**Shilpa Bansi Sable**

**9307180401**

[**sableshilpa1515@gmail.com**](mailto:sableshilpa1515@gmail.com)

**http://linkedin.com/in/shilpa-sable-919b2b143**



**CAREER OBJECTIVE**

**Objective:** My goal as a data Engineer is to keep growing in dynamic companies. I want to be in a role where I can use my technical skills, leadership qualities. I am excited to work with other who share my passion for solving problemsandmaking the most out of data analysis and business intelligence opportunities.



**CAREER SUMMARY**

* Overall 3.10 years of extensive experience in different domains in the IT industry as a Big Data Engineer in Hadoop Ecosystem Development and Pyspark-API's.
* Hands-on experience in Big Data Core components and ecosystems including Data Ingestion, Data Processing and Data Cleaning.
* Proficient in developing data extraction logic using Pyspark.
* Have good experience in writing SQL.
* Experience in building optimized big data pipeline using HDFS, Hive, Spark on cloud premises.
* Proficient in handling large a mount of Structured and semi-structured data within targeting data from multiple sources.
* Batch processing, integrating solutions and process structured and semi-structuredData.
* Well-versed in Spark Architecture, specializing in Spark-Core and Spark-SQL.
* Skilled in transferring data from RDBMS to HDFS and Hive tables using JDBC & ODBCDriverPyspark-API's.
* Experience in using Hadoop distribution like Cloudera.
* Experience in handling large tables with optimization techniques like partitioning, bucketing, and loading in Hive.
* Experience in designing Internal and External tables in Hive.
* Acquire and manipulate raw data and followed by performing analysis of processed data.
* Basic nowledgeofAmazonWebServices(AWS)likeS3,EC2,EMR,lambda,Glue,Crawler,Athena.
* Innovative Data Engineer with a passion to use data for business growth and improve decision
* **Platforms:** Windows, Linux



**TECHNOLOGY PROFICIENCY**

* **Languages:** Python, SQL
* **Big Data Ecosystem:** PySpark, Hive, Sqoop Hadoop (YARN, HDFS) Kafka
* **Development Tools:** PyCharm
* **Cloud Technologies:** AWS (S3, IAM, EC2, EMR, Redshift, Glue, Athena)
* **Relational data base:** MySQL
* **Source/Target Code Management:** Git, Git-Hub
* **Management Tool**: Jira.
* **Industry Group:** Healthcare , E-Commerce



**PROJECT**

**Project name: Capital Data Service**

**Project-2**

Description :- I worked on a project called Capital Data services in the finance domain. CDC is a

comprehensive data architecture that involves data ingestion, transformation, business computations,

and reporting. The project focuses on ingesting data from various sources, performing specific

transformations, and storing it in Hadoop structures like Hive. This enables the business team to analyze

financial transactions, assess risks, ensure compliance, and make data-driven decisions.

**ROLES AND RESPONSIBILITIES**

• Involved in import data from various RDBMS into HDFS using JDBC & ODBC drivers.

• Perform Validations on Imported Data, and merge it with Existing Data.

• Designed both Managed and External Hive Tables and Defined static and dynamic partition to optimize

performance.

• I Worked with various File Formats like Text File, Sequence File, ORC Files, Avro Files. Written Hive

Queries to analyze data for our business need.

• Write Hive DDL to Create Hive Table for Optimize Query Performance.

• Create Spark jobs for data cleaning, processing.

**Vistara Informatics Pvt. Ltd., Bangalore**



**ORGANIZATIONAL SCAN**

**May 2023 - Present**

**Data Engineer**

**Technologies Used:**

**Sqoop, HDFS, Hive, Pyspark, Hive, Oracle Database, MySQL, SFTP,**

**Health care System Data Analysis Domain:** Insurance

**Duration:** May 2023 Till time

**Roles & Responsibilities: Data Engineer – Trainee (Data Ingestion Framework Role)**

* In this framework, we fetch the data from different source and ingest the data into data lake
* For this we are directly connecting to their source system and we ingest to data into our Hive table. For some of the sources, we are using file formats for ingestion.
* For testing, we are taking some sample data and we ingest that data into our Hive table with Data Migration application.
* Then we load historical data batch wise into the cluster. Environment setup and Spark Session Creation.
* Data Extraction, Data Transformation, Handling Missing Value Data Type Conversion, Advanced
* Data Manipulation, Data Loading, Performance Tuning

**Augustus Healthcare India LLP, Pune**

**March 2021 – March 2023**

**Data Engineer**

Technology used my SQL server

**Roles & Responsibilities : Data Engineer**

* Import Data into the Hive from various Relational Databases (MySQL Server) into HDFS.
* Write Hive DDL and DML queries in Hive SQL.
* Perform optimization techniques to enhance performance Worked with various File Formats like Text File, Column oriented file various Compression formats like Snappy.
* Designed both Managed and External Hive Tables and Defined static and dynamic partitions as per requirement for optimized performance on production datasets.
* Join operations of large tables using optimization techniques in SQL.



**EDUCATIONAL QUALIFICATION**

**Education:**

* Bachelor of Pharmacy from Dr. Vedprakash Patil Pharmacy College, B.A.M.U -2020.
* Master of Business Administration in Information Technology Management from Jai Bhawani Institute of Management, B.A.M.U -2023.



**PERSONAL DETAILS**

* **Language:** English, Hindi ,Marathi
* **Address:** F 5/10, N-7, Ayodhyanagar, Cidco, Chhatrapati Sambhajinagar (Aurangabad) Maharashtra.-431003.

I hereby declare that the particulars of information and facts stated herein above are true, correct and complete to the best of my knowledge and belief.

**Place: (Shilpa Bansi Sable)**