ISTC for Cloud Computing: Center Overview

Greg Ganger & Phil Gibbons

ISTC-CC Retreat 2012

November 29, 2012

http://www.istc-cc.cmu.edu/

Why ISTC for Cloud Computing?

- Cloud computing has exploded onto the scenes
  - By 2020, most computing may be cloud-based
- Cloud computing promises huge benefits in…
  - resource efficiency and utilization
  - agility and productivity of application dev+deploy
  - system and service robustness
  - data sharing and end user access
- But, hype far ahead of capability & knowledge
  - huge need for research at many levels

ISTC-CC: Institutions & Faculty

- Carnegie Mellon University
  - Greg Ganger (PI), Dave Andersen, Guy Blelloch, Garth Gibson, Carlos Guestrin, Mor Harchol-Balter, Todd Mowry, Onur Mutlu, Priya Narasimhan, M. Satyanarayanan, Dan Siewiorek, Eric Xing
- Georgia Tech
  - Greg Eisenhower, Ada Gavrilovska, Ling Liu, Calton Pu, Karsten Schwan, Matthew Wolf, Sudha Yalamanchili
- Princeton University
  - Mike Freedman, Kai Li, Margaret Martonosi
- University of California at Berkeley
  - Anthony Joseph, Randy Katz, Ion Stoica
- Intel Labs
  - Phil Gibbons (PI), Michael Kaminsky, Mike Kozuch, Babu Pillai

Carnegie Mellon University Georgia Tech intel PRINCETON UNIVERSITY UC Berkeley

ISTC-CC: Intel Oversight & BoA

- Intel oversight
  - Wen-Hann Wang (Executive Sponsor)
  - Rich Uhlig (Managing Director)
  - Limor Fix (Director of UnCoR)
  - Chris Ramming (Director of UCO)
  - Jeff Parkhurst (Program Director)

- Board of Advisors (including Wen-Hann and Rich)
  - Jason Waxman (Gen. Mgr. Cloud Infra. Group, Intel)
  - Balint Fleischer (Gen. Mgr. Data Center Group, Intel)
  - Frans Kaashoek (Professor of CS&Eng, MIT)
  - Randy Bryant (Dean of School of CS, CMU)
  - Pradeep Khosla (Chancellor, UC San Diego)

