

Avinash Sable

Email: avinashsable43@gmail.com

Mobile: 9545493644

• **PROFESSIONAL SUMMARY:**

6.2 Years of overall experience.

- Around 5 years of experience in **Data Engineering and ETL** Projects
- Developed pyspark scripts and create the data frames and performed transformations on it.
- Involved in working on SparkSQL code as an alternative approach for Faster Data processing and better performance.
- Experience in Data Ingestion ,Transformation and performance tuning.
- Transferring data from RDBMS to HDFS and Hive tables by making use of pyspark.
- Injected and processed large amount of data from various structured and semi structured sources into HDFS(AWS cloud)
- Hands on experience on Amazon web services mainly S3,EC2,EMR,RDS,Redshift,Glue,Athena.
- Worked on QA support activities, test data creation and Unit testing activities.
- Gather and process raw data at scale and analyze processed data
- Actively participated in daily standup, biweekly scrum and project meetings
- Skills: Spark, PySpark, EC2, AWS Glue,Python,EC2,HDFS,SparkSQL,ETL

EDUCATIONAL QUALIFICATION:

- BE (CSE) Jawaharlal Nehru Engineering College, Aurangabad, Maharashtra **(2011 to 2015)**

PROFESSIONAL EXPERIENCE:

- Worked as a Senior Software Engineer for **Infosys Ltd** from **Oct 2021 to May 2023.(C2H)**
- Worked as a System Engineer for **MAN Energy Solutions (Volkswagen Group) (Permanent)** from **Sep 2019 to Aug 2021.**
- Worked for **Wipro Ltd.** from July 2018 to Aug 2019 as an ETL Analyst.**(C2H)**
- Worked at **Lokmat Media Pvt. Ltd** as Software Trainee. **April 2017 to April 2018.**

TECHNICAL SKILLS:

Operating System	Windows, UNIX
Database	SQLServer, Oracle.
Hadoop Ecosystem	HDFS,Hive
Apache Spark	Spark-Core, Data frames, SparkSQL, Pyspark
Programming Language	Python
Other Skills	Pycharm ,Jupyter Notebook,Informatica

Cloud Technologies	AWS(RDS,EMR,EC2,S3,Redshift,Glue,Athena)
Orchestration Tool	Airflow
Other Tool	JIRA, Git, GitHub

Assignments

- The details of the various assignments that I have handled are listed here in descending order

Company: **Infosys Ltd (Oct 2021 to May 2023)**

Domain: **Banking Domain**

PROJECT DETAILS

Project#1 (March 2022 to May 2023)

Title: Data Migration

Technology: DMS, Pyspark, Docker, Python, AWS (S3, Athena, Glue, Ec2)

Role: Data Engineer

Description: Data migration with various formats from raw s3 bucket to curated s3 bucket transformation of data according to ETL specifications and business logic.

Responsibilities:

- Experience in writing data extraction pyspark scripts with business logic.
- Experience in building optimized data pipeline using s3, glue, pyspark, Athena and redshift on cloud premises.
- Developed batch processing, integrate solutions and process structured and unstructured data.
- Having experience in spark architecture including spark core and spark SQL.
- Experience in data ingestion, transformation, and performance tuning.
- Responsible for applying data cleaning techniques and remove inaccurate and corrupted data to improve data quality which significantly improves data quality.
- Worked with parquet file formats and used various compression techniques to leverage the storage.
- Filtered data and performed multiple operations in order to get meaningful insights that help our client in decision making and analysis.
- Created pyspark scripts for validating the data which is migrated.
- Understanding the upstream source nature and work with business cases.

Project#2 (Oct 2021 to Feb 2022)

Project Title: Financial Intelligence Unit (FIU)

Role: Data Engineer

Environment: Hbase, HDFS, Spark, AWS S3,SparkSQL,Python

Project Details: Financial intelligence unit is the collection of information about the financial affairs of entities of interest, to understand the nature and capabilities, and predict their intentions. Generally, the term applies in the context of law enforcement and related activities.

