

Aditya Bagbande

PUNE - 411028

+91 87-88-011-123 # ADITYABAGBANDE5@GMAIL.COM

Summary

- 5+ years of experience in Data engineering and BigData Eco-System like Hive, Spark, Python and also in AWS S3, Glue, Airflow and Snowflake.
- Experience in working with RDBMS technologies such as MS-SQL, Writing SQL queries on SQL Server.
- Good Experience in closely working with client.
- Worked as Production Team member from Offshore.
- Experienced in AWS services for data engineering.
- Implementation for ETL processes to support analytics.
- Interface with customers, gathering requirements and delivering complete performance.
- Good team player with exceptional Leadership, Interpersonal, Communication and Analytical skills.

Skill Highlights

- **Technology** – AWS, Big Data, Airflow, Snowflake and Databricks
- **AWS Services** – S3, Glue, Athena, EMR
- **Domain** - Retail and Logistics
- **Bigdata Technologies** – PySpark, Spark, HDFS

Experience

Project #4

Data Engineer – June 2022 to Current

Cognizant, Pune

‘JPMC’ - We get the data from different sources few are near real time and the remaining are batch data (we consume data from s3 i.e., parquet files). We store this data in hive tables. So, all this is was done on on-prem. Now we are migrating the project on cloud. The process of data consuming will remain the same. The only change is migrating the models from python 2 to python 3.8 and make the model cloud ready. Once the model is deployed to cloud, we need to run the model and compare the data with on-prem data.

Responsibilities

- Analysing the data when loaded to hive tables. Updating the model code with the latest changes.
- Refactoring the code to make the model cloud ready.
- Comparing the data i.e. score comparison between onprem data and cloud data.

Technology Used

- Language: Pyspark
- Technology: AWS (S3, Athena, EMR)

Project #3

Data Engineer – March 2021- May 2022

Infosys, Pune

‘NIKE GSM’ – Collects data from Different vendor/factory groups which comes in the form of Parquet files. Here data is extracted and stored in the database. Some cleansing rules are performed on top of the data and after that it is integrated with necessary dimensions and move the data to snowflake. At the end this data is consumed by Tableau Dashboard to generate the Reports.

Responsibilities

- Working at client location to gather and understand requirements and parallelly working to fix the data pipelines.

- We get the data from different vendors we process the data by applying some cleansing rules and integrate with necessary dimensions.
- Mainly I work for the sourcing data and check whether the data is compatible for end users to test.
- Airflow DAG Simplifications and remediating the changes to the data products and parallelly doing the unit and regression testing so that the data can be used by the end user.

Technology Used

- Language: PySpark | Hive
- Techonolgy: Airflow | Snowflake | AWS

Project #2

Data Engineer – December 2018 – March 2021

Atos | Syntel, Pune

'GTM' - group gathers information about the customer details of the packages disbursed by FedEx group mainly in the US and Canada region and to Integrate data from all sources to provide complete, timely and reliable consumer analytics for planning, analysis and execution. To Understand the nature and capabilities and predict the intentions of customers to achieve business goals.

Responsibilities

- Worked with retail and logistics customers to understand requirements and provide exceptional service in data processing in Hadoop eco system.
- Understanding the upstream source nature and working with business cases.
- We get the data from different business teams and the data varies correspondingly.
- We get the data each day and some files we get each week.
- Designed and implemented Data Lake on AWS.
- Responsible for checking and modifying the existing Spark jobs.
- Responsible for extracting the schema and loading the data in hive to do transformations using files.

Technology Used

- Language: Spark | Python
- Techonolgy: Glue|Athena|CloudWatch|Boto3

Project #1

Data Engineer – December 2018 – March 2021

Atos | Syntel, Pune

'GTM - CLM' is complete suit for logistics business solution. FedEx GTM-CLM group gathers information about the customer details of the packages disbursed by FedEx group mainly in the US region, to understand the nature and capabilities and predict the intentions of customers. There is a application which is webbased. Web application which is SQE, used to administrate the system of logistics business.

Responsibilities

- We receive files from other teams, which are in encrypted file format.We run some scripts which are in shell and Perl scripts that decrypt the file and load the data in the external database.
- Responsible for writing new Perl-Scripts or to modify existing scripts according to client's requirement.
- Responsible for modifying the existing PISql procedures according to client's requirement.

Education

B.Tech/B.E-

Pune University (Computer Engineering) Marks : 65%

HSC (XII) -

Shivaji Mahavidhyala, Udgir Marks : 60%

SSC (X)-

Vidhya Vardhini High School, Udgir Marks : 85.09%

Accomplishments

- Won 'Best outgoing student' in SSC as well as Engineering for my exemplary performance.
- Won 'KUDOS' at Atos-Syntel for quality of work and my ability to understand issues and fix them quickly with zero defects.
- Research and survey paper "A Comparative Study and Analysis of Approaches towards Agriculture Supervision" accepted at the ICSPICC2016 conference under the IEEE Bombay section and published at IEEE Xplore

