SPANDAN BEMBY

spandanbemby.com ◇ github.com/spandanb
spandan.bemby@gmail.com ◇ (650) · 283 · 5376 ◇ San Francisco, CA

EXPERIENCE

Cruise Automation May 2017 - Present
Data Infrastrucure Engineer *San Francisco, CA*

- Led design and implementation of catalog service which provided RESTful access to location, encoding, schema, and other metadata about various datasets.
- Service provided fine-grained data access by time range and other dataset-specific attributes for a variety of use-cases, e.g. data-analytics, visualization.
- Designed the system to be generic and eventually allow producers to automatically push data into lake.
- QPS: 200req/s; P50: 40ms; P90: 100ms; P99: 2000ms
- Built using Python, Flask, and served via uwsgi and Nginx; metadata store was PostgreSQL.
- Previously, worked on the construction of the data lake, including ingestion, transcoding (ROS bag to Avro), storage, and indexing of data.
- Worked on the ingestion of various sources of data, including both real and simulation vehicle data.
- Datasets included sensor data, e.g. camera, lidar, as well as structured messages passed between processing nodes.
- Analyzed access patterns and frequently worked with data consumers to build indices for efficient data access.
- Sim pipeline was ingesting ∼100 TB data a day.
- Storage was first AWS S3 and then GCS.
- Build a time-series analysis tool using elasticsearch and grafana.

Insight Data Science September 2016 - December 2016
Data Engineering Fellow *Palo Alto, CA*

- Designed a platform to give runners real time feedback on how their instantaneous speeds compare with other users'.
- Utilized Spark for batch and real time processing, Elasticsearch for geospatial queries, and Kafka for ingestion.

University of Toronto May 2014 - September 2016
Research Assistant *Toronto, Canada*

- Created an advanced orchestration platform with support for: 1) arbitrary L2 networks (overlays), 2) software-defined networking 3) multiple public and private clouds, 4) containers, and 5) bare metal servers.
- Utilized linux-network stack, Open vSwitch, Ryu, Docker, OpenStack (Heat), Boto3, and Ansible.

EDUCATION

University of Toronto *September 2016*
M.A.Sc. in Computer Engineering
Thesis: Distributed orchestration of heterogeneous resource types over public and private clouds

SKILLS

Concepts Distributed data storage and processing, Relational databases (PostgreSQL), performance analysis
Languages Python
Technologies PostgresSQL, Docker, Kubernetes, AWS, Google Cloud