

Tanuj Huria

Certified Associate Developer

Mobile No: -8950449974 | E-mail: tanu.huria@gmail.com | D.O.B: 14/11/1993

Current Address: 6/14, Sector-14, Gurugram, Haryana, 122001

Career Objective:

To pursue a challenging and rewarding career that provides ample opportunities to learn, apply and enhance my knowledge and experience for the overall growth of the organization and self.

Profile Summary:

- AWS Cloud implementation experience on EC2, S3, Glue, Kinesis, Lambda, API Gateway, Secret Manager, SSM, EMR, SNS, Athena, Redshift and other services
- Data Pipeline: Data Lake on S3 and transforming data from glue.
- Real-time Pipeline: Kinesis stream & Lambda
- Spark Architecture and Components : Spark Core & Spark SQL with Data frames
- Data Analysis using Spark Core & Spark SQL & Spark Streaming.
- Spark SQL with MySQL (JDBC) source
- Spark RDD Actions & Transformations.
- Splunk dashboard and logs forwarder.
- BigData Technologies: Hadoop(CDH Distribution), Spark Ecosystem.
- Hadoop HighLevelLanguages: Hive2.3.2.
- Knowledge of end to end project work flow.

Experience Summary

- Worked as Software Developer in Tata Consultancy Services (Gurugram) (Apr 2016 – Feb 2019)
- Working as Associate in Cognizant (Gurugram) (Feb 2019 – Nov 2019)
- Working as Senior software engineer in Incendo(Gurugram)(Nov 2019 - Present)

Projects

Company name : TATA Consultancy Services (Gurugram)

Job profile : Software Developer

Project Name : Customer BigInSights Application Integration Hadoop/Spark

Environment: Hadoop YARN, Spark Core, Spark Streaming, Spark SQL, Scala, Python, Kafka, Hive, Sqoop, Amazon AWS, Cassandra, Oozie, Jenkins, Cloudera, Linux.

Skills: Python, SQL, PL/SQL, HiveQL, Unix, Shell Scripting, HDFS, YARN, Hive, Sqoop, Spark, Kafka, Zookeeper, Oozie, Yarn, Spark, Streaming, Spark SQL, Spark RDDs, AWS(EC2&EMR), Docker, kubernetes.

Team Size : 15

Description : The primary objective of this project is to integrate Hadoop (Big Data) with the Customer BigInSights Application to leverage the raw/processed data that the big data platform owns. It will provide an enriched customer experience by delivering customer insights, profile information and customer journey. This will allow us to prioritize conversations to drive value generation and a 360 degree view of an account member.

Company name : Cognizant Technology Services (Gurugram)
Job profile : Associate

Project Name : GRT Splunk Onboarding

Environment: Hadoop, Splunk Spark Core, Spark Streaming, Spark SQL, Kafka, Hive, Sqoop

Team Size :12

Description : The primary objective of this project is to ingest logs of different jobs(kafka,hive) in to splunk system. This will allow us to monitor our different jobs into single dashboard and debug in-case of any issue.

Company name : Incedo Technology Services (Gurugram)
Job profile : Senior software engineer

Project Name :BPD bank

Environment: AWS EC2, S3, Glue, Kinesis,Lambda , API Gateway , EMR , Secret Manager, SNS, Athena, Python,Pyspark,Airflow

Team Size :8

Description : The primary objective of this project is to ingest data from source system to Aws S3 by using different aws services(S3,spark, Kinesis,Glue). This will allow data science team to run algorithms on transformed data and produce targeted customers list to send back to client system.

Customer visualization dashboard using Amazon Quick sight.

Real-time pipeline to ingest data and produce insights using amazon kinesis and lambda function.

Role and Responsibilities : •

- Responsible for ingestion data from client server to AWS S3 data lake using python script.
- Data Quality scripts using pyspark to perform quality check on raw data and filter out data.
- Developed Glue scripts (pyspark) to create features as per the Data Science requirement.
- Developed python scripts to monitor incoming data and then send notification using AWS Glue, Lambda and SNS service.
- Developed python scripts to refresh Athena tables as per the requirement.
- Responsible for building real time pipeline for data ingestion and provide insights on the fly.
- Build dashboard related to different job structure
- Developed Spark scripts by using Pyspark shell commands as per the requirement.

- Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive.
- Developed Python scripts, UDFs using both Data frames/SQL/Data sets and RDD in Spark 2.x for Data Aggregation, queries .
- Experienced in performance tuning of Spark Applications for setting right Batch Interval time, correct level of Parallelism and memory tuning.
- Optimizing of existing algorithms in Hadoop using Spark Context, Spark-SQL, Data Frames and Pair RDD's.
- Performed advanced procedures like text analytics and processing, using the in-memory computing capabilities of Spark.
- Worked on Cluster of size 40 nodes.
- Analyzed the SQL scripts and designed the solution to implement using Spark
- Implemented schema extraction for Parquet and json file Formats in Hive.
- Implemented Partitioning, Dynamic Partitions, Buckets in HIVE.

Academics

- B-Tech(CSE) from Vaish College of Engineering with 75.08% (Year 2011-2015)
- XIIth - Passed from D.A.V Public School with 70.80% (Year 2011)
- Xth - Passed from D.A.V Public School with 71.5% (Year 2009)

Technical Proficiency

Programming Languages	Python, Pyspark
DBMS	MYSQL, Oracle, SQL Server
Operating System	Windows, Ubuntu

Interests

Music | Travel | Sports | Fitness