

# Ghanshyam Pimpalshende

## Associate Consultant

✉ ghanshyam.ap97@gmail.com ☎ 9158571227 📍 Pune, India

### PROFILE

---

Data Engineer with **4+** years of hands-on experience in **designing and developing ETL** data pipelines handling Big Data loads along with Data Warehousing, Data Modeling and developing critical applications using **AWS Cloud** services, **Pyspark, SQL, Python, Databricks**, Hive/Hadoop technology/language and Airflow orchestration.

Worked with top rated assets management vendors to develop optimized ETL data management pipelines configuring best investment assets under management and savings/retirement plans.

### PROFESSIONAL EXPERIENCE

---

#### Associate Consultant

KPMG International Ltd

11/2022 – present

**Project** (Phase 1 and Phase 2) : AWS Data Engineering and warehousing for US based Assets Management Firm.

**Description:** Developed centralized data warehouse with optimized data pipeline to achieve fully automated process to get raw data from source and ingest data in target tables within AWS Cloud using SQL, Python and Pyspark.

#### Responsibilities( Phase 2 of Project) :

- Build and developed ETL data pipelines using Pyspark, Python and redshift SQL to perform ingestion, cleansing, transformation and loading utilizing Pycharm and Jupyter IDEs.
- Applied SCD Type 1, Type 2, CDC approach to detect changes in a source database and deliver the changed data to a target database.
- Used multiple AWS services like S3, Glue, Redshift, Lambda to perform the operations.

#### Responsibilities( Phase 1 of Project) :

- Understanding the business processes and translating them into business requirement documents based on functional design specifications.
- Developing PySpark and Spark SQL code strategies on AWS Databricks clusters.
- Designing both Dataframes and views within Databricks notebooks following data profiling on raw files.
- Orchestrating Databricks jobs and workflows using AWS Glue and Airflow DAGs.
- Writing the final processed data to Amazon S3, making it readily available for downstream like PowerBI for data visualization.

#### SDET Engineer

Infosys Ltd

02/2021 – 09/2022

**Project** : Financial Intelligence Unit (FIU360)

**Description:** Enterprise data warehouse was set up and designed to provide support for various use cases including analytics, processing, storing, and reporting of voluminous changing data using SQL, S3, Glue, Pyspark, Python (OOPS).

#### Responsibilities:

- Understand the nature and characteristics of data and check data flow.
- Developed pipelines for importing data from various RDBMS, HBase, HDFS using spark which populates Hive Table.
- Perform data injection into HDFS and helped in data movement in multiple databases layer.
- Responsible for analyzing and cleansing raw data by performing hive queries.
- Designed Both Managed and External Table.
- Optimize code using pyspark for better performance within Glue jobs.
- At the end uploading the output data to output directory on S3.

**Project:** Legacy Data Migration

**Description:** Source( AWS S3 Data Lake) to Hadoop migration project. Data migration utility to migrate existing system into Hadoop (Big Data) from source system and perform aggregation operations within Hadoop Big Data framework using Pyspark, SQL and Python.

**Responsibilities :**

- Analyze and understand data migration requirements and business needs.
- Design and develop data models to extract information from various data sources.
- Efficiently load and transform data into Spark DataFrames for further processing.
- Perform comprehensive data cleaning and validation, ensuring accuracy and consistency before saving to an Amazon S3 bucket.
- Implement data pipelines to load cleaned and processed data from Amazon S3 to Amazon Redshift for further analysis and reporting.
- Monitor daily ETL (Extract, Transform, Load) jobs to ensure smooth operation and timely data availability.

## SKILLS

---

**Language & Technology :-**

Python, PySpark, SQL, Databricks, ETL, Spark, Hive/Hadoop

**Cloud Technology :-**

AMAZON WEB SERVICES ( Glue, Redshift, S3, Athena, Lambda, API Gateway )

**OS**

Linux, Windows

**Database**

Oracle, MySQL, MongoDB , IBM DB2

## EDUCATION

---

**Bachelor Of Engineering(B.E.)**

Government College Of Engineering, Aurangabad  
Passed with Distinction

2016 – 2020

## PUBLICATIONS

---

**Self-penned Sports and Political articles:**

The Hitavada Newspaper ( Weekly Twinkle Star Magazine)

Penned some intuitive articles on topics ranging from Sports and Politics in the 'Creative Youth' section of Magazine .

## AWARDS

---

**Narotam Sekhsaria Silver Jubilee Scholarships :**

Narotam Sekhsaria Foundation ( Ambuja Cements Group)

Got awarded with the scholarship for being CBSE District Topper(2016 Batch) and scoring 100/100 in Computer Science subject.