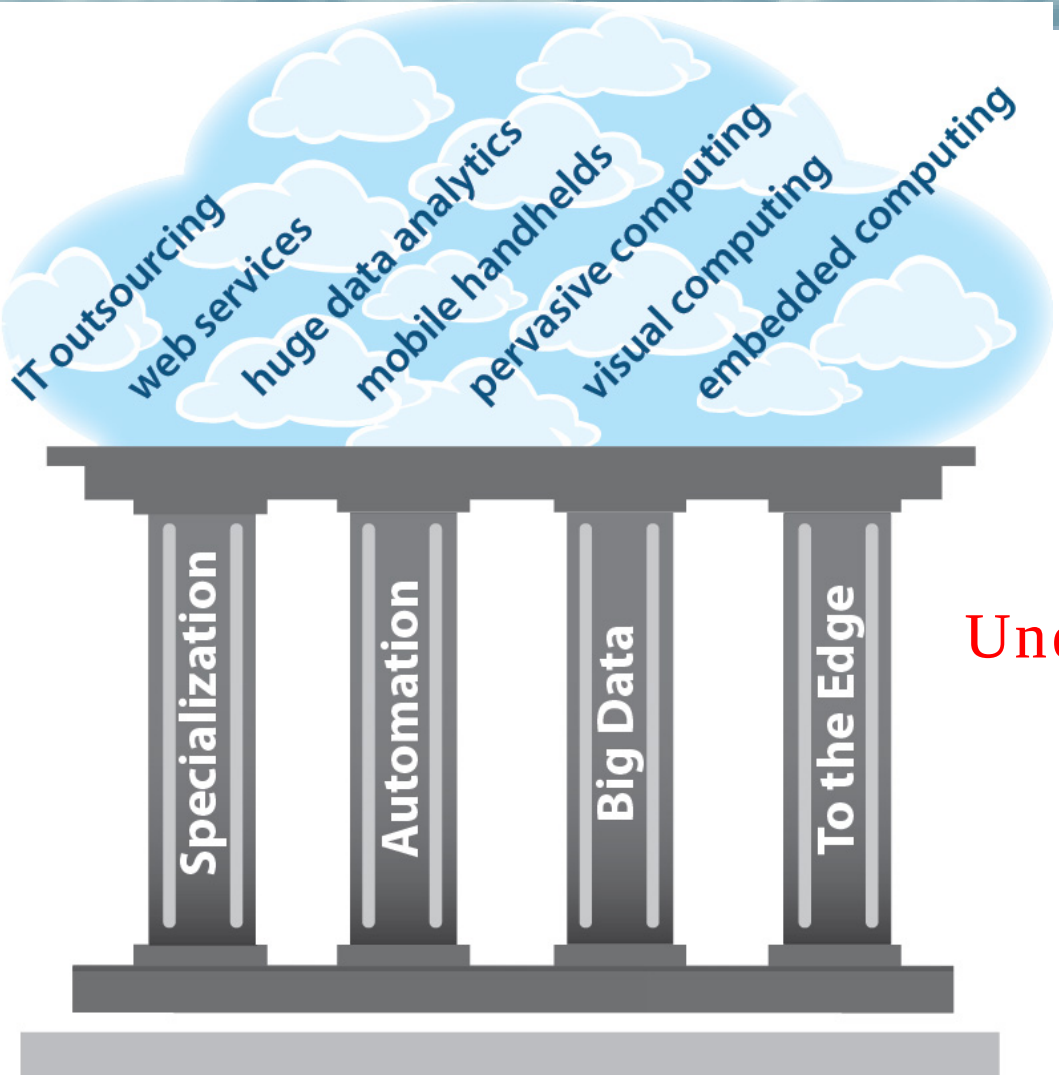
ISTC-CC Research Vision

Future Cloud Computing will:

- incorporate heterogeneous mixes of specialized platforms exploiting a variety of emerging technologies
- be dominated by big data analytics over stored and live data feeds
- encompass billions of edge devices through new paradigms for meshing clients and cloud
- rely on significant advances in automation to realize desired efficiency and productivity

ISTC-CC Pillars

ISTC-CC's Research Pillars



Underlying Infrastructure enabling the future of cloud computing

www.istc-cc.cmu.edu

Carnegie Mellon University
Georgia Tech
intel
PRINCETON UNIVERSITY
UC Berkeley

Automation Pillar

- Automation is crucial to cloud reaching potential
  - We suspect that no one here needs to be convinced of this…
- Management is very hard, but cloud makes it worse
  - Much larger scale
  - Much more varied mix of applications/activities
  - Much less pre-knowledge of applications
  - And, we're adding in platform specialization ☺
- Leaps forward needed on many fronts…
  - Diagnosis, scheduling, instrumentation, isolation, tuning, …

Automation Projects

- **A1: Adaptive sizing/scheduling of elastic services**
  - adaptive control over how many machines contribute to each service and work distribution among them
- **A2: Run-time validation of cloud software & patches**
  - improving productivity by automatically identifying problems in patches and other cloud software changes
- **A3: Problem diagnosis tools & techniques**
  - new scalable collection and analysis techniques to enable rapid, robust diagnosis of failures and performance problems for both providers and tenants (apps and services)
- **A4: Dynamic scheduling of heterogeneous mixes of frameworks and services**
  - new multi-level scheduling architectures enabling SW framework specialization (esp. Big Data) across specialized platforms

