

**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](mailto:ajay.gupta@wmich.edu)  
 276-3104

WiSe Lab @ WMU      Cloud Computing      2017      1  
[www.cs.wmich.edu/wise](http://www.cs.wmich.edu/wise)

---

---

---

---

---

---

---

---

**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  
[www.cs.wmich.edu/wise](http://www.cs.wmich.edu/wise)

---

---

---

---

---

---

---

---

**Welcome to Cloud and Scalable Computing 2017**

One of the goals in this course revolves 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

WiSe Lab @ WMU      Cloud Computing      2017      40  
[www.cs.wmich.edu/wise](http://www.cs.wmich.edu/wise)

---

---

---

---

---

---

---

---

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

- Over 20 definitions:
  - [http://cloudcomputing.sys-con.com/read/612375\\_p.htm](http://cloudcomputing.sys-con.com/read/612375_p.htm)
- A compromised definition©
  - "A Cloud is a type of parallel and distributed computing 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

WiSe Lab @ WMU      Cloud Computing      2017      41  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

---

---

**What Is this Class about?**

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

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

---

---

---

---

---

---

---

---

---

---

**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      43  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

---

---

## Do You Use “the Cloud”?



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

Cloud Computing 2017 44

---

---

---

---

---

---

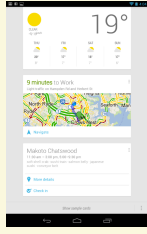
---

---

---

---

## 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

---

---

---

---

---

---

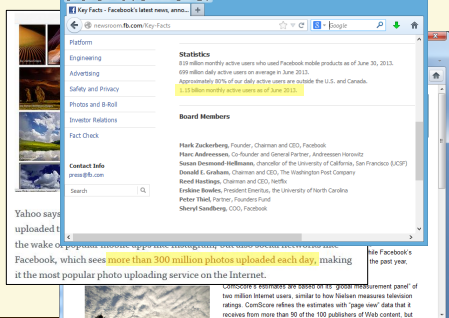
---

---

---

---

## The Cloud Enables Scale



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

Platform

- Engineering
- Advertising
- Safety and Privacy
- Photos and 3rd-Party
- Investor Relations
- Fact Check

Contact Info

press@fb.com

Search

Statistics

239 million monthly active users who used Facebook mobile products as of June 30, 2015.

600 million daily active users on average in June 2015.

Approximately 50% of our daily active users are outside the U.S. and Canada.

1.15 billion monthly active users as of June 2015.

Board Members

Mark Zuckerberg, Founder, Chairman and CEO, Facebook

Sheryl Sandberg, COO, Facebook

Facebook's the past year

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

Cloud Computing 2017 45

---

---

---

---

---

---

---


---

---

---

### 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

---

---

---

---

---

---

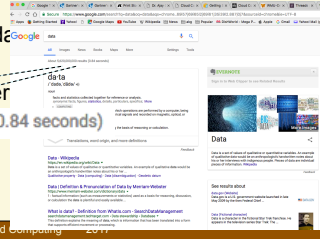
---

---

### 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



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

---

---

---

---

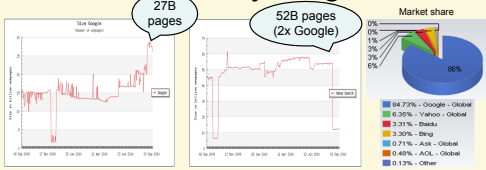
---

---

---

---

### 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?

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

---

---

---

---

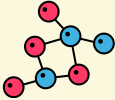
---

---

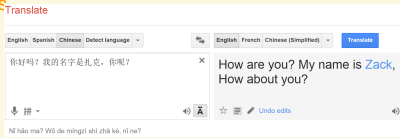
---

---

## 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

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

---

---

---

---

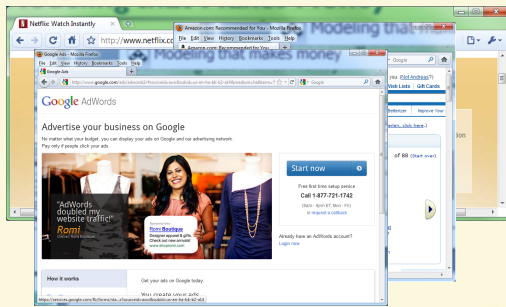
---

---

---

---

## Models are valuable



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

---

---

---

---

---

---

---

---

## 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      Cloud Computing    2017  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

### 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?

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

Cloud Computing 2017 54

---

---

---

---

---

---

---

---

### 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, ...

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

Cloud Computing 2017 55

---

---

---

---

---

---

---

---

## 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

WiSe Lab @ WMU      Cloud Computing    2017      56  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

## 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  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

## 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  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

### 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  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

### 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

WiSe Lab @ WMU Cloud Computing 2017 61  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

### Important Tip

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

WiSe Lab @ WMU Cloud Computing 2017 62  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---



**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

WiSe Lab @ WMU      Cloud Computing      2017      63  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

**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!

WiSe Lab @ WMU      Cloud Computing      2017      64  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

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

WiSe Lab @ WMU      Cloud Computing      2017      65  
www.cs.wmich.edu/wise

---

---

---

---


---

---

---

---

### 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)

WiSe Lab @ WMU      Cloud Computing      2017      66  
www.cs.wmich.edu/wise

---

---

---

---

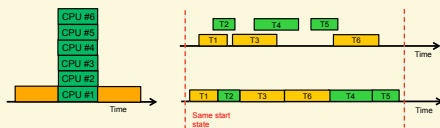
---

---

---

---

### 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      Cloud Computing      2017      67  
www.cs.wmich.edu/wise

---

---

---

---


---

---

---

---

### Programming the cloud



- How to program a building-sized computer?
  - MapReduce programming model; Hadoop
  - **Homework #nn?**: 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      Cloud Computing      2017      68  
www.cs.wmich.edu/wise

---

---

---

---



---

---

---

---

## Working with large data sets

[http://www.esa.int/gsp/visu/spacescience/esa\\_sdc](http://www.esa.int/gsp/visu/spacescience/esa_sdc)
Source: Bharath et al., WWW2009

- 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 #nnn**: SocialRank

WiSe Lab @ WMU  
www.cs.wmich.edu/wise
Cloud Computing 2017
69

---

---

---

---

---

---

---

---

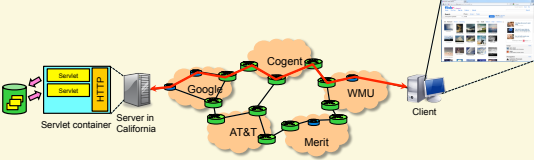
---

---

---

---

## Interacting with the cloud



- 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 #nnnn**: 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



- 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?



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

---

---

---

---

---

---

---

---