

CS5950 / CS6030
Cloud Computing
Summer II 2017

<http://www.cs.wmich.edu/gupta/teaching/cs6030/6030cloudSummII17/cs6030cloud.php>

Ajay Gupta
B239, CEAS
Computer Science Department
Western Michigan University

ajay.gupta@wmich.edu
276-3104

WiSe Lab @ WMU Cloud Computing 2017 1

Acknowledgements

- I have liberally borrowed these slides and material from a number of sources including
 - Web, AWS Educate
 - MIT, Harvard, UMD, UPenn, UCSD, UW, Clarkson, . . .
 - Amazon, Google, IBM, Apache, ManjraSoft, CloudBook, . . .
- Thanks to original authors including Ives, Dyer, Lin, Dean, Buyya, Ghemawat, Fanelli, Bisciglia, Kimball, Michels-Slettvet, . . .
- If I have missed any, its purely unintentional. My sincere appreciation to those authors and their creative mind.

WiSe Lab @ WMU Cloud Computing 2017 2



WiSe Lab @ WMU Cloud Computing 2017 3

Source: <http://www.free-pictures-photo.com/>

The Next Revolution in IT The Big Switch in IT

<ul style="list-style-type: none"> • Classical Computing <ul style="list-style-type: none"> – Buy & Own – Hardware, System Software, Applications often to meet peak needs. – Install, Configure, Test, Verify, Evaluate – Manage – .. – Finally, use it – \$\$\$\$\$....\$(High Cost) 	<ul style="list-style-type: none"> • Cloud Computing <ul style="list-style-type: none"> – Subscribe – Use – \$ - pay for what you use, based on QoS
---	--

Every 18 months?

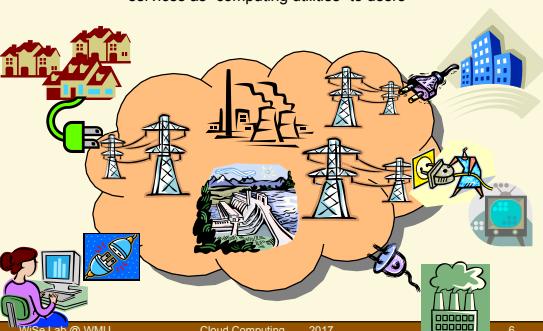
WiSe Lab @ WMU Cloud Computing 2017 4

Subscription-oriented & metered, Essential Utilities and Networks

<p>(1) Water</p> 	<p>Water Distribution Network</p> 
<p>(2) Electricity</p> 	<p>High Voltage</p>  <p>Power Grid</p> 
<p>(3) Gas</p> 	
<p>(4) Telephone</p> 	<p>Telecom Networks</p>  

WiSe Lab @ WMU www.cs.wmich.edu/wise 19 2017 5

Power Grid Inspiration for Computing?: Deliver IT services as "computing utilities" to users

