



Scalable Web Software

CS193S - Jan Jannink - 1/05/10

Administrative Stuff

- Jan Jannink: jannink@stanford.edu
- Aditya Acharya: acharyaa@cs.stanford.edu
- Ed Suh: edsuh@stanford.edu
- virtual office hours: phone / google talk
- web: <http://cs193s.stanford.edu/>
- Help us make the new course awesome

10 Years in the Trenches

- Transition from Student to Educator, Consultant to Angel Investor
- Unprecedented era of innovation
- Window of opportunity to share experiences
- Keeping content relevant wrt future

Goal: Think Big

- Build open source website as a team
- Use tools that support our processes
- Learn practices based on those tools
- Tackle simple and unsolved problems
- Get real world coding experience
- Scale from the laptop to the cloud

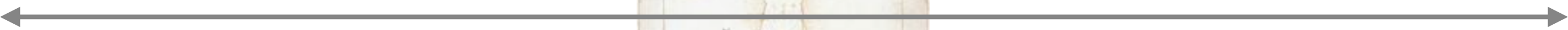
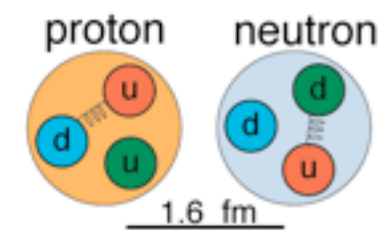
Weekly Syllabus

1. Scalability: (*Jan.*)
2. Agile Practices
3. Ecology / Mashups*
4. Browser / Client
5. Data / Server: (*Feb.*)
6. Security / Privacy
7. Analytics*
8. Cloud / Map-Reduce
9. Publish APIs: (*Mar.*)
*
10. Future

Projects

1. Set up Eclipse, GWT, SVN environment
2. Write basic unit tests & code (20%)
3. Set up Cruise Control & Ant
4. Main project - up to 3 coder team (50%)
5. Integrate with other teams (30%)

Scale

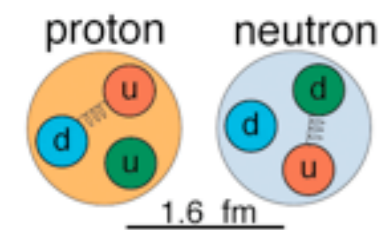


10^{-16} ...

10^{-4} 1 10^4

... 10^{26}

Scale



10^{-16} ...

10^{-4} 1 10^4

... 10^{26}

unaided humans operate in 1/5th the feasible range

Scale



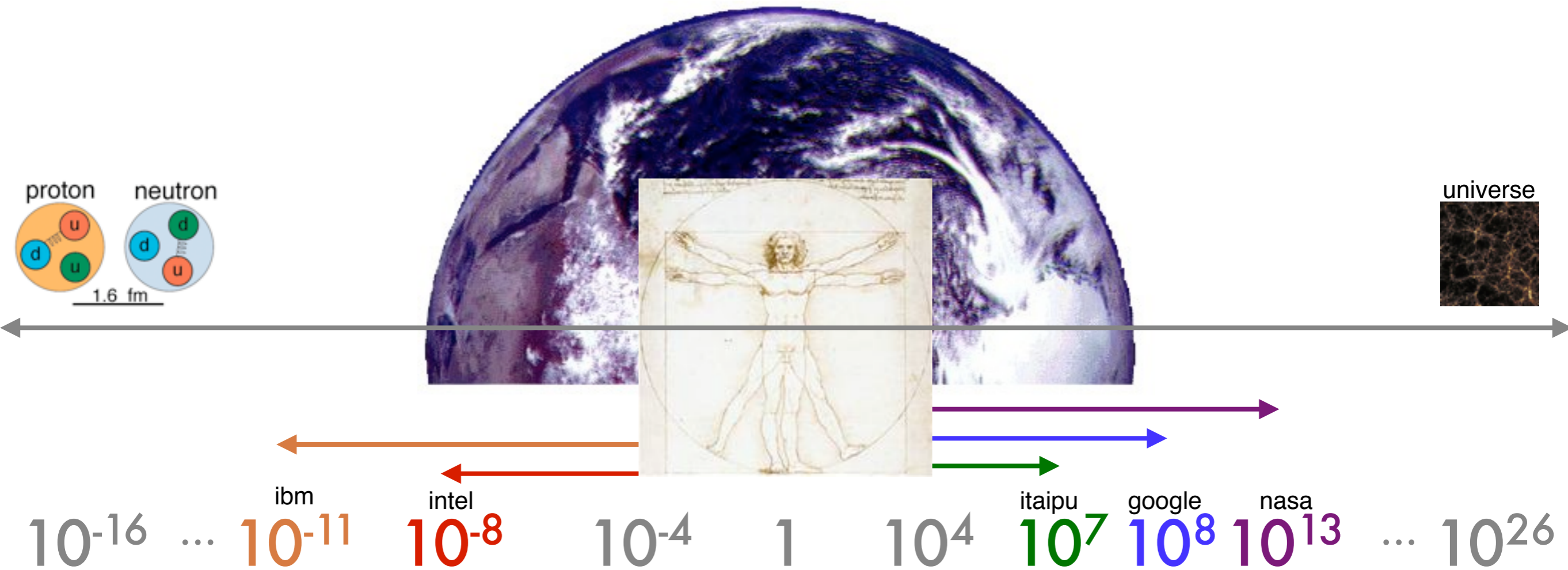
unaided humans operate in 1/5th the feasible range

