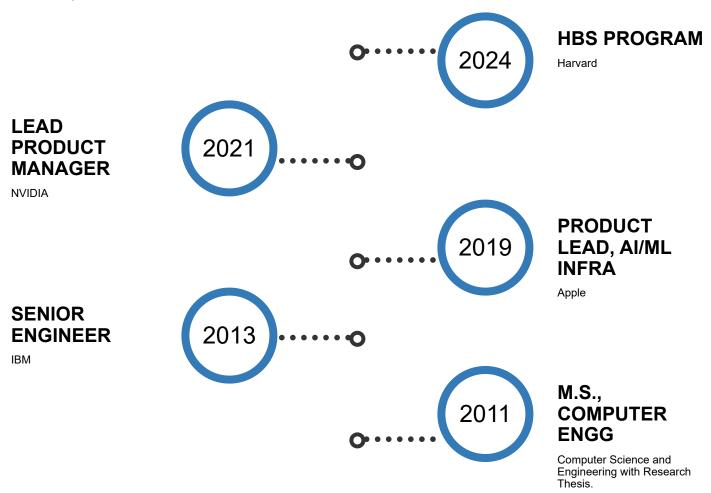
Introduction & Background

- Experienced Product Manager with 15 years in technology and business.
- · Built products and end-customer solutions that are currently LIVE at Apple, IBM and NVIDIA.
- Currently a Lead Product Manager in NVIDIA's Enterprise Product Group. Expertise in delivering technology-backed products with a focus on good engineering principles and clean code.
- Successful in large-scale enterprise roles across India and the USA, adept at navigating diverse go-to-market motions.
- M.S., North Carolina State University | EB1-A Green Card Holder (Einstein Visa)
- Harvard Business School Executive Education, Hybrid Program (Commencing December 2024)

Areas of experience & interest



Work Experience



Additional Information

NVIDIA:

- DGX Cloud MVP business model for Top-20 customers/Hyperscalar-level pre-training.
- Day 2 playbook for data manageibility and partnerships with CSP on ingress/egress and staging.
- 1000+ fine-tuning jobs on fine-tuning with DGX Cloud via Hugging Face front-end with pay-as-you-go B2B2B business model.
- Successful internal alignment on standardized API spec for GPU consumption in managed service mode via partners.
- 10+ model and data partners onboarded for NVIDIA Inference microservices (NIM) adoption.

Apple:

- Responsible for applying 2D Deep Learning models in Apple factories for Capex savings and yield improvements of Iphone Line of Business.
- Responsible for data movement to centralized data management and real time processing of model inference.
- Managed global teams for building and deploying Infra for always-ON inference platform using on-prem and CSP utilities.
- Responsible for building data and model drift observability platform and flows.
- Responsible for scaling the product from single Apple Line of Business to across all Apple products, resulting in \$50M+ savings for Apple manufacturing.

IBM:

- Responsible for setting up distributed data processing pipeline using Apache Spark on Hadoop Data lake.
- Responsible for working with Major CSPs to show value of IBM solution for fastest shuffle + sorting performance.
- · Led team to put a streaming and batch processing pipeline for enterprise use-case with Apache Spark.
- Worked with research to optimize IBM compute architecture for contemporary workloads like Spark in-memory processing and Deep learning workloads.
- Awarded patents in the process of working with research around distributed processing and helping take IBM's effort to US National Laboratories.
- Got eligibility for Einstein Visa and got Green Card in Einstein visa category due to work done with IBM engineering and research teams and results delivered to enhance US technology.

Rohit Taneja | rohit23taneja@gmail.com | +1-919-675-9274 SF Bay Area, California