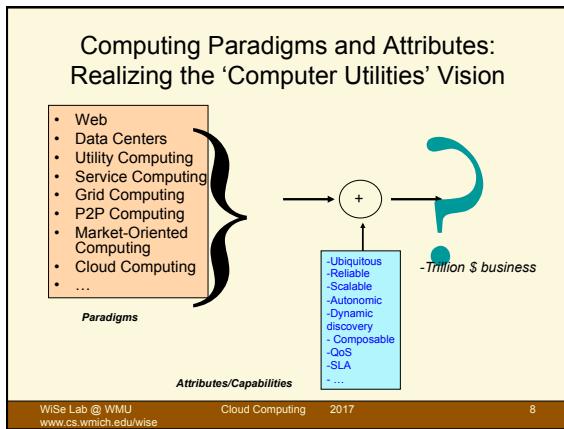


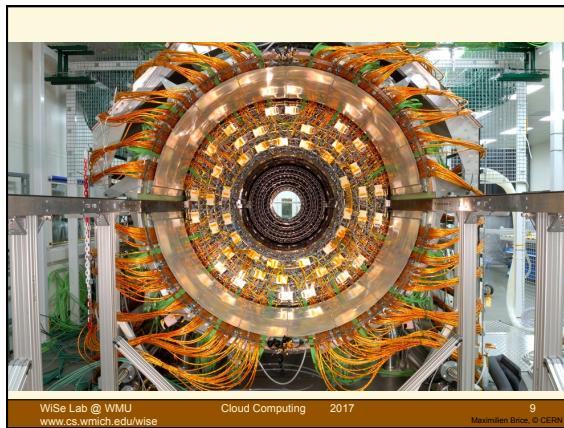
WiSe Lab @ WMU Cloud Computing 2017 6

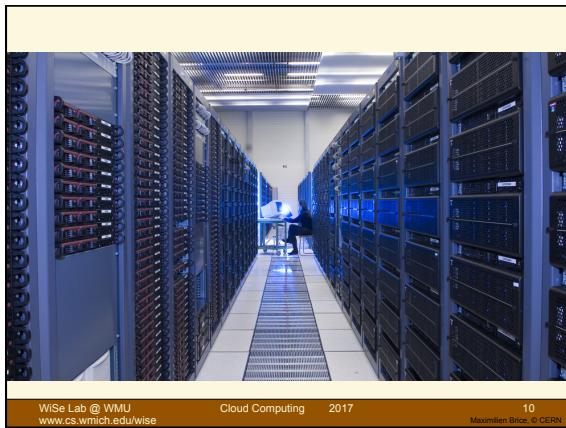
**“Computer Utilities” Vision:
Implications of the Internet**

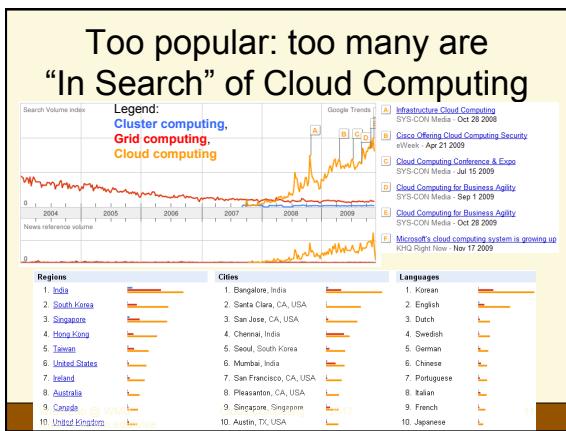
- 1969 – Leonard Kleinrock, ARPANET project
 - “As of now, computer networks are still in their infancy, but as they grow up and become sophisticated, we will probably see the spread of ‘computer utilities’, which, like present electric and telephone utilities, will service individual homes and offices across the country”
- Computers Redefined
 - 1984 – John Gage, Sun Microsystems
 - “The network is the computer”
 - 2008 – David Patterson, U. C. Berkeley
 - “The data center is the computer. There are dramatic differences between developing software for millions to use as a service versus distributing software for millions to run their PCs”
 - 2009 – “**The Cloud is the computer**” – anonymous...

