GitHub: github/dhananjaymehta
LinkedIn: dhananjaymehta

DHANANJAY MEHTA

(785) 340-5569
mehta.dhananjay28@gmail.com
San Francisco, California

EXPERIENCE

Data Engineer II — **Sojern Inc., CA** — **Jan 2019 - Apr 2020**

- Developed data pipeline on Google Cloud Platform to process batch and stream data using DataFlow, Pub/Sub, BigTable , Google Cloud Storage, and BigQuery
- Automated data pipelines with Apache Airflow for easy scheduling and deployment
- Improved data pipeline by implementing sanity checks using great_expectation, pandas, and BigQuery
- Reduced time for debugging and impact analysis by creating lineage graph for ETL jobs
- Created REST Api manager in python to integrate new demand-side platform(DSP), TheTradeDesk (TTD). This automated manual tasks like campaign setup, advertiser creation, and report scheduling by using TTD API
- Developed a data pipeline using Airflow, Google Cloud Storage, DataFlow, and BigQuery load ads performance and delivery click, conversion, impression) data reports from TTD
- Reduced complexity of accessing data from BigQuery by building a microservice using python, gRPC and ProtoBuf to provide ads performance data. Service used for automating ads campaigns, generating performance dashboards, etc.
- Optimized data quality and accessibility for SFDC data like campaigns, accounts, contacts by microservice using python, PostgreSQL, gRPC, Protobuf. This was used by the front-end, sales, and marketing teams

Data Engineer — **Hortonworks, Santa Clara, CA** — **Jun 2017 - Dec 2018**

- Developed a credit card fraud detection model using random forest regression to identify fraudulent payments. Scaled the model to handle 100B+ data points using AWS, Spark, Spark-ML, HDFS, Hive and hosted as REST endpoint on IBM DSX
- Automated deployment and maintenance of Hadoop cluster on AWS using python, shell scripts and Ambari
- Headed the integration of IBM DSX with HDP, this included HDP components such as Hive, Spark Livy, Zeppelin

Data Engineer (Fellow) — **Insight Data Science, Palo Alto, CA** — **Jan 2017 - May 2017**

- Built a web application “InVoiceSmart”, that analyzed purchase invoices for restaurants by creating purchase portfolio, generated purchase price spike alert, and predict purchase price for upcoming seasons
- The app was built on AWS using Kafka, Spark, Hive, S3, and Flask

Software Engineer — **Accenture, India** — **Jul 2012 – Jul 2015**

- Maintained the data pipeline for financial reporting for the client. Use shell script and JCL
- Maintained mainframe and IMS data warehouse and performed migration of data to IBM DB2
- Automated validation of migrated data using a shell script and stored procedures

EDUCATION

M.S. in Computer Science, Kansas State University, USA — **Aug 2015 - May 2017**
B.Tech. in Computer Science, Graphic Era University, India — **Aug 2008 - May 2012**

SKILLS

Languages	Python, SQL, Shell Scripting, Java
Databases	PostgreSQL, BigQuery, Hive, HBase, SqlAlchemy
Data Engineering	AWS, GCP, Spark, Hadoop, CloudSQL, Pub-Sub, Kafka, Elasticsearch
Tools	Git, Jenkins, Protobuf, Kubernetes, gRPC, great_expectations, Airflow
Other	Scikit-learn, Pandas, Jupyter, NLP, Flask, Hortonwork Data Platform

HONORS AND RECOGNITIONS

- **Founding president** of ACM Special Interest Group in Artificial Intelligence at Kansas State University(KSU)
- Nominated for **best teaching assistant** for the year 2015-16, only nominee from the college of engineering