Outline

• Quick recap of what we have learnt as part of this course
• How to prepare for the examination
What is Cloud Computing?

• Computing as a utility
  – Consumers pay based on their usage

• Cloud Computing characteristics
  – Illusion of infinite resources
  – No up-front cost
  – Fine-grained billing (e.g. hourly)

• Gartner: “Cloud computing is a style of computing where massively scalable IT-related capabilities are provided ‘as a service’ across the Internet to multiple external customers”
Cloud Computing - Services

• **Software as a Service – SaaS**
  – A way to access applications hosted on the web through your web browser

• **Platform as a Service – PaaS**
  – Provides a computing platform and a solution stack (e.g. LAMP) as a service

• **Infrastructure as a Service – IaaS**
  – Use of commodity computers, distributed across Internet, to perform parallel processing, distributed storage, indexing and mining of data
  – Virtualization

- **SaaS**
  - Facebook, Flikr, Myspace.com, Google maps API, Gmail

- **PaaS**
  - Google App Engine, Force.com, Hadoop, Azure, Heroku, etc

- **IaaS**
  - Amazon EC2, Rackspace, GoGrid, SciCloud, etc.
Cloud Computing - Themes

• Massively scalable
• On-demand & dynamic
• Only use what you need - Elastic
  – No upfront commitments, use on short term basis
• Accessible via Internet, location independent
• Transparent
  – Complexity concealed from users, virtualized, abstracted
• Service oriented
  – Easy to use Service Level Agreements
Cloud Computing Progress

• Short term and long term implications of cloud
• Economics of cloud users and cloud providers
• Challenges and opportunities offered by cloud computing
Cloud Providers

• Amazon Web Services
  – EC2, S3, EBS, Elastic Load Balancing, Amazon Auto Scale, Amazon CloudWatch, IAM, Elastic Beanstalk, CloudFormation, Simple Workflow Service, Elastic MapReduce

• Private Cloud enabling technologies
  – Eucalyptus
  – OpenStack
    • Worked with SciCloud

• PaaS
  – Google AppEngine
  – Windows Azure
Scaling Applications on the Cloud

• Two basic models of scaling
  – Vertical scaling, aka Scale-up
  – Horizontal scaling, aka Scale-out

• Scaling Enterprise Applications in the Cloud

• Load balancing
  – Types and algorithms

• Autoscaling
Economics of Cloud Providers

• Cloud Computing providers bring a shift from high reliability/availability servers to commodity servers
  – At least one failure per day in large datacenter

• Why?
  – Significant economic incentives
    • much lower per-server cost

• Caveat: User software has to adapt to failures
  – Very hard problem!

• Solution: Replicate data and computation
  – MapReduce & Distributed File System
MapReduce

• Programmers specify two functions:
  \[ \text{map } (k, v) \rightarrow <k', v'> \]
  \[ \text{reduce } (k', v') \rightarrow <k', v'> \]
  – All values with the same key are reduced together
• The execution framework handles everything else...
• Not quite...usually, programmers also specify:
  \[ \text{partition } (k', \text{number of partitions}) \rightarrow \text{partition for } k' \]
  – Often a simple hash of the key, e.g., hash(k') mod n
  – Divides up key space for parallel reduce operations
  \[ \text{combine } (k', v') \rightarrow <k', v'> \]
  – Mini-reducers that run in memory after the map phase
  – Used as an optimization to reduce network traffic
Hadoop Processing Model

• Create or allocate a cluster
• Put data onto the file system (HDFS)
  – Data is split into blocks
  – Replicated and stored in the cluster
• Run your job
  – Copy Map code to the allocated nodes
    • Move computation to data, not data to computation
  – Gather output of Map, sort and partition on key
  – Run Reduce tasks
• Results are stored in the HDFS
MapReduce Examples

- Distributed Grep
- Count of URL Access Frequency
- Reverse Web-Link Graph
- Inverted Index
- Distributed Sort
Synchronization in Hadoop

• Approach 1: turn synchronization into an ordering problem
  – Sort keys into correct order of computation
  – Partition key space so that each reducer gets the appropriate set of partial results
  – Hold state in reducer across multiple key-value pairs to perform computation
  – Illustrated by the “pairs” approach in calculating conditional probability of words

• Approach 2: construct data structures that “bring the pieces together”
  – Each reducer receives all the data it needs to complete the computation
  – Illustrated by the “stripes” approach
Research @ Mobile & Cloud Lab

- Scientific computing on the cloud
- Classification on how the algorithms can be adapted to MR
- Limitations of Hadoop MapReduce
  - Alternatives
- Mobile Cloud
  - Mobile Cloud Binding Models
This week in lab

• Complete the work from previous labs
Examination

• Exam I : Monday, 04.05.2015, 10:00 – 12:00 in room J. Liivi 2 - 122
• Exam II : Tuesday, 05.05.2015, 10:00 – 12:00 in room J. Liivi 2 - 202
Preparation for Examination

• One of the earlier exam paper is kept online
• Slides are mostly self sufficient
• References are mentioned for further reading
• Mainly focus at keywords
• Let us discuss some questions...
References

• Basics of Cloud Computing - Lectures
  https://courses.cs.ut.ee/2014/cloud/spring/Main/Lectures