|  |  |  |
| --- | --- | --- |
| TM Objective ***Versatile, motivational, client-facing, high-energy professional*** *with expertise in python, databases and various open source technologies holding experience with an organization of repute in* ***the IT industry.*** Skills  * PySpark, Python, Informatica * AWS Glue, EC2, S3, AWS Neptune, Step functions, Lambda, GCP services * Classification and regression algorithms, NLP, Modal optimization * SQL, S3, Athena, Redshift, AWS RDS, Graph Database * Kafka, HDFS, Sqoop, Hive * Windows, Linux * Git, TFS * Jupyter Notebook, Pycharm * ServiceNow, Jira * Agile | |  | | --- | | TANMAI MEHROTRADATA ENGINEER | Aws architect(+91) 9899972825 | [*tm81739@gmail.com*](file:///C:\Users\ya97\Documents\tm81739@gmail.com)[*www.linkedin.com/in/tanmai-mehrotra-123205a2/*](http://www.linkedin.com/in/tanmai-mehrotra-123205a2/) | <https://github.com/GitTanmai> |  LEAD COnSULTANT • ITC INfotech • JUL 2020 - PRESENTWorking in full capacity with Sun Life Financial as Data Engineer.Design and Develop ETL Processes in AWS Glue to migrate data from external sources like S3, Parquet/Text Files into AWS Redshift, AWS RDS and SQL server.Developed POC on resume ranking using ML classification algorithm and NLP techniques which would help HR team to a perform initial level of resume screening.Created POC for analyzing logs data using AWS Neptune(Graph database)  * **Tools/Technologies Used**: AWS Glue, Pyspark, Python, Step function, AWS Neptune, Redshift, Athena, RDS, Bitbucket  Associate • CIANS ANALYTICS • JUL 2019 – JUL 2020  * Developed Ingestion, Validation and Monitoring Framework using Python/PostgreSQL for generating rules, validating those and monitoring the file before uploading them into the system. * Handling complex json and performing data cleansing activities. * Working on ETL pipeline creation using pyspark from multiple data source like S3, Redshift, FTP servers and other cloud platform. * Simultaneously worked on multiple small projects on regression and classification in ML. * **Tools/Technologies Used**: Pyspark, Python, Linux, AWS, Git, Postgres, Kafka, Jupyter Notebook.  SySTEM ENGINEER • TATA CONSULTANCY SERVICES • JUL 2016 – JUL 2019  * Designing and creating ETL pipeline for well known pharmaceutical company Eli Lilly. * Created workflow using informatica to maintain the Enhanced Clinical Trial System data in data warehouse. * Major role in migration of Informatica 9.x to 10.1. * Debugging complex informatica mapping to resolve. * POC to import data from AWS S3 into Spark RDD and performed transformation and action on RDD’s. * Worked and learned great deal from AWS Cloud services like EC2, S3, VPC, Lambda, EBS. * Used SQL to perform to perform backend testing on database * **Tools/Technologies Used**: Informatica, Python, PySpark, AWS, Postgres, Oracle SQL, TFS, HP ALM QC, Crystal Reports  EducationB. TECH • JUNE 2016 • UPTU/SRMS ENGINEERING COLLEGE  * Got 73% Aggregate * Sentiment analysis of a live social media data stream. (Minor Project) - Using Hadoop, hive, python. * IT Club Website. (Major Project/Internship) and ERP system using Java, Oracle DB, HTML, CSS, JavaScript.  CERTIFICATES  * AWS Certified Solution Architect * Informatica Power Center Developer (Edureka Certified) * Big Data Hadoop Certification (Edureka Certified |