Automation at Retreat 2012

- Talks / Breakouts
  - “Performance Isolation and Fairness for Multi-Tenant Cloud Storage”, Mike Freedman
  - “Virtual Platforms: Hypervisor-Level Support for Increased Consolidation”, Ada Gavrilovska
  - “Stuff happens”, BoF Breakout on Problem Diagnosis
- Posters
  - vQuery: Tracking configuration + performance
  - Black-box localization of storage problems in parallel file systems
  - Diagnosing performance changes by comparing request flows
  - Dynamic power management for multi-tier data centers
  - CacheScale: saving cash by using less cache
  - alsched: Algebraic scheduling in heterogeneous clouds
  - Heterogeneity and dynamicity of clouds at scale: Google trace analysis
  - Problem localization for Hadoop using Draco
  - Mochi: Visual log-analysis based tools for debugging Hadoop
  - Guardrail: High fidelity checking of device drivers for I/O
  - JackRabbit: Improved Agility in Elastic Distributed Storage
  - Landslide: Systematic Dynamic Race Detection in Kernel Space
  - Oncilla – A GAS Run-time for Efficient Resource Partitioning in Data Centers
  - Exploratory testing at scale
  - Visual signatures for Hadoop diagnosis

Example Automation activities in posters

- **Different problem diagnosis approaches**
  - No single approach handles all scenarios… a collection of tools needed
  - One style exploits hardware/software counters
    - e.g., peer comparison to look for strange nodes
    - e.g., correlating performance metric changes with configuration changes
  - Another style exploits request flow tracing in distributed services
    - E.g., compare request flow graphs across two periods of time
    - E.g., find longest-latency paths in complex services-of-services
- **New approach to job scheduling in heterogeneous clouds**
  - Assigning the right resources to each of many different applications
  - Multi-level scheduling when the “apps” are parallel frameworks
- **New algorithms for elasticity of cloud services**
  - Enabling use in multi-tier services
  - Enabling use in stateful services, like Big Data storage

Big Data Pillar

Customer Database

~600 TB

HD Internet Video

12 EB/yr

Particle Physics

300 EB/yr

- Extracting insights from large datasets
  - "Analytics" or "Data-intensive computing"
  - Becoming critical in nearly every domain
    - likely to dominate future cloud data centers
- Need new programming/execution models
  - For productivity, efficiency, and agility
  - Resource efficient operation on shared, specialized infrastructures

10^6 MEGA 10^9 GIGA 10^{12} TERA 10^{15} PETA 10^{18} EXA

Estimating the Exaflood, Discovery Institute, January 2008
Amassing Digital Fortunes, a Digital Storage Study, Consumer Electronic Association, March 2008

Big Data Projects

- **B1: APIs and frameworks for advanced machine learning on Big Data**
  - new programming and system abstractions enabling efficiency and productivity for next-generation DISC
- **B2: Characterization and better programming of Big Data applications**
  - measuring, characterizing (benchmark creation), and building tools to enhance effectiveness of DISC application activities
- **B3: Online high ingest for Big Data cloud data stores**
  - exploring trade-offs and new approaches to supporting high-ingest data, including efficiency, consistency, and latency

Big Data at Retreat 2012

- Talks / Breakouts
  - 6 talks
    - Gibson, Guestrin, Ananthanarayanan, Stoica, Cipar, Willke
    - Describing different approaches to cloud support for Big Data computations
    - ... and one highlighting a success story of new collaboration with SAL
  - “Frameworks, frameworks everywhere!”, BoF Breakout

- Posters
  - GraphLab 2: Distributed graph-parallel computation on natural graphs
  - Big Machine Learning: needs and directions
  - Space-efficient, high-performance rank & select structures
  - LazyTables: Distributed data for machine learning
  - Scaling metadata in HDFS
  - Scaling Metadata Performance for POSIX Applications
  - Y-ray: Scalable real-time extraction of log file updates from VMs
  - A data-correlation-aware framework for sparse regression in the cloud
  - GraphChi: Large-scale graph computation on just a PC
  - PACMan: Coordinated memory caching for parallel analytics jobs
  - Performing cloud computation on a parallel file system
  - PLFS/HDFS: HPC applications on cloud storage

Specialization Pillar

Low power nodes

Many-core

Phase-change memory (PCM)

