UTTASARGA SINGH

Hartford, CT | (513) 394-9500 | uttasargasingh9067@gmail.com | linkedin.com/in/uttasargasingh | github.com/uttasarga9067

TECHNICAL SKILLS AND CERTIFICATIONS

Programming languages: Python, PySpark, Scala, SparkR, Spark SQL, R, Java, HTML, CSS, JSON, JavaScript, Unix/Linux Shell Scripts, C/C++.

Applications: Microsoft Azure, Azure Databricks, Apache Hadoop (HDFS, MapReduce), Apache Spark, Docker, Kubernetes, Apache Airflow, Dagster, Git, Amazon Redshift, Amazon S3, Amazon SQS, Amazon EC2, Quartz Scheduler, Oracle E-Business Suite 12.2.5, Oracle WebLogic 10.3.6, Oracle Cloud Fusion Accounting Hub Reporting Service, Tableau.

Database Systems: Oracle Enterprise Edition-12c, MySQL, Oracle DB, NoSQL, PostgreSQL.

Statistical Applications: Bayesian Inferences, R (Markov Chain Monte Carlo Sampling [MCMC], Just Another Gibbs Sampler [JAGS], Bayesian Average Sampling [BAS], Sparklyr, dplyr, ggplot2, tidyr), Python (Pandas, Matplotlib, NumPy, Scikit-learn, Seaborn, NetworkX, PyMC3).

Certifications: Databricks Academy Learning Paths, Azure Fundamentals AZ-900, OC-1Z0-1072-20.

EDUCATION

Master of Engineering, Computer Science, *University of Cincinnati* CGPA: 3.8/4.00 **Aug 2019 – Dec 2020**

Bachelor of Technology, Computer Engineering, *NMIMS's MPSTME* GPA: 3.02/4.00 **Aug 2012 – May 2016**

PROFESSIONAL EXPERIENCE

1. **Hartford Steam Boiler IIC –** *Hartford, CT, United States of America*

Data Engineer Associate **Aug 2020 – Present**

- Responsible for creating high-quality modeling datasets for Business Insights and Analytics Team, by developing and maintaining Big Data ETL pipelines in Azure Data Factory and Azure Databricks. **(PySpark, Scala, Spark SQL)**
- Develop modeling datasets by combining internal and external data assets from On-Prem Hosted SQL Servers, thus enabling rapid prototyping of ideas for feedback using Python.
- Implementing Data/ML Solutions on complex, high-volume, high-dimensionality data from various third-party data sources. (**ATTOM**, **HSPSL, MS FOOTPRINT)**
- Established a directed approach for Analytical Model Deployment with Hadoop based technologies, meanwhile documenting results using strong analytical and reasoning capabilities and communicating the work performed and results achieved at weekly sprint meeting through data storytelling and visualizations.
- Collaborating with Data Hunting Team for developing scalable data pipelines and automating the creation of Data profiling Documents, which is then used by Underwriting Team for further quotations of Deductibles of an Insurance Policy.
- Designed and engineered Data Engineering Projects independently, using wide variety of tools and software available for implementing solutions.
- Responsible for operational scoring, support, and maintenance of Farm EB (Equipment Breakdown), a machine learning model predicting the costs of breakage of various agricultural and farm equipment.

2. **TechSoup Global –** *San Francisco, CA, United States of America*

Data Engineer **Feb 2021 – July 2021**

- Developed a Data Warehouse in Snowflake's Data Cloud and implementing scalable data pipelines for enterprise level datasets.
- Collaborating with Data Architects and Business Analysts to research and resolve the nuances in the Data, meanwhile developing and deploying the pipelines for translating business requirements into data models and loading the essential data into the Warehouse.
- Developed SQL queries and presented results from Amazon Redshift, which enabled engineers to communicate actionable insights with TechSoup's top tier Customers and meet business needs.
- Created Design Documents and Data Flow Documents using draw.io, meanwhile implementing Quartz scheduler solution in JAVA (CTP Platform) to schedule and trigger CRON Jobs to perform Data Loading and Migration from Amazon Redshift (PostgreSQL Database) to Snowflake Datawarehouse.
- Standardized Data modeling and design frameworks for analysis, design, building, testing and maintenance
- Implemented data pipelines using Amazon SQS Message Queuing Service, Dagster ETL Tool.
- Responsible to develop and engineer maintainable code in Python, thus collaborating with various cross-functional teams across organization.

3. **Hartford Steam Boiler IIC –** *Hartford, CT, United States of America*

Machine Learning Intern **May 2020 – Nov 2020**

- Developed and deployed a Bayesian Hierarchical Model using Statistical and Machine Learning techniques in Spark R which helped in predicting 20% better Risk Scores than the traditional Large-Scale Machine Learning Model-**Location Risk Score** figures.
- Collaborated with Senior Data-Scientists to perform raw data mining by using statistical techniques and developed Machine Learning algorithms using PySpark to achieve seamless integration of Third-Party Data Sources.
- Implemented a Data-pipeline in Python and automated the development of Technical EDAs in Azure Databricks, which reduced the duration of Development by 85%, and articulated the technical challenges and solutions to design interactive Statistical dashboards to communicate actionable insights at all levels of the Organization.
- Solved complex industrial and technical problems in a re-insurance industry using advanced mathematical modeling and optimization techniques including big data preprocessing, features engineering, algorithmic selection and evaluation, model development and deployment.
- Performed A/B testing, data wrangling, presentations, reporting and dashboard development, working with Cross-functional teams in an Agile-Scrum Lite methodology focused on adaptive planning and extreme Programming to drive actionable insights.

4. **Accenture INC. –** *Mumbai, MH, India*

(Started as Associate Software Engineer, promoted as Software Engineer) **Jan 2017 – July 2019**

- Performed Database and Application Migration to Oracle Cloud while ensuring that the Database Architecture meets the Client's ongoing growth needs.
- Effectively communicated with Application Development Team in Agile Software Development methodology to develop database architectures and data models which helped in troubleshooting issues like scalability, replication, and performance.
- Developed database schema for databases on Oracle EBS version 12.2.5, hence managing large-scale FinTech data sources.
- Performed cloning of Oracle Applications and executed sanity testing for Development & Production environments using RMAN backup/recovery tool to troubleshoot procedural issues in data pipelines independently.
- Migrated and tuned Non-RAC Oracle Databases to RAC for enabling optimum use of resources across distributed clustered nodes during high traffic influx meanwhile ensuring the environment meets client's SLAs.

SUMMER INTERNSHIPS

ALSO Digital Private Limited, Mumbai, MH, India **May 2015 – July 2015**
Financial Analyst Intern

Reliance Communications, Mumbai, MH, India **May 2014 – July 2014**
Software Engineering Intern

ACADEMIC PROJECTS

Implemented Spark using Scala and Spark SQL for Complex Data Processing and Analysis.

- Developed a Python Application and implemented Spark using Scala and utilized Spark Core, Spark Streaming and Spark SQL API for analyzing TBs of datasets and performing Complex operations on Hadoop Distributed File System.
- Developed a Bash Shell Script which helps the User to automate the task to process and filter large chunks of data sets of the same extension by creating a directory and copying them in their respective directories.

Machine Learning Model for Loan Prediction.

- Implemented a Prediction Model in Python using Logistic Regression, Support Vector machines, Decision Tree Classifier and KNN's to predict whether a new individual will be granted a Loan based on various factors.
- Analyzed the Factors affecting the Loan granting process and processed a large test dataset to be compatible with the trained model.