** to automate **ETL processes**, seamlessly transferring data from **Amazon S3** to **Redshift**. Introduced data quality checks and performance improvements that reduced processing time by **30%** and ensured **99.9% data accuracy**. This solution empowered real-time analytics and reporting, providing actionable insights that supported key business decisions across multiple departments.

- **Building a Customer Segmentation Pipeline with Azure ML**
  Developed clustering models using **Azure ML Designer** to segment customers based on their purchase history and demographics. Utilized **k-means** and **DBSCAN** algorithms to identify distinct customer segments, resulting in a **10% increase in targeted marketing campaign effectiveness.**

- **Real-Time Analytics Pipeline: Kafka- Powered Data Processing**
  Built a high-performance data pipeline leveraging **Apache Kafka** and **REST APIs**, achieving a **throughput of 1 million events per second** with **40% lower latency**. Integrated **Generative AI models** to enhance analytics for IoT, finance, and customer behavior use cases. This solution reduced infrastructure costs by **30%** and enabled real-time, AI-driven decision-making, driving innovation and efficiency across the business.

- **Real-time Logistics Optimization for Amazon Delivery**
  Designed and implemented a real-time data pipeline that optimized Amazon delivery routes and schedules, leading to [quantifiable result, e.g., a 5% reduction in delivery times, a 3% increase in driver efficiency].