- Specialization is fundamental to efficiency
  - No single platform best for all application types
  - Called **division of labor** in sociology
- Cloud computing must embrace specialization
  - As well as consequent heterogeneity and change-over-time
  - Stark contrast to common cloud thinking
- New approaches needed to enable…
  - Effective mixes of targeted and general platform types
  - Nimble incorporation of new technologies and accelerators

Specialization Projects

- **S1: Specialized platforms of wimpy nodes**
  - exploring + extending range of apps for such platforms by overcoming OS limits, memory limits, and scalability issues
- **S2: Specializing heterogeneous many-core platforms for cloud computing**
  - exploring + extending range of apps for such platforms with new resource mgmt algorithms and portability abstractions
- **S3: Exploitation of NVM in cloud computing**
  - exploring how cloud apps can exploit NVM as well as needed system support and possible NVM specializations

Specialization at Retreat 2012

- Talks / Breakouts
  - “Architecting and Exploiting Asymmetry to Accelerate Bottlenecks in the Cloud”, Onur Mutlu
  - “We’re Still Having FAWN”, Dave Andersen
  - “Heterogeneous Parallelism and GPU Offloading: Optimization and Synchronization Challenges”, Margaret Martonosi
  - “Memory-Efficient GroupBy-Aggregate using Compressed Buffer Trees”, Hrishi Amur

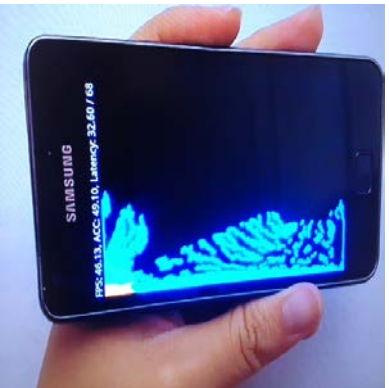
  - “There’s a processor in there?”, BoF Breakout
  - “An architectural melting pot”, BoF Breakout

