

GAURAB KUMAR JHA

Staff Engineer | Data Analytics Engineer



🏢 Western Digital
✉ Email: gaurabkjha@gmail.com
📱 Mobile: +91 9923832293
🔗 GitHub: github.com/gaurabjha

OVERVIEW

To get a challenging position where I can put all my experience and knowledge to give a significant growth to the organization.

PROFESSIONAL SUMMARY

- Almost 8 years of experience as JAVA, Python & Big Data Developer
- 5+ years of experience in Big Data & Cloud Development – Batch and NRT Processing of the clinical data, SSD/HDD Test Data
- Fair Knowledge on Java, Python, C#, ruby scripting and relational databases, SQL(Structured Query Language) etc.
- Designed and Contributed to the design and implantation of various applications and tools.
- Interviewed, Mentored, evaluated and graduated new hires and interns in the team.
- Good Knowledge on Dependency management, Source Control and comprehension tool Such as Apache Maven, Git etc.
- Participated in peer code review to maintain high coding standards.
- Good team player with excellent communication skills, interpersonal skills with problem solving, trouble-shooting capabilities.
- Handled responsibility of Release and Deployment manager for pushing the code into production environment
- Worked on Cloud Storage & AWS – EMR, EC2, GCP/GKE

SCHOLASTIC PROFILE

2010–14	B.Tech (ECE)	West Bengal University	8.40 CGPA
2008–10	Higher Secondary (PCM) – 12th	Kendriya Vidyalaya (CBSE)	78.2%
2000–08	Secondary - 10th	VMS (ICSE)	88.2%

TECHNICAL SKILLS

Languages: JAVA, Python, C#

Front End: HTML, JavaScript

Framework: Apache (Crunch, Spark, Kafka, Hive), Spring Boot

Hadoop Ecosystem: MapReduce, HDFS, YARN, Oozie, HBase

IDE's: JetBrains IntelliJ, Eclipse, PyCharm SQL Developer, Visual Studio Code

Serialization Format: AVRO, JSON, CSV, Parquet

Database: MySQL, Oracle11g, Oracle12c, HBase

Build Tools & Source Control: Apache Maven, GitHub, BitBucket

CI/CD & Code Review Tools: Jenkins, Crucible, Nexus, Spinnaker, Splunk

Cloud Infrastructure Provider: GCP, AWS (S3), Cloudera

Other Tools: Node Management (CHEF/Ansible), Kubernetes, Mesosphere, Cloudera Manager CDH6, Confluent, Putty, Hue, Airflow

PROJECTS

#1 Western Digital | Sandisk

May 2020 - Present

Environment: Java, JDK 11, Python, Spark, Confluent Kafka, Kubernetes, Apache Airflow, GKE/GCP etc.

Role: Staff Engineer | Data Analytics Engineer

Responsibilities:

- Mesosphere to GKE Migration
 - Migrate Mesosphere to GKE for Dev/Stage/Production Environment
 - Install, Maintain Tune Confluent Bitnami Kafka using Bitnami Helm Chart on the Google Kubernetes Engine
 - Manage the Operation Team to keep the environments up and running.
- Create and Maintain Spark Pipelines for Report Generator.
- Create Airflow Dags for Backfill Automation.
- Design and implement Airflow Dags to run the Model on the data and score them for the HDD Testing and Manufacturing team.
- Maintain the Badge processing via Airflow/Python , and Streaming via Kafka Streams/Java.
- Add features to the kafka streams (ETL) as per the request from the data science and the Manufacturing testing team.
- Build multiple Benchmarking Tools for functional and Load Testing of the systems.
- Part of the Architectural forum, which discusses and Develops IT solution for the

issues and use case requested across the organization.

- Participate in the hiring efforts.

#2 Oracle Cerner Corporation (Healthcare Solutions, Services)

August 2018 – May 2022

Environment: Java, JDK 1.8, Crunch, Spark, Kafka, HDFS, HBase, Oozie, CHEF etc.

Role: Java | Big Data Developer

Responsibilities:

- Batch ETL (Extract Transform Load) of client's clinical data by deploying multiple data pipeline implemented in Spark (PySpark) and Apache Crunch (MapReduce Jobs).
- Perform Near Real Time (NRT) Processing and Transformation of clinical data using Apache Kafka and Apache Storm – deploying storm Topologies for each client.
- Perform data cleansing, transform and normalization of clinical data ingested into HDFS from various data sources – serializing it using AVRO serialization format and storing the data it into HBase Tables.
- Configure & deploy Oozie workflow using CHEF to Schedule MapR/Crunch Jobs.
- Wrote Unit and Integration Tests using Mockito, JUnit, Scala testing frameworks.
- Worked on various language e.g. Ruby, Shell to write script for functional testing of the functionality. Also, Setup different pipeline in Jenkins using Jenkins file and groovy script.
- Generate smoke test data, and run an integration test to check – if the data is processed correctly

#3 NTT Data

February 2017 – August 2018

Client: Owens & Minor (healthcare logistics)

Environment: Hive, Hbase, Storm, Maven, Kafka, ZooKeeper, Linux, Splunk, Apache Crunch and Oracle12c

Role: Application Development Consultant

Responsibilities:

- Created Hive tables to migrate and load large sets of data in Big-data Ecosystem.
- Created Shell Scripts to implement the business logic and automate the data loading
- Wrote Apache Crunch programs to run Business logic to load the data into the system.
- Worked with job schedulers like Rundeck and Oozie
- Providing data quality and availability to BI team
- Implemented unit test cases
- High-speed querying.
- Being a part of a POC effort to help build new dataset.
- Involved in the bug fixing and enhancements.

#4 Tech Mahindra

October 2014 – January 2017

Client: AT&T Inc. USA (Telecom)

Environment: Java 1.7, SQL, PL/SQL, Servlets, JSP, Spring, Hibernate, Eclipse, Putty, SOAP UI, SQL Developer, Toad and Oracle Database (10g and 11g)

Role: Software Engineer

Responsibilities:

- Create/modify/update Web modules of the myAT&T web application using the spring and Hibernate framework.
- Address to defects raised by testing teams – .jsp updating and stitching.
- Maintaining code changes and updating based on client requirements.
- Development mainly based on core and advanced java.
- Enterprise Document Delivery – development of notification sender application from the AT&T web application to end customers.
- Create/update XMLs and XSDs for proper DOM management.
- Handling defects and exceptions highlighted by the testing teams.
- Code changes and updating the application based on client requirements.
- Working on Web Services, Core Java, JEE code changes based on requirements.
- Updating and maintaining the database of client customers and application.

ACHIVEMENTS & AWARDS

- *Awarded recognition for a successful and seamless migration of Kafka/Kafka streams from mesosphere to GKE*
- *“1st Place in Q4 2020 Hackathon at Cerner for Automations of the deployment and running the workflow on AWS EMR.”*
- *“NOTT Award for Extra Ordinary Performance in Q2 2019 and Q2 2020”, and few organizational awards for helping setting up new team within the organization.*
setting up an on-boarding project for new hires in the team, which gives them the exposure to technical stack which the team uses.
- *“Bravo” for commendable contribution to the myATT project.*
- *Awarded with Golden Certificate, Technical Topper of the Badge and Technical Topper of the ELITE (Training Program of TechM) 2014.*

This is to certify that all the information given above is true to the best of my knowledge