Datacenter Simulation Methodologies
Introduction

Tamara Silbergleit Lehman, Qiuyun Wang, Seyed Majid Zahedi
and Benjamin C. Lee

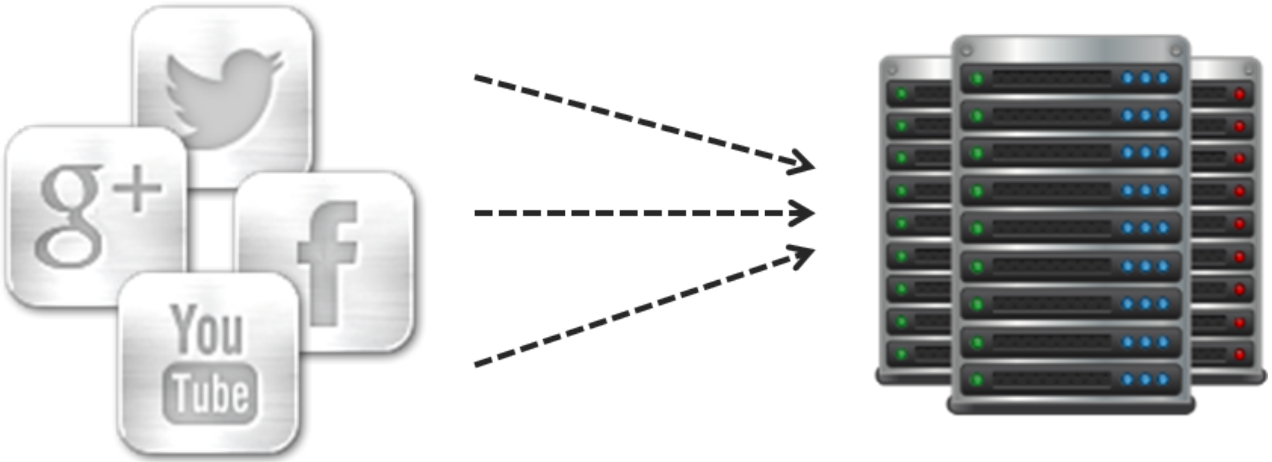
This work is supported by NSF grants CCF-1149252, CCF-1337215, and STARnet, a Semiconductor Research Corporation Program, sponsored by MARCO and DARPA.

Future of Computer Architecture

- Methodology supports general-purpose design
- Research bifurcates into mobile, datacenter systems

