Adwiteey Mauriya

https://amauriya.com
+351 927 728 594
send@amauriya.com

amauriya.com

ABOUT ME

A former computational physicist who enjoys working with data, familiar with gathering, cleaning and organizing data for use by technical and non-technical personnel. Advanced understanding of statistical, quantitative and other analytical techniques. Detail-oriented, methodical and enterprising with strong focus on devising and running effective processes. Result oriented, data driven and a strong desire to expand the current skill-set with a preference for working in a fast- paced environment. Successfully led projects in academia and industry from a vague idea to scientific publications independently as well as with a team.

EDUCATION

- **Ph.D.** 2015 - 2020 in **Technological Physics Engineering** at Instituto Superior Tecnico, Lisbon, Portugal.
- **B.S.-M.S Dual degree,** 2009 - 2014 with **Physics Majors** at Indian Institute of Science Education Research, Pune, India.

TECHNICAL SKILLS

Advanced: Python, SQL, Scala, Fortran, Bash

Intermediate: R, Matlab, C, C++, Kotlin

Elementary: Ruby

SOFT SKILLS

Curious, life long learner, team player.

Experienced in presenting results to expert and non-expert audiences in formal and informal settings.

Languages: English(Fluent), Hindi(Native), Portuguese(A2)

TECHNOLOGIES

ETL Tools: Apache Hadoop, Apache Spark, HDFS, Apache Airflow, Kubernetes, Docker, MPI, MapReduce

Analytics Tools: Scikit-learn, Keras, Tensorflow, Tableau, SRSS, PowerBI, Databricks, Zeppelin Notebook

Computing Platforms: Snowflake, Google Cloud Platform, AWS, Hadoop, Supercomputers

IDEs and source code management: IntellijIdea, Visual Studio, Git, SVN

Project management tools: Sharepoint, Jira, Microsoft Excel, Confluence

KEY ACHIEVEMENTS

- Top 1% in most difficult engineering exam in India (IIT-JEE) 2009.
- Received competitive **KVPY** fellowship in 2009(for 5 years), funded by Dept. of Science and Tech and Government of India.
- Received prestigious **FCT** fellowship (includes tuition fees + living expenses for 4 years) and Eurofusion research grants for participating in experiments and traveling to multiple laboratories in EU.
- Invited talk at Max Planck Institute for Plasma Physics, Garching, Germany

WORK EXPERIENCE

Data Engineer, Tripadvisor.Inc

11-2020-present

Currently maintaining and monitoring the datasets stored in multiple data lakes and warehouses (e.g. GCP, Snowflake and Hadoop clusters) of customer data platform and enterprise data teams by optimising the currently existing ETL and creating new if it is required. ETL saves cost of storage and processing of more than 10 millions USD annually by the entire team. The constructed datasets are in BigQuery, Hive and Snowflake which involved redesigning of warehouses, building data pipelines and optimising the underlying queries, accommodating the new data model, and liaising with stake holders. Improvements and performance was assessed in a form of dashboards and charts to have a bird eye view. I have created series of automated scripts to monitor ETL of various business segments in the company. I advocate the code review standards to have clean clear and consistent business data. I have provided support for onboarding of new data engineers in the team and the company.

Business Data Analyst (FreeLancer), NetJets.Inc 10-2019-07-2020

Achieved 80% success in pilot training warehouse automation by building multiple reports and liaised with SMEs to identify discrepancies for troubleshooting and corrective action. We planned and organized workshop, producing 1.1 times the goal of qualified leads. Synthesized current business intelligence data to produce SRSS and Tableau reports and polished dashboards, highlighting findings and recommending changes using scrum framework and Kanban method. Developed flight school KPI metrics derived from raw company data to track improvements in gross margin of training department of the company.

Independent Project (Disruptions mitigation modeling in Tokamak using Neural Networks) 06-2019-12-2019

The purpose of this project was to create a tool/work-flow to predict the disruptions in Tokamak (Future Nuclear Fusion reactor) using neural networks by processing the measured experimental data. Tokamak has more than 200 probes and it produces few Terabytes of data with every experiment. Experimental data is stored in a noSQL database. Data is stored in binary format and can be accessed using MDSPlus server. I set up PySpark, Apache Spark and Hadoop in a workstation. Created a python library to create RDD from the NETCDF data. I used the ML library using PySpark and also run tensorflow on a Spark clustor.

Visiting Researcher at University of Oxford and Culham Center for Fusion Energy, U.K. 08-2016-09-2018

Developed numerical tools to study complex physical phenomena in nuclear fusion reactors. It required parallel programming in Fortran with MPI and Python. It is scaled on supercomputers which has CPUs and GPUs. Modeled the fusion experiments with an underlying theoretical model by developing the required numerical tools in Python. It required in-depth knowledge of mathematics, statistics, programming, and physics. Developed a Python framework to import fusion experiments data in multiphysics codes for various kinds of physics study. It also visualizes the output and fully modular.

Project Scientist at Institute for Plasma Research, India 06-2014-12-2014

I developed an algorithm and checked the time complexity(efficiency) of 1-D Grad-Shafranov equation to implement it in a parallel HPC code. Merged fix boundary equilibrium with free boundary method and patched it in a Fortran library which is integrated with a multiphysics code.

Linux System Administrator (Volunteer) at Soft matter simulation Laboratory, IISER Pune India 09-2012-04-2013

Provided IT support to a research group to resolve issues related with Linux operating systems, setting up scientific softwares, libraries and debugging High Performance Computing codes. I consistently coded multiple batch jobs in Bash and Python to automate tasks.

PROJECTS

- Python framework to access data from a national lab database.
- Twitter sentiment analysis using Pyspark
- Disruptions mitigation modeling of a Tokamak experimental using Neural Networks
- Workflow to analyse the experimental data and run programs in supercomputer.
- Python post processing library to visualize turbulence simulation data.
- Parallel code development for turbulence simulation study
- Fortran source code generator written in Python3.
- Python Tkinter GUI for parsing json
- Web scrapping using python

RELEVANT CERTIFICATIONS

- Machine learning (1 course, Coursera)
- Data Science using Python and SQL (7 courses, Coursera)
- Hadoop fundamentals and programming (5 courses, Coursera)
- Functional programming with Scala (3 courses, Coursera)
- Web design (2 courses, Free Code Camp)
- SRSS reporting (1 course, Udemy)