Written examination in Cloud Computing

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	Σ	Grade
Maximum points:	8	6	5	7	10	12	6	10	9	5	8	4	90	—
Achieved points:														

Question 1)

Maximum points: 8

Imagine you have trained a pigeon to carry a USB flash memory drive with a storage capacity of 64 GB. The pigeon can fly with an average speed of 72 km/h.

For what range of distance has the pigeon a better data rate than a computer network whose data rate (excluding overhead) is 250 Mbps?
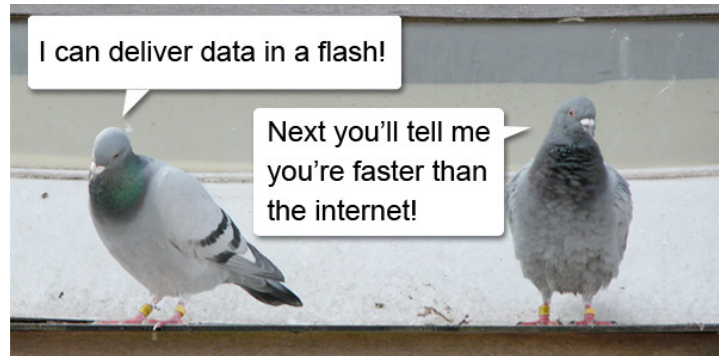

Image source: http://www.usb-flashdrive.co.uk

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)

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

a) For exercise sheet 4, you implemented a grapical remote desktop solution for a Linux instance. Which protocol did you use to implement the remote desktop solution with Linux?

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 5, 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 5, 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: 10

For exercise sheet 5, you implemented with the infrastructure services of the Amazon Web Services a highly available High Throughput Cluster of virtual web servers. One part of this task was to attach an EBS volumes as persistent storage to each web server instance. All storage volumes contained identical data.

But not for all possible scenarios, it is the optimal approach, when all instances have EBS volumes with identical content attached.

For exercise sheet 6, you investigated other approaches and summarized their advantages and drawbacks. The focus of your investigation were the aspects availability, cost, throughput, scalability and complexity of the different approaches.

Explain one of the approaches you investigated and summarize its advantages and drawbacks. Focus on the aspects availability, cost, throughput, scalability and complexity. Name the services and software solutions you used. Maybe a diagram is helpful too.

Question 5 – Additional Page)

Maximum points: 10

Question 6)

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 parcel 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: 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) By which approach does High Availability Clustering achieve its objective?

e) Explain the difference between Shared Nothing and Shared Disk Clusters.

f) Explain the main difference between a SAN (Storage Area Network) and a NAS (Network Attached Storage).

Question 8)

Maximum points: 10

For exercise sheet 10, you implemented a MPI program, which calculates π via Monte Carlo simulation. Several options exist to solve this task. Two of them are:

- The master node creates a specified amount of random numbers and distributes them to the nodes. In this case, additional nodes just reduce the required computing time.
- Each node creates a specified amount of random numbers. In this case, additional nodes will not reduce required computing time but the result will be closer to π.

Implement the solution you created for exercise sheet 10 in pseudo-code. Use comments to explain what operations are carried out for what reason.

These MPI functions may be useful for your implementation.

- `MPI_Bcast();`
- `MPI_Comm_rank();`
- `MPI_Comm_size();`
- `MPI_Finalize();`
- `MPI_Get_processor_name();`
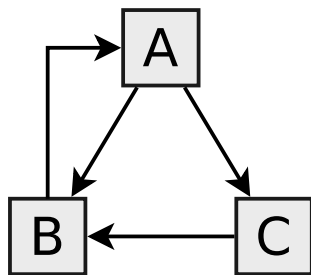- `MPI_Init();`
- `MPI_Reduce();`

Question 9)

Maximum points: 9

- PR_p = PageRank of a web page p
- $L_{IN}(p)$ = Set of documents, which refer to $p \implies$ 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
A	1		1,28125		1,1494140625		1,127166748
B	1		1,09375		1,19921875		1,1918029785
C	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 12, you implemented a Private Cloud storage service, which implements the S3 API. Which one of the existing solutions did you use?

Question 11)

Points: ....

Maximum points: 1+1+1+1+1+1+1+1=8

a) What is the difference between emulation and virtualization?

b) How works application virtualization?

c) What is the function of the Virtual Machine Monitor (VMM)?

d) Where runs the Virtual Machine Monitor (VMM)?
   ☐ The VMM runs *hosted* as an application in the host operating system.
   ☐ The VMM runs *bare metal* and replaces the host operating system.

e) Can all physical hardware resources be virtualized when full virtualization is used?
   If this is not possible, give an example where it does not work.

f) Where runs the hypervisor when paravirtualization is used?
   ☐ The hypervisor runs *hosted* as an application in the host operating system.
   ☐ The hypervisor runs *bare metal* and replaces the host operating system.

g) Why is for paravirtualization a host operating system required?

h) Name a drawback of operating system-level virtualization (containers/jails).

Question 12)

Maximum points: 4

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