Name - Prasanth M	**Company** - Accenture
Contact No: +91 8247084782	**Current Location -** Bangalore
Email Id: mprasanth021196@gmail.com	

Summary

- IT Professional with overall 4 years of Experience in systems development, databases & analytics, comprehensive experience as Data engineer.
- Expertise in Big Data, Hadoop, SQL, ETL and various frameworks in Hadoop such as Pig, Hive, Sqoop, Flume, Zoo keeper, Oozie, Kafka, HBase, HDFS, YARN, & HDP 2.x.
- Excellent knowledge of Hadoop architecture and on Hadoop Daemons such as Name node, Data Node and Expertise in Hadoop ecosystem such as HDFS, Resource Manager, Node Manager and Map Reduce programming Paradigm.
- Good understanding of RDD operations in Apache Spark i.e., Transformations &Actions, Persistence, Caching, Accumulators, Broadcast Variables.
- Strong experience and knowledge of real time data analytics using Spark Streaming, Flume and in configuring producers and consumers on Apache Kafka.
- Expert in working with Hive data warehouse tool creating Internal and External hive tables, data distribution by implementing partitioning and bucketing, writing and optimizing the HQL queries.
- Experience in importing and exporting data using Sqoop to HDFS from Relational Database Systems.
- Experience in designing both time driven and data driven automated workflows using Oozie.
- Created Hive Internal and External tables and loaded the data in to tables and query data using HQL.
- Load and transform large sets of structured, semi structured and unstructured data.
- Proficient in working with Databases like DB2, Oracle.
- Experience in Daily production support to monitor and trouble shoot Hadoop/ Hive Jobs.
- Have knowledge on Spark concepts like RDD's, Dataframe have worked on a POC by using the Azure Databricks (Spark), To load the data into Dataframe and performing the transformations on top of Dataframe and Saving the data into Azure SQL Table.
- Ability to work effectively and efficiently in a team and individually with excellent interpersonal, technical and communication skills.
- Ability to work in fast-moving, multi-disciplinary working teams and a committed team player with willingness to quickly adapt to new environments and technologies.

Key Domain and Technical Knowledge

- **Domain**: Travel & Hospitality
- **Technical**: Apache Hadoop, Hive, Sqoop, Kafka, Storm, flume, Pig, Scala, Python, Hortonworks Data Platform, MapReduce, HQL, SQL, Shell scripting.
- **Programming:** SQL, Scala, core java and Python basics
- **ETL Tools:** Informatica, PDI, Teradata
- **BI Tools:** Qlik, Cognos
- **Databases:** DB2, MySQL

Academic Qualification

- Bachelor of Technology in Computer Science and Engineering from GITAM University, Visakhapatnam.

COGNIZANT:

Work Experience

- Worked as a "**Programmer Analyst**" for Cognizant Technology Solutions Pvt Ltd, Bangalore.

Project 1:

Client: **Travelport**   **Period**: June 2018 – June 2020

Domain: Travel & Hospitality

Project: Travelport BI Consolidation

Description:

Travelport's vision is to transform their Analytics and Information Management delivery platform, lower total cost of ownership and enhance business value to their customer. They wanted to track all business intelligence development activities across the Organization.

Role and Responsibilities:

- Understanding business logic of all PDI applications.
- Loading source data from different sources like compressed files (.gz), Storm files into HDFS and hive persistent tables using Map-Reduce program in PDI tool.
- Scheduling PDI events using ESP Scheduler.
- Loaded data into Hive by using Sqoop and used Hive QL to analyze the partitioned and bucketed data, executed Hive queries on Parquet tables stored in Hive to perform data analysis to meet the business specification logic.
- Importing and exporting data using Sqoop framework.

- Created technical design documentations.
- Optimized MapReduce jobs to use HDFS efficiently by using various compression mechanisms
- Creating Internal & External Hive tables, loading the data and writing Hive queries which will run internally in MapReduce.
- Implemented Partitioning, Dynamic Partitions and Bucketing in Hive for efficient data access
- Developed Hive queries for creating foundation tables from stage data
- Involved in loading data from LINUX file system to HDFS.
- Created and scheduled events in ESP for running the PDI Jobs.
- Loaded data from legacy systems i.e., Galileo (1G), Apollo (1V) cores data into Worldspan (1P) and have also performed the archival process.
- Used orc tables as stage tables for faster processing.
- Checking Ambari portal to see the cluster's health, jobs status and logs.

