Abhishek Gupta

Email – abgupta.456@gmail.com

Contact – 91-9782135409

Professional Synopsis
Big Data Engineer with 3 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 Deloitte USI.   • Have worked in Hadoop Ecosystem, Web applications and scalable micro services.   • Design and build data processing pipelines using tools and frameworks in the Hadoop ecosystem   • Involved in projects on Batch processing using Hadoop as well as Real-Time streaming, Data Ingestion, Transformation and processing.   • Good hands on SQL and experienced in NoSQL database such as MongoDB   • Working in Agile methodology 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, MapReduce, Spark, Hive, Kafka
Web Tech.	Restful, Jersey, Flask, JMeter
Databases	MySQL, MongoDB, Amazon Redshift
Tools & Tech.	AWS, Eclipse, Git bash, Docker Container, Bamboo, HP Fortify, Jupyter, Anaconda

Key Projects

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

Deloitte has developed some very reliable predictive models based on AI and ML for a US based Health Insurance major over a period of 5 years. Now Deloitte 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 client with below tasks as -

- Creating the data pipeline in Apache Spark, Scala, Python and to ingest reference table data from different source into the cloud system.
- Develop APIs for DQR engine and transformation for curated layer
- Creating Python to automate the object creation tasks and metadata for object creation task.
- Designing and developing the complete data pipeline in Apache 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 data pipeline logic for same load to provide data feed to downstream application as needed based on Slowly Changing Dimensions concepts.
- Design and develop framework to handle spark jobs as per previous job feedback.
- Develop deployment strategy and Create Bamboo plans with multi-tenant and multi environment support.
- Create a daily claim decision suggestion system through Python Flask based microservice and interact with the Mainframe process provide recommendation based on different Machine learning prediction model
- Design and implement validation and transformation framework to validate the payload and transform it to required model.
- Create test plans for performance testing using JMeter to test the load on microservices and load balancers performance

Technologies: Hive, Spark, Scala, Python, Git, Kafka, Bamboo, Flask, JMeter, Elastic search

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.
- Use KDB integration to gain access real time containers.
- 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: MapReduce, Kafka, Hive, Spark, KDB, 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

Airtel CRD: (Jan,15 – feb,16 @ Internship at Data Infosys Ltd.)

Scope of this project is to provide analytical service for the enhancement of Telecommunication recommendation engine using customer billing data according static area plot. My task is to -

- Develop MR jobs for Batch processes to make report per hope time & Ingest data using sqoop from RDMS and excel workbooks and implementation hive table depending on read/write scenarios.
- Perform Design, develop and generate reports that measure and tie marketing performance to business results.

Technologies: MapReduce, Sqoop, Hive, Spark, NumPy, Jupiter notebook,AWS

Employment Details	
Employment with Deloitte USI as **Big Data Consultant**	Dec,17 – Current
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 Excellence in Hadoop and Map Reduce by Cloudera (MOOC). • 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). • Certificate of Big Data and Hadoop Essentials by Udemy • Certificate of Git and GIthub by Lynda. • J2EE and Android Enterprise certificate by Hewlett-Packard.
Presentations & POC's	• Data Duplication solution in Big Data • Data Lake Design Model • Security of Applications