

Abhishek Mote

Chicago, IL • moteabhishek94@gmail.com • H-1B • Open to relocation and travel • +18478345564

• <https://www.linkedin.com/in/abhishekmate/>

EXECUTIVE SUMMARY

Experienced IT professional with a proven track record of over 7 years, adept at successfully implementing both waterfall and Agile methodologies in challenging environments. Specializing in Hadoop/ Big Data ecosystem, ETL Data pipeline, Data Lake storage analysis, Data Warehousing, Data processing and Cloud Engineering.

EDUCATION

Northwestern University, Evanston

MSIT

GPA: 3.77

Relevant coursework: Data Analytics, Managerial Analytics, Statistics, Data Analytics using R, Python, Applied Data Science, Computer Systems, Data Science for Business Intelligence, Enterprise Data Science, and Data Warehousing.

Rajiv Gandhi Technical University, India

B.Tech Mechanical Engineering

GPA: 3.0

PROFESSIONAL EXPERIENCE

WellDyneRx - Chicago, IL

WellDyne is a **Pharmacy Benefit Management Company**. The objective of the project was to move the data from on-prem to Azure Data Lake as a part of Lift and shift of the application.

Oct'21 - Present

Data Engineer

- Designed and trained **Machine Learning Models**, including **Logistic Regression**, **XGB** and **random forest** to predict patient outcome on electronic health record (**EHR**) and other health data
- Integrated data from various sources, including **EHR's**, lab results, wearable devices into **Azure Data Lake Gen 2**, pre-processed and cleaned the data to ensure high-quality input for ML models.
- Deployed the models using **Azure ML** service and implemented continuous monitoring and retraining, conducted extensive testing to tune **model accuracy** and reduce prediction latency.
- Collaborated with **DevOps** teams to design **AWS** architecture, resulting in a 30% boost in file intake speed and a 15% reduction in infrastructure costs.
- Employed **SQL Server Management Studio** for extracting and exporting on-premises data, collaborating closely with the **Azure Data Factory** team to ensure seamless data movement.
- Conducted a comprehensive assessment of on-premises **SQL Server** databases, defining migration requirements and goals.
- Migrated on-prem data from **SQL** server using **SSMS** to **Azure Data Lake Gen 2**, utilizing **Azure Data Factory** resulting 30% improvement in data accessibility and efficiency.
- Utilized **Spark SQL** and **DataFrames** to execute complex queries and ensure optimal performance in processing large datasets.
- Collaborated with cross-functional teams to ensure seamless integration of cloud-based data platforms, including AWS and Azure, and readiness to adapt to platforms like **Redshift**.
- Provisioned **servers** using **Azure**, including managing volumes, implementing **Azure Content Delivery Network (CDN)**, deploying **Virtual Machines (VMs)**, setting up **Azure Database (Azure SQL)**, configuring security groups, and employing auto-scaling and load balancers (Application Gateways) within the designated virtual private connection.
- Employed **Azure Synapse Analytics** (formerly SQL Data Warehouse) to examine partitioned and bucketed data, deriving various metrics for dashboard reporting
- Implemented data governance policies and procedures, ensuring data quality, consistency, and security throughout the organization.
- Leveraged **Azure Synapse Analytics** to execute complex **SQL queries** on processed data, ensuring optimal query performance.
- Efficiently cleaned and transformed data using **Power BI**, used **DAX queries**, crafted custom **measures**, designed dashboard leading to improved data resulting in 15% reduction in HR operational costs.
- Proficient in Azure **DevOps** and CI/CD pipelines, utilized GitHub Workflow.

Amazon Development Center India – India

Largest **E-Commerce Company**. The objective of the project was to extract actionable insights from various facets using robust Data Engineering methodologies.

Aug'18 - Jan'21

Sr. Data Analyst

- Imported data from various sources such as **S3** and **blob storage**, transformed them into **Spark Data Frames** for further analysis.
- Worked collaboratively with **AWS** services including **EC2**, **S3**, **Route 53**, **Lambda**, and **Cloud Formation** to establish a robust data infrastructure.
- Experience in optimizing **Spark Jobs** and **SQL Queries**, improved query performance using spark optimization techniques.
- Utilized **PySpark**, **S3**, and **Python** to ensure the seamless operation of data pipelines, allowing 24/7 availability for data processing.
- Write Spark code in **Python**, utilizing **Spark SQL** and **Data Frames** for aggregation purposes, ensuring the generation of actionable insights.
- Created **EMR Clusters** to write spark jobs for data transformation and store them into **S3 buckets** for further analysis.
- Authored data pipeline using **AWS Glue**, **Glue catalog**, **S3**, and **AWS Redshift** serverless to move the data from raw s3 to final s3.
- Wrote **PySpark** codes to transform data as per business logic and saved the final output in parquet format for read-heavy operations
- Worked on **AWS Redshift**, created on-demand tables, and worked in **Redshift Query editor** to create **SQL** queries and views
- Responsible for creating on-demand tables on S3 files using **Lambda Functions** and **AWS Glue** using Python and PySpark
- Created a Hive schema with bucketing and partitioning for speed improvement.
- Selected and configured IAM tools, defined user roles based on the responsibilities, and implemented multi-factor authentication for enhanced security.
- Implemented **IAM access controls** for team members with necessary access without compromising security and established the principle of least privilege.
- Mentored junior team members, providing technical guidance and support to facilitate their growth and development within the organization.

Axis Bank – India

Aug'16 - Jul'18

Banking and Finance Company. I worked on ensuring the integrity of data foundations and enabled business executives and data scientists in efficient decision making.

Data Analyst/ SQL Developer

- Responsible for defining, developing, and communicating key metrics and business trends to partner and management teams.
- Worked with stakeholders, including business users, data architects, and developers, to gather requirements and design SQL solutions that meet their needs.
- Created monitors, alarms, notifications and logs for Lambda functions, **Glue Jobs**, **EC2**
- Transformed the data received from source systems using Scala and Python as per business transformations, saved the data into the files to load into **HIVE**.
- Developed shell scripts to automate the daily jobs, weekly jobs and Spark scripts to import large files from Amazon S3 buckets.
- Used **SQL** complex queries and subqueries, Advanced Excel for complete data preparation and data analysis
- Performed data analysis and maintenance on information store in **MySQL**, **MS SQL server**, **MongoDB** database.
- Played a pivotal role in building and updating **SQL** and **SSRS reports**, and successfully tested and implemented reports into **Microsoft SSRS** reporting services through the Source Control tool, **Bitbucket**, as part of the BI Reporting team's responsibilities.

SKILLS

Functional: Agile, JIRA, SDLC, Data Manipulation, Requirement Gathering, Data Visualization, Exploratory Data Analysis

Technical: SQL, Python, Power BI, SSRS, Bitbucket, Terraform, R, PostGreSql, SSMS, Machine Learning, Azure Data Factory, Data Lake storage Gen2, Azure Data Bricks, Azure Pipelines, EC2, DynamoDB, S3, Redshift, Cosmos DB, Git Hub, CI/CD

CERTIFICATIONS

AZ-900 Microsoft Certified Azure Fundamentals – Udemy

Databricks Lake House Fundamental – Databricks

SQL for Data Analysis and Business Intelligence

Certified Scrum Master (CSM) – Scrum Alliance

Certified Scrum Product Owner (CSPO) – Scrum Alliance

PMI Agile Certified Practitioner – LinkedIn/PMI