Specialization at Retreat 2012

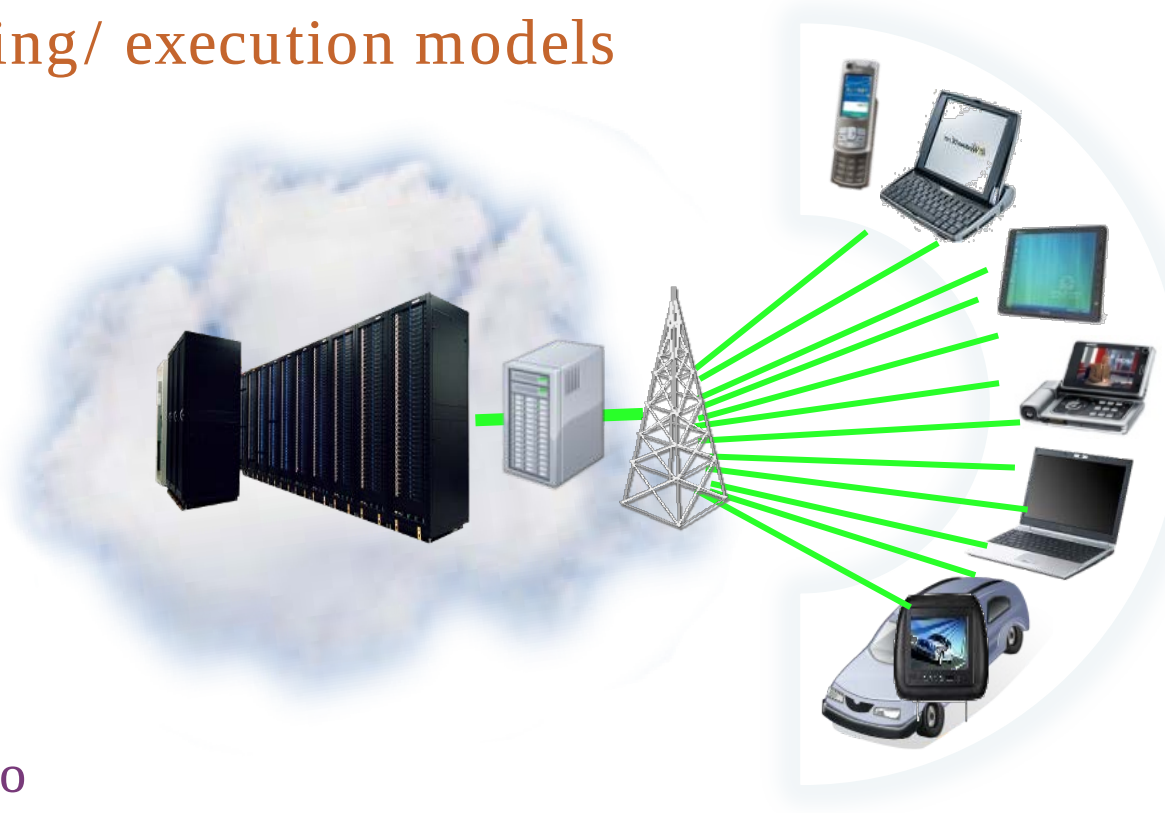
- Posters
  - **Techniques for Data Mapping and Buffering to Exploit Asymmetry in Multi-Level Cell (Phase Change) Memory**
  - Hashfilter: Making Bloomfilter Even More Compact and Deletable
  - MinBD: Minimally-Buffered Deflection Routing for Energy-Efficient Interconnect
  - Application-to-Core Mapping Policies to Reduce Memory Interference in Multi-Core Systems
  - MemC3: Compact and Concurrent MemCache with Dumber Caching and Smarter Hashing
  - **Row Buffer Locality Aware Caching Policies for Hybrid Memories**
  - Tiered-Latency DRAM: A Low Latency and Low Cost DRAM Architecture
  - Base-Delta-Immediate Compression: Practical Data Compression for On-chip Caches
  - Linearly Compressed Pages: A Main Memory Compression Framework with Low Complexity and Low Latency
  - HAT: Heterogeneous Adaptive Throttling for On-Chip Networks
  - **Staged Memory Scheduling**
  - A Case for Subarray-Level Parallelism (SALP) in DRAM
  - Analyzing and Optimizing GPU Communication and Computation
  - Memory-Efficient GroupBy-Aggregate using Compressed Buffer Trees
  - RAIDR: Retention-Aware Intelligent DRAM Refresh
  - MISE: Providing Performance Predictability in Shared Main Memory Systems
  - RowClone: In-DRAM Copy and Initialization of Bulk Data
  - **Identifying and Mitigating Memory Resource Contention to Accelerate Data-Intensive Workloads**

To the Edge Pillar

- Edge devices will participate in cloud activities
  - Serving as bridge to physical world (sense/actuate)
  - Enhancing interactivity despite location / connectivity
- Need new programming/ execution models
  - For adaptive cloud
    + edge cooperation



Cloudlet demo



To the Edge Projects

- **E1: Adaptive work division among clients, edge servers, and core cloud**
  - new programming frameworks and system architectures, including edge-local cloud resources, for dynamic exploitation of cloud resources to enable rich edge device experiences
  - example app: interactive perception

- **E2: Efficient, client-effective use of limited wide-area bandwidth**
  - new communication approaches for reducing reliance on expensive WAN-uplink bandwidth, including exploitation of new de-duplication and replication/consistency models

To the Edge at Retreat 2012

- Talks / Breakouts
  - "Beyond Offload: Cloudlets for Large-Scale Video Upload", Pieter Simeons
  - "Livin' on the edge", BoF Breakout
- Posters
  - Eiger: Stronger Semantics for Low-Latency Geo-Replicated Storage
    - Builds on their SOSP'11 paper (COPS)
  - Egalitarian Paxos
    - Removes serial bottleneck in Fault Tolerance thru State Machine replication
  - Rapid VM Synthesis for Cloudlet
    - Reduces critical path for offload from mobile to cloudlet
  - Scalable Crowd-Sourcing of Video from Mobile Devices
    - Using cloudlets for UPLOAD
  - Intelligent Virtual Appliance Delivery over Wide Area Networks

