

VIKAS MALIK

TEL: +91 8800759173 | E-MAIL: IMVIKAS10101@GMAIL.COM

AN OVERVIEW

- A dynamic, dedicated and responsible AWS certified data engineer professional with total 9 years of experience in IT, IOT and Product engineering.
- Have over 5 years of experience as big data developer and architect with very good exposure on technologies like Spark, Scala, Hive, Sqoop, HBase, Oozie, Map Reduce, HCatalog, Pig, Zookeeper, Kafka.
- Gained Experience in AWS S3, EC2, Redshift Data warehouse, Redshift Spectrum, Glue, Athena serverless architecture.
- Have created multiple pipelines to load and transformed data from different database to HDFS and S3.
- Have worked on versatile domain and module like Banking, BFSI, Customer matching module, loyalty program, point of sales program, GIS, GPS tracking etc.
- Served as the Scrum Master with responsibilities of Capacity Planning, Velocity Planning, Facilitating Sprint Ceremonies and bridging between Product Owner and the Scrum Team.
- Possess a perseverance, flexible & detail-oriented attitude.
- Significant abilities in carrying out analysis; finding gaps in process as well as devising a road map for enhancing various applications.
- An effective communicator with excellent relationship building & interpersonal skills.
- A quick learner with the ability to work under pressure and meet deadlines record.

WORK EXPERIENCE

- Working with **Citi Bank** as a **Lead Data Engineer** since May-2020.
- Worked 4.3 years with **Harman Connected Services** (acquired Fishbowl), a company of Samsung as a **Lead Big Data Engineer** from March-2016 to May-2020.
- Worked 4.2 years with **MapMyIndia**, brand of CE Info Systems Ltd. as Senior Software Engineer from December-2011 to February-2016.

ACADEMIA

- Master of Computer Application from IEC College of Engineering and technology (Greater Noida), Uttar Pradesh technical university with 70.0% marks in 2011.
- Bachelor of Computer Application from D.A.V. College, C.C.S. University with 70.2% marks in 2008.

SKILL SET

Data Processing Frameworks	Spark, MapReduce, HDFS, YARN, Oozie, Pig, Hive, Drill, Sqoop, Zookeeper
Language/Scripting	Java, Scala, Shell Scripting
Database	Hbase, RedShift, MySQL, SqlServer
Cloud Technologies	Amazon S3, EMR, IAM, Athena, Glue, Redshift Spectrum
Domain Knowledge	Banking, Restaurants POS, GIS, Customer matching module

PROJECT EXECUTED

Project 1: **Global Data Repository - GDR2**
Client: CitiCorp, Tenure: Jun-20 to Till date
Designation: Lead Data Engineer

Technology: Spark, Scala, Hive, Hbase, Sqoop, Drill, IBM DB2, Yarn, Zookeeper, Parquet

Client Description: Citicorp is a diversified financial services holding company that provides a broad range of financial services to consumer and corporate customers. The Company services include investment banking, retail brokerage, corporate banking, and cash management products and services.

Key Responsibility Areas:

1. Analyze and understand the legacy code written in Abinitio Java and Talend for migration to spark.
2. Used various spark Transformations and Actions for cleansing the input data and processing raw file.
3. Involved in writing custom spark programs using scala API for data processing.
4. The hive tables are created as per requirement were Internal or External tables defined with appropriate static, dynamic partitions and bucketing, intended for efficiency.
5. Load and transform large sets of structured, semi structured data using hive.
6. Used Spark and Spark-SQL to read the parquet data and create the tables in hive using the Scala API for generating the complex fact table for reporting.
7. Developed Hive queries for the analysts.
8. Experienced in using the spark application master to monitor the spark jobs and capture the logs for the spark jobs.
9. Implemented Spark using Scala and Spark SQL for faster testing and processing of data.
10. Implemented Spark using Scala and utilizing Data frames and Spark SQL API for faster processing of data.
11. Worked on using the Airflow as scheduling tool for spark and sqoop jobs
12. Used bitbucket and teamcity for version control, code deployment through CICD pipeline.

Project 2: **Fishbowl Data Platform**

Client: Fishbowl, Tenure: March' 18 to May' 20

Designation: Lead Data Engineer

Technology: Spark, Scala, SparkSql, Amazon S3, EMR, IAM, Athena, Glue, Redshift Spectrum

Client Description: Fishbowl is the leading data, marketing, and analytic solution provider uniquely serving restaurants. Helping clients to increase the frequency of visit of their guests and increase the amount spent by guests while dining on each visit to the restaurants in the space of quick service and fast casual restaurant.

A product to offer our clients an easy way to get insights into their guest activities. This will be instrumental in asking data questions in English and get answer in form of a report or dashboard instantly for taking immediate actions to integrate business risks.

