Vinay Mehta

hello@vinaym.com
Palo Alto, CA

Work

Lambda Labs
Product Marketing Manager
10/2021 - 07/2022

- Led the launch of a GPU laptop built by RAZER and marketed towards deep learning engineers. Extremely successful product launch resulting in 3x sales, 40% increase in services revenue, and first coverage for the company by major tech news outlets: [1, 2].
- Ran a quarter long project to develop messaging. Munged data with pandas from unnormalized SQL databases to generate product line's lifetime order history and customer base. Targeted those customers with a persona survey and conducted follow-up interviews to understand customer usage patterns. Identified opportunity to update pricing strategy by reducing hardware price, introducing support plans, and increasing the margins on factory dual-boot installs.
- Project managed a cross-functional group across the company to beta test production validation hardware and OS images, design and implement a new brand direction (landing page and packaging), proactively identify support issues, and establish manufacturing and shipping procedures.
- Ultimately recognized that the technical aspects of my work were the most fulfilling, and that product marketing was not going to be a good long-term fit for me.

Flex Logix
Technical Marketing Manager
01/2020 - 08/2021

- Interdisciplinary role leading technical marketing and GTM strategy for the InferX accelerators for convolutional neural networks.
- Consolidated Excel based customer performance estimators into a Python script with a maintainable YAML config of the accelerator architecture. Reduced the turnaround time for estimating a customer model's performance from two weeks to 24 hours.
- Ran design sprints for marketing presentations, blog posts, and web seminars at industry trade shows (Linley Processor Conf., AI HW Summit) and via paid media channels.

Lyft
Hardware Engineer
08/2018 - 01/2020

- Rose to directly responsible individual in the team that designed, built, and verified the next-generation compute hardware for Lyft's self driving system.
- Demonstrated an alternate C++ inference pipeline with on-GPU HEVC decoding (NVDEC), quantized FP16 models (TensorRT), and concurrent execution of batch=1 inferences on a single GPU (Triton Inference). Resulted in adoption of TensorRT by the perception team, HEVC encoding by the camera team, and initiated a push by the systems engineering team to create requirements allowing for allocating multiple cameras to a single GPU.
- Explored different system architectures balancing fault tolerance, reliability, and performance goals with mass, volume, thermal, and cost constraints; achieved cross team consensus from mechanical, thermal, systems, and product teams.
- Drove product lifecycle from RFQ and vendor selection, to final integration of validated and verified system into vehicle.
- Directly managed two summer interns evaluating inference accelerators.

SpaceX
Software Intern
05/2016 - 08/2016

- Built the Java backend for a system that queried and downloaded part files from the central parts repository onto 3D printers in the additive manufacturing lab.
- Maintained the interface and functionality of the Engineering Change Notice form, adding support for different Falcon 9 Blocks and Variants.

Argosy International
Operations Intern
05/2015 - 08/2015

- Built a spreadsheet model to determine the location of a new aluminum honeycomb processing facility, aggregating by ZIP code data on wages, utility rates, commercial real estate property values, and predicted shipping costs from overseas factory and on to existing customers within North America.
- Board of directors ultimately elected to build the facility in Alabama, following the top-1 recommendation of the model.

Education

Columbia University
B.S. Computer Engineering
Class of 2018, GPA: 3.8 / 4.0

Software Engineering

Programming Languages

Python	Expert
SQL	Proficient
C	Proficient
Java	Proficient