TWINKLE JAIN

Boston, MA ⋄ 857.707.8421 ⋄ jain.t@northeastern.edu ⋄ linkedin.com/in/jaintwinkle ⋄ jaintwinkle.com

SUMMARY

A fundamental problem in the performance of distributed environments is the management of limited resources. A central goal of current research work is to optimize resource utilization for HPC and Big Data in an application-transparent manner (no modification of the application). This has been successful for: a) a low-overhead checkpoint-restart architecture for CUDA; b) an assessment of existing speculative execution for detecting and handling straggler tasks in Spark; and c) an analysis of the role of heartbeat arrival in straggler detection in Hadoop.

EDUCATION

Northeastern University, Boston, MA — Fall 2016 – Present
Ph.D. in Computer Science (switched from MS in Fall 2017) — (GPA: 3.7/4.0)
Thesis advisor: Prof. Gene Cooperman
Relevant Courses: Computer Systems, Algorithms, Virtualization, Compiler

Jai Narain Vyas University (currently M.B.M. University), Jodhpur, India — Aug 2012 - Oct 2015
Master of Computer Applications, First Class with Distinction — (equivalent GPA: 4.0/4.0)
Relevant Courses: Algorithms, Data Structures, Computer Architecture, Operating Systems

WORK EXPERIENCE

Northeastern University — Sep 2016 – Present
Graduate Research Assistant — Boston, MA
- **Projects**: Fix single point of failure of the ROS master, CRAC: a split-process-based architecture to checkpoint CUDA, CRAC-M: a flexible split-process design to support checkpointing with multiple lower halves (CUDA and MPI).
- **Teaching Assistant**: Computer Systems (CS 5600) in Fall 2022 and Intensive Computer Systems (CS 7600) in Spring 2021, and Spring 2023.

MemVerge, Inc. — May – Aug 2022
Research Intern (Remote) — San Francisco, CA
- Developed and maintained MANA, an MPI-Agnostic Network Agnostic checkpoint-restart tool for MPI applications.
- Troubleshot memory corruption-related issues in MANA to support at least three scientific HPC applications contributing 20% of the total machine hours at National Energy Research Scientific Computing Center (NERSC) sites.

IBM TJ Watson Research Center — Jun – Aug 2021
Research Intern (Remote) — Yorktown Heights, NY
- Analyzed existing resiliency support and proposed an improved framework in Ray, a distributed execution platform.
- Demonstrated a 5% improvement in runtime; averted crashes caused by configuration-related memory overflow.

Inria — May – Jul 2019 & Feb – Aug 2020
Research Visitor — Nantes, France
- Evaluated and improved speculative execution implementation (to detect and handle stragglers) in Hadoop and Spark.
- Published findings across multiple conference papers (see publication section).

Mentor Graphics (Siemens EDA) — May – Aug 2018
System Engineer Intern — Waltham, MA
- Developed a Distributed Multi-Threaded CheckPointing (DMTCP) plugin to restart an optimized checkpointed build as a debug build; reduced time taken in debugging an in-house large-scale application by 75% (see patent, below).

Stratus Technologies — May – Aug 2017
Platform Engineer Intern — Maynard, MA
- Assessed performance of COarse-grained LOck-stepping (COLO) technique on QEMU for fault-tolerance in servers.

SELECTED ACADEMIC PROJECTS

Northeastern University
Compiler for a small Programming Language — Sep – Dec 2018
- Built a complete compiler in SML-NJ language for Andrew Appel's Tiger programming language in a team of two.

Decrease Down-time in Live Process Migration — Nov – Dec 2016
- Decreased downtime by 80% in process migration from one host to another by prioritizing memory-page sending order.

Jai Narain Vyas University
Maze Traversing Robot — Oct 2014 – Jan 2015
- Designed and built a wireless, camera-driven robot that identifies the path in a maze. Programmed a microcontroller to communicate via an RF module for navigation input (see Pixelate award).

AWARDS

- Represented Northeastern University and Khoury College of Computer Sciences at the 2019 Grace Hopper Celebration of Women in Computing conference.
- Led team of four; 3rd place among 33 teams in "Pixelate" at Techfest 2015, Bombay (Asia's Largest Technical Festival).
- Won four 1st-place and three 2nd-place awards at state-level technical events at Encarta 2013, Jodhpur, Rajasthan.

PATENT

Software Checkpoint-Restoration between distinctly compiled executables (Granted in Aug 2022)
Twinkle Jain, Vipul Kulshrestha, Kenneth W. Crouch (US11429379B2) Siemens Industry Software Inc.

SELECTED PUBLICATIONS

Full Papers
- "Towards an effective Speculative Execution in Spark," **T. Jain** and S. Ibrahim (under review). (VLDB 2023)
- "On the (In)Accuracy of Stragglers Detection in Hadoop," T. Lambert, **T. Jain**, and S. Ibrahim. (FGCS 2023)
- "CRAC: checkpoint-restart architecture for CUDA with streams and UVM," **T. Jain** and G. Cooperman. (SC 2020)

Short Papers
- "Stragglers' Detection in Big Data Analytic Systems: The Impact of Heartbeat Arrival," T. Lambert, S. Ibrahim, **T. Jain** and D. Guyon. (CCGrid 2022)
- "Transparent Checkpointing for OpenGL Applications on GPUs," D. Hou, J. Gan, Y. Li, Y. El Idrissi Yazami, and **T. Jain** (SuperCheck 2021)
- "Checkpointing SPAdes for Metagenome Assembly: Transparency versus Performance in Production," **T. Jain** and J. Wang. (SuperCheck 2021)

Posters/Presentations
- Stragglers Detection in Big Data Analytic Systems: The Impact of Heartbeat Arrival (CCGrid 2022)
- Checkpoint/Restart of MPI applications over the Cray GNI Network via Proxies (MUG 2018)
- DMTCP: Fixing the single point of failure of the ROS master (ROSCon 2017)

ADDITIONAL RESEARCH EXPERIENCE

Invited Talks
- MANA/CRAC: A Preprocessor for Safe Transparent Checkpointing of MPI and CUDA (EuroMPI/USA 2022)
- C/R Requirements for CUDA Applications at NERSC (Checkpoint/Restart Requirements Gathering Workshop 2022)

Reviewer
- Program committee member and reviewer for SuperCheck 2021, SuperCheck-SC 2021, and SuperCheck-SC 2022.
- Reviewer for IEEE International Conference on Robotics and Automation (ICRA) 2021 and High Performance Computing (HiPC) 2020.

TECHNICAL KNOWLEDGE

Languages and APIs:	C/C++ (CUDA/MPI/POSIX), Python (numpy/matplotlib/pandas).
Distributed Frameworks:	SLURM, Hadoop, Spark, Yarn, HDFS.
Deployment Tools:	Kubernetes, OpenShift, Docker.
Tools:	GDB, Git, Shell, LaTeX, VS Code.