ISTC-CC: Cross-Institution Collaboration

	Project	Faculty
S1	Specialized Platforms of Wimpy Nodes	**Andersen[C], Schwan[G], Kaminsky[I], Kozuch[I], Pillai[I]**
S2	Specialized Platforms of Heterogeneous Many-Cores	**Mutlu[C], Mowry[C], Gavrilovska[G], Schwan[G], Yalamanchili[G], Martonosi[P], Gibbons[I], Kozuch[I]**
S3	Exploring the Disruptive Impact of NVM on Cloud Computing	**Andersen[C], Mutlu[C], Schwan[G], Kaminsky[I], Gibbons[I]**
A1	Mechanisms and Scheduling Algorithms for Elastic Sizing of Scalable Services	**Harchol-Balter[C], Ganger[C], Pu[G], Kozuch[I]**
A2	Run-time Validation of Cloud Software	**Mowry[C], Gibbons[I], Kozuch[I]**
A3	Problem Diagnosis	**Ganger[C], Narasimhan[C], Schwan[G], Eisenhauer[G], Liu[G], Wolf[G]**
A4	Robust Mixed-framework Scheduling of Data Center Resources	**Ganger[C], Stoica[B], Joseph[B], Katz[B], Kozuch[I]**
B1	Scalable and Efficient DISC for advanced machine learning algorithms	**Guestrin[C], Blelloch[C], Ganger[C], Gibson[C], Xing[C], Stoica[B], Gibbons[I]**
B2	Characterization, Classification and Better Programming of Big Data Apps	**Gibson[C], Ganger[C], Schwan[G], Pu[G], Katz[B], Blelloch[C], Gibbons[I], Kozuch[I]**
B3	Flexible and Efficient Ingest-intensive Semi-Structured Cloud Data Stores	**Gibson[C], Ganger[C], Schwan[G]**
E1	Cloud-Assisted Mobile Client Computations	**Satya[C], Siewiorek[C], Liu[G], Schwan[G], Gavrilovska[G], Martonosi[P], Pillai[I]**
E2	Wide-Area Replication, Consistency, and Deduplication	**Andersen[C], Freedman[P], Li[P], Kaminsky[I]**

ISTC-CC: Year 1 Publication Highlights

- SOSP (Oct'11) – 3 papers
- SOCC (Oct'11) – 4 papers
- SC (Nov'11) – 2 papers
- MICRO (Dec'11) – 6 papers
- ASPLOS (Mar'12) – 2 papers
- NSDI (Apr'12) – 3 papers
- EuroSys (Apr'12) – 5 papers
- CLOUD (Jun'12) – 3 papers
- HotCloud (Jun'12) – 6 papers
- ISCA (Jun'12) – 3 papers
- SPAA (Jun'12) – 3 papers
- IGCC (Jun'12) – 2 papers
- Usenix ATC (Jun'12) – 2 papers
- PACT (Sep'12) – 3 papers (accepted)
- OSDI (Oct'12) – 3 papers (accepted)

68 published papers highlighted in ISTC-CC Newsletter for Year 1 (July'11- June'12)

ISTC-CC: Year 1 Honors

Highlights of Year 1 Honors

- National Academy of Engineering (Li) [joining Katz, Siewiorek]
- ACM Fellow (Blelloch) [now 7 in all + 6 IEEE Fellows]
- IEEE Young Computer Architect Award (Mutlu)
- Sloan Foundation Fellow (Andersen)
- SIGOPS Hall of Fame paper (Gibson, Katz)
- SIGMOBILE Outstanding Contribution Award (Satya, Siewiorek)
- PECASE (Freedman)
- IEEE Technical Achievement Award (Liu)
- Endowed Professorship (Ganger, Martonosi)
- Multiple Best Paper awards, Joulesort wins, etc

