Distributed Data-Intensive Systems

CSCE 438
Texas A&M University
Department of Computer Science and Engineering
Instructor: Prof. James Caverlee

Introduction / Administrivia
January 14, 2014
Distributed Objects Programming?
• Principles of distributed computing and programming with current paradigms, protocols, and application programming interfaces including Sockets, RMI, CORBA, IDL, Servlets, Web Services; security issues with public/private keys, digital signatures, forms and GUI based applications with multi-tier components, database connectivity and storing/streaming data structured using XML.
This course

• Introduction to the principles of distributed computing and programming in the context of the emerging data-centric view of computing (popularized by recent discussions of Big Data and Cloud Computing).
What do we mean by distributed computing?
Pros and Cons of Distribution?
Four Big Course Themes

• Peer-to-peer systems
• Web-based systems
• Cloud computing
• Crowd-powered systems (crowdsourcing)
Peer-to-Peer Systems
The Web
Cloud computing
IBM’s Big Plans for Cloud Computing

By QUENTIN HARDY

Google Joins a Heavyweight Competition in Cloud Computing
Ambition is an impressive thing, particularly when a desire for world domination is combined with existential survival.

Four heavyweight tech companies are translating that ambition into investments in their cloud computing services: IBM, Microsoft, Amazon and Google are all expected to spend more than $1 billion annually on their global networks in the coming years.

(NYTimes Dec 4, 2013)
Crowd-powered Systems
Crowdsourcing

Make tasks available for anyone online to complete

Quickly access a large user pool, collect data, and compensate users

Example: NASA Clickworkers

100k+ volunteers identified Mars craters from space photographs (14k hours of free work)

Aggregate results “virtually indistinguishable” from expert geologists
Soylent: A Word Processor with a Crowd Inside

This paper introduces architectural and interaction patterns that span many levels of conceptual and practical activity. Authoring tools offer help with pragmatism. We thus present Soylent, a word processing suite the Find-Fix-Verify crowd programming pat

tness, cost, wait time, and work time for edits.

Shortn

This paper introduces architectural and interaction patterns that span many levels of conceptual and practical activity. Authoring tools offer help with pragmatism. We thus present Soylent, a word processing suite the Find-Fix-Verify crowd programming pat

tness, cost, wait time, and work time for edits.

Crowdproof

Intuitive, but they didn’t let people be able to control computers efficiently developed by software companies, unless they know how. If one who knows nothing about programming needs to click through 100 buttons to complete a job everyday, the only thing she can do is simply to click through those buttons by hand every time. But if she happens to be a computer programmer, there is a little chance that she can write a program that automates everything. Why is there only a little chance? In fact, each GUI application is a big black

The Human Macro

Write a request:

Find Creative Commons figure for paragraph
2009 challenge to find 10 red weather balloons at unknown locations in the USA
Winner gets $40k
Free language education for the world

Sign up with Facebook

Sign up with email

Duolingo offers courses in:

- Spanish
- English
- French
- German
- Portuguese
- Italian

Looking for courses in other languages?
The Language Incubator allows the community to create new courses for Duolingo. See other languages or help start one.
Blend of Theory and Practice

• Let’s learn some theoretical foundations
  • e.g., Consistent hashing, bloom filters, …

• **BUT**, let’s get our hands dirty
  • Build a P2P app, build a crowdsourcing app
  • Write MapReduce jobs on Amazon Web Services
  • …
What is the course not?

- Theory of distributed computing
- Low-level distributed systems programming
- Networking class
- OS-level
- Web development
Syllabus

• http://courses.cse.tamu.edu/caverlee/csce438/
To Do

- Piazza
- Start reading!
- See you Thursday