

SHUBHAM BHUMKAR

Mobile: +919503368768

Email: shubhambhumkar25@gmail.com**Career Objective**

To take a challenging role in the organization and give efficient and effective solutions that would help the organization to achieve the best solution in business which would increase its productivity in the market.

Experience Summary

- Currently working with **Clairvoyant India Pvt. Ltd. Pune.**
- Working in the IT industry for 6.3 years in Big Data Development / Cloud Technology.
- Have experience in building ETL pipelines on On Premises & AWS Cloud
- Data migration experience from RDBMS to On-Prem & On-Prem to Cloud
- Quick learner and good performer both in team and independent job environments.

Technical knowledge

| | |
|---|--|
| Big Data Technologies | Spark-SQL, Spark Core, Spark Functions, Hive, HDFS |
| Language API | Python, Java-8 |
| Cloud - Amazon Web Services | S3, Lambda, Glue, Athena, Step Functions, Redshift, CloudFormation |
| Other Skills | SQL, Unix/Linux |
| Bug reporting tools & Version Controls | JIRA, GitHub, GitLab. |
| Hands On | Scala, Spark Streaming |

Organizations

- FIGmd India Private Limited, Pune
 - Duration – 8 January, 2018 to 20 March, 2021
 - Designation – Associate Software Developer
- Clairvoyant India Private Limited, Pune (An EXL Company)
 - Duration – 13 May, 2019 to Current
 - Designation – Senior Software Engineer

Project #1**Project Name:** NFL – Fan DB**Organization:** National Football League, NY, USA**Project Domain:** Sports & Analytics**Duration:** February 2023 to till date

Tools / Technologies: PySpark, SQL, S3, AWS Glue, Athena, Azure repos

Role: Data Engineer

Description:

This project involved processing sports ticketing, fan and analytics data into the S3 data warehouse. To create an ETL pipeline which will source the data from various source systems like parquet files, S3 buckets or Athena tables.

Roles & Responsibilities:

- Creating the Glue jobs which will extract the data from Athena tables.
- Apply the business logic and filtering over this data from the SQL files present in S3.
- Handling the duplicates and creating snapshots of the target tables.
- Loading the processed data in data warehouse databases in Athena or S3.
- Create and run the crawlers to generate the tables if not available.
- Perform basic unit testing once data is loaded in DW.
- Deploy, monitor and maintain the live running modules.

Project #2

Project Name: SVB-ODP

Organization: Silicon Valley Bank, USA

Project Domain: Banking

Duration: April 2022 to 31st Dec 2022

Tools / Technologies: PySpark, SQL (Oracle), S3, AWS Glue, AWS Lambda, Athena, DynamoDB, Step Function, GitLab

Role: Data Engineer

Description:

This project involves creating ETL ingestion & foundation layers across S3 and Redshift. Another use case which involved creating step functions having Lambda and Glue jobs which will act as component phases for ex. Pre-update processes, merging of tables, unloading of data from S3 to Redshift, Cataloging and partitioning and post-update processes.

Role and Responsibilities:

- Extracting the data from sources like S3, Oracle, MS-SQL
- Load this data in Raw Data Layer
- Creating lambda functions and Glue job to transform the S3 File data.
- Transformation includes masking, filtering, and applying DQ checks,
- Transformed data will be loaded into the Foundation layer.
- Query Athena for basic level validations

Project #3

Project Name: Polaris ADMC **Organization:** PayPal Corp.

Project Domain: Payments Gateway

Duration: March 2019 to March 2022

Role: Spark-Big Data Developer / ETL through Spark

Tools/Technologies: Java-8, Scala, Apache Spark (Spark-SQL, Spark-Core), Hive, HBase, HDFS, Google Cloud Platform, Oracle SQL Developer, Intelli-J IDEA, Unix/Linux.

Description:

Project involves generating the different types of flatten reports by joining multiple tables, doing column level transformations as per the requirements and creation of data.

Role and Responsibilities:

- Creating the ETL Pipelines to generate different categories of reporting.
- This includes, reading the data from different sources, Ex. From a File, from a RDBMS
 - (Oracle), From a HDFS Location or From a Hive Table
- Process this data, do respective joins which are required to flatten, do transformations as per the need and load into targets.
- Target included mostly Hive tables, Oracle tables and Parquet Files in HDFS.
- Writing the unit test cases for the modules.
- Calculate the performance for the concerned jobs and do optimizations wherever applicable.
- Doing the production deployments
- Proof of Concepts for data migration from on prem to cloud (GCP DataProc / GCS Buckets)

Project #4

Title: JANUS

Client: US Healthcare and CMS Registries

Role: Big Data Developer

Duration: January 2018 to March 2019

Environment: Scala-Spark, Hive, Hadoop Ecosystem, HDFS, SQL, GitHub, Linux.

Role and Responsibilities:

- Building frameworks, functions and utilities in Scala using Apache Spark include, Dataframe operations, RDD operations, SparkSQL, handling Spark sessions, ScalaTest and SBT.
- Writing the ScalaTest scripts and frameworks to test the modules and code working.
- Firing hive queries to access data stored in HDFS,
- POC on processing data within AWS S3 Or GCP Data-Proc Buckets to HDFS and HDFS to Local Environment

Educational Qualifications

- Bachelor of Engineering from Shivaji University, Kolhapur in the year 2016 with 58%.
- H.S.C. from Latur Divisional Board in the year 2012 with 69%.
- S.S.C. from Pune Divisional Board in the year 2010 with 83%.

Personal Information

- Name: Shubham Prakash Bhumkar
- Date of Birth: 25-12-1994
- Gender: Male
- Address: A-708, Vision Ace, Tathawade, Pune 411-033
- LinkedIn: <https://www.linkedin.com/in/shubham-bhumkar-153b23135>