Kesav Kumar J S

Software Engineer – Big Data engineering, Data Analytics, Data Science

Email: jskesavkumar@gmail.com **|Mobile**: +918870024013

LinkedIn: https://www.linkedin.com/in/kesav-kumar-5ba674150/ **|GitHub**: https://github.com/kesavsivakumar

Kaggle: https://www.kaggle.com/kesavsivakumar |**Dockship**: https://dockship.io/author/kesavs-294

Big Data Engineer, a Google cloud practitioner and a **data science enthusiast** with **3+ years** of experience in developing ETL applications using pyspark &hive, optimizing query performance, and providing technology support as a consultant in running data intensive applications following Agile methodologies.

Skills

Backend	**Languages**: Python, shell scripting	**Frameworks**: Pyspark, Pandas, Keras
Databases	**RDBMS**: MySQL , Big Query	**No SQL and Big data** : HiveDB, Teradata, IBM- cloudant ,Big table
Cloud Technologies	Google cloud (GC) compute engine , GC data proc, GC data flow, GC Big query, GC cloud SQL,GC cloud storage , IBM node red, IBM Watson	
Version Control	Git	
Deployment Tools	Jenkins-CI/CD ,XL release, Docker	
Project Management	Rally, JIRA -Agile Methodology	
Fundamental Concepts	Big Data architecture, Data Structure & Algorithms, Data analytics, Design Principles, System Design, Machine & deep learning	

Professional Experience

Big Data Engineer, Tata Consultancy Services(Bengaluru) |JULY 2021–Present

Enhancements & Optimization

- Improved the through put of ETL application by identifying poorly partitioned table and repartitioning them.
- Built a small python feature for parsing HQL (hive query language) files to extract the queries and trigger them via spark instead of conventional MapR jobs – published this story internally for which I was recognized as Contextual Masters
- Worked on a POC to main the impacting database tables during migration of ETL application from Teradata to Cornerstone Big Data. Created an utility to parse/scan the project repository and identify the impacted tables

Security & Authentication

- Briefly worked on an utility to encrypt(PGP encryption) & transfer the extracts via Secure file transfer system

Scalability & Fault tolerance

- Built custom pipeline to compartmentalize long running jobs into separate small job. The ETL framework could run 3x resource intensive jobs (from 100 to 400) at a time using the pipeline
- Developed a alert system (pyspark utility) to identify the long running jobs in the spark resource queue and re-run jobs which got terminated due to resource exhaustion.

Leadership & Support

- Facilitated technical discussions/ meetings regarding the migration of project from on premise to google cloud
- Provided technical support to team members in fixing bugs and closing L3 and L2 incidents

Data Analyst , Tata Consultancy Services (Chennai)|Mar 2021–May 2021

- Analyzed Payment UIs to identify the key features and eliminate unwanted & less of use data to improve data collection and increase customer satisfaction.

Education

- **B.Tech. Computer Science and Engineering ,** SASTRA Deemed to be University . CGPA -7.77/10.00
  2017–2021

Certifications

- NPTEL Certification – Programming in python
- NPTEL Certification – Introduction to programming in C
- SQL for Data Science- University of California, Davis – 2020
- Data Science Foundations - level 1 – issued by IBM
- Deep Learning Specialization – issued by deeplearning.ai
- Machine Learning with Python – issued by Cognitive class AI
- Architecting with Google Cloud Platform Specialization

Publications, Projects & Blogs

- **Stacked Dark COVID-Net:** A Multi-class Multi-label Classification Approach for Diagnosing COVID-19 Using Chest X-Ray Images, published in International Conference on Recent Trends in Image Processing and Pattern Recognition
  DOI : doi.org/10.1007/978-3-031-07005-1_7
- **Forecasting power output of wind turbines-IBM Hack Challenge – JULY 2020.** A Long Short Term Memory (LSTM) neural network was built on keras for predicting the wind speed and wind direction .The neural network was deployed with a flask web app using heroku PaaS. https://github.com/kesavsivakumar/SBSPS-Challenge-953-Predicting-the-energy-output-of-wind-turbine-based-on-weather-condition-
- **Identifying the optimal entropy parameters using Farmland fertility algorithm (FFA) and** employing a suitable deep learning/machine learning model to classify ictal & non ictal seizure
  https://github.com/kesavsivakumar/optimal_entropy_paramters_for_EEG_signal_classification-FarmLandFertility_algorithm-
  https://www.kaggle.com/code/kesavsivakumar/wavlet-decomp-entropy-features-of-coefficients
- **Introduction to PySpark : a quintessential tool for BigData analytics -**
  https://dockship.io/articles/61115ec0f1b0710d0f928521/introduction-to-pyspark-:-a-quintessential-tool-for-bigdata-analytics-
- **Batch Renormalization in Image Denoising** https://dockship.io/articles/5fe2f8a25a4490141dab0b31/batch-renormalization-in-image-denoising-