Key Responsibility Areas:

1. Extracting, interpreting and analyzing data to identify key metrics and transform raw data into meaningful, actionable information.
2. Building data pipelines using data factory to cleanse, standardize, ingest, transform data and enrich the database.
3. Developing Spark code using Scala to process and create dimension/fact tables.
4. Creating pipelines and data-sets to orchestrate data.
5. Writing hive queries to populate data for aggregate layer
6. Bring in data from different systems and covert input data into standard tabular format

Project 3: **Guest Analytics ETL 3.0**

Client: Fishbowl, Tenure: Mar'16 to Dec`18

Designation: Lead Hadoop Developer

Technology: MapReduce, HDFS, YARN, Oozie, Pig, Hive, Drill, Zookeeper, Sqoop, Kafka, Spark, Weka

Description: Fishbowl helps restaurants leverage data to drive predictable sales growth by using Tubule, which assessing data from MSSQL and hive but query response was slow. To improve that they need new UI interface to show guest analytics with faster response. For that this ETL was developed where we use drill as query engine and pig for data transformation.

GA ETL provides following advantages.

1. Automated data pipeline from 2.0 to new 3.0 databases.

2. Use of parquet data format for faster query.

3. Use of drill for faster queries.

Key Responsibility Areas:

1. Clustering of customer data using MapReduce, Weka based on purchase spend, frequency etc.

2. Write script in Pig, Oozie, Hive for data transformation and loading.

3. Creation of drill view on top of hive external tables.

4. Manage Customer matching module (CMM) and Loyalty data load.

5. Requirement analysis for new table structure.

6. Write web services to fetch data from drill.

7. Assigning and reviewing the jobs done by team members.

Achievement: Setup a new team of Big data and trained them in Oozie, pig, hive.

Project 4: **Map Data and Route Creator**

Client: MapMyIndia, Tenure: Dec'15 to Feb'16

Designation: Hadoop Developer

Technology: HDFS, YARN, Oozie, Pig, Hive, Sqoop

Description: The product is developed to enhance the map data by analyzing the incoming GPS data from mobile or vehicle tracking devices. This product provides below functionalities.

1. Automated data pipeline from MySQL to HDFS.

2. Aggregate input data based on sources and compare that from existing data with calling api in MapReduce.

3. Insert new map data in hive for data engineer team to inspect and integrate.

Key Responsibility Areas:

1. Develop automated data pipeline from MySQL to hive partitioned table using Oozie, Sqoop, Hive scripts.

2. Writing MapReduce program to aggregate data based on user-id and travel path and compare that data to existing data and generate new route data.

3. Send automated report and job status through email to data engineering team.

Project 5: **VEHICLE TRACKING SYSTEM 4.0 AND 3.0**

Client: MapMyIndia, Tenure: Dec'14 to Feb'16, July'12 to Dec'12

Designation: Software Developer

Technology: Java, spring, Hibernate, JQuery, Ajax, MySQL, JBOSS 6, XMPP, Multithreading, Socket programming, Collections

Description: The product is developed to support the transport operations in India to minimize the process and operations that are manual and labor intensive and maximize the transport usage and accuracy.

VTS provides following facilities.

1. Real-time tracking of vehicles and employees.

2. Vehicle and driver allocation

3. Send SMS and E-mail to transport manager in case of geo-fence entry and exit, daily vehicle running reports.

4. Alert the transport manager in case of vehicle theft.

5. Generate Reports.

6. Admin module to configure the access level for various clients and vehicles.

Key Responsibility Areas:

1. Requirement analysis for reports and real time view for vehicle.

2. Preparation of web design of the application.

3. Writing Use Case Scenarios.

4. Support in developing database design and logic.

5. Develop reports and real time view for vehicle tracking.

Project 6: **Transport Management System**

Client: Jet Airways, Tenure: Jan'13 to Jan'14

Designation: Software Engineer

Technology: Java, Spring, Hibernate, YUI, Ajax, HTML 5, MySQL, JBOSS, Multithreading, Socket programming, Collections

Description: The system is developed to support the Jet Airways employee transport operations in India to minimize the process and operation which are manual and labor intensive and maximize the transport usage and accuracy.

TMS provides the following facilities.

1. Real-time tracking of vehicles and employees.
2. Uploading and generate the routing support in form of trips.
3. Vehicle and driver allocation
4. Clubbing, de-clubbing employee from trips or generate ad-hoc trips.
5. Send SMS and E-mail to employee, driver and transport manager according to the work flow.
6. Alert the base station in potential delay of trip completion.
7. Generate Reports.

Key Responsibility Areas:

1. Gathering requirements from client.
2. Requirement analysis for application development.
3. Preparation of web design of the application.
4. Writing Use Case Scenarios.
5. Developing database design and logic.

PERSONAL DOSSIER

Date of Birth: 01-04-1988

Marital Status: Married

Present Address: A-259, Sector-31, Noida

I hereby declare that all the information given above is correct to the best of my knowledge.