Prateek Smith Patra

Cloud Data Engineer

Being a Cloud Data Engineer i have ability to work with unstructured data coming from various Big Data sources, to make decisions through technology of AWS, Data Science, Big Data & Spark tool's for the organizations and Data Integration Using ETL tools.

https://leetcode.com/prateek_smith/

✉ iamprateeksmith@gmail.com

🔲 linkedin.com/in/prateek-smith-patra-76a3031b5

📱 +91-9337208588

⭘ github.com/prateeksmith99

WORK EXPERIENCE

Big Data Integration Full Stack Internship
Cognizant

02/2022 - 07/2022   *Bengaluru, Karnataka, India*

- Developed and maintained large scale Data Warehousing and Data Lake layers.
- Develop, construct, test and maintain architecture and large data processing and ETL pipelines Using Informatica PowerCenter & IICS.
- Build a Synchronization task, Replication task, setting up External ID's in Salesforce, Upserting Data & Creating a mapping configuration task by adding parameters to a mapping.
- Build a trusted data pipeline with a modernized ETL tool on a cloud-native insight platform with IBM DataStage.
- Transformed huge volumes of data efficiently using Abinitio ETL Software Tool.
- Data Integration Using Talend.
- Worked on Data engineering pipelines using Apache Airflow.

Data Scientist Internship
Technocolabs Softwares Pvt. Ltd.

09/2021 - 11/2021   *Indore, Madhya Pradesh, India*

- **Project Team Lead Work:** Build a machine learning model that will predict if a user will skip a song or not based on the given information about the user's previous actions during a listening session along with acoustic features of the previous songs.

Computer Vision Developer Internship
iNeuron.ai

07/2021 - 12/2021   *Bengaluru, Karnataka, India*

- Developed an application using Image Processing, Deep Learning & Computer Vision as an Intern with iNeuron for Intelligent Radiologist Assistant (Knee MRI) project under Computer Vision Developer Technology.

EDUCATION

BTECH - Computer Science (CSE)
Silicon Institute of Technology (SIT)

05/2018 - 06/2022, 8.4%   *BBSR, ODISHA, INDIA*

HIGHER SECONDARY – Science
DAV Public School KNG

05/2017 - 06/2018, 67.87%   *BBSR, ODISHA, INDIA*

SECONDARY
DAV Public School KNG

05/2015 - 06/2016, 84%   *BBSR, ODISHA, INDIA*

SKILL'S

- **Programming Language -** PYTHON, SCALA, JAVA, C
- **Big Data with Hadoop and Spark -** HDFS, ZooKeeper, Hive, HBase, NoSQL, Oozie, Flume, Sqoop, Spark & PySpark, PySpark RDD, Spark Streaming, PySpark SQL, Kafka, SparkR, SparkSQL, MLlib, and GraphX
- **ETL Tools -** Informatica PowerCenter, IICS, DataStage, Talend, Abinitio
- **DevOps Tools -** Docker, Kubernetes, Ansible, Terraform, Jenkins, AWS Elastic Beanstalk, GKE, Git & GitHub, Linux, UNIX & Shell Scripting, XML & JSON
- **Data Science & Web Framework -** ML, DL, CV, NLP | Django, Flask, Flasgger, StreamLit
- **Cloud, Visualization & Database Technology -** **AWS**, Azure, GCP, Salesforce, Snowflake, Databricks, Postman (API) | PowerBI, Tableau, Apache Splunk | Oracle, MySQL, PL/SQL, NoSQL (MongoDB)

ACHIEVEMENT'S

Google Cloud Certificates
https://www.qwiklabs.com/public_profiles/6db0ffde-6dd4-4c28-8b21-f375bb3cb34e

Microsoft Certified: Azure Data Fundamentals
https://www.credly.com/badges/d19f7893-d18c-42b1-a7f8-d47bdef896f5

Microsoft Certified: Azure AI Fundamentals
https://www.credly.com/badges/c9a38420-256c-436f-86f3-76405582d8b7

PROJECT'S ACCOMPLISHED

1 - Build an ETL service pipeline to load data incrementally from Amazon S3 to Amazon Redshift using AWS Glue

2 - Health Care Data Warehouse System using Informatica PowerCenter ETL Tool

The purpose of the software requirements document is to systematically capture requirements for the project and the system "Blue Care Health Plan Inc." to be developed. The primary objective of the project is to create a database with required data validations as mentioned. The database is in turn used for generating welcome letters in xml format for each subscriber (This module is to analyze the Subscriber information provided, transformed and load them into the health care Data Warehouse) & Implemented SCD Type 1, Type 2 and Type 6.

3 - Retail Data Warehouse ETL Using Informatica PowerCenter, IICS, Talend, IBM DataStage & Abinitio

- Performed Expression, Router, Filter, Aggregator, Joiner, Rank, Sequence, and Lookup Transformation before loading data into the target location (Flat Files/Oracle Database/Data WareHouse/Salesforce CRM Object's) & performed Synchronization task & Replication task by setting up External ID's in Salesforce using IICS Data Integration Service's.

4 - Anti-Spoofing Face Recognition System using OAK-D and DepthAI

5 - Real Time Facial Emotion Detection Application Using CNN

6 - Fashion Recommendation System Using Content-based filtering

7 – Imported data from various data sources and created a DWH using Talend

8 - Processing NYSE (New York Stock Exchange) data using Hive

9 - Generate movie recommendations using Spark MLlib

10 - Churn the logs of NASA Kennedy Space Center WWW server usingSpark

11 - Build real-time analytics dashboard for an e-commerce companyusing Apache Spark, Kafka, Spark Streaming, Node.js, Socket.IO and Highcharts

LANGUAGE'S

English – Full Professional Proficient
Hindi – Full Professional Proficient