

KSHITIJ MALOO

BUSINESS INTELLIGENCE, DATA WAREHOUSING & DATA ENGINEERING

Technical Lead - Domain experience in verticals such as Pricing, Banking, Logistics, Pharma and Automobile

📞: 9928931891

 www.linkedin.com/in/kshitij-maloo-49143359/

✉: Kshitijmaloo@gmail.com



Executive Profile

- Dynamic and motivated IT professional with around 8 years of experience as a Big Data Engineer with expertise in designing data intensive applications using **Hadoop Ecosystem, Big Data Analytical , Cloud Data engineering , Data Warehouse / Data Mart, Data Visualization , Reporting , and Data Quality solutions .**
- Result-oriented Professional with **nearly 8 years** of experience in multiple phases of **Data Warehousing** and **Cloud ETL Methodologies & Implementation** in IT Industry
- Worked on **Data Extraction, Transformation & Loading** for providing ETL solutions
- Expertise in Build data architecture for ingestion, processing, and surfacing of data for largescale applications. 📊 Proficiency in languages like **SQL, Python, Spark**
- Experience with cleaning, aggregating, and pre-processing data from varied sources
- Experience in Hadoop led development of enterprise level solutions utilizing Hadoop components such as Apache Spark, Hive, HDFS, Hive, YARN, Kafka.
- 4+ years of expertise in AWS Cloud Technologies for big Data solution.
- Mastery of relevant technical skills, including SQL, data modeling, schema design, data warehouse, BI reporting tools (e.g. Tableau), scripting for automation
- Experience **with cleaning, aggregating, and pre-processing data from varied sources** with proficiency in **languages like SQL, Python, Spark**
- Expertise in data modeling and transformation of large-scale data sources using SQL, Spark
- Profound experience in Creating, debugging, scheduling and monitoring Jobs with **Apache Airflow and Azkabaan**. Expertise in Designing solution for Airflow for Orchestration.
- Skills in relevant technical skills, including **SQL, data modeling, schema design, data warehouse, BI reporting tools (e.g. Tableau, Superset, Quicksight)**, scripting for automation
- **Experience in performance tuning for improving the system response time**
- Drove the **creation** of designs and plans for the integration for all data warehouse technical components; administered data design, tool design, data extracts/transforms, networks and hardware selections.
- **Proficient** in the creation & maintenance of **database/application architecture** and standards following best practices for physical structures, logical databases, and user management
- Impacted the **data warehouse architecture** and reduced the TAT
- Established healthy relations with **clients & stakeholders** to provide support & gain understanding of **business process** and provided most suitable **data solution**
- Customer-focused attitude and desire to interface directly with end-user clients



Key Impact Areas

Database Design/Architecture / ETL Processes Development



Project Management



AWS



Big Data & Spark



Data Extraction & Transformation



Onsite-offshore coordination



Stakeholder Engagements



Team Building & Leadership



Team Management



Apache Airflow



Python



Soft Skills

Communicator



Innovator



Collaborator



Thinker



Intuitive



☰ Technical Skills

- **AWS Cloud:** EMR, Redshift, Amazon Ec2, AWS S3, MWAA, Elastic Load Balancer, Auto Scaling Groups, AWS Lambda, SNS, SES, Amazon RDS
- Data Modeling & Mining, Elastic Search(ELK Stack)
- **Big Data Technologies:** Hadoop, HDFS, HIVE, Yarn, Apache Spark, Sparklib, Mapreduce
- **BI Tools:** Tableau, Superset, Quicksight, Power BI
- **Languages:** Python, HTML, CSS, Spark, pyspark
- **Databases:** Informatica, IBM DB2, SqlServer, MySQL, Oracle, PostgreSQL, DynamoDB, Snowflake
- **Source Control Tool:** Subversion SVN, Mantis Bug Tracker, Assembla, Jira, Gitlab, Bitbucket
- **ETL Tool:** Talend, Informatica, Dataiku
- **Other Tools:** Microsoft Access, VBA, Visio, Oracle DBA Studio, Developer 2000, IBM Data Studio



Certifications

- ATCS Business Analytic Certification of various Project



Education & Credentials

- **Bachelor of Technology** from Swami Keshwanand Institute of Technology, Jaipur (RTU Kota in **2014**)

📅 Career Timeline



Professional Experience

Since Oct'21: Paytm Payments Bank (Technical Lead Lead)

Key Result Areas:

- Responsible for Designing end to end architecture over AWS Cloud and maintaining quality reference data in source by performing operations such as cleaning, transformation and ensuring Integrity in a relational environment. Worked as Solution architect for Bank-DWH.
- Designed, Architect and Developed many Internal Frameworks like BI Reporting Tool, Compute Module for Internal Reporting and data sharing.
- Responsible for many Inhouse project development and expertise in Managing a Team
- Developed and integrated Superset with On-premised Hadoop cluster For Enhanced BI Reporting for users.
- Designed and developed Security Framework to provide fine grained access to objects in AWS S3 using AWS Lambda, DynamoDB. Set up and worked on Ranger authentication principals to establish secure network communication on cluster and testing of HDFS, Hive, Pig and MapReduce to access cluster for new users.
- Performed end- to-end Architecture & implementation assessment of various AWS services like Amazon EMR, Redshift, S3. Used Kafka Batch Stream for Ingestion from various Data Sources.
- Used AWS EMR to transform and move large amounts of data into and out of other AWS data stores and databases, such as Amazon Simple Storage Service (Amazon S3). Developed Architecture at Cloud for Bank - DWH team.
- Creating Lambda functions with Boto3 to automate the Airflow DAG's Generation process with User input as JSON Only.

