

## Anurag Agarwal

[anuragagarwal712@gmail.com](mailto:anuragagarwal712@gmail.com)

+91-7248804947

+91-7017094257

### WORK EXPERIENCE

- Overall, 3+ years of IT experience around 2.5 Years into Big Data Technology.
- Attaining a daily scrum call with team members and discussing the task on a particular sprint
- In every sprint we have assigned tasks that we are completing as per timeline
- Developed PySpark programs and applied various transformations as per requirement in glue jobs
- Creating automation data pipeline using aws services like s3, lambda, glue, step functions for data processing and to make data ready for data scientist
- Weekly call with client for their requirements and further enhancement
- Job Monitoring
- Import data from s3 raw bucket and execute lambda job with the help of trigger and developed glue jobs for data cleaning and data masking
- Build an automation tool that was further used to validate source and target data
- Work with the teammates to understand and support them in importing data from source to s3
- Injected and processed large amount of data from various structured and semi-structured sources into s3 (AWS Cloud).
- Have good knowledge in Hadoop distribution like Cloudera.
- Have good knowledge in transferring data from RDBMS to HDFS and Hive tables using PySpark API's.
- Have good knowledge in creating tables, partitioning, bucketing, loading in Hive
- Ability to work effectively with associate at all levels within the organization

## TECHNOLOGY SKILLS

<b>Operating System:</b>	Windows, Unix.
<b>Hadoop Ecosystems:</b>	HDFS, MapReduce, Hive, YARN
<b>Framework:</b>	PySpark, Hadoop
<b>IDE:</b>	PyCharm, Jupyter Notebook
<b>Cloud platforms:</b>	Amazon Web Services (AWS)
<b>Programming Language:</b>	Python and Java
<b>Relational Database:</b>	Oracle
<b>Ticketing Tool:</b>	JIRA
<b>Version Control Tool:</b>	GIT & GITHUB
<b>SDLC Methodology:</b>	Agile

## ORGANIZATIONAL SCAN

**Organization:** Zielotech Software Pvt. Ltd.

**Designation:** Data Engineer

**Duration:** July 2019 - Present

## PROFESSIONAL EXPERIENCE

**Project #1: Bank Data Analysis**

**Duration:** March-2021 - Present

### Roles & Responsibilities:

- Build automated data pipeline for data scientist to save their time and smoothly working with data scientist to create **ECR image** and to get desired output using aws **sagemaker and step functions services**.
- Importing batch data in s3 buckets of aws and creating backup of that data as per dates using **lambda function**.
- Responsible for applying data cleaning techniques to correct and remove in-accurate and corrupted data, to improve data quality which significantly improves productivity.
- Applying various **PySpark transformation** to cleaned and validate data in glue jobs.
- Writing desired output in **s3** bucket
- Filtered data and performed multiple operations in-order to get meaningful insides that help our client in decision making and analysis.
- Developed **Glue jobs** as per new requirements
- Building **Docker image** in **sagemaker** for required and requested libraries for data scientist based upon code

## Project #2: Health care data Analysis

Duration: May-2020 to Mar-2021

### Roles & Responsibilities:

- Importing data in s3 buckets created and take date wise backup of that data file.
- Developed lambda triggers jobs for source upstream data so that specific glue job will automatically started for processing.
- Developed Glue jobs to load data from data file to staging database table.
- Applying various **pyspark** transformation to cleaned and validate data in glue jobs.
- Applying **spark SQL** to segregate data in master database table and error database table which is valid and invalid as per constraints respectively.
- Developed Glue jobs as per client requirements and updating values in table as per action taken using **spark SQL**
- Required tables storing into **RDS database system**
- Testing code and finalize the job.
- Exporting output in RDS table which is extracted as per business logic and storing it in **S3** location.

### EDUCATION

Master of Technology in Computer Science & Engineering | Percentage 81% | College of Engineering Roorkee (COER) Roorkee | December '2017

### DECLARATION

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

**Anurag Agarwal**