**CHAITRA TADAGA**

[chaitradataengineer1@gmail.com](mailto:chaitradataengineer1@gmail.com) (469) 839-3924 Milford, CT

# DATA SCIENTIST/DATA ENGINEER

**Summary**

Data Scientist with 5+ years of experience in building data intensive application, tracking challenging

architectural and scalability problems, collecting and storing data. Well-acquainted with Big-Data, Data Warehouses, Analytics, Machine Learning models and Wrangling.

# SKILLS

* Python ● TensorFlow ● Hive
* SQL/NoSQL ● Scala ● Data Analysis
* Amazon Web Services ● PySpark ● GitHub
* ETL ● Hadoop ● Agile/Scrum
* Machine Learning ● Tableau ● Java

# WORK EXPERIENCE

**Radiance Technologies** – Data Scientist

August 2022 – Present Palo alto, CA

* Improved data quality & performance by 40% by analysing, verifying and modifying SAS & Python scripts.
* Gathered, organized & analysed 2016, 2017, 2018 marketing data using Python, Hive, SAS, ML
* Created and presented models for global sales revenue using the above mentioned, which resulted in a 50% increase in domestic sales.
* Was able to discover that on average people are 25% more likely to buy fish on rainy days.
* The marketing team increased fish sales by 40% using this information.
* Create a report and dashboard in Tableau using Pivot tables and chartings.

**VISA** - Data Engineer

March 2022 – August 2022 Palo alto, CA

* Designed and implemented a real-time data pipeline to process semi-structured data by integrating millions of records from multiple data sources using PySpark.
* Led the migration from SQL to Postgres SQL using Tusker migration tool which increased the performance by 14%.
* Involved in analysis of end user requirements and business rules based on given documentation and worked closely with tech leads and Business analytics in understanding the current system.
* Responsible for generating actionable insights from complex data to drive real business results for various applications teams.
* Also involved in to coordinate the stakeholders with the detailed progress report of the product. Report was built using the Tableau.

**Agile data Inc** - Data Scientist

June 2021 – December 2021 Katy, TX

* Worked on **Spark** improving the performance and optimization of the existing algorithms in Hadoop using Spark Context, **Spark-SQL**, Data Frame, Pair **RDD**'s
* Applied a combination of SVC, Logistic Regression, Linear SVC, XGBoost and Logistic Regression CV model on a large customer feedback text dataset anddeveloped an LSTM and BERT deep learning model for topic classification.
* Worked on Ingesting data by going through cleansing and transformations and leveraging **AWS Lambda**, **AWS Glue** and Step Functions.
* Implemented usage of Amazon **EMR** for processing Big Data across a Hadoop Cluster of virtual servers on Amazon Elastic Compute Cloud (**EC2**) and Amazon Simple Storage Service (S3).
* Developed a python script to hit **REST API**’s and extract data to AWS **S3.**
* Conducted **ETL** Data Integration, Cleansing, and Transformations using **AWS Spark script.**
* Deployed the project on Amazon **EMR** with S3 connectivity for setting a backup storage.
* Ensured code is of the highest quality throughout the SWDLC (Software Development Life Cycle) from start to ﬁnish.
* Worked on **AWS Data Pipeline** to configure data loads from **S3** to into **Redshift.**
* Involved in creating **Jenkins** jobs for **CI/CD** using git, Maven and Bash scripting.

**SM search system LLP** – Data engineer

October 2015 – December 2020 Bangalore, IN

* Experienced in loading the raw data into **Spark Data Frames** and validate the data.
* Experienced in converting the **validated RDDs into Data frames** for further processing.
* Implemented the **Spark SQL code** logic to join multiple data frames to generate application specific aggregated results.
* Experienced in fine tuning the jobs for better performance in the **production cluster space**.
* Imported data from **AWS S3** into **Spark RDD,** Performed transformations and actions on **RDDs**.
* Designed appropriate **partitioning/bucketing schema** to allow faster data retrieval during analysis.
* Continuous monitoring and managing the **Hadoop cluster** through **Cloudera Manager.**
* Experience working with various File-Formats (Parquet, Avro, dat & JSON) and Compressions (Gzip).
* Responsible for generating actionable insights from complex data to drive real business results for various applications teams and worked in **Agile Methodology** projects extensively.
* Worked with **SCRUM** team in delivering agreed user stories on time for every Sprint.
* Implemented **UNIX scripts** to define the use case workflow and to process the data files and **automate the jobs**.

# EDUCATION

* Master of Science in Data Science January 2020 – December 2021 University of New Haven CT, USA
* Bachelor of Engineering in Computer Science Engineering August 2011 – May 2015 APS College of Engineering Bangalore, India