Abhishek Gupta

Email/skype – abgupta.456@gmail.com

Mob – 91-9782135409

Email/skype – abgupta.456@gmail.com

Mob – 91-9782135409

Professional Synopsis
Data Engineer with 5 Years of IT experience in problem solving, developing and maintaining products with high business impact. Currently working as Data Engineer with Back-end development capabilities in Walmart. ● Have worked in GCP, Hadoop Ecosystem, Web applications and scalable micro services. ● Design and build data processing pipelines using tools and frameworks in the Hadoop ecosystem ● Good hands-on SQL and experienced in NoSQL database such as MongoDB ● Working in Agile methodology, coding standards, VCS, CI/CD tools and test-driven development. ● Able to work independently with minimal supervision.

Skill Set	
Core	Data Structure, oops, Python, Java
Hadoop Eco-system	Hadoop, Spark, Hive, Kafka, Airflow, GCP
Web Tech.	Restful, Jersey, Flask, JMeter
Databases	MySQL, MongoDB
Tools & Tech.	Eclipse, Git bash, Docker Container, Bamboo, HP Fortify, Jupyter, Anaconda

Employment Details	
Employment with Walmart Labs, India as **SDE 3**	Dec,19 - Current
Employment with Deloitte USI as **Big Data Consultant**	Dec,17 – Dec,19
Employment with Tata Consultancy Services Ltd. as **IT Developer** (Innovation Lab)	Jul,16 – Nov,17

Educational Background	
B.E. (CSE)	Jaipur Engineering College & Research Centre (JECRC)
Certifications & Achievements	● Certificate of Digital: Apache Spark-BFS Intermediate by Tata Consultancy Services ● Honor Code Certificate on scoring 94% in Big Data with Apache Spark from University of California (MOOC). ● GCP certified Data Engineer
Presentations POC's Blogs	● Security of Applications ● Cloud Native solution ● Spark on Kubernetes ● Dynamic Schema in Stream or batch

Key Projects

Governance Strategy :   (Dec,17 – Currently@ Walmart Labs)

My tasks are to design and develop the ETL systems for various data sources over which reporting dashboards are built:

1. Connect with the stakeholders and understand the long-term needs
2. Design cloud native solution and priorities the features in product
3. Design, model and develop Cloud native process and tools for CICD, ETL and reporting.
4. Analysis, design and develop cloud native ETL pipelines.
5. Design standards and develop reusable components for saving development time for ETL solutions.
6. Conversion of manual reporting to adhoc reporting solution and eliminate unnecessary BI tools to save costs.

Technologies: Spark, Hive, Big Query, Looper, Concord, GCP, Airflow, Automic, Erwin

Cognitive Cloud Steward: (Dec,17 – Dec,19 @ Deloitte USI)

Deloitte has developed some very reliable predictive models for US based insurer majors and aimed to commercialize those applications to other insurers by setting up a cloud based commercial data lake where data will be ingested from different sources and different insurers vendors. The data in end will be consumed by ML models to predict the industry trends. My responsibilities involved to handle Anthem and PepsiCo users with below responsibilities -

- Designing and developing the generic complete data pipeline in Spark, Scala and Python to convert the data from commercial data lake into semi-structured format in JSON and provide it for ML Data Model based applications.
- Developed the loose coupled data quality rules and transformation engine data pipeline for all the insurers.
- Create spark job submission framework to create dynamic spark submit using previous job statistics which handle different data loads provided by clients.
- Develop deployment strategy with Bamboo for multi-tenant and multi environment.
- Create a real time recommendation system for fraud detection through Python Flask based microservice and interact with the different Machine learning models.
- Maintain reporting framework for incremental data using user specific data rules and SQL queries.
- Create test plans for performance testing using JMeter to test the load on microservices and load balancers performance and automate this process using Bamboo.
- Interact with data scientist and BA's to understand the requirements and design related changes.

Technologies: Hive, Spark, Scala, Python, Git, Kafka, Bamboo, Flask, JMeter, Bonobo, Airflow

---

Gen Log Data Lake: (July,16 – Nov,17 @ TCS)

Morgan Stanly have a data lake where it collects the various trading web applications data for different environments. Impact of this project is to maintain data lake from collected from various web application's QA, Prod and other environments. Responsibilities -
- Developed real-time data ingestion layer using Kafka and Kafka Connect in Apex application and spark application.
- Fuzzy logic enhancement with pipeline of ML interaction logic like Euclidian.
- Developed queries required to interact with HIVE tables. And write filter using Pyspark
- Perform analyses and provide interpretation, guidance and recommendations on marketing performance and market data to aid decision making.

Technologies: Kafka, Hive, Spark, Tableau, Jupyter notebook

Security Arch. Remediation Application Enhancement: (Feb,17 – June,17 @ TCS)

Scope of this project is to add Restful web services and remediate cross site scripting, SQL injections, OS injection and other security loopholes. My task is to -
- Analyze, design and create service using Jersey library and handle DAO layer using Spring.
- Remove the XSS bugs using ESAPI and java encoding libraries and use the prepared type in Business Logics (Spring JDBC template) to remove sql injection bugs.
- Optimize the code using generic design pattern of java.
- HP Fortify job integration to provide better one click solution.

Technologies: Spring, Apache cxf, Apache jersey, Service Security, Sybase DB, Cucumber