Scale



unaided humans operate in 1/5th the feasible range
technology has effectively doubled our reach

Scale



unaided humans operate in 1/5th the feasible range
technology has effectively doubled our reach

Why Scalability Matters

- The Internet is the ultimate megaphone
- Good word of mouth results in growth
- 'Hockey stick' growth is exponential
- Many software processes are brittle
- Investors expect big returns

Napster Example

- Most popular IM client at its peak
- Incredibly scalable architecture
 - ~200 servers supporting up to millions of simultaneous users
- Provided illusion of ubiquity
- Multiple redundant dev teams
- Major release schedule impacts

Conceptual Framework

- Contrast business goals & science
- Visual perspective of innovation
- Motivate agile practices

Sample Competitor Slide

Feature 2

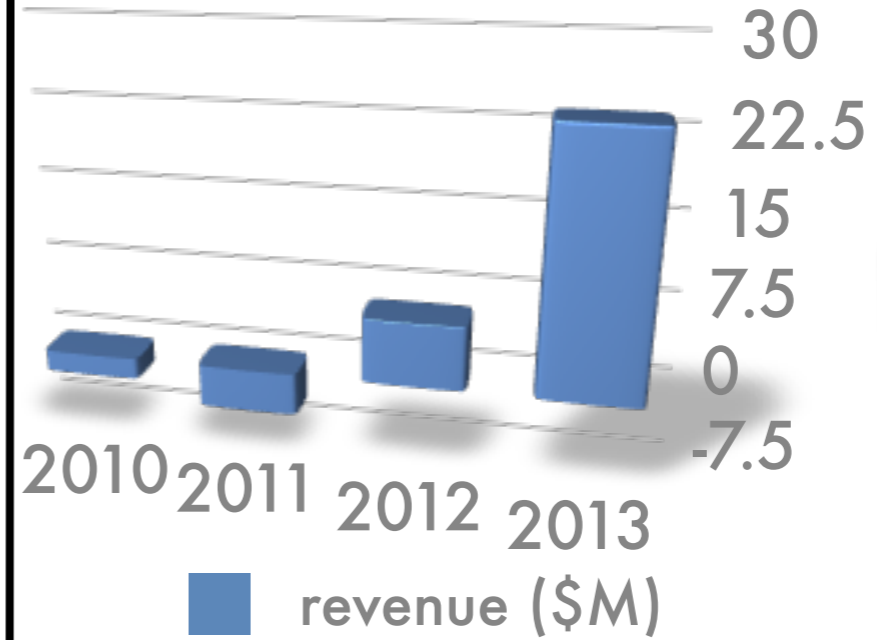


Feature 1



Sample Competitor Slide

Feature 2



Feature 1



Innovation

What We
Know We
Know

Innovation

What We Don't Know We Don't Know

What We
Know We
Know

Innovation

What We Don't Know We Don't Know

What We Don't Know We Know

What We Know We Don't Know

What We Know We Know

Innovation

What We Don't Know We Don't Know

What We
Want to
Achieve

What We Don't Know We Know

What We Know We Don't Know

What We
Know We
Know

Innovation

What We Don't Know We Don't Know

What We
Want to
Achieve

What We Don't Know We Know

What We Know We Don't Know

What We
Know We
Know

Innovation

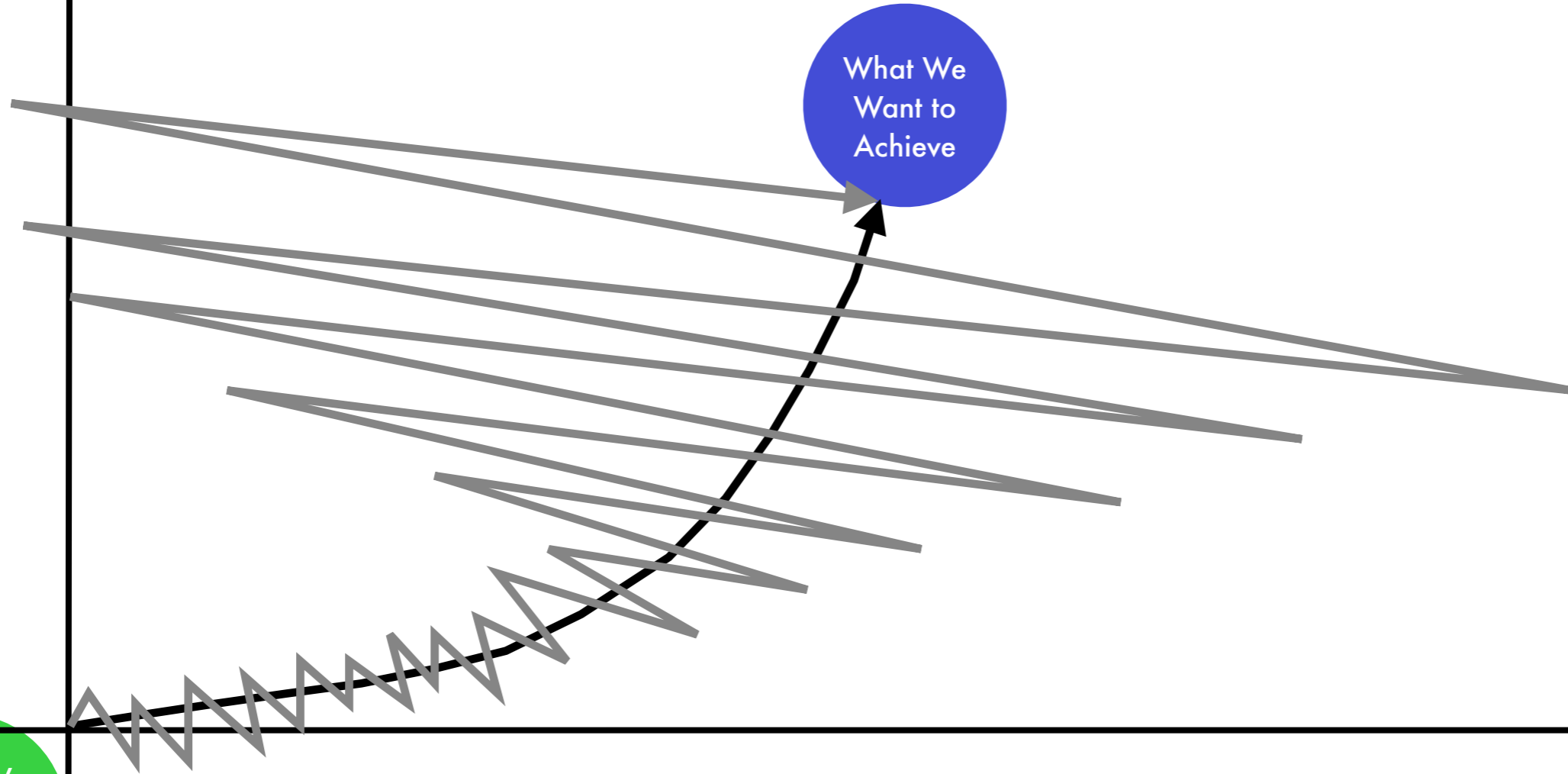
What We Don't Know We Don't Know

What We Know We Don't Know

What We Want to Achieve

