

Biswajit Sahu

Bengaluru, Karnataka, India



biswajit.146@gmail.com



+917019840640/+919438910342



<https://www.linkedin.com/in/biswajit-sahu-3a736397/>

Summary

Working as a Senior Hadoop/Spark Developer, with Innova Solutions. Having total 5+ years of experience in IT industry.

I am well versed in Hadoop ecosystem and related technologies. Expertise in Building Pipeline, for loading Batch and Streaming data.

Currently working as a Spark Developer with Big Data/Hadoop technologies like HDFS, HIVE, PIG, Sqoop, Spark, Scala, Python, Kafka, AWS, Airflow, SQL, NOSQL, Python, StreamSets, Shell scripting, Control-M, Druid, SuperSet.

Experience



Senior Data Engineer

Innova Solutions

Nov 2021 - Present (10 months +)

Working for a US based Health Care Client. Managing a huge medical data of patients from different zones.

Working for to get the real time prediction for patient based on input data from the source system, also to get the Auto Review of the claim process, admission and other medical related decisions.

Roles and Responsibilities:

- Develop application to load medical data into Druid system.
- Develop Real time application for to get prediction in real time/near real time.
- Developed an Auto Review process to get the flow automatically streamed line as soon as any new patient admitted.
- Responsible for end-to-end process/steps flow of the project
- Automate the job run using scheduling tool
- Create clusters in AWS to run the Spark job with scheduling.
- Developed a different set of queries, usable for front end team.
- CI/CD using Jenkins.



Senior Software Engineer

Lumen Technologies (Formerly known as CenturyLink)

Jul 2017 - Nov 2021 (52 months)

Focus of this project is to create a Data Lake to save all the different data from different sources in HDFS. The data is a Historical and Streaming data. Also, we are working in file-based Ingestion, which we getting from different sources.

The primary phase of the project is to load the data from different sources and make it available for the other downstream project/users. Also making data available for Data Science team for Business.

Roles and Responsibilities:

- Preparing DDL and creating Hive tables as per the requirement.
- Creating Control-M jobs for job scheduling.
- Loading data from streaming source to the HBase/Hive using Kafka and Spark streaming.
- Loading data into Hive tables using Sqoop.
- Preparing all project related documents like Checklist's, TDD.
- Actively participation in any defect resolution and get it resolved within ETA.
- Driving the scrum calls in the absence of Scrum master.
- Actively driving the working session in which different; joins with their facing issues in project

Education

Silicon Institute of Technology, Bhubaneswar

Bachelor's Degree, Bachelor of Technology in Electrical & Electronics.

Aug 2011 – Jul 2015

Skills

Big Data • Hadoop • Hive • Sqoop • Apache Spark • Scala • MapReduce • BMC Control-M
• Jira • Agile Methodologies

Honors & Awards



GEM Award

Apr 2022