DEVESH SOLANKI

+91-8867941471 • devesh.acc.99@gmail.com • in/deveshsolanki

SUMMARY

Collaborative Data Engineer with over 3 years of experience in designing, optimizing, and deploying large-scale data processing systems. Expert in leveraging big data technologies like Apache Spark, Flink, Hadoop, and Kafka to drive business transformation and deliver actionable insights. Proven track record in reducing operational costs, enhancing system performance, and ensuring the reliability of real-time data pipelines. Experienced in cloud migrations from on-premise to AWS and GCP, with a passion to achieve strategic goals.

WORK EXPERIENCE

Software Engineer, Verse Innovation — July 2021 - Present

- **Data Pipeline Development:** Developed and deployed large-scale data processing systems, including Spark-based batch jobs and Flink streaming applications. Managed pipelines that processed 5TBs+ data daily, improving query performance by 20% and enabling real-time analytics with sub-second latency.
  - developed live-calling feature, processing user activity events and cross-referencing historical data in GCS to enhance real-time notifications on user online and busy statuses.
  - helped maintain in-house DB(obelix) which was a user-profile DB and served as base DB for notification tool.
  - created contact-sync pipeline, pushing events into a master dataset, improving follow notifications and recommendations, with data stored in Redis for API use.
  - created real time tables on GCP BigQuery to enable efficient querying, improving data accessibility and reducing dependency on live databases for analytics use case
  - created bloomFilters which store key values from delta tables and used for optimized recommendation processing
- **Cloud Migration & Optimization:** Worked on the migration from on-prem Kafka to AWS Kafka, followed by AWS to GCP migration. Achieved more than 30% reduction in cloud costs and enhanced system performance and reliability.
- **Cost Optimization:** Reduced EMR service costs by 30-40% through performance tuning and configuration optimizations in Spark and Flink jobs. Executed POCs for GCP services to ensure smooth migration while maintaining system efficiency.
- **Automation & Innovation:** Automated complex data processes using Python and shell scripting, reducing manual effort by 50%. Integrated Apache Superset and WhyLabs for advanced monitoring and observability, improving data quality and system visibility.
- **System Performance & Monitoring:** Optimized Spark and Flink jobs, resulting in significant operational cost savings. Implemented monitoring and alert systems with Grafana and OpsGenie, ensuring fault tolerance and reducing incident response times.
- **Collaboration:** Worked closely with product, business and analytics teams to meet data requirements, increasing data coverage by 40% and enabling more comprehensive analytics.

EDUCATION

B.Tech in Informtion Sciences, DSCE — 2017 - 2021

- Best Project in 2019
- 4th Place in Hackathon in JSSATE organised by Honeywell

TECHNICAL SKILLS

- **Big Data Ecosystem**: Apache Spark, Flink, Kafka, Hadoop, Airflow
- **Cloud Technologies**: GCP(Dataproc, BigQuery, GCS), AWS(Kinesis, Athena, Lambda, Firehose)
- **Programming Languages**: Java, Python, SQL, Scala, C/C++
- **Database Technologies**: Redis, Druid
- **Data Visualization**: Grafana, Superset