# Written examination in Distributed Systems

July 15th 2014

Last name:
First name:
Student number:
I confirm with my signature that I will process the written examination alone and that I feel healthy and capable to participate this examination.
I am aware, that from the moment, when I receive the written examination, I am a participant of this examination and I will be graded.
Signature:

- Provide on all sheets (including the cover sheet) your *last name*, *first name* and *student number*.
- Use the provided sheets. Own paper must *not* be used.
- Place your *ID card* and your *student ID card* on your table.
- You are allowed to use a *self prepared*, *single sided DIN-A4 sheet* in the exam. Only *handwritten originals* are allowed, but no copies.
- You are allowed to use a non-programmable calculator.
- Answers, written with pencil or red pen are *not* accepted.
- Time limit: 90 minutes
- Turn off your mobile phones!

#### **Result:**

Question:	1	2	3	4	5	6	7	8	9	10	11	12	13	Σ	Grade
Maximum points:	4	6	5	7	7	12	6	4	9	5	14	5	6	90	
Achieved points:															

Student number:

# Question 1)

Points: .....

Maximum points: 4

How long does it take to transfer 7.5 TB via a 1 Gbps (= 1,000 Mbps) Ethernet?

#### Question 2)

Points: .....

Maximum points: 1+1+1+1+1+1=6

a) What is the central statement of Moore's law?

b) What is the Von Neumann bottleneck?

- c) How can the Von Neumann bottleneck be weakened?
- d) What is the central statement of Amdahl's law?

- e) Which important factor is ignored by Amdahl's law?
- f) What is the central statement of Gustafson's law? (Highlight the difference against Amdahl's law.)

# Question 3)

Points: .....

Maximum points: 1+1+1+1+1=5

a) Describe the shared memory architecture in just a few words.

b) Name two challenges of shared memory architectures.

c) What is the difference between asymmetric and symmetric multiprocessing (SMP)?

d) Describe the distributed memory architecture in just a few words.

e) Name a drawback of distributed memory architectures.

# Question 4)

Points: .....

Maximum points: 1+2+2+1+1=7

- a) For exercise sheet 7, you implemented a remote desktop solution for a Linux instance and a Microsoft Windows instance both. Name a protocol, you used to implement a graphical remote desktop solution.
- b) If you create a cluster of virtual server instances in EC2, you can distribute the instances over multiple regions. Give an advantage and a drawback of this method.

c) If you create a cluster of virtual server instances in EC2, you can distribute the instances over multiple availability zones. Give an advantage and a drawback of this method.

- d) For exercise sheet 8, you implemented with the infrastructure services of the Amazon Web Services, a highly available High Throughput Cluster of virtual web servers. Which web server software did you use?
- e) For exercise sheet 8, you implemented with the infrastructure services of the Amazon Web Services, a highly available High Throughput Cluster of virtual web servers. The web server data was stored in EBS volumes. Which Linux file system did you deploy on the EBS volumes?

## Question 5)

Points: .....

Maximum points: 1+1+1+1+1+1+1=7

- a) In which category of cloud services is human creativity offered for low cost or donated from volunteers?
- b) Why is the term "Cloud Operating System" misleading?
- c) In which category of cloud services can customers run virtual server instances and even realize virtual data centers?
- d) What is a PaaS, and what can customers do with it?
- e) What do customers need to use software services?
- f) What is the main difference between Public and Private Cloud services?
- g) What is a Hybrid Cloud?

# Question 6)

Points: .....

Maximum points: 3+7+2=12

Your local time in Frankfurt am Main is Monday 09:00 (UTC+1). You need to copy 3 TB of data into the storage service S3. You have two options:

- Scenario 1: You immediately start at 09:00 (UTC+1) to upload the 3 TB of data to S3 via the internet. Consider the data rate between your computer and S3 is 100 Mbit/s.
- Scenario 2: You use the AWS Import/Export service. Therefore you copy the data to a HDD, which is connected via USB 3.0. The transfer rate (for write) is 125 MB/s.

After you copied the data, you pack the HDD into a parcel and send it via a package delivery company to Amazon. DHL, UPS and FedEx can deliver a parcel from Frankfurt am Main in less than 24 hours to most places in Europe.

You need 15 Minutes to put the HDD into a parcel and another 15 Minutes to bring the parcel to the branch office of a package delivery company.

The patcel must arrive at the branch office of the package delivery company no later than 16:30 (UTC+1) to arrive at Amazon in Ireland at 9:00 (UTC) the next working day.

An Amazon employee needs to copy the data from the HDD to the S3 service. The transfer rate of the HDD (for read) is 150 MB/s.

Consider 3 hours additional overhead for the in-house mail at Amazon to ship the HDD to the correct employee.

Calculate...

- a) for the first scenario, how long it takes until the data is copied to S3.
- b) for the second scenario, how long it takes until the data is copied to S3.
- c) the data rate of the second scenario.

(For all subtasks, the calculation steps must be visible.)

# Question 6 – Additional Page)

Maximum points: 3+7+2=12

### Question 7)

Points: .....

Maximum points: 2+4=6

Company X runs 8,000 computer workplaces.

- Scenario 1: Fat clients (PC)
  - Electical power rating per desktop: 350 watts
  - Electical power rating per screen: 80 watts
- Scenario 2: Thin clients
  - Electical power rating per thin client: 40 watts
  - Electical power rating per screen: 80 watts
  - Electical power rating per server blade: 400 watts
  - Each server blade has enought resources to run 50 virtual desktops

Calculate for both scenarios the electricity costs per year for 24/7 operation when the electricity price is  $0,28 \in$  per kWh?

