**SANTOSH KUMAR THAMMINENI**

**Mob:- +91 8125733409**

**Email:- contactsantoshkumar9@gmail.com**

*Professional Summary*

- 4 + Years' experience in Big Data Development and Ecosystem like Map Reduce, HDFS, Hive, Pig, Hbase, Zookeeper, Sqoop, Spark Frameworks
- Experience in Spark (Spark/Scala)
- Experience in coding in Scala, Unix, Python, SQL
- Experience in building end to end Ingestion Framework using **Python/Scala** for ingesting data from OLTP systems to Hadoop Datalake which takes care of metadata captivity, change data capture, Email Alerts, Job failure and recovery.
- Strong experience in troubleshooting spark applications and various performance considerations to take for efficient memory handling of long running spark applications
- Experience in using Hive Query Language for Data Analytics
- Experience in working with Apache Sqoop to import and export data to and from HDFS and Hive
- Experience in developing the workflows in oozie.
- Proficient in designing and querying the NoSQL databases like HBase.
- Working experience on integrating different eco-systems like HBase - Hive and Hbase -PIG.
- Excellent customer management/resolution, problem solving and debugging skills.

**Education Profile**

- M. Tech from Hasvita Institute of Engineering and Technology, Hyderabad in 2014

- B. Tech From K.S Raju Institute of Technology and Sciences, Hyderabad in 2012

**Job Profile**

- Working as a Big Data Developer/Data Engineer from May 2017 – Till Now
- Worked as a Technical Writer from May 2015 – April 2017

*Skills Profile*

| Operating System | Windows, Linux |
|---|---|
| Database | MYSQL, Oracle, SQL Server |
| Hadoop Technologies | HDFS, Hive, Sqoop, Hbase, Spark(python/Scala), Hortonworks |
| Languages | Scala, Unix, Python, SQL |
| IDE | Eclipse, Spyder, IntelliJ, PyCharm |
| Version Control | GIT |
| Cloud Technologies | AWS, Azure |

## PROJECT 1

Company      Ernst & Young
Client          Private Bank
Role            Consultant/Data Engineer
Environment   HDFS, Hive, Spark Core, Spark SQL, Scala, Python, Airflow
OS              Windows
Team            6
Duration        April 2021 – Till Now
Location        Hyderabad

### Description:

Worked as a Big Data Engineer in the Risk Management Team where the Bank wanted to store, process & manage the huge amount of data in day-to-day operations collected from various sources. The system majorly checks the credibility of the customer & looks for the credit risks.

### Responsibilities:

- Worked on Feature planning, architectural designing for every PI planning.
- Developing applications with streams and lambda expressions through functional programming.
- Optimizing of existing internal algorithms in Hadoop using Spark Context, Spark-SQL, DataFrames and Datasets
- Involving in writing complex queries and optimizing hive and presto to process the data to achieve faster indexing and made the system more scalable. Taking ownership for the model and pre-processing module of the project and data ingestion pipelines.
- Involved in creating partitioning and bucketing of HIVE tables.
- Designing project structure and migrating to GitHub along with Bitbucket which enables effective maintainability of source code Brining best practices of Agile methodology covering all areas for SDLC including the iterative and incremental development lifecycle.
- Automating package build and deployment using Jenkins and shell scripts and automating CI/CD pipeline which makes development team life easy. Mentoring new joiners in the team and making them productive.
- Effective usage of CICD tools like Jenkins, Git, etc
- Participating in design discussions, suggesting cutting edge design and implementing them.
- Developing, Implementing, and maintaining applications.

## PROJECT 2

Company      Flooredge Solutions Private Limited
Client          Tech Mahindra
Role            Big Data Developer
Environment   Hadoop, Spark SQL, Hive, Shell Scripting, Python, Crontab, Spyder, Hortonworks
OS              Windows
Team            6
Duration        Mar 2019 – March 2021
Location        Bangalore

**Description:**

American corporation that supplies equipment, services and software to enable the manufacture of semiconductor (integrated circuit) chips for electronics. Aim of the project is to create an Enterprise Data lake for Operational Data from heterogeneous sources viz. Oracle, SAP, SQL Server, SharePoint, Custom files to a common Hadoop Platform

**Responsibilities:**

- Part of Building End to End Ingestion Framework using Python, Sqoop, Hive and Impala for ingesting data from Oracle to Hadoop which takes care of appropriate typecasting of datatypes data types through a data dictionary, change data capture, Email Alerts, Job failure and recovery. MySQL is used for capturing the running metadata.
- Implemented framework to capture data from csv files to HDFS
- Worked on creating SQL for Transformation of data

**PROJECT 3**

| | |
|---|---|
| Company | Flooredge Solutions Private Limited |
| Client | Tech Mahindra |
| Role | Big Data Developer |
| Environment | Hadoop, Spark, Hive, Spark, Sqoop, Teradata, Oracle, SQL, Shell scripting |
| OS | Windows |
| Team | 5 |
| Duration | Aug 2017 – Mar 2019 |
| Location | Bangalore |

**Description:**

A Multinational Ecommerce company want to maintain the transactions for the analytical purpose. The aim of this project is to process both batch and real time data where multiple source of channels has been involved. This project involves working in multiple tracks such as migrating the data from Oracle to Hadoop,

**Responsibilities:**

- Worked on transformation using spark SQL
- Creating tables in different stages such as Core, Stage, Semantic which involves loading data from source to stage and stage to core and core to semantic
- Created ETL process using Apple Framework
- Applied transformations in Hive tables
- Automated the Jobs using crontab
- Performed the Data load testing Using Shell Scripts