

Sumant
pratik@concept-inc.com
Salt Lake City, UT

Delivery manager transitioning into **Big data engineer / Cloud Data Analyst** role after completing MS in computer systems from Arizona state university in Dec 2022. **Certified GCP Architect** with experience in Data analysis, ML & big data processing

Skills

Big Data processing/visualization – Hadoop, HDFS, Spark, Hive, Sqoop, Tableau

Cloud – GCP Cloud Functions, pub / Sub, Dataproc, Composer, Datastore, Dialogflow, **AWS** - S3, Lambda, SQS, EMR, EC2

Databases – SQL, Oracle, SQL Server, Hive, Cassandra, Postgres, Mongo DB, Big Query, DynamoDB

Programming – Python, SQL, Java, VC++, C++, C, Javascript, Docker, Microservices, Git, Jenkins, Agile

ML Algorithms – Regression, Naive Bayes classifier, SVM, Decision Tree, Random Forest, Ensemble, KNN, Deep Learning, CNN, RNN, LSTM, K means, **Packages** – Numpy, Pandas, Scikit-learn, Keras, Tensorflow, PyTorch, Lightgbm, XGBoost, Matplotlib, Nltk

Certifications, Publications and Awards

- **Google Cloud Certified Professional Cloud Architect (2023), Udacity Nanodegree in Machine Learning (2018)**
- Cloud data pipeline automation framework 2020 - <https://doi.org/10.30534/ijatcse/2020/200942020>
- IBM Eminence & Excellence award (2014, 2011), Infosys Power Programmer award (2004, 2003), PMI 2010, CSQA 2003

Work Experience

Cloud Engineer, Arizona State University, Phoenix, Aug/2021 onwards (**AWS Cloud**)

- Implemented an image recognition service on AWS using **Python flask** Web server on **EC2**, **SQS queues** to communicate between web and app tier, App tier using **ML based image recognition** and Data tier using **S3**. The App tier used **AMI based launch template** which was **auto-scaled** based on CPU Utilisation.
- Redeveloped the same application using **AWS lambda** for **serverless autoscaling** and **DynamoDB**.
- **Assessed 50 teams**, for the above assignments, based on **demo** by project teams and **code review**.
- Assisted Cloud computing Course preparation, grading homework, quizzes, exams and address student concerns
- Prototyped escape room gaming based corporate training to improve effectiveness.

Sr. Cloud Data Engineer, MRO Corp., India, Sep/2018 - Apr/2020 (**GCP, AWS, Big data analysis, Healthcare**)

- Clients: American Academy of Ophthalmology, American College of Rheumatology, American Board of Family Medicine
- Launched a clinical quality measurement and reporting **SaaS** product on **GCP** that supports 120M patients.
- Redesigned the product in consultation with **Google** tech team at **Mountain View, CA** resulting in 40% savings.
- Extracted clinical elements from unstructured health reports of various EHRs using **Python regex-based parsers**.
- Created **Hive** data lake with performance optimization using **bucketing & partitioning** to store the patient visit records.
- Ingested data from **MySQL** to Hive using **Sqoop**. Transformed visit records to patient records and history using **Scala** programs on **Spark** and persisted the result in **Cassandra**.
- Developed provider quality (QPP) scorecards in **Scala** using **Spark SQL, Data Frames and Cassandra**.
- Investigated the accuracy of the scorecards by performing **data analysis** using **Google Big Query** and Python.
- Set up and Deployed **Docker containers** to break up monolithic app into **microservices**, improving developer workflow, increasing scalability.
- Executed **microservices** in **python** by **reading scorecard data from Cassandra** to be consumed by the front end.
- Worked on **Container management** using **Docker** by writing Docker files and set up the automated build on Docker HUB and installed and configured **Kubernetes cluster** for the microservices.
- Designed and published a [data pipeline automation framework](#) on GCP to achieve the objectives of Budget management, Data pipeline prioritization and Reuse. This resulted in **optimized job scheduling and 30% more revenue**.
- Utilized **various GCP services** like Google cloud function, Cloud scheduler, Cloud Pub/Sub, Buckets (Storage), Cloud Composer, Airflow, Big table, Docker, Cloud Dataflow, Cloud Dataproc and Service Accounts for the data pipeline automation framework.
- Wrote and maintained automated, fault tolerant multi-tier GCP environments spanning across multiple availability zones using **Terraform scripts**.
- Experience and knowledge of release management process, Git and Jenkins.