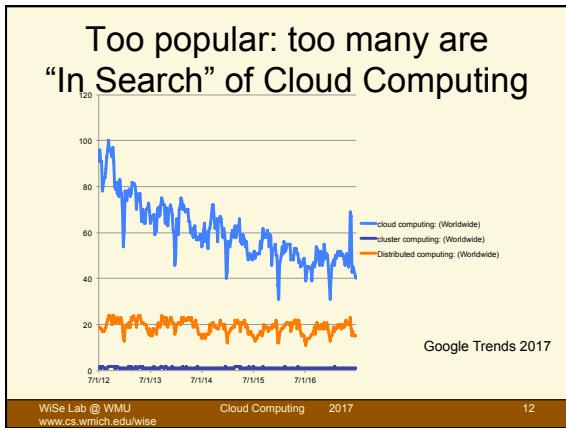
WiSe Lab @ WMU Cloud Computing 2017 7

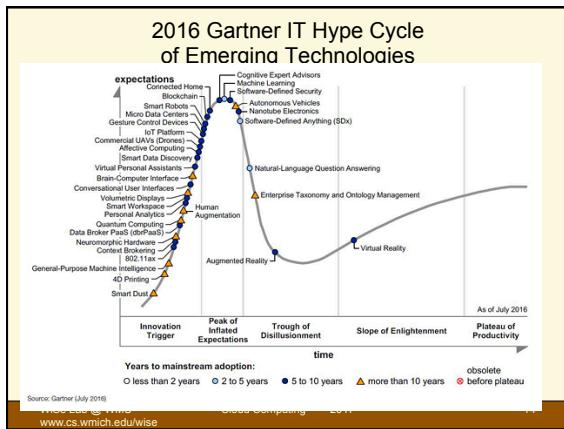
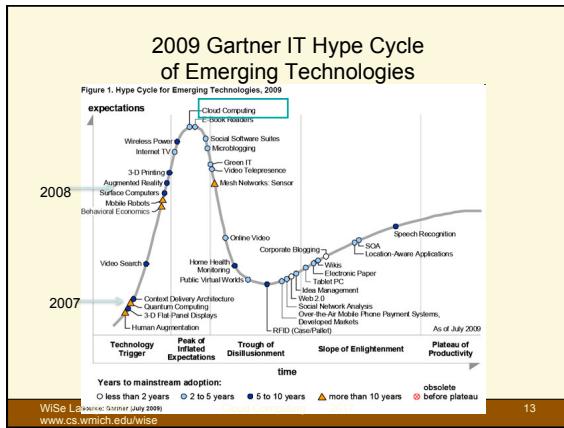












Top 10 for 2017

- Applied AI & Advanced Machine Learning
- Intelligent Apps
- Intelligent Things
- Virtual and Augmented Reality
- Digital Twins
- Blockchain & Distributed Ledger
- Conversational Systems
- Mesh App & Service Architecture
- Digital technology Platforms
- Adaptive Security Architecture

Gartner

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

16

Defining Clouds: There are many views for what is cloud computing?

- Over 20 definitions:
 - http://cloudcomputing.sys-con.com/read/612375_p.htm
- A compromised definition^②
 - "A Cloud is a type of parallel and distributed system consisting of a collection of inter-connected and **virtualised** computers that are **dynamically provisioned** and presented as one or more unified computing resources based on **service-level agreements** established through **negotiation** between the service provider and consumers."
- Keywords: Virtualisation (VMs), Dynamic Provisioning (negotiation and SLAs), and Web 2.0 access interface

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

17

Cloud Services

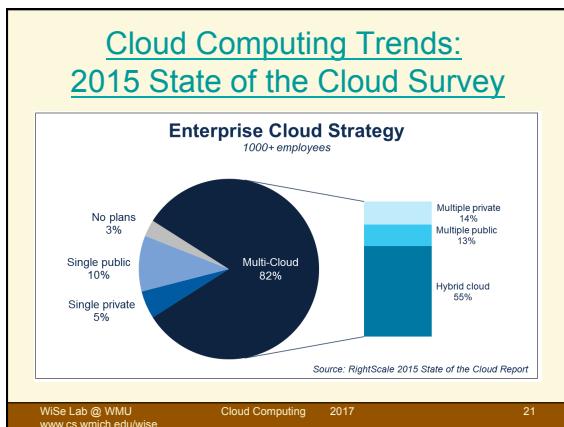
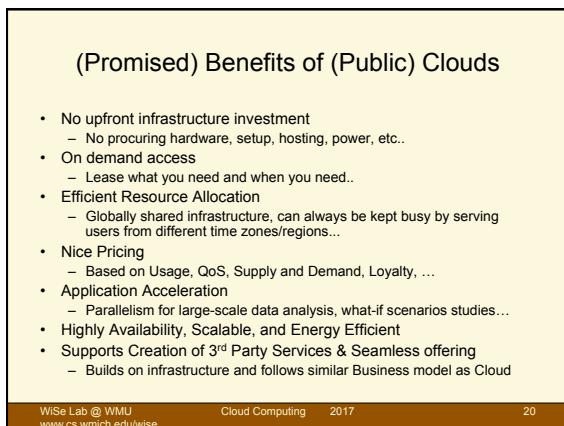
- Infrastructure as a Service (IaaS)
 - CPU, Storage: Amazon.com, Nirvanix, GoGrid....
- Platform as a Service (PaaS)
 - Google App Engine, Microsoft Azure, Manjrasoft Aneka..
- Software as a Service (SaaS)
 - SalesForce.Com

