

DATTA TELE

SENIOR DATA ENGINEER

CONTACT

datta.tele@gmail.com 
409-300-8045 
Austin, TX 
[LinkedIn](#) 
[Github](#) 

EDUCATION

MASTER OF SCIENCE
INDUSTRIAL ENGINEERING
Lamar University
Aug 2017
GPA: 3.73

SKILLS

- Programming: Python (Pandas, Scipy, NLTK), SQL, JavaScript, Jinja2, HTML, CSS, Familiar: Scala
- Databases and Cloud: MySQL, PostgreSQL, Redshift, MS-SQL Server, NoSQL, MongoDB, Elastic Search, GraphQL, AWS, AWS-lambda, S3, Google Cloud Platform, Digital Ocean, Azure
- Visualization: Plotly, Matplotlib, Seaborn, Bokeh, D3.js, leaflet.js, open-source visualization tools.
- ML Framework: Scikit-Learn, TextBlob, TensorFlow, Theano, EWMA, ARIMA, Quantopian
- Development Tools: Django, Flask, Dash, Fast-api, tornado, jQuery, Bootstraps, React.js, Familiar: Node.js
- Server & Deployment: Nginx, Docker, Docker-Machine, Kubernetes,
- Big Data Analytics Tools: Spark, Hive, Pig, Solr, Airflow, Dataflow, Tableau, Mode Analytics, Snowflake, Dremio, Salesforce, Google Analytics.

CAREER OBJECTIVE

• Senior Data Engineer with 4+ years of experience in data profiling, data mining, statistical analysis, natural Language processing, mapping, predictive modeling, and web application development. • Proficient in Python, SQL, database development, machine learning, and web programming. • Passionate in team building and believes in lean development with agile/scrum methodology.

WORK EXPERIENCE

SENIOR DATA ENGINEER

Vida Capital / Jul 2018 - current / Austin, Texas

- Designed and implemented a real-time data pipelines to process semi-structured data by integrating 40+ million unstructured raw data from 10+ data sources into Data Lake
- Created a cross platform file transfer software to empower internal users that automated data profiling, id-reconciliation, and performing data uploads from on-premise file server to S3 that saved over \$2,000,000+/year by reducing manual operation 10+ hours workload to 5 min per task
- Created internal python library used to parse and reformat data from external vendors and clients reducing the error rate in the data pipeline by 14
- Designed architecture and built centralized ID system algorithm to interconnect micro-services with vendors data sources to identify unique linkage between unstructured data format that reduced cost \$1,200,000+ by avoiding duplicate request calls with vendors services
- Built secure and automated robust Data Migration service from on-premise database (MSSQL-Server) to internal databases (Postgres RDS) for daily transactions update of millions of rows of data resulting in an annual cost savings of over \$480,000
- Created ELT service to store processed data in various data format from business portals and CRM platforms into columnar parquet format in S3, relational analytics data in Postgres and Redshift, and unstructured data format in MongoDB that saved over \$8,000 per month cloud service cost
- Architected and developed analytical application platform to support multi-billion-dollar business with day-to-day operations by reducing inefficiency with minimum supervisions and automated self-resolution any uncritical level issues
- Managed and partnered with client data science team to support \$4,000,000/year predictive analytical project with planning, development and proactive manner introduced frequent improvements to predictive model
- Orchestrated distributed workload architecture to support concurrent REST-API calls with up-time of 99.9% to process high volumes data loads and user activity using tornado framework, Apache Arrow and PySpark

- Optimized SQL queries that resulted in reduction in cron job task schedule from every 12 hours to every 30 min with reduction for data write operation
- Led a team of 4 developers, 2 data scientist, 2 analyst and over 10 contractors

DATA MANAGER

Vida Capital / Mar 2018 - Jun 2018 / Austin, Texas

- Architected and built cloud infrastructure from replacing traditional on-premise MC-Access DB and file sharing servers into scalable AWS based environment with monthly operational cost saving of over \$12,000 dollars
- Developed centralized platform for data processing, application development, and reporting which reduced manual workload by 33% monthly

RESEARCH ASSISTANT

Lamar University / Feb 2016 - Jan 2018 / Beaumont, Texas

- Created a database from 1+ million unstructured datasets and 100+ features by conceptualizing safety records and PPE standard in industries
- Built Multinomial Naive Bayes Classifiers with performance rate of around 85% to categorize responsible causes for accidents and safety requirements in work environments
- Optimized performance timing for training model from 12 hours to 1-2 hours
- Improved an overall accuracy of the predictive model by 30.7
- Identified 84% classification errors in data collection process with regression analysis and proposed recommendations to American Bureau of Shipping (ABS)

DATA ANALYST

Walks / Nov 2017 - Jan 2018 / Austin, Texas

- Defined business KPI metrics with a diverse group of stakeholders and identified data processing gaps by saving overall 20+ manual hours per month
- Enhanced data collection process in sales and ticketing process from different vendor channels with data augmentation that resulted in monthly +40 hours workload reduction
- Introduced unique ID hashing technique that resolved data integrity process issues across stakeholders and reduced error rate by 3
- Developed reporting dashboards using Tableau to conducted analysis of travel expenses trend with respect location and seasons that impacted optimization per tour guide cost over \$2,000
- Designed data web scraper functions to extract travel trends and tourism related conversation in social platform from consumers that increased customer engagement by over 20
- Orchestrated sentiment analysis (NLP) on 11,000+ traveler feedback and developed user stories for customer flow