DRAFT

RAPIDS

Microsoft

Microsoft Azure

CLOUD-BASED DATA SCIENCE AT THE SPEED OF THOUGHT USING RAPIDS - THE OPEN GPU DATA SCIENCE ECOSYSTEM

AGENDA

1. Introduction (not hands-on) - 20 min
   - Speaker Introduction
   - Why RAPIDS
   - Component Overview
   - Getting Connected to VMs
   - *Break - 5 min*

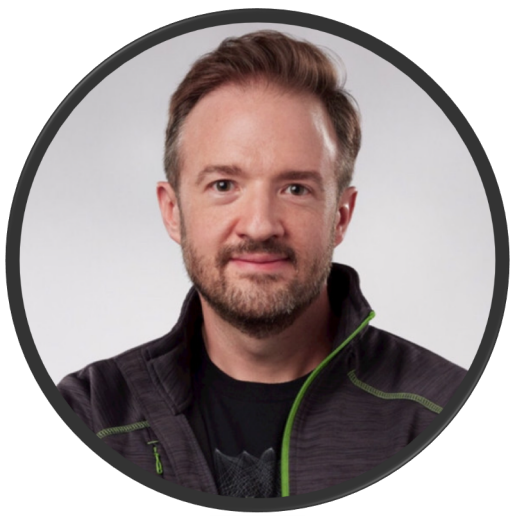
2. Tutorial (hands-on)
   - Classification of astronomical sources - 45 min
   - *Break - 5 min*
   - Finding commonalities within populations - 45 min
   - *Break - 5 min*
   - Cyber flagging anomalous network communications - 45 min
   - *Break - 5 min*

3. Conclusion (not hands-on)
   - DASK Example (Multi-GPU
   - Roadmap
   - Conclusion

SPEAKER INTRODUCTION

Joshua Patterson is a General Manager of Data Science at NVIDIA leading engineering for RAPIDS.AI, and a former White House Presidential Innovation Fellow. Prior to NVIDIA, Josh worked with leading experts across public sector, private sector, and academia to build a next-generation cyber defense platform. His current passions are graph analytics, machine learning, and large-scale system design. Josh also loves storytelling with data and creating interactive data visualizations. Josh holds a B.A. in economics from the University of North Carolina at Chapel Hill and an M.A. in economics from the University of South Carolina Moore School of Business.

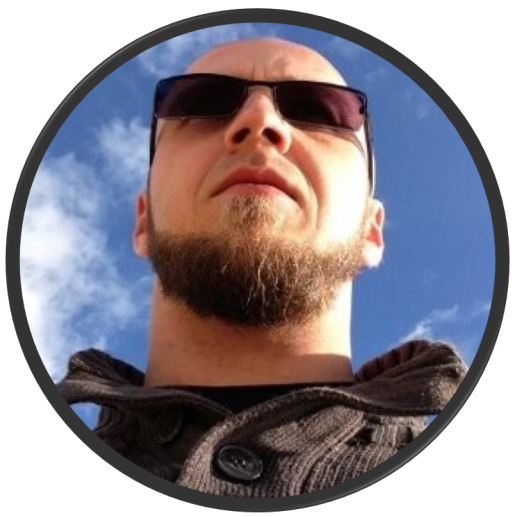
Brad Rees is a Manager in the AI Infrastructure group at NVIDIA and lead of the RAPIDS cuGraph team. Brad has been designing, implementing, and supporting a variety of advanced software and hardware systems for over 30 years. Brad specializes in complex analytic systems, primarily using graph analytic techniques for social and cyber network analysis. His technical interests are in HPC, machine learning, deep learning, and graph. Brad has a Ph.D. in Computer Science from the Florida Institute of Technology.



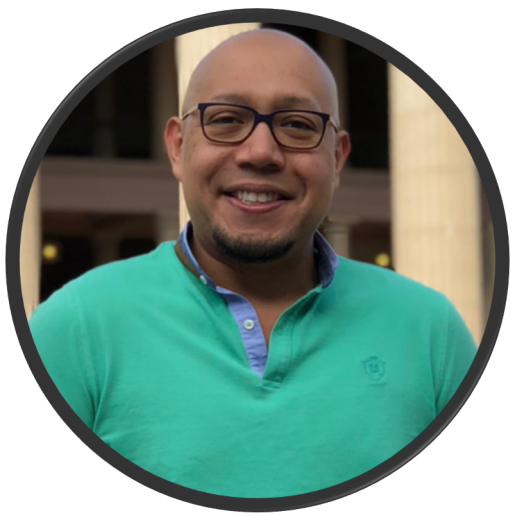
Bartley Richardson, PhD is a Senior Data Scientist in AI Infrastructure (RAPIDS) at NVIDIA. He leads a team that researches and applies GPU-accelerated ML and DL to help solve today's information security and cybersecurity challenges. Prior to joining NVIDIA, Bartley was a technical lead and performer on multiple DARPA research projects, where he applied data science and machine learning algorithms at-scale to solve large cybersecurity problems. His primary research areas involve NLP and sequence-based methods applied to cyber network datasets as well as cross-domain applications of machine and deep learning solutions. Bartley holds a PhD in Computer Science and Engineering from the University of Cincinnati with a focus on loosely- and un-structured logical query optimization. His BS is in Computer Engineering with a focus on software design and AI.

Keith Kraus is a Senior Engineer of Applied Solutions Engineering at NVIDIA and lead the RAPIDS cuDF team. At NVIDIA, Keith's focus is on building GPU-accelerated solutions around data engineering, analytics, and visualization. Prior to working for NVIDIA, Keith did extensive data engineering, systems engineering, and data visualization work in the cybersecurity domain focused on building a GPU-accelerated big data solution for advanced threat detection and cyber-hunting capabilities. Keith graduated from Stevens Institute of Technology with a BEng in computer engineering and an MEng in networked information systems.

Corey Nolet is a Data Scientist & Senior Engineer on the RAPIDS cuML team at NVIDIA, where he focuses on building and scaling machine learning algorithms to support extreme data loads at light speed. Prior to working at NVIDIA, Corey spent over a decade building massive-scale exploratory data science & real-time analytics platforms for HPC environments in the defense industry. Corey holds Bs. & Ms. degrees in Computer Science. He is also working towards his PhD in the same discipline, focused on scaling unsupervised machine learning algorithms. Corey has a passion for making sense of the world through data.



Tom Drabas is a Senior Data Scientist at Microsoft. Tom has over 15 years of international experience working in airline, telecommunication and technology industries. He holds PhD in airline operations research field from the University of New South Wales. During his time at Microsoft he has published multiple books and authored a video series on data science, machine learning and distributed computing in PySpark. Tom's research interests include parallel, deep learning and machine learning algorithms and their applications.



Juan-Arturo Herrera is a Senior Data Scientist at Microsoft where he has contributed to Customer Churn analysis, Windows ecosystem driver pattern mining, and Windows updates driven my ML models: device targeting optimization, update adoption speed forecast, as well network cost optimization. Prior to working for Microsoft, Juan-Arturo work worked in Adaptive Technologies Inc. modelling interventions as an optimization problem, as well as in the Centre for Complex Sciences (C3) collaborating on the prediction of emerging diseases. He received his Ph.D. from the Applied Mathematics and Systems Research Institute (IIMAS) in the National Autonomous University of Mexico (UNAM). His research interests include multi-objective optimization, genetic algorithms, machine learning and large-scale data mining problems.