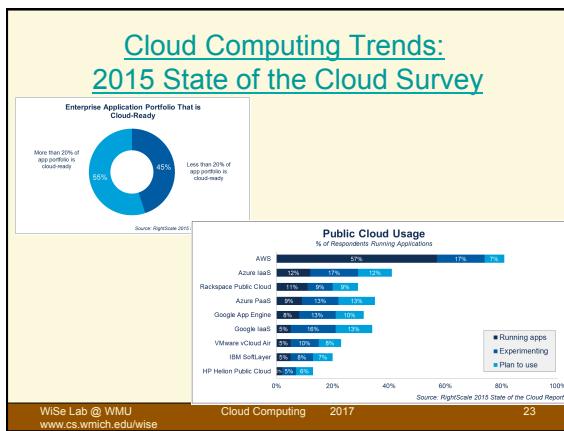
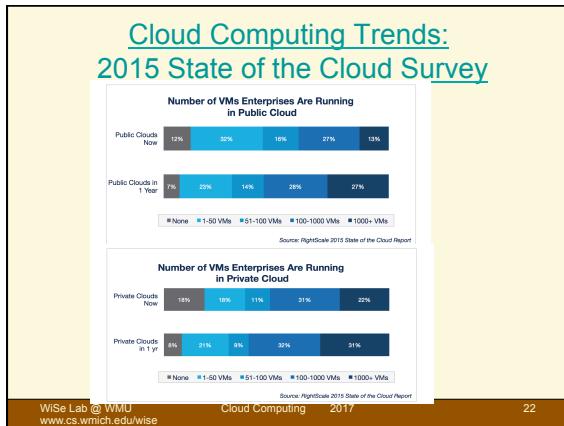
Software as a Service (SaaS)
Platform as a Service (PaaS)
Infrastructure as a Service (IaaS)

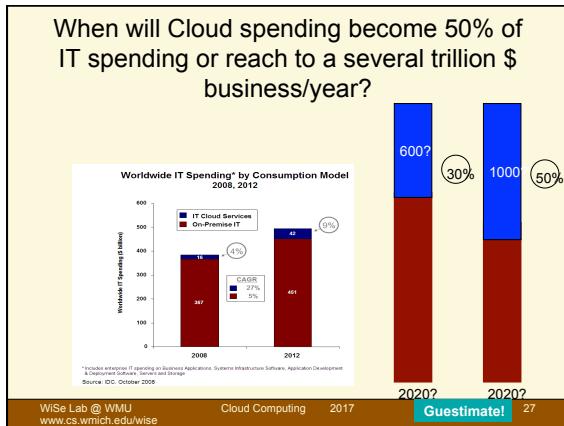
WiSe Lab @ WMU
www.cs.wmich.edu/wise

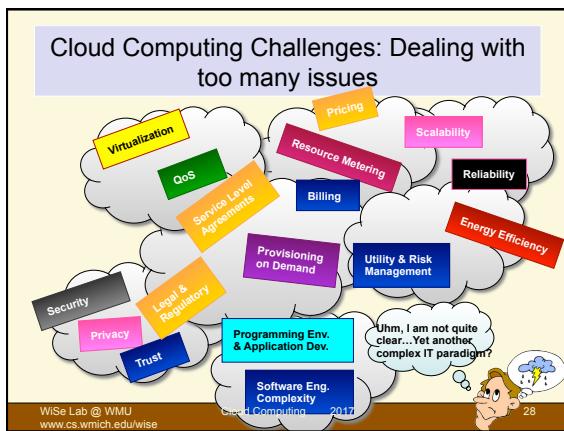
Cloud Computing 2017

18









What next?

- more on market oriented and technical view of cloud computing... and then onto basics...
- Course info

WiSe Lab @ WMU www.cs.wmich.edu/wise Cloud Computing 2017 29

What is the course about?

