

Prajeeth Emanuel

pvethanayaga@cs.stonybrook.edu ☎ (631) 202-9716 📍 New York, NY in [prajeeth-emanuel](https://prajeeth-emanuel.github.io)

EDUCATION

Stony Brook University

M.S. Computer Science

Graduation May, 2022

Stony Brook, NY

- Selected Coursework: Introduction to Computer Vision (CSE527), Introduction to Robotics (CSE525)
- Advised by Prof. Fusheng Wang - Research Project: Scalable Reachability Queries on Spatial-temporal Datasets.

IIIT Hyderabad

B.Tech. Computer Science

2012 - 2016

Hyderabad, IN

- Selected Coursework: Operating Systems, Mathematics III, Principles of Programming Languages, Advanced Computer Networks, Artificial Intelligence, Database Systems, Distributed Systems, Statistical Methods in AI

WORK EXPERIENCE

Apple Inc.

Software Engineer

Jan. 2019 – October 2020

- Designed the logic & algorithms on the web crawling system (real-time & batch streaming) used in Apple Maps.
- Developed terabyte-scale daily ETL at Apple for processing/indexing for classifying relevant webpages using MapReduce, Spark & Java.
- Designed and implemented the data architecture for indexing & querying billions of webpages with a focus on design structures that are adaptable to changes in the source data or business processes.
- Wrote several memos used to guide Apple Map's data quality practices in web discovery. Identified points at which data audits, unit tests, and documentation can be most effective.
- Primary analyst to help build the relevance system and identify improvements for information retrieval.
- Researched and improved algorithms to serve the ops teams with precise topical web pages.
- Worked extensively with Matplotlib within Python for analytical studies to build Jupyter Notebook dashboards.
- Deployed high-performance systems with reliable monitoring and logging practices in close coordination with the SRE team.

Qalaxia Inc.

Software Engineer, Data

Jul 2016 – Dec 2018

- Designed and implemented large stateful batch and real-time data pipelines to mine large-scale datasets using Spark, Hive, and Java.
- Built high-performance ETL pipelines and data warehousing systems to enable data science workflows.
- Primarily responsible for maintaining the critical pipelines that support analytic systems and A/B experiments.
- Developed data models for optimal storage and retrieval to meet critical product and business requirements.

Grants

- **NSF SBIR** (2018) [#1843326](#) - Skill-aware query engine for K12 Classrooms: Develop new skill-aware AI technology that automatically maps student's raw queries into topics, sub-topics, and intent.

SKILLS & INTERESTS

- **Languages:** Java, Scala, Python, C++
- **Distributed Systems Frameworks:** Spark, Hadoop, Kafka
- **Workflow Scheduling:** Airflow, Jenkins
- **Monitoring Tools:** Sentry, Grafana
- **Datastores:** MySQL, MongoDB, Elasticsearch, Cassandra, Redis
- **Infrastructure:** AWS, GCP