

YOGESH TEMKAR

Big Data Engineer
Aurangabad, India
yogesh.temkar333@gmail.com
09545111211

PROFILE

Experienced Sr Data Engineer skilled in optimizing data pipelines, automating workflows, and enhancing data infrastructure for informed decision-making. Proven ability to identify process improvements and ensure data security. Committed to delivering innovative data solutions.

SKILLS

Big data - PySpark, SQL, Hadoop, HDFS, Hive,
Kafka, Airflow, ETL,

Database - MySQL, SQL Server, MongoDB

Operating System - Windows, Linux and MacOS

Cloud Technologies - AWS (ECS, RDS, EMR, S3,
Athena, Glue, Lambda)

Programming Languages - Python, SCALA and R

Ticketing Tool - JIRA

EMPLOYMENT HISTORY

Sr. Data Engineer, Rishabh Software Private Limited (Wolters Kluwer), Pune

MARCH 2023 – PRESENT

Project Name: Basecone

Role: Sr. Data Engineer

Tech Stack - Python, PySpark, SQL, HDFS, Hive, Airflow, AWS services (Glue, S3, , Athena, IAM, Redshift)

The purpose of the project was to have a cloud-based invoice processing and document management in realtime data migration. It helped organizations manage and process invoices more efficiently, reduce manual data entry, and enhance overall financial workflow automation.

Client works on cloud based inventory processing product named Basecone throughout Europe.

- Basecone is a cloud based invoice processing and document managing, Which automated invoice capture, digital approval workflows, and integration with accounting and financial systems.
- Also made end to end data pipelines and orchestrated them with Apache Airflow, improving data flow efficiency and enabling real-time data processing for better decision-making.
- Conducted a thorough review of existing data pipelines, identifying bottlenecks and performance issues. Subsequently, optimized critical pipelines, resulting in a significant reduction in data processing time and resource utilization.
- Participated in a data migration project to transition from a legacy data warehousing solution to a modern, scalable data platform using Trino, ensuring minimal downtime, data consistency, and improved query performance for analytical workloads.
- Experience in dealing with various file formats like CSV, Parquet, ORC, Avro.
- Possesses a strong understanding of Agile methodologies and has extensive experience working within Agile development environments, demonstrating expertise in Software Development Life Cycle (SDLC) processes.

Data Engineer - Specialist, LTIMindtree, Pune

JANUARY 2021 – FEBRUARY 2023

Project Name: Privacy Durable

Role: Data Engineer

Tech Stack - PySpark, SQL, HDFS, Hive, Airflow, AWS services (Glue, S3, IAM, Redshift, EC2)

The purpose of the project is to monitor flow of data project in Realtime. Data migration utility to migrate on server data to data warehouse. And further that data is utilized in different analysis and report generation.

Client works for renowned hotel chain which works throughout globe.

- Privacy durable was a project which mainly focused on CRM data of a Hotel chain to have better offers and customised services.
- CRM data is consists of all the data related to location, travel history and category based on expenses done in the past.
- The generated data at their end, later pushed into the big data pipeline for cleaning, processing and Business analysis.
- The data flows in streaming manner using Kafka and NiFi.
- Have responsibility to Write the data extraction logic for the application.

Data Engineer, Crest Informatics Private Limited, Hyderabad

DECEMBER 2017 – DECEMBER 2020

Project Name: Data Flagship

Role: Data Engineer

Tech Stack - PySpark, SQL, HDFS, Hive, Airflow, AWS services (Glue, S3, IAM, Redshift, EC2)

The purpose of the project is to do the ingestion of the AIS data that is present in mainframe system to Hive table. It's Mainframe to HDFS ingestion project in Realtime. Data migration utility to migrate on server data to data warehouse. And further that data is utilized in different analysis.

Client works for Korean Government's smart city executions in different cities of South Korea.

- Data Flagship is project which mainly focus on port management of a Yeosu City port.
- AIS data is consists of all the data related to vessel ships, tugboat and pilot ships with information such as geo location and active hours.
- These ports generate the AIS data at their end & push that into the big data pipeline for cleaning, processing, and Business analysis.
- The data comes in batch manner and updated to server every day at midnight.
- We must create end to end data pipeline for AIS data using HDFS, Apache Spark, MySQL and Apache Hive.
- Comes up with missing scenarios and get them clarified with BA.
- Implement Data Validation, Quality Checks, Profiling.

Marketing Executive , Vivo Smartphones Pvt Ltd, Pune

JANUARY 2016 – NOVEMBER 2017

- Served as a thoughtful advisor, working with Sales and Marketing professionals.
- Introduced scheduled targeted email marketing programs.
- Worked alongside marketing specialists and graphic designers.
- Assisted in the creation of graphic materials for the use of the marketing department.
- Launched New products and managed those events.

EDUCATION

MBA(Marketing), Deogiri Institute of Engineering and Management Studies, Aurangabad

JULY 2015 – JULY 2017

Graduated with grade A and University level recognition for final project.

Bachelor of Computer Science, MGM Dr. GY Pathrikar college of CS and IT, Aurangabad

FEBRUARY 2013 – OCTOBER 2015

Graduated with First Division having 68%.

COURSES

Big Data (Python, Spark, Scala, SQL, AWS), Simplilearn

JULY 2018 – DECEMBER 2018

DECLARATION

I hereby declare that the information mentioned above is true to the best of my knowledge.

(YOGESH SOPAN TEMKAR)