- Coded Spark Frameworks to load, transform data, fix defects like SCD 2 date chaining, cleaning up duplicates. Developed reusable framework to be leveraged for future migrations that automates ETL from RDBMS systems to the Data Lake utilizing Spark Data Sources and Hive data objects.
- Conducted Data blending, Data preparation using Quicksight for monitoring Data pipelines.
- Integrated Apache Airflow & MWAA with AWS to monitor multi-stage workflows with the tasks running on Amazon EMR and Ec2. Developed Airflow for On-premise Hadoop cluster Jobs Orchestrations.
- Worked with HIVE data warehouse infrastructure-creating tables, data distribution by implementing partitioning and bucketing, writing and optimizing the HQL queries
- Developing ETL pipelines in and out of data warehouse using combination of Python, Spark, SQL.



Previous Experience

Jan 21-Oct21 : Atria India Pvt Ltd , Gurgoan as Project Lead

July'20-Jan 21: (On Site Paytm Payments Bank, Noida), TEKsystems India Pvt Ltd : Noida as Sr. BI Engineer.

Dec'17-July'20: Pratham Software – Jaipur as Senior Data Engineer

Jun'14 - Dec'17: ATCS Inc., Jaipur as Software Engineer

Key Result Areas:

- Conducting **direct discussion** with the Functional/End Users to analyze systems requirements for developing technical specification using Accenture Tools
- Preparing ETL high level technical design document; designing and supporting high level ETL solutions working with IT/Business stakeholders
- Contributing in end-end execution of projects from planning to execution primarily on ETL, Reporting, Data Warehousing & Database Architecture; implementing the project plans within preset budgets & deadlines
- Architecting complex ETL Packages with dynamic and parameterized connections enabled for heterogeneous sources such as relational databases, flat files and excel files
- Designing and creating mappings, source qualifier queries, sessions and workflows as per the design
- Modifying existing mappings for enhancements to implement new business requirements
- Implementing **new data source integration**, addressing **techno-functional issues** and interacting with **stakeholders & users** for resolution during production deployment
- Collaborating with business users on requirements, objectives and measures for architecting data warehouse solutions and integrating technical components
- Extending support & guiding Business Users for the existing system and appropriate **data marts** to fetch the required information
- Liaising with **Client Touchpoints** across the organization and ensuring the **implementation, training & adoption** of organization's proprietary solutions
- Mentoring, **training & monitoring** the performance of team members to ensure efficiency in process operations and meeting of individual & group targets

Major Projects

Fresh Lake: It's internal Project for DWH Migration to AWS Cloud.

- Ownership of that Migration with architecture design, Development and migration.
- Worked on Development & Design of many frameworks for Movement of data and Tasks from one platform to another.

Zeus: its Internal Product that do BI Reporting from On-premises data.

Technologies & Tools: Superset, Spark, PostgreSQL, Redshift, Hadoop

- Owner for Creating solution for Internal Users to enhance BI Capabilities of Bank.
- Developed and managed the architecture and solution for Bank Use cases with Superset.

Bank DWH – worked on Bank datawarehouse system to generate multiple reports to feed to multiple stakeholders.

Technologies – Python, AWS Redshift, S3, Hive, cloudera Hadoop

Glassbox: Its product to serve pharmaceuticals business to get insight on their Revenue & feedback.

Technologies & Tools: Airflow, Redshift, Matillion, Python, snowflake, SQL, pyspark, dataiku

- Worked on client handling (requirement gathering, client Communication)
- Migration of legacy system to new on demand technology stack.
- Managing data engineering functions including, but not limited to: data extract, transformation, loading, integration in support of enterprise data infrastructures— data warehouse, operational data stores and master data management.

Pricing Catalyst B2B: It's a product to serve the business for B2B Market and to provide price breakup and details to the Customer and to enhance the pricing using the older data.

Technologies & Tools: Python, PySpark, SQL, SqlServer, Dataiku, SqlServer

- Managing data engineering functions including, but not limited to: data extract, transformation, loading, integration in support of enterprise data infrastructures – data warehouse, operational data stores and master data management

Category Strategy: An application to create actionable business strategies for assortment and Pricing. (Retail Solution)

Technologies & Tools: Python, Pyspark, Sql, Sqlserver, Dataiku, Sql Server

- Worked on client handling (requirement gathering, client Communication)
- Developed data architecture design to enable analysts to perform targeted customer analysis.

Pricing Catalyst (Food): Client is an American food service, facilities, and uniform services provider to clients in fields of education, health care, business, corrections, and leisure. Catalyst Pricing Tool is designed to enhance and support Client's end-to-end pricing process

Technologies & Tools: Python, Pyspark, Sql, Sqlserver, Dataiku, Sql Server

- Determines data retention and archiving policies and managed data engineering functions including, but not limited to: data extract, transformation, loading, integration in support of enterprise data infrastructures – data warehouse, operational data stores and master data management.

Service Contract - Daimler: Its Mercedes Benz Financial System Reporting tool to analysis of Current Market and Forecasting the upcoming Market Data.

Technologies & Tools: Power BI, Excel, IBM DB2, Informatica, Sqlserver

- Automated forecasting system to forecast the Revenue and Expense for MB as per their current market situation

Asset-Daimler: Application tool for selling contract on the vehicle for all countries

Technologies & Tools: Power BI, Excel, IBM DB2, Mysql

- Review of Code of all developers and worked on to create application User Friendly to enhance the business

ASPECT-Daimler: Daimler application for Analysis of After sales business and for Suggesting Business strategy for ongoing Business

Technologies & Tools: Power BI, Excel, IBM DB2