

Balaji

PROFESSIONAL SUMMARY:

- Having 4.8 years of experience in analysis, design, development & maintenance of quality applications in an Big data process by using Hadoop Framework and spark technologies
- Worked on Hadoop, Hive scripts, Sqoop, SparkSQL, Pyspark and AIRFLOW for Scheduling Jobs.
- Expert in working with Hive data warehouse tool-creating tables, data distribution by implementing partitioning and bucketing, writing and optimizing the HiveQL queries.
- Experience in importing and exporting data using Apache Sqoop from Relational Data Base Systems to HDFS and vice-versa
- Hands on experience in setting up workflows using Apache Airflow for managing and scheduling Spark jobs
- Worked on reading multiple data formats on HDFS using SparkSQL
- Load the data into Spark RDD and do in memory data Computation to generate the Output response.
- Exploring with Spark improving the performance and optimization of the existing Data in Hadoop using Spark Context, Spark -SQL, Data Frame, Spark RDD's.
- Implemented Spark framework using SparkSQL for faster testing and processing of data.
- Hands on experience on SPARK. Creating the Data Frames handle in SPARK with Python.
- Experience with Agile Development model right from the requirement gathering to Deployment and production support.
- Involved in daily SCRUM meetings to discuss the development/progress and was active in making scrum meetings more productive.
- Adapt and adhere to industry standards while working with multi-cultural teams.
- Developing, managing & tracking Project plans with periodic reporting to senior management.

Tools: -

- HDFS, Hive, Sqoop, oozie, Pyspark, AWS EMR, AWS Athena, AWS S3, AWS Glue, IAM, CloudWatch

Experience:

- Currently working with MindTree, Bangalore as a Software Engineer from Mar 2018 to Dec 2022.

PROJECT-3

- Project: Lewis Stores
- Client: Costco
- Role: Hadoop Developer
- Duration: May 2021 – Dec 2022
- Tools: HDFS, HIVE, SQOOP, OOZIE, Pyspark, Aws

Description:

- Purpose of this project is to store huge amount of data and to perform Data Analysis on the Retail data available from a retail

store. One of the key things we focus at John Lewis is provide a unique and personalized customer experience when the user shops at our website or using our mobile application. This means understanding customer's likes and dislikes, shopping patterns. Data from these data points could be structured and semi-structured in few cases. The data set contains retail transaction of store and all these data is collected, aggregated and will be performing data analysis in order to help the retail chain which customers are purchasing the products and the buying patterns of the product and help the retail chain to improve its operations.

Responsibilities:

- Developing Spark Programs using Python API's and generated reports.
- Load the data into Spark RDD and do in memory data computations to generate the output response.
- Written the Apache HIVE scripts to process the HDFS data.
- Created Hive tables to store the processed results in a tabular format.
- Moved all log files generated by various network devices into HDFS location.
- Worked on developing the oozie scripts to automate the jobs.
- Worked on Apache falcon to trigger the jobs.
- Developed the SQOOP scripts in order to make the interaction between HIVE and MySQL Database
- Involved in the requirement analysis phase.

PROJECT-2

- Client: Production Sales Analyze
- Role : Data Engineer
- Duration : Jan 2020-APR 2021
- Tools: HDFS, Aws s3, Lambda, Glue, Hive, Athena, PySpark

Description

To analyze historical sales information and generate reports on monthly basis and compare sales between months, compare sales of different product categories, analyses growth and decline in sales, generating weekly, monthly and yearly report for the same. Doing forecast for future sales on also comparing the target achieved or not for weekly and yearly basis. When selling products in order to prevent products stock out due to the uncertainty events, enterprises usually reserves a certain amount of safety stock. The amount of safety stock is directly related to inventory costs and no. of sales. Determining the amount of future safety stock as per the previous sales report

Roles & Responsibilities:

- Involved in converting Hive queries into Spark transformations using Spark Transformations by using Python.
- Created Glue jobs and Executed, Fetched the results for data comparison.
- Worked on AWS Glue Data Catalog, Crawlers
- Importing data from multiple sources such as AWS S3 into Spark Data Frame
- Hands-on experience working with Amazon Web Services like S3, Athena, CloudWatch, IAM, Glue,
- Extracting features from data sets using Spark SQL
- Developed spark scripts by using Python as per requirements.
- Worked on writing Queries in Athena and compared data between hive and spark
- Worked on migration work from Hadoop to Spark and AWS environment.
- Worked on Spark performance tuning which implemented in spark code.
- Implemented **Spark** using **Spark SQL** for faster processing of Structured data
- Tested and reported defects in an **Agile** Methodology perspective.

PROJECT-1

- Project : Product information Management
- Client: HP

- Role : Hadoop Developer
- Duration : April 2018- DEC 2019
- Tools: EC1,S3,HDFS,Hive,Airflow

Description:

Improve the revenue and margin performance of your business by making better, faster merchandising decisions.
Analysing raw data, drawing conclusions & developing recommendations Writing SQL scripts to manipulate data for data loads and extracts.

Roles &Responsibilities:

- Written the Apache HIVE scripts to process the HDFS data.
- Created Hive tables to store the processed results in a tabular format
- Moved all log files generated by various network devices into HDFS location
- Moved all crawl data flat files generated from various systems to HDFS for further Processing.