

Saumitra Chaskar

saumitrachaskar@gmail.com | www.linkedin.com/in/saumitrachaskar | +1 (217) 904 0764

EDUCATION

University of Illinois Urbana Champaign

Master of Computer Science

Relevant Coursework: Databases, Advanced Data Management, Machine Learning and Data Systems

Expected Dec 2024

GPA: 4.0/4.0

Sardar Patel Institute of Technology

Bachelor of Tech Computer Engineering

Relevant Coursework: Operating Systems, Distributed Systems, Communication Networks, Algorithms, Computer Security

Graduated May 2021

CGPA: 9.69/10

SKILLS

- **Programming Languages:** Python, SQL, PL/SQL, Shell Scripting, Javascript, C++, Java
- **Databases/Cloud Platforms:** Hive, Postgres, GraphQL, MySQL, MongoDB, Elasticsearch, Databricks, GCP, Azure, Azure DevOps
- **Frameworks:** Hadoop, Spark, Kafka, Istio, Django, Flask, Pandas, Angular, Kubernetes
- **Other Tools:** Linux, Git, Maven, Informatica ETL, Jupyter, Tableau, Power BI, SAP BOBJ, JIRA, Confluence

PROFESSIONAL EXPERIENCE

Homesite Insurance

Madison, WI

Data Engineer Intern

Jun 2024 - Present

- Modernized and optimized a Python-based insurance rating tool using BigQuery achieving a 30x improvement in overall runtime
- Composed a CI/CD pipeline for deploying the tool on GCP using GCS, Pub/Sub, Functions, and Cloud Run, eliminating manual touchpoints. Wrote comprehensive testing scripts to validate data integrity, significantly increasing data reliability

National Center for Supercomputing Applications

Champaign, IL

Research Assistant

Jan 2024 - Present

- Developed a virtual imaging system for augmenting PET and CT scan image datasets by enabling imaging simulations at scale and reducing the overall simulation runtime by 4x
- Created scalable data pipelines using Apache Airflow to orchestrate legacy imaging scan simulation programs on an HPC K8s cluster

VMware

Bangalore, India

Data Engineer I

May 2023 - Jul 2023

- Developed Tableau dashboards for tracking the success of VMware's perpetual to SaaS transformation journey by following agile methodologies, reducing the conception to completion time of dashboards by almost 50%
- Headed a cross-functional team of 3 developers for delivering quarterly releases for data products and reports about VMware customer and sales datahubs, handling more than 50 analytics products

Member of Technical Staff

Aug 2021 - Apr 2023

- Migrated legacy Postgres and HANA procedures to Kubernetes-backed Spark jobs and optimized and modularised code reducing the overall runtimes by 30%
- Authored configurable and automated ETL pipelines using Python scripting and spark to pull in data from 10+ source systems to a Hive data warehouse, eliminating manual intervention and code redundancy
- Assumed the role of Data Product Owner for Customer Discovery Tool and oversaw data quality and usage, documented the data product capabilities on Confluence thus enabling cross-team consumption and self-serve analytics through the Presto query engine

Software Development Intern

Jan 2021 - Jul 2021

- Architected and implemented a DSL-based orchestrator to automate API request-response chaining testing scenarios, reducing testing time by 25% and saving 20 developer hours per sprint
- Built a tool for cataloging open-source software used by applications across the VMware IT landscape, integrated the tool in the DevSecOps pipeline, and extended its usage for vulnerability management and alerting

PROJECTS

Zerogya

- Implemented a hospital inventory management and medical records storage system using Azure Cosmos, Apps, MLaaS, and Blobs
- Integrated ML-aided diagnostic services powered by the data shared by a network of hospitals through Federated Machine Learning models in TensorFlow, achieving 92% accuracy on breast cancer tumor classification using the approach mentioned above

IdentifyEasy

- Prototyped an entirely distributed and decentralized PII management wallet for streamlining KYC processes
- Implemented data control and ownership policies using Ethereum for permissions and access tracking of the documents and InterPlanetary File System (IPFS) for document storage and presented the proof of concept by deploying the application on a network of 20 computers acting as IPFS nodes

VM Scaler

- Developed Powershell scripts for autoscaling cloud applications using Azure VM Scale Sets, Azure Functions, and ARM templates
- Eliminated human intervention in manual scaling, application upgrades, and failure recovery using Python scripts for monitoring and alerting and Powershell for managing Azure resources