ISTC-CC: Amplifying Funding

New Amplifying Funding awarded in Year 1

- Greg Ganger – NSF
- Ada Gavrilovska and Karsten Schwan – NSF, Samsung
- Garth Gibson – NSF, NSA, LANL (multiple awards)
- Mor Harchol-Balter – NSF
- Onur Mutlu – NSF
- M. Satyanarayanan and Dan Sieworek – NSF
- Ion Stoica, Anthony Joseph, Randy Katz – NSF Expedition
- Sudhakar Yalamanchili – NVIDIA, AMD

- Consortia: AMPlab (Berkeley), CERCS (GA Tech), PDL (CMU)
- Fellowships: Intel, Facebook, NSF, NSERC, CRA, etc.

+ Considerable Amplifying Funding carrying over into Year 1

=> Intel support for 25 students, yet working with 68

Agenda for Rest of Today

[10:30-11:15] Keynote: Balint Fleischer (GM, Intel Data Center Group)

[11:15-11:45] Poster previews (30 secs each) - Babu Pillai (Intel Labs) 

[11:45-1:15] Lunch /Poster session #1

[1:15-2:45] Poster session #2

[2:45-3:00] Break

[3:00-4:30] Research Talks session #1

- Garth Gibson (CMU), "Convergence of BigData Infrastructure for HPC & Internet Services"
- Carlos Guestrin (CMU/Washington), "GraphLab2: A Distributed Abstraction for..."
- Ganesh Ananthanarayanan (UC Berkeley), "PACMan: Coordinated Memory ..."

[4:30-4:45] Break

[4:45-6:15] Research Talks session #2

- Onur Mutlu (CMU), "Architecting and Exploiting Asymmetry to Accelerate Bottlenecks ..."
- Ada Gavrilovska (GA Tech), "Virtual Platforms: Hypervisor-level Support for Increased..."
- Dave Andersen (CMU), "We're Still Having FAWN"

[6:15-6:30] Group Photo  **then Transportation to dinner**

[6:30-9:00] Reception / Dinner at Pittsburgh Athletic Club

Agenda for Tomorrow

[8:30-9:00] Breakfast / Welcome

[9:00-9:45] Keynote: Das Kamout (Intel IT Cloud Lead)

[9:45-10:45] Research Talks session #3

- Ion Stoica (Berkeley), “Spark, Shark, and BlinkDB”
- Mike Freedman (Princeton), “Performance Isolation and Fairness for Multi-Tenant...”

[10:45-11:00] Break

[11:00-noon] Research Talks session #4

- Margaret Martonosi (Princeton), “Heterogeneous Parallelism and GPU Offloading...”
- Pieter Simoens (CMU), “Beyond Offload: Cloudlets for Large-Scale Video Upload”

[noon-12:20] Minutes of Madness - Michael Kaminsky (Intel Labs)

[12:20-1:45] Lunch / Birds-of-a-Feather Breakouts - Mike Kozuch (Intel Labs)

[1:45-3:15] Research Talks session #5

- Jim Cipar (CMU), "Exploiting Data Staleness for High-performance Machine Learning"
- Hrishi Amur (GA Tech), “Memory-Efficient GroupBy-Aggregate using Compressed...”
- Ted Willke (Intel SAL), “GraphBuilder: A SAL/ISTC-CC Collaboration Story”

[3:15-4:15] All-Hands Brainstorming - Jeff Parkhurst (Intel)

Concluding Thoughts

Retreat Goal: Benefit the research projects

- Community building
- Brainstorming/feedback (but not ARs ☺) on:
  - ideas & approaches for tackling the research challenges
  - jump-start collaborations
  - synergies & connections
- Learn, share & have fun

Who to see about...

- Poster Previews/Sessions: see Babu Pillai
- Minutes of Madness: see Michael Kaminsky
- Birds-of-a-Feather Breakouts: see Mike Kozuch
- Logistics questions/issues: see Jennifer Gabig