Suraj G Bankar

Data Engineer

Aurangabad, Maharashtra
surajkbani@gmail.com
| +91 8999931749 |

PROFESSIONAL OBJECTIVE

To work in an organization where I can contribute and enhance my knowledge and growth.To enhance my working capabilities, professional skills, business efficiencies and to serve my organization in best possible way with sheer determination and commitment to realize the organizational objectives.

SKILLS

Python, SparkSQL, MySQL,Postgresql,Pyspark,Linux, Oracle, HDFS, MapReduce, YARN, Glue, Athena,Redshift, AWS EC2, S3, RDS, DMS, EMR cluster, Version control systems(Github), Cloud, Jira,

WORK HISTORY

Arrow bench solutions Pvt. Ltd, Bangalore - *Data Engineer*
Aug 2019 - PRESENT

- Around 3 years of experience in different domains in the IT industry and mainly in Hadoop Ecosystem Development and Spark API's and Cloud service (Amazon Web Services).
- Experience of designing and building analytical data warehouse including using ETL and other data tool
- In depth understanding of Spark Architecture including SparkCore, Spark-SQL, DataFrames .
- Experience in working on file compression techniques as well as various file formats like ORC, Parquet, AVRO
- Resolve JIRA tickets raised by Client
- Experience in basic Linux Scripting.
- Have a hands-on experience with Amazon web services mainly S3, EC2, IAM roles, and DMS.
- Perform tasks such as writing scripts, writing SQL queries, etc.
- Develop batch processing, integration solutions, and process structured and Unstructured Data.
- Strong problem solving & technical skills coupled with clear decision-making.

PROJECTS

2. DATA WAREHOUSING

Duration: Aug 2021 to Till Date

Designation: Associate Data Engineer

Roles & Responsibilities

- Design and develop ETL process in AWS glue to migrate data from external sources like S3, orc/parquet/text files into AWS redshift.
- By Using AWS Glue we performed transformations and actions.
- Data Extraction, aggregations and consolidation data with in AWS Glue using pyspark
- Used AWS Glue catalog with crawler to get the data from s3 to into redshift
- Written pyspark Queries for Data Analysis to meet the business requirements.
- Responsible for analyzing and cleansing raw data by performing Glue Jobs.
- Optimize code using PySpark for better performance.
- For processing, we used PySpark, and for storage AWS S3.
- Monitoring the job, if any problem occurs solving that problem.
- At the end upload the output data to the output directory in S3 to Redshift.
- Monitoring Job Daily, Fixing Production Bugs, and Raising Ticket for Production activity..

1. RETAIL STORE DATA PIPELINE

Duration: Jan 2020 to July 2021

Designation: Associate Data Engineer

Roles & Responsibilities

- Retail stores , serving the customers worldwide with a range of products like groceries, home decor, furniture, electronics.
- These stores generate the sales data at their POS (Point of sale) & get pushed into the big data pipeline for cleaning, processing and Business analysis.
- Basically, we have created an End-to-End data pipeline for Retail stores data using HDFS,Apache Pyspark,mySql And PowerBi.
- We have to filter data and perform multiple operations in order to get meaningful insights that will help the client in decision making and analysis.
- We have written logic in Pyspark using Pycharm IDEA.
- Have responsibility to Write the data extraction logic for the application.
- Have responsibility to write efficient code as per required and requested.

EDUCATION

Mgm's Jawaharlal Nehru Engineering, Aurangabad
B.E
Year 2012 to 2016

Government College of Arts and Science, Aurangabad
HSC in Science
Year 2010 to 2012

Godavari public school, Aurangabad
SSC year 2010

PERSONAL DETAILS

Date Of Birth: 3rd Jan 1995

Gender: Male

Languages Known: English, Hindi, Marathi

Nationality: Indian