What We Don't Know We Know

What We Know We Know



Effective Processes

- Open (accretive)
- Agile (adaptive & fast)
- Social (soft power)
- Iterative (divide & conquer)
- Sustainable (institutional memory)

Course Topics

- Scalability Science
 - people, processes
- Server & Network Scalability
 - data, caching, partitioning, replication
- Client & User Scalability
 - UI, usability, engagement, revenue

Tools to Try this Quarter

- Development: eclipse, junit, svn / git
- Deployment: cruise control, ant, trac
- Client: gwt, gears, yslow
- Server: memcached, mysql, hadoop
- Cloud: aws, app engine, cdn

Possible Projects

- Login Module
- Email Module
- Data Model
- User Profile
- User Linking
- Sharing data
- Multimedia embeds
- Facebook Connect
- Twitter Output
- LinkedIn API
- SEO tests
- Advertising

First Task

- set up google plugin for eclipse
- set up svn
- (extra credit: find easy way to use git)
 - i.e., explain install, use, maintaining
- get appengine account
- test out the sample gwt application

Q & A Topics

- Open Source license for our software
 - GPL, BSD
- What do we want to build?
- Some startup scalability experiences
 - MySimon, GigaBeat, Napster, imeem

Worth Checking Out

- Powers of Ten
 - <http://www.powersof10.com/>
- The Nature of Technology
 - W. Brian Arthur
- Google plugin for Eclipse
 - <http://code.google.com/eclipse/>

My Background

- Addicted to stepping out of my comfort zone
- 4 continents before finding the Bay Area
- Stanford CS PhD
- Repeat World Masters Games
Synchronized Diving Champion
- Father / Mentor / Advisor / Investor