

RISHI TAMRAKAR

Big Data Engineer

📍 Pune, India

☎ +91-8959630508

✉ rishi.tamrakar@yahoo.com

🌐 linkedin.com/in/rishitamrakar/

WORK EXPERIENCE

Clairvoyant (Data Engineer II :- Apr 2019 till date)

Tata Consultancy Services

(Big Data and ETL Developer :- Mar 2016 till Mar 2019)

ACHIEVEMENTS

- Awarded with **Clairvoyant "WFH Warriors Award"** for hard work, sincerity and commitment towards the work during Covid 19 .
- Received **Star of the Month** award, multiples **On the Spot** awards & appreciated by Senior Management for good work .
- Received appreciation from Customer, offshore and onsite management team for automating manual process.

STRENGTHS

- Quick learner
- Multi tasking
- Adaptive
- Flexible Timing

CERTIFICATIONS

- Hadoop 101 (BD0111EN) Certificate from IBM COGNITIVE CLASS.ai provided by Bigdata University.
- Acquired competencies in SQL, Hadoop and Informatica Power Center to benefit the project.

EXPERIENCE SUMMARY

- Experienced Big Data Developer with close to 4.6 years of industry experience. Working on projects in **Banking & Financial Services** industries, performing **ETL operations** with primary focus in **Data Cleansing, Data Profiling, Data Ingestion, Data Transforming and Data Loading**.
- Experience in all phases of Software life-cycle including **Requirements gathering, Analysis, Design, Development, Implementation, Enhancement, Tuning, Maintenance and Production support** .
- Experience in Working in Big Data/ETL Project and Technologies like **Apache Spark, Hadoop, HDFS, Hive, HBase, Hive Acid, Impala, SQL, Oracle, Teradata, SQL Developer, Informatica Power Center, Unix, Python, Core Java**.
- Good understanding of Data Warehousing Concepts such as **Data Lakes, Data Mart, Slowly Changing Dimensions**, Dimensions and Fact tables, Surrogate Key.
- 3+Years of Experience in designing and developing solution for processing large set of data using **Spark, Hive, Hadoop, HDFS**.
- Experience in **Data Modeling (Relational, dimensional, Star and Snowflake Schema), Data analysis, Data Lake, implementations of Data warehousing, fact table, dimensions**.
- Hands on in **Git, Bitbucket** and **Jira** for QA and production deployment of Hadoop project.
- Sound knowledge of **Internal/External tables, Partitioning , Bucketing, Joins** and **Optimizations techniques**.
- Expertise in writing **optimize spark code**, using **Hive Transactional tables** and **HBase Tables** for merging latest snapshots with history data sets.

CORE SKILLS



SQL & Data Warehousing



Spark with Java/Python

OTHER TECHNOLOGICAL PROFICIENCY

Devops Tools

- Bitbucket,
- DropZone
- Jira
- Git
- Putty & WinSCP

IDE:

- IntelliJ Idea 11.0.4
- SQL Developer 18.3.0
- HDP

ETL Tools:

- Informatica
- SQOOP

BIG Data Skills:

- **Apache Hadoop:** HDFS, Map Reduce, Hive, SQL, Sqoop, Spark, Java, Python ,Hbase

MAJOR PROJECT GLIMPSES

Apr 2019 - Polaris

EDUCATION

- **Bachelor of Engineering** with CGPA 7.2 RGPV, Bhopal
- **Class of XII** with 80.4% from M.P Board
- **Class of Xth** with 72% from M.P Board

Till date

PayPal

The purpose of this phase of the project is to replace the existing pipelines of transactions systems to a new platform Operational Data Source (ODS), which is a Hadoop based scalable solution for visualizing 360 degree insight of a transaction lifecycle at NRT using common data storage for PayPal and all its subsidiaries and at the same time serve the reporting requirement for the downstream like Merchant Reporting, Declined Analysis etc.

Team Size: 7

Roles and Responsibilities

- Working as Big Data Developer to deliver Hadoop based solution for the subprojects under PayPal Polaris.
- Designing and developing Big Data pipeline with focus on collecting, storing, processing and analyzing huge sets of data from multiple data sources.
- Converting existing ETL data pipelines into Spark.
- Support business decisions with ad hoc analysis as needed.
- Reconciling the data at each step to mitigate any loss of observations by using our data reconciliation framework.
- Develop big data pipelines supporting continuous updates and inserts in Hadoop region.
- Automate big data pipeline using Control-M scheduling tool.
- Created HBase Tables and partitioned table in Hive to maintain historical data.
- Created hive queries and Spark Code to perform SCD on business data.
- Code optimization to ensure better performance
- Coded Complex Logic in Hive and Spark with several Joins , Aggregate according to business requirement.
- Worked on ACID tables as well for Update support.
- Involved in documentation of all the important and relevant information needed for deployment in PROD environment.
- Helped in Production fixes. Followed by End to End Devops Model for Deployment Tracking.

Environment/Tools

Spark, Hadoop, Hive, HBase, SQL, Putty, WinSCP, Git, Java, Unix

Sep 2017 -
Mar 2019

Anti Money Laundering (AML)

Pittsburgh National Corporation Bank (PNC)

The purpose of this project is to develop Hadoop based scalable solution for analyzing suspicious customers and suspicious transactions and replicate functionality of Mantas system and its provided rules along with oracle db to hadoop platform.

Team Size: 16

Roles and Responsibilities

- Designing and developing Big Data pipeline using pySpark to collect data from multiple sources and store in hive external tables.
- Converting existing ETL data pipelines into pySpark.
- Support business decisions with ad hoc analysis as needed.
- Consume incremental and upsert data from oracle tables and ingest it to Hadoop ecosystem.
- Create hive Partitioned and bucketed tables to store data in different layers.
- Coded Complex Logic in Hive and Spark with several Joins , Aggregate according to business requirement.
- Involved in documentation for PROD environment and worked on production fixes as required.

Environment/Tools

Spark, Hadoop, Hive, Impala, SQL, Putty, WinSCP, Git, Python, Unix, sqoop

Mar 2016 -

Enterprise Risk Hub (ERH)

Aug 2017

Pittsburgh National Corporation Bank (PNC)

Enterprise Risk Hub, a comprise data warehouse which act as a single point of access for multiple enterprise risk related data. Data from multiple sources and different areas will be standardize based on archer system to serve various reporting requirement for downstream.

Team Size: 5

Roles and Responsibilities

- Designing and developing ETL Informatica mappings and workflows for data processing.
- Implementing SCD1 and SCD2 as per the business requirements.
- Writing complex SQL queries with joins, aggregate functions etc.
- Code optimization to ensure better performance.
- Support business decisions with ad hoc analysis as needed.
- Involved in documentation of all the important and relevant information needed for deployment in PROD environment.
- Logging Defects and co-ordinate with Test team for Defect fixing, Defect retesting and closure.
- Work on production fixes as required.

Environment/Tools

Informatica Power Center, Teradata, Oracle, CA7, SQL, Putty, WinSCP, Unix, Artifactory

DECLARATION

I hereby declare that the information given above is true to the best of my knowledge & belief.

Rishi Tamrakar