Resume

Name: Santhosh B K   **E-Mail: iamsanthoshbk@gmail.com**

Designation: Data Engineer   **Mobile: +91 9741732696**

Objective

To deliver my expertise with the client business to enhance the Operating System to incorporate the IT requirement with business process to achieve an optimal environment. A good learner with upcoming needs and a good team player.

Career Snap Shot

- Overall 4+ years of experience in IT and relevant into Bigdata 3.2 years with strong technical skills in My SQL, Python, Hadoop, Spark, PySpark (using Big Data & Hadoop development using HDFS, Hive, MapReduce, Sqoop, HBase etc.)
- Basic Knowledge in AWS, Lambda, Airflow, EMR, EC2, Databricks. Including in designing, developing applications using Hadoop Ecosystem components.
- Self-Motivated, able to work independently and in teams, handle multiple projects paying attention to details, accomplished problem solving, analytical Skills, Performing Task during the Project within decided Time Frame

Educational Qualification

- **B.E in Civil Engineering** from Dr.S.M.C.E in 2016.

Professional Experience

- Currently working as a Data engineer for JBG INFOSOLUTIONS PVT. LTD from 06/07/2019 to continuing.

KEY SKILLS

- **Programming Language**: Python, Scala
- **Operating System / ERP Version**:- Windows 10, Ubuntu (Linux), MAC
- **Tools / DB /Packages/ Framework /ERP Components**:- Spark, PySpark, Python, Hive, HDFS, MapReduce, MySQL, AWS.
- **Functional:** Good communication & Inter-personal skills with multi-tasking talents and comfortable working in a team environment with tight schedules and meet deadlines

Key Responsibilities

- ➢ Experience in Hadoop ecosystem & Big-Data Analytics.
- ➢ Experienced in Spark ecosystem components like Spark batch, Spark streaming, SparkSQL, and Spark-core.
- ➢ Experienced in using HDFS, Map Reduce, Hive, Airflow, HBase and Cloud era Manager.
- ➢ Good knowledge of Hadoop Architecture and Yarn Architecture.
- ➢ Good knowledge and Understanding of Spark-standalone-Architecture and Spark on-Yarn Architecture.

- ➢ Experienced in working Hive data warehouse system, developing Data pipeline, implementing complex business logic and optimizing business logic.
- ➢ Basic Knowledge and Understanding AWS (S3, Athena, EC2, EMR).
- ➢ Experienced Hadoop Developer, have a strong background with file distribution systems in a big- data. Understands the complex processing needs of big data and have experience in developing codes and modules to address those needs.
- ➢ Knowledge of Map Reduce programs to perform analysis and MR unit jobs for testing.
- ➢ Strong Programming skills Python, Object-oriented programming.
- ➢ Excellent communication, organizational skills, attention to detail & the ability to meet tight deadlines.

Project Details

Project :-2

Client	Lloyds Bank
Role	Data Engineer
Duration	November 2020 to till Now

Client Information:

With overall millions of customers like RRTS (ora-std) MTG, PCL, CRD, customers and a finance, data optima. Analytics is a critical component of Lloyds Bank business. Bank collects huge amount of usage component of Lloyds Bank business. Lloyds Bank collects huge amount of usage information about all their products, services and sites such as business data, Financial Data, product and device specific information and other similar data and that will be consumed for Business Analytics and build Big Data Solutions for various Lloyds Bank Business units. Compliance, business and marketing to improve the services, measure. The collected data is aggregated and analyzed for adhere legal and analyze the usage and to help to find services and advertising more relevant to the customers.

Technology & Tools used:

Python, MySQL, Hadoop, Hive, Spark, PySpark, Yarn, PyCharm, AWS (EMR, S3, IAM, data bricks), Windows, Linux.

Roles and Responsibilities:

- Data is pulling from Oracle using Sqoop to hive tables.
- Reading data from source (hive).
- Understand business story and analyze functional specifications &design feasible solutions.
- Using PySpark for processing and Applying Analytics on data per client Requirement.
- Developed code to write the logic data into Hive External Tables.
- Involved in setting up the properties for Apache Airflow.

- Involved in configuring and creating Autosys schedule.
- Involved in various log files to fix the issues during unit testing and performance tuning & involved in developing the PySpark.

Project:-1

Client	MTN
Role	Data Engineer
Duration	July 2019 to November 2020

Client Information:

Millions and Billions of users under the MTN, they have products like Telecom, the data analytics is a critical component of MTN-Business. It collects huge amount of usage information about all their products, services and sites such as Complete Calls and Call Records, product and device specific information and other similar data and that will be consumed for Business Analytics and build Big Data Solutions for various Toronto-Dominion Business units. The collected data is aggregated and analyzed for adhere legal compliance, business and marketing to improve the services, measure and analyze the usage and to help to find services and advertising more relevant to the customers.

Technology & Tools used:

Python, MySQL, Hadoop, Hive, Spark, PySpark, Yarn, PyCharm, Windows, Linux.

Role and Responsibility:

- Injection data from Oracle to Bigdata layer at Hive partition tables.
- Loading data from Hive tables to Python Spark for Processing Data.
- Understand business story and analyze functional specifications & design feasible solutions.
- Based on business logic at applying cleaning the data, applying transformations and analytics with Python spark.
- Developed and implementing of code to write the logic data in to Hive External Tables.
- Involved in various log files to fix the issues during unit testing and performance tuning.
- Involved in developing the python spark.