MTC:

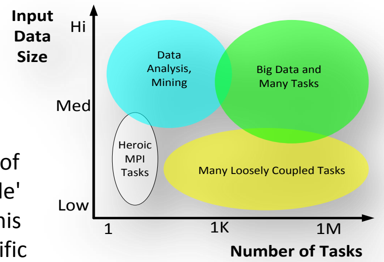
Many-Task Computing
http://datasys.cs.iit.edu/

Illinois Institute of Technology
Computer Science Department
Data-Intensive Distributed Systems Laboratory

Exascale computers will enable the unravelling of significant scientific mysteries. There are many domains (e.g. weather modelling, national security, drug discovery) that will achieve revolutionary advancements due to exascale computing. Predictions are that supercomputers will reach exascales by 2019, and will be composed of millions of compute nodes aggregating billions of threads of execution. The exascales computing era will bring new fundamental challenges in how we build, manage, and program computing systems. Our decades-old approaches (e.g. parallel file systems, MPI) will have to be radically changed to support the coming wave of extreme-scale general purpose parallel computing.

A New Vision

Professor Ioan Raicu of the Data-Intensive Distributed Systems (DataSys) Laboratory has proposed the Many-Task Computing (MTC) paradigm, which aims to bridge the gap between high-performance computing (HPC) and high-throughput computing (HTC). Although MTC is not a replacement for HPC, the MTC paradigm can address many of the HPC shortcomings at extreme scales (e.g. reliability, programmability) for a large class of applications. Professor Raicu's work addresses the particular challenge, set forth by the Department of Energy Advanced Scientific Computing Research program, of "fulfilling the science potential of emerging multi-core computing systems and other novel 'extreme-scale' computing architectures, which will require significant modifications to today's tools and techniques". This work addresses "advanced hardware and software architectures for exascale computing systems, scientific data management and analysis at scale, and scalable and fault tolerant operating and runtime".



Proposed Work

The DataSys Laboratory is working on developing both the theoretical and practical aspects of building efficient and scalable support for Many-Task Computing at exascales, through three research activities:

1. Design, analyse, and implement a distributed data-aware execution fabric supporting HPC/MTC
2. Integrate the distributed data-aware execution fabric with the Swift parallel programming system
3. Evaluate work with real applications at pre-exascales, as well as simulations at exascales

The execution fabric will be fault tolerant by having all compute nodes participate in the job submission and handling process; work stealing will be used to achieve efficient distributed load balancing. The fabric would guarantee job execution and dependencies, and would rely on an underlying scalable distributed storage system for inter-process communication. Data-aware scheduling would maximize data locality by scheduling computational tasks close to the data. Computations will be overlapped with I/O to reduce wasted resources and hide latencies. The fabric will support both MTC workloads and HPC. The fabric will be elastic, allowing it to grown and shrink in resource usage based on the application demand. The fabric will also support compact task representation to alleviate task submission bottlenecks for common patterns (e.g. "for each x do y"). The execution fabric will be integrated with several other projects, including FusionFS (fusion distributed file system), D^3 (direct distributed data-structure), and Swift (parallel programming system). The work will be evaluated with many applications (e.g. bioinformatics, medicine, pharmaceuticals, astronomy, physics, climate modelling, economics, and analytics) through the Swift project collaboration on the largest high-end computing (HEC) systems (see table on the right).

Field	Description	Characteristics	Status
Astronomy	Creation of montages from many digital images	Many 1-core tasks, much communication, complex dependencies	Experimental
Astronomy	Stacking of cutouts from digital sky surveys	Many 1-core tasks, much communication	Experimental
Biochemistry*	Analysis of mass-spectrometer data for post-translational protein modifications	10,000-100 million jobs for proteomic searches using custom serial codes	In development
Biochemistry*	Protein structure prediction using iterative fixing algorithm; exploring other biomolecular interactions	Hundreds to thousands of 1- to 1,000-core simulations and data analysis	Operational
Biochemistry*	Identification of drug targets via computational docking/screening	Up to 1 million 1-core docking operations	Operational
Bioinformatics*	Metagenome modeling	Thousands of 1-core integer programming problems	In development
Business economics	Mining of large text corpora to study media bias	Analysis and comparison of over 70 million text files of news articles	In development
Climate science	Ensemble climate model runs and analysis of output data	Tens to hundreds of 100- to 1,000-core simulations	Experimental
Economics*	Generation of response surfaces for various economic models	1,000 to 1 million 1-core runs (10,000 typical), then data analysis	Operational
Neuroscience*	Analysis of functional MRI datasets	Comparison of images; connectivity analysis with structural equation modeling, 100,000+ tasks	Operational
Radiology	Training of computer-aided diagnosis algorithms	Comparison of images; many tasks, much communication	In development
Radiology	Image processing and brain mapping for neuro-surgical planning research	Execution of MPI application in parallel	In development

Note: Asterisks indicate applications being run on Argonne National Laboratory's Blue Gene/P (Intrepid) and/or the TeraGrid Sun Constellation at the University of Texas at Austin (Ranger).

Impact

This high impact work is transformative due to its radical distributed architecture for compute management, departing from traditional centralized job management systems. This work is controversial as it forces an outside the box thinking and changes the application landscape of HEC dramatically from what it has been over the past decades. The Many-Task Computing paradigm will make exascale computing more tractable, touching virtually all disciplines in HEC, fuelling scientific discovery and economic development at the national level. It will extend the knowledgebase beyond HEC into commodity systems as the fastest machines, tools, and techniques generally become mainstream systems. It will open doors for a large new class of applications to run on HEC.

DataSys Data-Intensive Distributed Systems Laboratory
ILLINOIS INSTITUTE OF TECHNOLOGY
Argonne NATIONAL LABORATORY
THE UNIVERSITY OF CHICAGO