Figures: **[Google]**, **[Apple]**

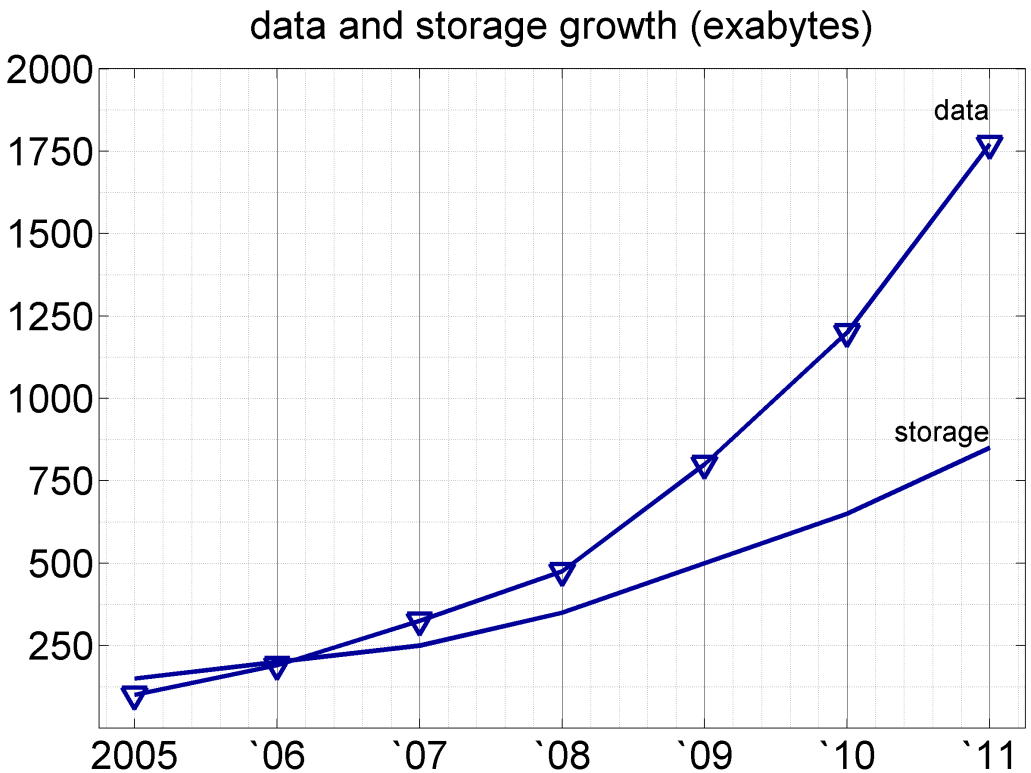
Cloud Computing Applications



- Data volumes are growing exponentially
- Cloud applications are diversifying rapidly
- Computing capability must grow

