Contact no. :
+91 702 046 9342

Vishwajeet Dabholkar

vishwajeetdabholkar@gmail.com
www.linkedin.com/in/vishwajeetdabholkar5002

EMPLOYMENT

Enterprise Solutions Engineer   **Singlestore**   **March 2023 - Present**

- Conceptualized and implemented various Proof of Concepts (POCs) utilizing SingleStore in diverse data analytics environments, emphasizing its flexibility and potential for data engineering solutions.
- Successfully completed a POC leveraging Arcion for data ingestion from Oracle to SingleStore, thereby demonstrating adaptability and capability in Data Migration.
- Utilized ChatGPT for building an interactive demo to exhibit SingleStore's comprehensive capabilities in data analytics, showcasing strong skills in using AI in data engineering contexts.
- Developed a complete application using Python and SQL, demonstrating SingleStore's full-text search functionality (match and against), effectively applied for duplicate detection. Showcased proficiency in applying advanced SQL concepts and Python programming.
- Proficient in AWS cloud platform, consistently applying AWS services to ensure reliable and scalable data engineering solutions.

Analyst, Data Engineer   **Fresh Gravity India Pvt.Ltd**   **April 2022 – Feb 2023**

- Created a MDM data pipeline for a SaaS platform for Reltio MDM using AWS Lambda, RDS, S3, severless framework API Gateway.
- Migrated existing python repositories from pip to poetry based package dependency manager.
- Worked on creating a ingestion platform using Azure Data factories.
- Created APIs for creating ADF linked service, dataset, pipeline creation & execution and pipeline status.
- Worked on migrating data from remote SQL sever to S3 using pyspark and AWS glue.

Big Data Engineer   **Freelance**   **Jan 2022 – March 2022**

Deviation report for measurements team(Jan 2022 - March 2022)
- Created a fully automated flow to query BigQuery using bigquery CLI into pandas data frame.
- Wrote a logic to showcase first level of analysis for measurements team for sources such as email, push notification, banner and sms.
- Used pandas and html to model and showcase report on a mail.
- Wrote a code to send the generated report to respective stakeholders.

Data ingestion pipeline optimization(Jan 2022 - March 2022)
- Handled high volume data in TBs daily across different verticals and optimized measurements pipeline to generate insights of customer and campaigns lifecycle.
- Optimized and revamped existing architecture and code to save daily computation time by 10 hours.

Big Data Engineer   **Lumiq.ai**   **March 2021 – Dec 2021**

API Integration of Customer 360 Data Lake with various sources (July2021 - Dec 2021)
- Creating AWS Lambda APIs to send data from Data Lake to various consumers
- Created a framework to query AWS Athena, take the result and convert it into a Data Frame and make customized JSON response
- Created a framework to query AWS RDS, take the result and convert it into a Data Frame and make customized JSON response
- Hosting Lambda APIs on AWS API Gateway
- Testing of API working on postman.

Customer 360 Data Lake (March 2021 - July 2021).

- Creating data lake for a insurance company using AWS services and PySpark.
- Ingesting data from multiple sources as the client has variety of systems in place for collecting data
- Creating APIs to ingest data from source into Data Lake
- Used Syniti DB replication tool to ingest data in near real-time from one source.
- Data lakes major ingestion was done by creating automated data pipelines using PySpark and Glue Jobs
- DataLake is created using S3 as storage.
- All the latest data is ingested in RDS for faster querying of data for reports.
- Data Lake in S3 has all the data, historical as well as latest.

Data Engineer, Associate   **Tata Consultancy Services Ltd.**   **Sep 2019 – March 2021**

Cash analytic and development (July 2020 – March 2021).

- Creating and managing SQL queries for complex views and procedures to load data. Using MemSQL's pipeline to load data from S3 bucket into memsql tables.
- Created a data migration pipeline for- Ingesting data stored in a parquet file at S3 buckets into memsql distributed database using pyspark.
- Created a Data Migration pipeline to transfer local excel file to S3 bucket in CSV format and ingesting that CSV file into memsql DB using PySpark.

Next Generation Data Lake (Oct 2019 – June 2020)

- Worked with a team of Data Engineers and Solution architects to create a new optimized data lake with AWS S3, AWS EMR, AWS Glue, AWS Athena, PySpark, Python, and Pandas. I did a POC using HUDI for Upsert operations on S3 bucket. The main objective of project was to create a Data Lake entirely based of AWS services for the Pay-as-you-go model.

Personal Projects:

PySpark Based ETL tool :

- Engineered a comprehensive ETL tool using Python, PySpark, and Flask to streamline data migration between CSV, SQL Server, PostgreSQL, and Oracle databases.
- The core of data migration was developed using PySpark, demonstrating the application of big data processing libraries for complex tasks.

ADDITIONAL EXPERIENCE AND AWARDS

- One of the youngest TCSer to get place in **Customer Hall of Fame.**
- **Recognition by MD's of John Deere US, Germany and India, MD and owner of Ghatge Patil Industries.**
- **Winner** of National level paper presentation for topic **"Cloud Computing".**
- **2nd Rank in Capgemini Tech challenge 2020 in PySpark** domain among 2849 participants.
- Filed patent for **"Automated Product Data Manager for Foundry Products".**
- **Champion** for SingleStore DB community hub, also chosen for their **first ambassador program**

Languages and Technologies

- ChatGPT Prompts;
- Python; SQL; Scala;
- PySpark; SparkSQL; MemSQL; Pandas; boto3;
- Azure data factory, Data-bricks; REST APIs
- AWS Glue, RDS, EMR, S3, EC2, API Gateway, KMS ; Syniti DB replication;
- Jupyter notebook; Git-hub; PyCharm; Postman;
- Big-query, Azkaban;