Roles and Responsibilities

- Understand the nature and characteristics of data and check how data flow
- Involved in import data from various RDBMS, HBase, HDFS using spark which populates Hive Table.
- Designed Both Managed and External Table
- Perform data injection into HDFS and helped in data movement in multiple databases layer
- Analyze data by running hive queries
- Responsible for analyzing and cleansing raw data by performing hive queries.
- Optimize code using pyspark for better performance.
- Ingest flat files like Delimited.
- We used for processing as pyspark, spark SQL and for storage AWS S3.

Company: MAN Energy Solutions Pvt. Ltd. (Sep 2019 to Aug 2021)

Domain: Production/Purchase

Technologies & Tools: Informatica, JIRA, SQL Server , Agile.

Project#3

Title: Corporate Actions Gateway

Environment: PySpark, Spark, AWS(S3,EC2), SparkSQL, Python,Agile.

Role: Data Engineer

Description: Scope of this project was to enrich announced corporate events with security details and make available the event details to end users.

Responsibilities:

Roles and Responsibilities

- Understand the nature and characteristics of data and check how data flow
- Involved in import data from various RDBMS, HBase, HDFS using spark which populates Hive Table.
- Designed Both Managed and External Table

- Perform data injection into HDFS and helped in data movement in multiple databases layer
- Analyze data by running hive queries
- Responsible for analyzing and cleansing raw data by performing hive queries.
- Optimize code using pyspark for better performance.
- Ingest flat files like Delimited.
- We used for processing as pyspark, spark SQL and for storage AWS S3.

Project#4

Project Title: Data Mart Creation(Aug 2019 to Dec 2019)

Role: ETL Analyst

Project Details: Created data mart for different department inhouse as per ETL specification and business logic.

Roles and Responsibilities

Responsibilities:

- Developed Test Plans, Test Cases, and Test Scripts for in QA.
- Managed and conducted System testing, Integration testing and Functional testing and generated defect summary reports
- Prepared status summary reports with details of executed, passed and failed test cases
- Participated in regular project status meetings related to testing
- Tracked the defects using JIRA

Company: Wipro Ltd. (July 2018 to August 2019)

Domain: FINANCE

Technologies & Tools: Oracle, SQL Server, HP ALM 12

Project#5

Title: Fiduciary

Environment : SIT (Appian, Oracle, Jira, Agile)

Role: Description: Scope of this project was to extract the data from different sources and arrange them in data ingestion for data monitoring and report generation. (Appian)

Responsibilities:

- Developed Test Plans, Test Cases, and Test Scripts for SIT and QA.
- Managed and conducted System testing, Integration testing and Functional testing and generated defect summary reports
- Prepared status summary reports with details of executed, passed and failed test cases
- Participated in regular project status meetings related to testing
- Tracked the defects using JIR

Project#6

Project Name: ETL GFB Migration (Application Upgradation Testing)

Environment: Informatica, Oracle12c, Microsoft SQL Server.

Role: ETL Tester

Description:

The main objective of the project is to migrate the applications from current Legacy environment (Informatica 9.5.1) to GFB environment (Informatica 9.6.1). It includes fixing the standards according to GFB Environment and making it compatible with it. Data was extracted from flat files, tables as Source, transformed and loaded to Staging Area and Staging Area to Data Warehouse.

Responsibilities:

- Developed Test Plans, Test Cases, and Test Scripts for SIT and QA
- Tracked the defects using HP ALM 12 tool and generated defect summary reports
- Prepared status summary reports with details of executed, passed and failed test cases
- Participated in regular project status meetings related to testing
- Interacted with developers, Business & Management Teams and End Users

Company: Lokmat Media Pvt Ltd. (April 2017 to April 2018)

Domain: MEDIA

Technologies & Tools: Putty, FTP, Page Creator.

Project#7

Environment: Windows, Page Creator, Putty, FTP

Description:

Tested the different applications involved in print media page creations.

Tested individual end application for different role-based UI.

Responsibilities:

- Developed Test Plans, Test Cases, and Test Scripts for SIT and QA
- Prepared status summary reports with details of executed, passed and failed test cases
- Participated in regular project status meetings related to testing
- Interacted with developers, Business & Management Teams and End Users

DECLARATION:

I hereby declared that above-mentioned information is correct up to my best knowledge and I bear responsibility for the correctness of the above-mentioned particulars.

Location: Pune

(Avinash Sable)