SANTHOSH BADAM

Hyderabad | **+91 8125937810 | santhoshbadam04@gmail.com**

Objective:

To work on big data problems at scale and contribute to overall growth of the company through data analytics.

PROFESSIONAL SUMMARY:

- **Hadoop/ Spark developer** with **3+ years** of designing and implementing Bigdata pipelines using MapReduce, Hive, Sqoop, Spark, Oozie , Airflow and Zookeeper.
- Good knowledge of Hadoop architecture and various components such as HDFS, Resource Manager, Node Manager and Data node.
- Experience in importing and exporting data using **Sqoop** from relational databases to HDFS and vice versa.
- Experience in creating Dataframes from Cloud/ HDFS files and processing using Apache **Spark(Python/ PySpark).**
- Ability to create/ query hive queries from metastore using **SparkSql**.
- Experience in Job scheduling and monitoring tools like **Oozie.**
- Working knowledge of using HDFS CLI and its associated commands.
- Hands-on experience working on **NoSQL** databases including **HBase, MongoDB** and its integration with **Hadoop cluster.**
- Querying experience in Relational databases - **MYSQL** and **PostGres.**.
- Part of the teams who followed **Software Development Life Cycle (SDLC)** using **Waterfall** and **Agile**, **Scrum** models.

Experience:

Organization**: VCM Technologies Pvt Ltd, Hyderabad**

Duration: **April 2019 - Present**

PROJECTS:

Big Data Engineer (June 2021 - Present)
Client: Yupp TV

- Ingest viewership data from S3 and perform various transformations on the data using PySpark.
- Create hive tables on top of processed data(insights).
- Come-up with analytical queries to provide insights on viewership data.
- Provide on-call support for production issues and triage them until closure.
- Use the JIRA tool for tracking issues.

Environment: Hadoop, Spark, Hive, Python, AWS, S3 and EMR

Software Developer (July 2020 - May 2021)
Client: Technology Credit Union

- Worked on building scalable big data pipelines using the Hadoop ecosystem of tools.
- Involved in the loading of structured and unstructured data into HDFS from multiple sources.
- Loaded data from MySQL, a relational database to HDFS on a regular basis using Sqoop Import/Export.
- Worked on creating hive tables for querying for downstream teams.
- Responsible for implementing data pipelines using PySpark.
- Used Spark-SQL to Load data into Hive tables and Written queries to fetch data from these tables.
- Worked on loading CSV/TXT/AVRO/PARQUET files using Python language in Spark Framework and processed the data by creating Spark Data frame and RDD and saving the file in parquet format in HDFS to load into the fact table using ORC Reader.
- Understanding the Business Requirement in detail, providing the solutions.
- Good experience in automating checks using Python Scripts.

Environment: Python, Spark, Spark-Sql, Hive, Sqoop, MySQL, Jenkins

---

Data Analyst   (June 2019 - May 2020)
Client: Technology Credit Union

- Worked on generating business reports in MySQL and providing them to multiple stakeholders.
- Hands-on experience using visualization tools - excel.
- Gather business requirements from multiple teams and refine them upon feasibility study.
- Provide daily, weekly and monthly reports on customers to alert the business stakeholders.
- Test and validate backend data manually and report if there are any issues.
- Automate manual checks using python, shell scripting.
- Used the JIRA tool for tracking all issues.

Environment: MySQL, Python, Linux, Excel and JIRA

COMPETENCIES:

Programming/Scripting Languages: Python, Java
Frameworks: Hadoop(HDFS, Map Reduce), Spark, Pyspark, Hive
Relational Databases:MYSQL, PostgreSQL
Data Ingestion: Sqoop
Cloud Technologies: AWS, Databricks

EDUCATION:

Masters, Structural Engineering (2014-16)
 Sri Chaitanya Technical Campus
Bachelors, Civil Engineering (2009-13)
 Kakatiya Institute of Technology & Sciences