- Datacenters dissipate tens of megawatts

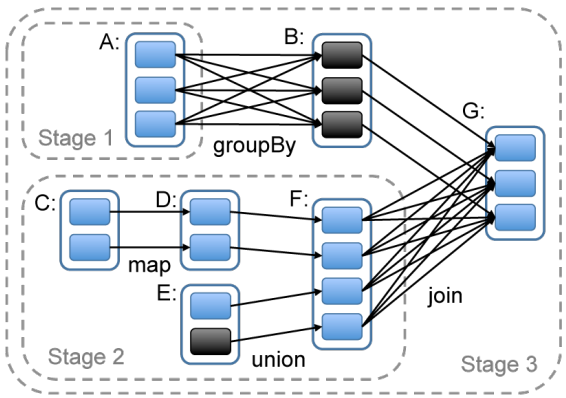
Data Deluge



- data ↑ 1.5×/yr
- storage ↑ 1.4×/yr
- datacenter demand ↑

Chart: **[IDC'08]**

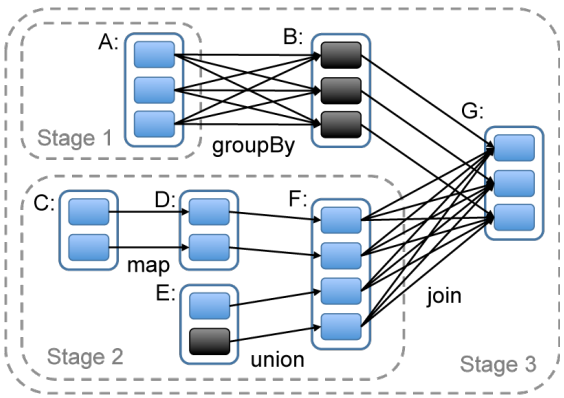
Big Data Domains

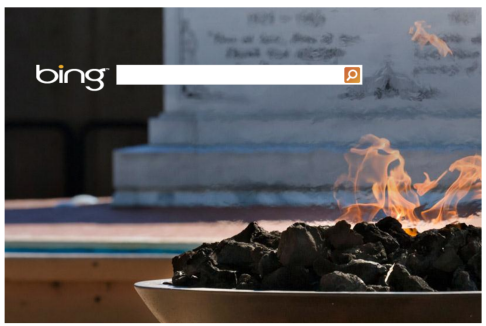


- **Machine Learning**
  MapReduce, Spark, GraphX, GraphLab

Images: **[memached]**, **[Microsoft]**, **[Amazon]**

Big Data Domains

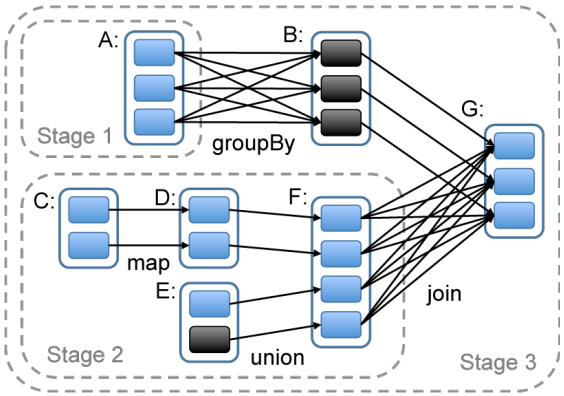


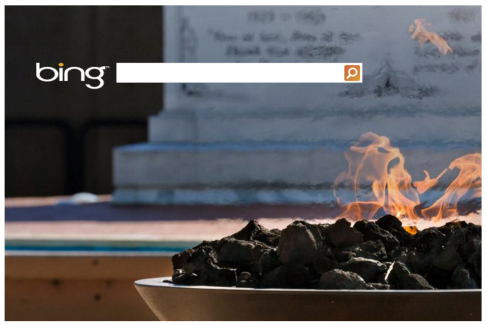


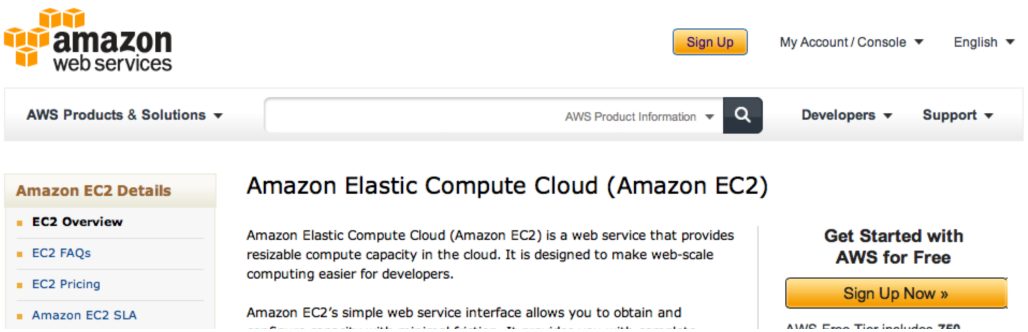
- **Machine Learning**
  MapReduce, Spark, GraphX, GraphLab
- **Search & Retrieval**
  Bing, Google, Solr

Images: **[memached]**, **[Microsoft]**, **[Amazon]**

Big Data Domains







- **Machine Learning**
  MapReduce, Spark, GraphX, GraphLab
- **Search & Retrieval**
  Bing, Google, Solr
- **Cloud Computing**
  EC2, Azure, AppEngine

Images: **[memached]**, **[Microsoft]**, **[Amazon]**

Big Computing and its Challenges

Big data demands big computing, yet we face challenges...

- Architecture Design
- Systems Management
- Research Coordination

Design Challenges

Pursue performance and energy efficiency

Design processors

- Microarchitecture (big versus small)
- Heterogeneity (big and small)

Design memory systems

- Technologies (DRAM, PCM, MRAM, etc.)
- Interfaces (DDRx, LP-DDRx, buffers, etc.)

Design communication

- Inter-processor
- Processor-memory
- Processor-accelerator

Duke UNIVERSITY

Management Challenges

Pursue performance and fairness

Allocation

- What resources are demanded?
- Analyze utility, preferences

Scheduling

- When are resources demanded?
- Analyze phase behavior

Sharing

- Which tasks to co-locate?
- What is each task's share?

Coordination Challenges

Anticipate management risk during design

- Design hardware, manager concurrently

Design for manageability

- Identify hardware to ease allocation
- Organize system to ease scheduling
- Reduce variance in task performance

Tutorial Objectives

By the end of the tutorial, participants will be able to...

- Deploy a full-system, cycle-accurate simulator
- Simulate processor and memory systems
- Simulate datacenter workload
- Design processor, memory systems for datacenter workloads

Duke UNIVERSITY

Tutorial Approach

We describe our practice and experience

- Describe experience in datacenter research as architects
- Describe strategies for integrating disparate frameworks
- Present a coherent methodology

We provide breadth and highlight existing frameworks

- Draw on related tutorials on simulators, applications
- Specify minimum steps for end-to-end experiments
- Refer to other materials for depth

Duke
UNIVERSITY

Tutorial Schedule

Time	Topic
09:00 - 09:30	Introduction
09:30 - 10:30	**Setting up MARSSx86 and DRAMSim2**
10:30 - 11:00	Break
11:00 - 12:00	Spark simulation
12:00 - 13:00	Lunch
13:00 - 13:30	Spark continued
13:30 - 14:30	GraphLab simulation
14:30 - 15:00	Break
15:00 - 16:15	Web search simulation
16:15 - 17:00	Case studies