- Application programming using clouds
 - “toy and simple” programs
- System issues in realizing clouds
 - Alternatives, installation, ...
- MapReduce: the “back-end” of cloud computing
 - Batch-oriented processing of large datasets
- Ajax: the “front-end” of cloud computing
 - Highly-interactive Web-based applications
- Computing “in the clouds”
 - Amazon’s EC2/S3 as an example of utility computing

Amazon Web Services

- Elastic Compute Cloud (EC2)
 - Rent computing resources by the hour
 - Basic unit of accounting = instance-hour
 - Additional costs for bandwidth
- Simple Storage Service (S3)
 - Persistent storage
 - Charge by the GB/month
 - Additional costs for bandwidth
- Elastic Map/Reduce
- You’ll be using EC2/S3/... for course assignments!

This course is not for you...

- If you’re not genuinely interested in the topic
- If you’re not ready to do programming
- If you’re scared of reading new material and presenting
- If you’re not open to thinking about computing in new ways
- If you can’t cope with uncertainty, unpredictability, poor documentation, and immature software
- If you can’t put in the time

**Otherwise, this will be a richly rewarding course!
And we will all learn this exciting area together...**



Cloud Computing Zen

- Don't get frustrated (take a deep breath)...
 - This is bleeding edge technology
 - Those W\$#!T@F! moments
- Be patient...
 - This is the third time I'm teaching this course
- Be flexible...
 - There will be unanticipated issues along the way
- Be constructive...
 - Tell me how I can make everyone's experience better

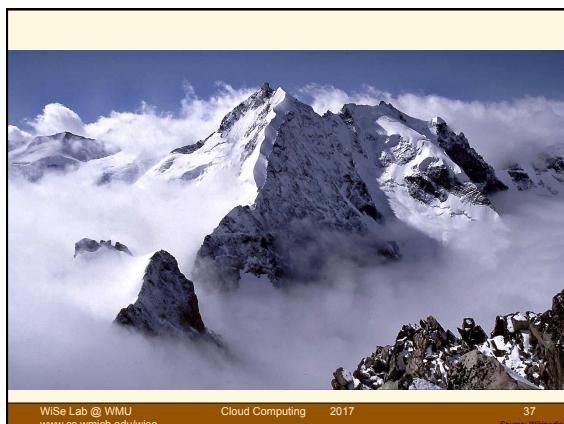
WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

34







Things to go over...

- [Course syllabus](#)
- Assignments and deliverables
- Projects
- Amazon EC2/S3/...
- AWS Certification

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

38

What next?

- more on market oriented and technical view of cloud computing... and then onto basics...

Welcome to Cloud and Scalable Computing 2017

Our goals in this course revolve around **scale** – of data, users, complexity:

- Understand **cloud computing** – how it lets us build **global-scale** services on shared hardware, why it's hard, how to use it
- Understand “**big data**” and using clusters of machines to handle data analysis at scale

Defining Clouds: There are many views for what is cloud computing?

- Over 20 definitions:
 - http://cloudcomputing.sys-con.com/read/612375_p.htm
- A compromised definition^①
 - "A Cloud is a type of parallel and distributed system consisting of a collection of inter-connected and **virtualised** computers that are **dynamically provisioned** and presented as one or more unified computing resources based on **service-level agreements** established through **negotiation** between the service provider and consumers."
- Keywords: Virtualisation (VMs), Dynamic Provisioning (negotiation and SLAs), and Web 3.0 access interface

What Is this Class about?

<p>"Cloud" computing</p> <ul style="list-style-type: none"> - The substance behind the hype - How "elasticity" and shared infrastructure are useful - The hard problems that need to be tackled by software - The different layers and services - How to build your own dynamic, "cloud hosted" software as a service - And how to use "platform as a service" tools to analyze Big Data 	<p>"Scalable" computing</p> <ul style="list-style-type: none"> - The issues in building global-scale services - Techniques for harnessing the power of thousands of machines to analyze data - The core ideas that span most tools, whether SQL, MapReduce, ... - Strengths and limitations of those approaches - Fundamental design patterns, operation primitives, algorithms
---	---