Delivery Manager, IBM, India, Aug/2006 - Dec/2017 (**Data Analysis, Management**)

- Clients: Nisource, USA, Air Canada, Vodafone Ireland, Vodafone Czech
- Optimized take off fuel load for a major airline using **regression analysis, Pandas, numpy, Scikit-learn**. This involved analysing 500 flight parameters and ensuring there is no increase in emergency landing requests.
- **Reduced incidents by 40%** and improved **SLA Adherence to 98%** for 900 people practice by analysing **metrics** using **IBM Cognos Analytics**.

- Provide historical trends using **IBM AMS Analytics Platform** for contract renewal and resolution. Participated in the team which was successful in **renewing the account contract** for 7 more years.
- Managed development and maintenance of a **Telecom domain** portfolio with diverse technologies on hybrid cloud - J2EE, VC++, .Net, Oracle ERP, Siebel CRM, Oracle Data warehouse, SQL, PL/SQL

Senior Java Developer, Infosys Limited, India, Oct/2000 - Nov/2005 (OOP, J2EE)

- Clients: The travel corporation, UK (Brands - Trafalgar Tours, Insight Vacations)
- Provided technical leadership for a centralized, web-based ERP solution on J2EE platform for a UK tours and travel client using **J2EE, Oracle 9i App Server, Oracle database, PL/SQL, JavaScript**
- Modelled **relational Data model** on Oracle Database to represent the business entities.
- Built Java **POJO data access** layer, Business logic objects and **Struts framework**-based frontend.

Founder Consultant, Eagle Software, Nottingham UK, Jan/2000 - Oct/2000 (OOP, VC++)

- Clients: Chronous Richardson, UK
- Delivered a **VC++ Model-View-Controller** controller to plan the production of batches in a rubber manufacturing plant.

Software Developer, Novasoft Technologies, Edinburgh UK, Jan/1999 - Dec/1999 (OOP, VC++)

- Clients: Sky TV UK
- C++ based software for user service configuration

Module Leader, Infosys Limited, India, Jan/1997 - Dec/1998 (OOP, C++)

- Clients: DelaRue, UK
- Coded a payments system and a cash handling device driver in C++ for retail client in UK

Recent Data Analysis Projects

- Participated in **Google Universal Image embedding 2022 Kaggle competition** and scored at par with **top 10 competitors**. The model can represent 17K type of objects from multiple domains (Location, Products) at instance level ("Statue of Liberty"). Designed a creative **Tensorflow** based **Deep Learning** model - **Open AI CLIP + Dropout + Dense(64-D) + Arcface + Softmax**
- Predicted **house prices using regression and ensemble algorithms** (GradientBoosting, LGBM, Ridge, Lightgbm, Voting, Stacking) in a **Kaggle Competition and scored within top 25%**. Made effective use of **data pre-processing techniques** such as Removing Outliers, Skew Adjustments, Natural order / categorical encoding, Feature engineering, Data Imputation, Data Standardisation.
- Forecasted stock prices by employing **LSTM and RNN with 96.5% accuracy**
- **CNN based** Object recognition, Customer segmentation with GaussianMixture clustering, PCA
- Predicted an individual's probability of involvement in shooting incident using a **SGDRegressor** based scoring system, with 95% accuracy. Projected crimes in Chicago using **Holt-Winters and Arima Time Series**.
- Authored an expository paper on implementing **differential privacy in deep learning**

Education

- Arizona State University, **Masters, Computer Systems**, 3.8/4, 08/21–12/22 (Calculus, Distribution, Cloud, Statistical ML, ANN)
- CDAC India, **Bachelors, Big data analysis**, Grade A, 2/18-8/18 (SQL, Hadoop, Hive, Pig, Scala, Python, Analytics, Tableau, ML)
- CDAC India, **Bachelors, Advanced Computing**, Grade A, 8/1996-1/1997
- Pune University India, **Bachelors, Electronics engineering**, Grade 1st class, 6/1990 – 6/1994