Manish Charu

(444) 444-4444 | mancharu@gmail.com | [LinkedIn](#) | [Github](#) | Plano, TX

SUMMARY

Experienced data engineer with over 12+ years of experience in designing and building scalable data applications. Proficient in building scalable data pipelines for both batch and streaming applications, leveraging expertise in Spark, Kafka, Hadoop, AWS, and NoSQL databases with a proven track record of optimizing performance and maximizing resource utilization. Experienced in designing data models for both SQL and NoSQL databases, with a strong leadership background in delivering successful data-driven projects across cross-functional teams that deliver actionable insights and drive business value.

SKILLS

Big Data Technologies	**:** Spark, Kafka, Snowflake, Databricks, Hadoop, Big Data, Map Reduce, Hive, Pig, Delta Lake, Apache Iceberg, Airflow, Astronomer, Presto, Athena, ETL, Data Modelling, Data Pipelines.
Cloud Platforms	**:** AWS, EC2, S3, SQS, SNS, ECS, EMR, Lambda, RDS, GCP, Cloud Technologies.
Backend Frameworks	**:** Spring Boot, REST APIs, Microservices, Event Driven Architecture, Distributed Systems.
Databases and Storage	**:** SQL, MySQL, PostgreSQL, NoSQL, HBase, Cassandra, Elasticsearch, Solr, Redis.
CI/CD and DevOps	**:** Jenkins, Git, Docker, Kubernetes, CI/CD.

WORK EXPERIENCE

Lead Data Engineer   *March 2023 – Present*

Programming Languages   **:** Java, Scala, Python.

Capital One - Richmond, VA

- Redesigned and optimized customer core elevation service, enabling seamless batch and real-time updates for millions of customers, resulting in an impressive 36X reduction in runtime from 72 hours to a mere 2 hours.
- Ideated & implemented a data driven testing framework for customer-core platform which improved developer productivity by 3x.
  - Mentored and helped the team in design decisions, optimizing and debugging complex spark & spring boot applications.

Lead Software Engineer   *May 2021 – Feb 2023*

Demandbase - San Francisco, CA

- Designed and implemented scalable data pipelines using Java, Scala, Spark & AWS and deployed them as scheduled jobs on Airflow/Astronomer which handles millions of events per hour.
- Redesigned Site Analytics platform by deprecating snowflake usage using latest apache iceberg, reducing the monthly cloud costs by 20%.
- Designed and implemented Intent data pipeline from scratch, leveraging latest Delta Lake architecture with Spark, EMR, SQS, and SNS. Processed millions to billions of intent records, resulting in a 4% increase in new customers and improved customer retention.
- Ideated, designed and implemented multithreaded HBase benchmark framework for monitoring & upgrading AWS HBase cluster and collaborated with DevOps teams in upgrading existing HBase clusters which improved the performance of all data pipelines by 4X.
- Supervised and mentored a team of junior developers, conducted code reviews and provided guidance and direction in design, implementations and resolution of highly complex technical problems resulting in a 15% increase in the speed of feature delivery, 75% increase in the data coverage and 5% increase in data accuracy at Demandbase.

Principal Software Engineer   *May 2019 – May 2021*

InsideView, A Demandbase Company - Hyderabad, India
- Implemented CMAAS (Company Matcher as a Service) using Java, Servlets, Spring Boot, ML models, Spark and deployed it as a scalable microservice on ECS, handling high concurrency leading to increased customer acquisitions and higher adoption rate.
- Led ideation and collaborated with cross-functional teams to upgrade EMR infrastructure, refactor the codebase, and implement crucial optimizations such as idle timeout and auto scaling on EMR clusters, resulting in an impressive 30% reduction in EMR cloud costs across the organization.
- Implemented a reporting & monitoring framework for Data Engineering, leveraging Spark, Kafka, HBase, and Grafana resulting in a 30% decrease in production issues.
- Implemented a data warehouse on top of Amazon S3 using Spark, Apache Iceberg, Athena, and Kafka, which assisted data scientists and data analysts in extracting meaningful insights. This, in turn, led to a 10% increase in new customers.

Senior Software Engineer   *April 2012 – May 2019*

InsideView, A Demandbase Company - Hyderabad, India
- Designed and implemented Data Platform-1.0 from scratch using HDFS, Hive, Pig, Oozie, Java, HBase and deployed it on HDP(Hortonworks Data Platform) resulting in an 80X performance improvement across all pipelines.
- Revamped and refactored legacy ETL/Data Pipelines, leveraging Spark DataFrames in Scala. Optimized Contact Dedupe, URL Validation, Company Transformation, and Address Standardization services, resulting in an outstanding 150X performance improvement.
- Architected and implemented a multi-node Kafka cluster including backup & recovery of metadata, enhancing data processing performance by 4X for the entire Data Engineering team.
- Designed and implemented a distributed Idgen web service capable of processing millions of records, resulting in a significant increase in productivity for internal customer-facing teams.

Software Engineer   *June 2010 – April 2012 United Online, Inc - Hyderabad, India*
- Implemented various backend modules for FTD websites using Java and Spring.

EDUCATION

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, Hyderabad, India   *June 2010*

Bachelor of Technology, Computer Science and Engineering

Relevant Coursework: Distributed Systems, Design & Analysis of Algorithms, Database Management Systems, Data Warehousing & Data Mining, Computer Networks, Information Security, Neural Networks

ACCOMPLISHMENTS

- Runner-up of InsideView Hackathon contest for 'URL Validation Project', which was later adopted as a core product feature within the company.
- Received the 'Best Employee' award in the 'We Innovate With Intention' category at Demandbase in 2022. ● Secured an all India 230th rank in the prestigious ISRO exam held in April - 2011.