EVENT SCHEDULE

8:30-8:50	Breakfast & Registration
8:50-9:00	Welcome
9:00-9:40	Orran Krieger, VMware
9:40-10:20	Harry Li, Facebook
10:20-10:30	Break
10:30-11:10	Adam Silberstein, Yahoo!
11:10-11:50	Joseph Hellerstein, Google, Inc.
11:50-1:00	Lunch
1:00-1:40	Jeff Hammerbacher, Cloudera
1:40-2:20	Andy Pavlo, Brown University
2:20-2:35	Break
2:35-3:15	Srikanth Kandula, Microsoft Research
3:15-3:55	Ch. Papamanthou, Brown University
3:55-4:00	Closing Remarks
4:00-5:30	Reception

DIRECTIONS TO THE CIT BUILDING

*From either I-95 South or I-95 North, take Exit 22A, Downtown, Memorial Boulevard.
*At the fifth traffic light, turn left onto College Street. Cross the bridge and proceed up the hill following College Street to the end.
*At the top of the hill, where College Street intersects with Prospect Street, you will see Brown's Van Wickle Gates. Turn left onto Prospect Street.
*At the next light, turn right onto Waterman Street. At the second light, turn right onto Brook Street. The visitor parking lot will be on your left immediately after turning onto Brook Street.
*The CIT – a large red and tan brick building – is across the street from the visitor parking lot, on the corner of Waterman and Brooke; the entrance is on the other side, facing the quadrangle.
*Event registration is on the third floor.

PARKING

Brown has a Visitor Parking Lot across the street from the CIT on Brook Street; this lot charges an hourly fee.

RSVP to abt@cs.brown.edu by May 6, 2010

The Industrial Partners Program (IPP) provides a formal mechanism for interactions between industry and the Brown Department of Computer Science. Member companies benefit from superior visibility in the Department, exclusive access to event/interview space in the CIT and assistance with the recruiting process; students benefit from specific information about opportunities for summer internships and full-time employment. Opportunities are also available for partners to form research collaborations with our faculty. The Department wishes to thank our industrial partners:

Premier Partner
Adobe

Affiliates
Apple
Facebook
Google
GTECH
Microsoft
NetApp
Oracle
VMware

Start-Up Supporter
LIPIX, Inc.

Individuals
Paul Edelman, Edelman & Associates
Robert Khoury, Worldwide Financial Industry Recruiting Services

To learn more about the Industrial Partners Program, contact:
Amy Tarbox, Program Manager Telephone: 401-863-7610 abt@cs.brown.edu

Brown University
Department of Computer Science
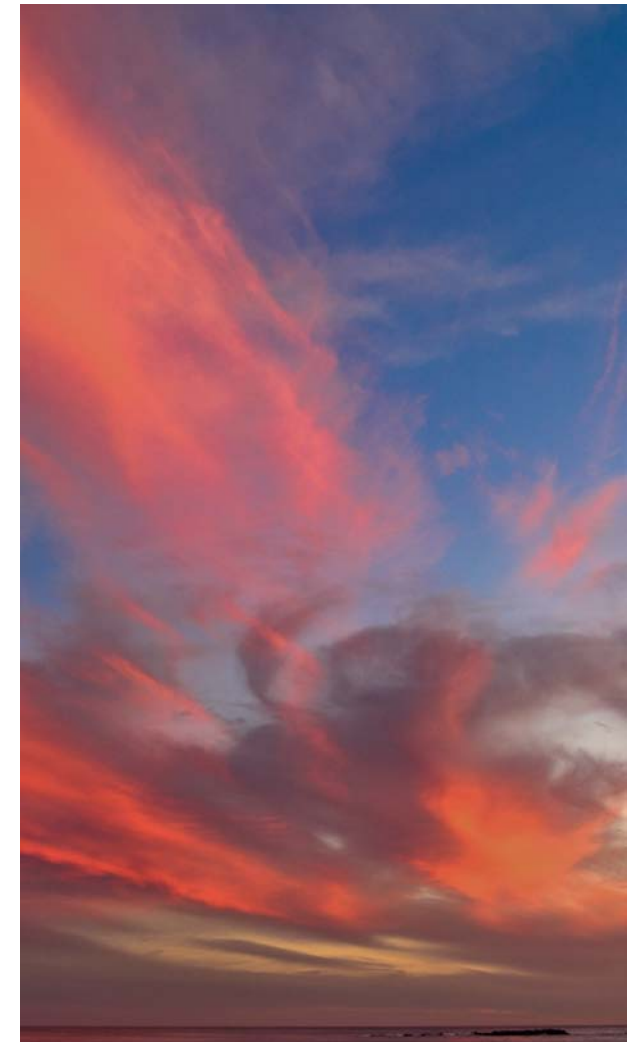Industrial Partners Program

42nd Symposium

Cloud Computing

Thursday, May 13th, 2010
8:30 AM - 5:30 PM * Room 368
Watson Center for Information Technology
Hosted by: Rodrigo Fonseca

ORRAN KRIEGER, VMware
Cloud Computing

Cloud computing and Virtualization are fundamentally transforming the computer industry. In this talk I will give a quick overview of VMware's overall vision for cloud computing, and then focus on VMware's approach to Infrastructure as a Service (IaaS). I will discuss first the fundamental abstractions necessary in an IaaS offering to make it of broad utility, then describe some of the technical challenges we see in implementing such an offering, and then discuss some of the long-term challenges that will require fundamental research to solve.