Project 2:

Client: **Anthem, INC**.   **Period**: June 2020 – Current

Domain: Health Care

Project: Anthem

Description:

Anthem, Inc., is a provider of health insurance in the United States. It is the largest for-profit managed health care company in the Blue Cross Blue Shield Association. As of 2018, the company had approximately 40 million members. The company operates as Anthem Blue Cross in California, where it has about 800,000 customers and is the largest health insurer. It operates as Empire BlueCross BlueShield in New York State and as Anthem Blue Cross and Blue Shield in 10 states.

Role and Responsibilities:

- Experience in extracting appropriate features from data sets in order to handle bad, null, partial records using Spark SQL.
- Experienced in working with spark eco system using Spark SQL and Scala queries on different formats like Text file, CSV file.
- Responsible in performing map, flatMap, join, aggregations, filter, and other transformations on the datasets using Spark.
- Implemented Spark using scala and Spark SQL for faster testing and processing of data.
- Involved in working with CSV, PARQUET files in Spark.

Accenture:

- Working as a "**Application Development Analyst**" for Accenture Solutions Pvt Ltd, Bangalore.

Project 1:
Period: May 2021 to May 2022
Client: CALSAWS Data Analytics (Health and Public Services)
CalSAWS stands for **California Statewide Automated Welfare System**. The CalSAWS Project is dedicated to hosting excellent case management for eligibility workers to efficiently distribute benefits to participants. They will house all 58 California Counties and replace the existing C-IV, LRS, and CalWIN systems. The Project is working towards Data migration and Data analytics on top of raw dataset which is present in the source application.

Technologies : PySpark, AWS S3, EMR, Athena, Glue, Oracle database.

- Extracted ODI queries and analysed it.
- Converted SQL to PySpark code for different reports.
- Analysed user requirement involve in various meetings.
- Used window function to optimize the PySpark codes.
- Added additional logic to derive new columns as per the new business requirement.
- Used PySpark framework to process the data.
- Pushed processed S3 data to MySql RDS environment for business consumption.
- Used Spark API's to optimize the PySpark code.
- Worked on EMR cluster for data processing.
- Writing SQL queries for business for getting insight of data.
- Imported data into datalake using various spark submit commands.
- Used orchestration for the end to end run of the application.
- Performed basic Unit Testing.

Programming Skills: PySpark, SQL,Git.
Tools: Amazon Workspace, Pycharm, AWS S3, EMR, Athena, Glue, SQL Developer, Toad, ODI, Bitbucket, Git, Jira, MySql, ICEDQ.

Project 2:
Period: **May** 2022 – Present
Client: Seattle Police Department (SPD)
The Seattle Police Department is the principal law enforcement agency of the city of Seattle, Washington, United States, except for the campus of the University of Washington, which is under the responsibility of its own police department.

Technologies : Amazon Workspace, Pycharm, AWS S3, Athena, AWS Glue Bitbucket, CI/CD Pipeline, AWS CloudFormation

Programming Skills: PySpark, SQL, Git.

- We will establish a JDBC connection to the DAP 1.2 (source) using AWS Glue Connection. We will then run the corresponding Glue jobs to fetch the data from DAP 1.2 and store it in RAW bucket. It is truncate & load.
- We will run another series of Glue jobs in which we apply all the required transformations, filter the data and store it in curated bucket in parquet format in s3. Similarly, we apply additional transformations and store it in star bucket in parquet format in s3. Finally, In the reporting layer we will aggregate by bringing two or more star tables together and finally store it in reporting bucket in parquet format in s3.
- The crawlers of raw, curated, star and reporting s3 buckets are run to get the tables created in Athena. We can perform interactive queries on them to fetch the count, run data analytics and to reconcile data with existing environment.
- Athena connects to tableau via a JDBC driver once the connection is established, they access the tables present in Athena to perform data analysis and business intelligence using the data visualization tool and then the dashboards which are created will be shared to the business users.

Certifications / Professional Awards:

- Certified in cognizant Big Data certification.
- Certified in cognizant Apache Hadoop and Hadoop Admin certification.
- Certified in cognizant SQl and Teradata certification.
- Certified in Cognizant Informatica Certification.