**SUBHAM KUMAR SAHOO Data Engineer**

**E-mail:** sksplay7@gmail.com | **Ph:** +91 7682025131 | **Location:** Bhubaneswar, Odisha

**TECHNICAL SKILLS**

**Programming languages:** Python, Shell scripting

**Databases and Data warehousing:** MySQL, DB2, Data Warehousing

**Big data frameworks:** Spark

**Cloud:** AWS Lambda, S3, Glue, Athena, IAM, RDS, DMS, API Gateway, EC2

**Web development:** Flask, HTML, CSS

**Source control and documentation:** Git, Github actions, Jira, Confluence

**WORK EXPERIENCE**

**Principal Global Services – Data Engineer Nov 2020 - Present**

**Hadoop divestiture**

* Crafted **AWS** pipeline to query data from **Athena** catalogue in another account using **Lambda function** and get the data daily as files into **S3 bucket**.
* Built process to **FTP** files to on-premise data servers to be used by Informatica workflows.
* Built **Athena** tables for these data using **Glue crawlers** to enable direct access to business stakeholder. Deployed pipeline using **AWS CDK**.
* Worked on modifying **Informatica workflows** to load these data into **Data Warehouse**.

**PCS data acquisition**

* Worked on requirement gathering and creating data dictionaries.
* Worked on **development, QA testing, UAT and scheduling** of Informatica workflows to load operational data to Domain and DW tables using full compare and delta load pattern.
* Created data validation scripts using **python** which reduced the time and efforts required, making the validation process **3X faster**.
* Created automatic code generation **Python** script to create the workflows without any manual effort reducing time required by **15X.**
* **Built shell script** to send automated email to stakeholders about load statistics of multiple workflows, **saving around 1-hour** efforts per day and ensuring proper data accuracy.

**Genesys cloud**

* Worked on pipeline to get data from Genesys API using python scripts hosted on **AWS EC2** instance. Scheduled the script using **cron** job.
* Created **Informatica workflows** to load DW tables from files with different load patterns.
* Researched on any **Q&A** from stakeholders and provided **production support**.

**AWS Lambda function deployment using container – POC**

* Containerized and deployed lambda function code to AWS using docker, ECR.

**OTHER PROJECTS (Blog -** https://subham-sahoo.medium.com/**)**

**Change Data Capture – Continuous Replication on AWS**

* Created pipeline to continuously capture any data changes in **RDS** MySQL database table using **DMS** and create files with update information in **S3** bucket.
* Used **lambda** function to trigger **Glue** jobs (with **PySpark** code) to read these files and create final files with updated changes.

**EDUCATION**

**Indira Gandhi Institute of technology**  **2016-2020**

Bachelor of Technology (Electrical): 8.29 CGPA

**ACHIEVEMENTS**

* 2 spot awards at my organization for good collaboration and performance.
* 3 gold medals in Wushu state championships and represented Odisha in Wushu Nationals 2019.
* AIR-1 National Engineering Olympiad 3.0.