HY559
Infrastructure Technologies for Large-Scale Service-Oriented Systems

Kostas Magoutis
magoutis@ics.forth.gr
http://www.ics.forth.gr/~magoutis

HY-559 Spring 2013
Course administration

• Meeting Thu/Fri 11-1pm at H.208

• Office hours by appt

• Tutorials will be occasionally held Fri 11-1am
  – News posted on web site and announced by email

• Communication
  – Send email to hy559@csd.uoc.gr to reach staff
  – For course-wide discussion
    • Send email “subscribe hy559-list” to majordomo@csd.uoc.gr
    • Subscribe to Google group hy559-s2013
Course administration

- Web site: http://www.csd.uoc.gr/~hy559

- Teaching assistant (office hours by appt)
  - Antonis Papaioannou

- Grade breakdown
  - Project 50%, final exam 40%, class participation 10%
Course administration

- Projects: Form teams of 3 and pick topics early
- Each team will report status to class bi-weekly
- Will use Microsoft Azure Cloud
Course overview

• Define services

• Types of services (software vs. hybrid)

• Features of a service
  – Description
  – Registration
  – Location
  – Invocation
Types of services

• Consulting engagements
  – IT optimization (servers, storage, etc)
  – Information lifecycle management, tech refresh

• Managed services
  – Traditional (pre-2008) model
  – Cloud computing

• Outsourcing
  – Keep own infrastructure or use provider’s?
  – Tough deal (due diligence, price 10-yr engagement)
Service-access software technologies

• Remote procedure calls, distributed objects
  – CORBA
  – Java RMI
  – Sun RPC
  – Web Services / SOAP
Virtualization technologies

- Virtual machines
- Virtual networks
- Virtual storage
Large-scale software systems

• Overall view

• Scalable persistence (tier 3)

• Session (soft-state) caching (tier 2)

• Orthogonal issues
  – High availability
  – Service-level agreements
Scalability goals and requirements

• Expandability
  – Increase system size (capacity) as needed

• Performance
  – Increase linearly with system size

• Availability
  – Survive failures gracefully

• Manageability
  – React to changes automatically
Amazon service-oriented architecture

Rendering service may construct its response by sending requests to over 150 other services

Stateless but may use caching

Each service in the call chain must obey performance contract

HY-559 Spring 2013
Amazon e-commerce platform

- World-wide operation, several data centers
- Tens of millions of customers at peak times
- Strict operational requirements
  - Performance, reliability, efficiency, continuous growth
- Reliability, availability amongst most important reqs
  - Dependent upon how application state is managed
- Need “always-writable” store, despite failures
Infrastructure management

• Data center management technologies
  – Microsoft Autopilot
  – HP SmartFrog

• Configuration and management databases (CMDBs)

• ITIL processes

• Business-IT dependency discovery and use

• The role of the human factor