Questio			Points:
a) Google Clou	-	ements	
IaaS	$\Box$ PaaS	$\Box$ SaaS	
b) Amazon S3	implements		
$\Box$ IaaS	$\Box$ PaaS	$\Box$ SaaS	
c) Google App	Engine imple	ments	
$\Box$ IaaS	🗆 PaaS	$\Box$ SaaS	
d) Amazon EC	2 implements.		
	PaaS	$\Box$ SaaS	
e) AppScale in	plements		
	□ PaaS	$\Box$ SaaS	
f) Google Clou	ıd Storage imp	plements	
	PaaS	$\Box$ SaaS	
g) Google Com	npute Engine i	mplements	
$\Box$ IaaS	🗌 PaaS	$\Box$ SaaS	
h) Microsoft O	ffice 365 imple	ements	
	PaaS	$\Box$ SaaS	

First name:

Student number:

Last name:

#### Question 9)

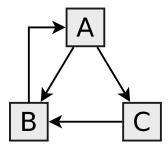
Points: .....

Maximum points: 9

- $PR_p$  = PageRank of a web page p
- $L_{IN}(p) = \text{Set of documents}$ , which refer to  $p \Longrightarrow$  incoming links
- $L_{OUT}(p) =$  Set of documents, to which p refers  $\implies$  outgoing links
- d = damping factor between 0 and 1

$$PR(p) = (1-d) + d * \sum_{p_i \in L_{IN}(p)} \frac{PR(p_i)}{\text{amount } L_{OUT}(p_i)}$$

Calculate the missing iterations of the PageRank algorithm for the given example scenario with d = 0.75.



	0	1	2	3	4	5	PR
А	1		1,28125		1,1494140625		1,127166748
В	1		1,09375		1,19921875		1,1918029785
С	1		0,625		$0,\!6513671875$		0,6810302734

# Question 10)

Points: .....

Maximum points: 1+1+1+1+1=5

a) What is WSDL and for what purpose is it used?

b) What is UDDI and for what purpose is it used?

c) Explain the difference between UDDI and WS-Inspection.

d) Describe the difference between the theoretical implementation of SOAP web services and the way, SOAP web services operate in practice.

e) For exercise sheet 11, you implemented a Private Cloud storage service, which implements the S3 API. Which one of the existing solutions did you use?

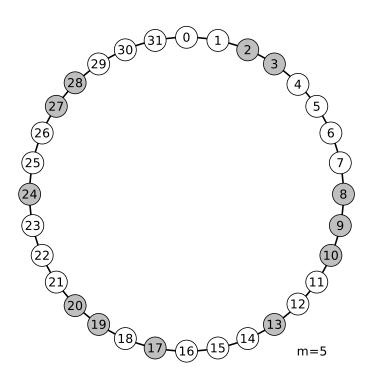
Student number:

# Question 11)

Points: .....

Maximum points: 1+1+1+10+1=14

- a) What is the drawback of linear search in the Chord ring?
- b) What way of searching in the Chord ring is preferred?
- c) To which node n gets a key k assigned to?
  - □ Direct predecessor
  - $\Box$  First node (starting from ID 1) without any keys assigned yet
  - $\Box$  The node with the same ID as the key
  - □ Direct successor
- d) Calculate the Finger Table values of node n = 22 and insert the correct values into the provided Finger Table.



Finger Table of node n = 22

Entry	Start	Node
1		
2		
3		
4		
5		

The table has 5 entries, because m contains the length of the ID in bits and m = 5

The Start value of entry i of the table on node n is  $(n + 2^{i-1}) \mod 2^m$ 

The Node value of entry i points to the first node, which follows to n at a distance of at least  $2^{i-1}$ 

e) Which node is responsible for the key (resource) with ID 11 ?

Last name:	First name:	Student number:
Question 1	2)	Points:
Maximum points: 5		
Only a single answer is co	prrect in each subquest	ion.
a) Centralized services	exist in	
$\Box$ Centralized P2P	□ Pure P2P	$\Box$ Hybrid P2P
b) No centralized servi	ces exist with	
Centralized P2P	$\Box$ Pure P2P	$\Box$ Hybrid P2P
c) A central point of a	ttack exists with	
$\Box$ Centralized P2P	$\Box$ Pure P2P	$\Box$ Hybrid P2P
d) Which architecture	causes the biggest net	work overhead?
$\Box$ Centralized P2P	$\Box$ Pure P2P	$\Box$ Hybrid 0P2P
e) Which architecture	causes the lowest netw	vork overhead?
$\Box$ Centralized P2P	$\Box$ Pure P2P	$\Box$ Hybrid P2P
f) Which architecture	implements a kind of o	dynamic, centralized service?
$\Box$ Centralized P2P	$\Box$ Pure P2P	$\Box$ Hybrid P2P
g) Napster (1999 - 200	1) implemented	
$\Box$ Centralized P2P	$\Box$ Pure P2P	$\Box$ Hybrid P2P
h) Which architecture	implements Ultrapeers	s (= Supernodes)?
$\Box$ Centralized P2P	$\Box$ Pure P2P	$\Box$ Hybrid P2P
i) Gnutella v0.4 imple	ments	
$\Box$ Centralized P2P	$\Box$ Pure P2P	$\Box$ Hybrid P2P
j) Gnutella v0.6 imple	ments	
$\Box$ Centralized P2P	$\Box$ Pure P2P	$\Box$ Hybrid P2P

# Question 13)

Points: .....

Maximum points: 1+1+1+1+1+1=6

a) Give a short definition of Cluster Computing.

- b) What is a Cluster of Workstations?
- c) How can the availability of a system be calculated?
- d) Explain the main difference between a SAN (Storage Area Network) and a NAS (Network Attached Storage).

- e) What is a Beowulf Cluster?
- f) What is a Wulfpack Cluster?