WiSe Lab @ WMU Cloud Computing 2017
www.cs.wmich.edu/wise 42

in a Nutshell

- How do we **build effective data-centric applications**, and serve them to the entire Internet?
 - You've learned procedural programming on a single machine – we'll look at data-centric programming across thousands of machines
 - We'll understand the issues in breaking up problems, global coordination, failures, and so on
 - We'll study many of the algorithms used by real Internet services
- How do we **take advantage of "the cloud"** – the vision of computing as a utility (like the power grid)?
 - You'll understand what lies underneath the cloud computing hype, and how to use the cloud
 - You'll build real Web projects hosted "on the cloud"

WiSe Lab @ WMU Cloud Computing 2017
www.cs.wmich.edu/wise 43

Do You Use "the Cloud"?



WiSe Lab @ WMU Cloud Computing 2017
www.cs.wmich.edu/wise 44

The Cloud in Your Pocket

Google Now is an example of the Cloud and Big Data, all in one
Hosted on Google's platforms in their data centers
Records your activities, runs large-scale analysis to make predictions!

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

45

The Cloud Enables Scale

Key Facts - Facebook's latest news, announcements and media resources - Mozilla Firefox

Platform

Statistics
830 million monthly active users who used Facebook mobile products as of June 2013.
699 million daily active users on average in June 2013.
Approximately 80% of our daily active users are outside the U.S. and Canada.

Board Members

Hank Zuckerberg, Founder, Chairman and CEO, Facebook
Mark Andreesen, Co-founder and General Partner, Andreessen Horowitz
David Hayes, Vice Chairman and President, Facebook, San Francisco (UCF)
Donald E. Graham, Chairman and CEO, The Washington Post Company
Reed Hastings, Chairman and CEO, Netflix
Ernest J. Moniz, Secretary, U.S. Department of Energy, The University of North Carolina
Peter Thiel, Partner, Founders Fund
Sheryl Sandberg, COO, Facebook

Yahoo says uploaded 1.4 billion photos in the wake of popular photo app Instagram, surpassing YouTube's record for most video uploaded in a day. Facebook, which sees more than 300 million photos uploaded each day, making it the most popular photo uploading service on the Internet.

Facebook's photo upload numbers are based on its "global measurement panel" of two million Internet users, similar to how Nielsen measures television audiences. Facebook says 1.4 billion photos is equivalent to the total amount of data it receives from more than 90 of the 100 publishers of Web content, but

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

46

The Cloud Enables Data-centric Computing

- Trend towards data-centric computing aka "big data"
- Today's currency on the Internet is data!
 - You "pay" for using Google, Facebook, etc. by letting them record your every action, link, search, etc.
- But data's value is not just economic:
 - It allows us to better answer questions, understand what's important, validate hypotheses about social interactions, ...
 - Example: Online Social Network research

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

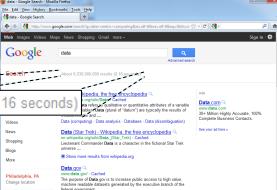
47

What kinds of data might we use?

- We all know that Google gives good search answers because it has a lot of data...
 - ... but what data how?

Let's consider...

About 6,030,000,000 results (0.16 seconds)



Is size everything?



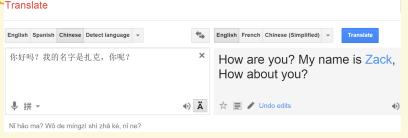
- Will the search engine with the most data have the best results?
 - Not necessarily - consider the above example from 2010
 - Search engines can crawl different sets of pages, and can rank them differently
 - What does this mean in the long term?

Sources: searchenginewatch.com, 6/20/10; http://www.mashable.com/2010/06/23/google-bing-data/

How does it work?