HARRY LI, Facebook
Doing What Works: Facebook's Innovation at Scale

Every computer scientist appreciates a contribution that is fast, efficient, and elegant. If a design has these properties, we sometimes say that it is the right solution. In practice, finding the right solution is a dark art at best, and more often, a hole down which many of us purists fall. Building systems that work well enough for now and leaving room to re-evaluate later is the more pragmatic approach. In this talk, I will describe two innovations that have enabled Facebook to scale with a user base that grows exponentially. One innovation can be thought of as "the right way" and the other as "it works well enough." Both have been critical to Facebook's success.

ADAM SILBERSTEIN, Yahoo!
Yahoo! Cloud Serving Benchmark (YCSB)

A prominent trend in cloud computing is the explosion in the number of systems targeting data serving. Unlike MapReduce systems, which target large-scale data analysis, cloud data serving systems support simple query types insert/update/get/delete).

One of the challenges we face at Yahoo!, and in the community at large, is that there are a large number of cloud serving systems, and it is hard to do an apples-to-apples comparison among them. In this talk I will discuss the open-source Yahoo! Cloud Serving Benchmark (YCSB). We have developed YCSB to facilitate comparison among cloud serving systems. I will describe our core benchmark, and experimental results from 4 systems and. I will also discuss some of the challenging comparison axes that remain open problems.

JOSEPH L. HELLERSTEIN, Google, Inc.
The Role of Quantitative Models in Building Scalable Cloud Infrastructures

Planetary scale cloud computing requires scalable infrastructures for compute, storage, network, and services that support programming models such as Map Reduce. There are many design choices that arise in the construction of cloud infrastructures. These design choices must be evaluated in terms of their impact on QoS considerations such as throughput, latency, and jitter as well as the consumption of power, compute, storage, and network bandwidth. This talk discusses ways in which Google uses quantitative models to evaluate design decisions for cloud infrastructures.

JEFF HAMMERBACHER, Cloudera
Evolving a New Analytical Platform: What Works and What's Missing

At Cloudera, we augment existing analytical platforms with some new tools for data management and analysis. In this talk, I'll share some experiences of what has worked across industries and workloads, and what new software components might help complete a new analytical platform.

ANDY PAVLO, Brown University
MapReduce and Parallel DBMSs: Together at Last

The MapReduce (MR) paradigm is heralded as a revolutionary new platform for large-scale, massively parallel data access. Some proponents claim that the extreme scalability of MR will relegate relational database management systems (DBMS) to the status legacy technology. In this talk, however, I will discuss the results from our recent benchmark study that suggest that using MR systems to perform tasks that are best suited for DBMSs yields less than satisfactory results [PPR+09]. This leads us to conclude that MR is more akin to an Extract-Transform-Load (ETL) system than a DBMS, as it is quickly able to load and analyze large amounts of data in an ad-hoc manner [SAD+10]. I will also discuss how the DBMS community has embraced MR technologies in the last year, and what features of DBMSs are being incorporated into popular open-source MR implementations.

SRIKANTH KANDULA, Microsoft Research
Combating Stragglers in MapReduce Networks

The phase structure of map-reduce jobs and resource sharing in cluster networks make task scheduling challenging. In particular, a few laggards can tremendously prolong job completion. This talk describes ClusterCull, an add-on to the job scheduler. ClusterCull culls stragglers based on their causes using resource-aware algorithms. From real-time progress reports, ClusterCull detects stragglers early in their lifetime, and takes appropriate action based on their causes. Early action frees up resources that can be used by subsequent tasks and expedites the job overall. Both trace-driven simulations and deployment in a 12K server cluster indicate that ClusterCull improves job completion time by an additional 3.1x over the existing state-of-the-art approach.

CHARALAMPOS PAPAMANTHOU, Brown University
Dynamic Provable Data Possession

Cloud storage is increasingly adopted by companies and individuals in order to save operating and maintenance costs. However, as remote severs may lose or modify data due to errors or malicious attacks, it is important to develop efficient methods that provide strong assurance of the integrity of the outsourced data. This talk presents Dynamic Provable Data Possession, a method that can be used to check the integrity of dynamic data stored in the cloud with minimal communication complexity.