# HARSH VICTOR CHALLA

+16025662038 ⋄ hvictorc@asu.edu ⋄ LinkedIn.com/in/harsh-challa/ ⋄ GitHub.com/Harshchalla⋄ Tempe, AZ (Open to Relocation)

Skilled graduate student in Big Data, Cloud, and Software Development with hands-on experience in Python, SQL, and AWS. Proven ability to deliver scalable solutions and data driven insights, eager to apply expertise in a collaborative environment.

## EDUCATION

**Arizona State University, Tempe, AZ, USA** August 2023 - May 2025
Master of Science in Computer Science GPA: 4/4

**Visvesvaraya Technological University, Bengaluru, KA, India** August 2019 - May 2023
Bachelor of Engineering in Computer Science GPA: 3.6/4

***Relevant Courses:*** *Data Structures, Big Data Analytics, Artificial Intelligence and Machine Learning, Data Mining, Data Communication, Distributed Database Systems, Data Visualization, Statistical Machine Learning, Cloud Computing.*

## EXPERIENCE

**Arizona State University** November 2023 - Present
*Business and HR Analyst* *Tempe, AZ, USA*

- Engineered VBA-embedded Python scripts that automated the manual process of formatting accounts data, achieving a 97.22% reduction in processing time from 12 hours to 20 minutes.
- Extracted, analyzed, and visualized reports from Workday and PeopleSoft in Tableau to enhance data-driven decision-making.

**Tequed Labs** September 2022 - September 2022
*Machine Learning Intern* *Bengaluru, KA, India*

- Led data cleaning and preprocessing initiatives for the house price prediction project, utilizing Python to handle missing values, outlier detection, and feature engineering, resulting in a model with 20% improved pricing accuracy in a tight timeline.
- Orchestrated and optimized data pipelines using scikit-learn, incorporating GridSearchCV for hyperparameter tuning, achieving price prediction accuracy of 91%.

**Cube Logistics** October 2021 - December 2021
*Data Analyst Intern* *Bengaluru, KA, India*

- Enhanced SQL query performance by optimizing SARGability, implementing composite indexing, and restructuring queries, resulting in a 50% reduction in execution times and improved efficiency in a Snowflake-based data warehouse.
- Conducted comprehensive exploratory data analysis on large-scale logistics datasets, uncovering critical KPIs that were strategically integrated into Tableau dashboards, driving enhanced decision-making.

## PROJECTS

**Sports Data Processing Pipeline|*Python, PySpark, SQL, AWS S3, Apache Spark***

- Developed a data pipeline leveraging Apache Spark to process cricket match data, handling large-scale data ingestion from Amazon S3, transforming and validating schemas, and optimizing performance for fast querying and analytics.
- Uncovered key insights on player performance and match trends using SQL analytics and visualizations.

**Real-Time Stock Market Data Pipeline|*Python, SQL, Apache Kafka, AWS S3, Glue, Athena***

- Developed a real-time data processing pipeline using Apache Kafka for streaming stock market data, AWS S3 for storage, and AWS Glue to transform raw stock market data into a queryable format and automate the loading process.
- Leveraged AWS Athena to perform ad-hoc queries on the processed data for analysis and insights. Integrated Python scripts to streamline the ingestion and transformation processes, enabling end-to-end automation.

**Outlier Detection and Statistical Modeling for Sensitivity-Driven Data Analysis|*Python, Statistical Analysis***

- Leveraged statistical methods like ANOVA for data preprocessing and classification, while developing and implementing outlier detection models such as MAD and Boxplot, achieving high F-scores (up to 0.99) in detection across diverse datasets.

**Elastic Image Recognition Service on AWS|*Python, AWS EC2, ECR, S3, Lambda, SQS, Auto-scaling***

- Engineered scalable cloud-based image recognition services using AWS, implementing an advanced auto-scaling algorithm.

## TECHNICAL SKILLS

| | |
|---|---|
| **Programming Languages** | Python, SQL, Shell Scripting. |
| **Frameworks** | PySpark, Flask, Numpy, Pandas, PyTorch, Keras, TensorFlow, Boto3, Scikit-learn, SQLAlchemy. |
| **Data Technologies** | MySQL, PostgreSQL, DynamoDB, Databricks, Snowflake, Apache Spark, Apache Kafka, Apache Spark, Apache Hadoop, Apache Zookeeper, Docker. |
| **AWS Cloud** | AWS EC2, S3, SQS, Lambda, Glue, Athena, QuickSight. |
| **Skills** | System Design, Distributed Systems, Data Structures, Design patterns, ETL, Data Engineering, Data Science, Data Analytics, Deep Learning, Machine Learning, NLP, Generative AI. |

**Oracle Certified - Cloud Infrastructure 2024 Generative AI Professional**
**AWS Certified Academy Graduate - AWS Academy Cloud Foundations**