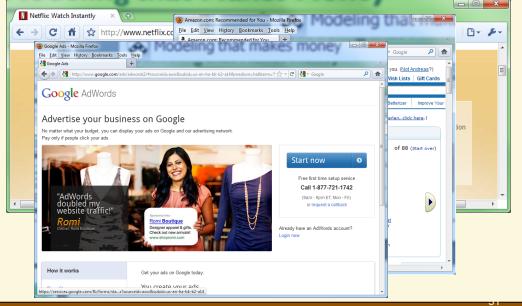


- Google pre-crawls and analyzes millions of images, sentences, searches, etc. – builds models



- Another example: Google & Bing Translate
 - Model how phrases in one language map to phrases in another language, by cross-comparing many, many translated docs

Models are valuable



WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

51

Data-centric computing is pervasive

- Today, Google and Friends aren't the only "Big Data" players
 - Not just Google & friends - banks, financial firms, academia, the government, companies, military, startups, ...
 - All need to store and analyze huge data volumes
- This is being enabled with a new generation of hardware "hosting" services – "the cloud" – and new, more scalable programming models

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

52

Should We Fear the Cloud and Big Data?



WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

53

The Cloud and Big Data Are Coming. Here!
Let's Understand How to Use Them for Good

- What does the technology let us do, and what remains hard?
- How do we ensure privacy, security, etc.?
- As we have ideas for useful, fun, helpful services, how do we make them happen?

Goals of the Course – Why You Should Be Here

- Understand what's underneath the Cloud
 - How does it work? What are its strengths? Its shortcomings?
 - Technologies: MapReduce, KVS/NoSQL, DHT, Ajax, XML, ...
- Understand the underlying principles
 - How do you build something that is so scalable, robust, etc.?
 - Lots of clever algorithms needed - very different environment!
- Be able to use the right approach when designing new protocols and web systems
 - How would you go about building the next Facebook?
 - Need to scale, be efficient, avoid failures, prevent intrusions, ...

Goals of the Course – Why You Should Be Here

- Gain practical experience with cloud technologies
 - Often, the best way to understand it is to build one yourself
 - In this course, ideally you should build a cloud-based application similar to mini-Facebook
- Understand the impact on society
 - Vulnerabilities, privacy concerns, data survivability, ...
 - Need to understand the current state of the technology!
- Anticipate what's possible in the future

Prerequisites and workload

- Necessary skills:**
 - Good **Java programming skills** (CS 3310 or equiv.)
 - No need to be a 'hacker'
 - But, if you don't have a lot of programming experience yet, you should be willing to invest the necessary time, esp. at the beginning
 - Some familiarity with Linux command line
 - Co-requisite: Foundation Courses (Algorithms, Data Structures, Systems...)
 - A willingness to "push the envelope"
 - Example: We may be using Node.js + Express this year, so you'll need to learn a bit of CSS and JavaScript
- Workload:**
 - Final group projects that integrate pieces from the class learning
 - Midterm and/or final exam

WiSe Lab @ WMU Cloud Computing 2017 57

Readings

This field is too recent, and too fast-moving to truly have a good textbook yet

- One required textbook
- AWS Certified Solutions Architect
- four reference books:
 - "Hadoop: The Definitive Guide, 3rd edition"
 - by Tom White (O'Reilly)
 - Cloud Computing for Machine Learning . By Kai Hwang
 - "Data Intensive Text Processing with MapReduce"
 - by Jimmy Lin & Chris Dyer
 - Mastering Cloud Computing: Foundations and Applications Programming by Raj Buyya

Supplementary handouts and Web references

- Refer to course web page:
<https://cs.wmich.edu/gupta/teaching/cs6030/6030cloudSumm17/TopicsCovered.html#SubmitInstrsCovered.html>
- Handouts, slides, example code, will be published there

WiSe Lab @ WMU Cloud Computing 2017 58

Policies: Collaboration

- All assignments must be done **individually**
 - Only exception is the final project (teams of two allowed)
 - All the code you submit has to be your own
 - Only exception: Code we have provided or explicitly authorized
 - WMU's Code of Academic Integrity applies
 - No cheating, plagiarism, fabrication, multiple submissions, gaining an unfair advantage, or facilitating (!) academic dishonesty
 - It's not worth it!! Penalties can be severe: **Zero tolerance policy** to ensure fairness
 - We will use various tools to actively look for cheating
 - These tools work: We have caught several cases in the past

WiSe Lab @ WMU Cloud Computing 2017 60

Policies: Collaboration

- Can we work on assignments together? Yes No
- Can I discuss the assignment with others (in general terms)? Yes No
- Can I use code I copied from the web? Yes No
- Can I ask questions about the assignments on web? Yes No
- I just happened to leave my svn password on my table, and XYZ just happened to find it. Will I be penalized for this? Yes No

Important Tip

- Please start your homeworks early!!!
- Please start your homeworks early!!!

Expected 'payoff'

- You will acquire a set of skills that is in very high demand right now
 - At Google, Facebook, and at many other places
 - Your projects should be useful in interviews
- You will gain interesting insights
- You will have a good basis for other courses or your research

A disclaimer...

- This is a "bleeding edge" course!
 - WMU is one of a handful of places offering these topics
 - The subject of this course is still evolving: no established curriculum, no classical textbooks yet
- Some of the material in the course will result in hair loss
 - Debugging distributed code is hard!
- We will be using some immature technology
 - We will do the best we can to smooth over the bugs
- I hope it will be a fun course, though...
 - ... and an interesting one!

WRAPPING UP: A SKETCH OF THE CLASS BY TOPIC & ASSIGNMENT



Cloud basics

- What is 'the cloud'?
 - Types of clouds; xaaS; utility computing
- What kinds of services does it provide?
 - AWS, EC2, EBS, S3, SimpleDB, key-value stores, ...
 - Amazon has donated free AWS usage credits for your homeworks
 - Case studies of cloud-based services
 - Cloud-based image search (Ajax, Node.js, SimpleDB, EC2, AWS)

Technical challenges

- What are key challenges in cloud computing?
 - Scale, concurrency, consistency, security, availability, ...
 - Speed of light (!), unreliable machines, ...
- What can you do about these challenges?
 - Concurrency control, weak consistency, locking, replication, ...
 - Special types of extremely scalable algorithms

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

67

Programming the cloud

- How to program a building-sized computer?
 - MapReduce programming model; Hadoop
 - **Homework #2?**: Geocoding with MapReduce and Google Maps
 - Beyond MapReduce: LINQ, Hive, PigLatin, XQuery, ...
- How to store/process petabytes of data?
Cloud file systems: HDFS, replication

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

68

Working with large data sets

- Useful algorithms and hypothesis testing:
 - How does YouTube recommend movies to you?
 - How does Google find the relevant pages for your search?
 - How does Goggles recognize images?
 - How can you (automatically) tell which emails are spam?
 - How does Facebook find out who your friends might be?
 - **Homework #3:** SocialRank

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

69

Interacting with the cloud

The diagram shows a network topology where a 'Client' computer is connected to a 'WMU' cloud. This cloud is interconnected with other clouds: 'Cogent', 'Google', 'AT&T', and 'Merit'. Within the WMU cloud, there are two servers labeled 'Servlet' and 'Servlet' with a 'Java' icon. A legend at the bottom left identifies these components.

- How do customers interact with the cloud?
 - Internet and web protocols: TCP, DNS, HTTP...
 - How to build a web server; Java servlets
 - Dynamic content: Ajax, Node.js; "Web 2.0"
 - Homework #4:** Social network visualizer
- How are cloud services interconnected?
 - Data interchange; XML; mash-ups

WiSe Lab @ WMU
www.cs.wmich.edu/wise Cloud Computing 2017 70

Putting it all together

A screenshot of a web browser displaying a Facebook-like application. The page shows a user profile for 'Mark Zuckerberg' with a profile picture, status update, and a news feed of posts from friends like 'John Smith' and 'Jane Doe'.

- Final project: Example build your own 'mini-Facebook' like app
 - Web frontend using GWT/Ajax, hosted on AWS
 - Cloud-based backend, using Amazon SimpleDB
 - Data analysis using MapReduce

WiSe Lab @ WMU
www.cs.wmich.edu/wise Cloud Computing 2017 71

Any questions?

A cartoon emoji of a thinking person with a yellow face, two question marks above their head, and a hand pointing to their chin in a thoughtful pose.

WiSe Lab @ WMU
www.cs.wmich.edu/wise Cloud Computing 2017 72

Stay tuned



Next time you will learn about:
Kinds of clouds; utility computing

WiSe Lab @ WMU
www.cs.wmich.edu/wise

Cloud Computing 2017

73
