Architecture of Information Systems (232011): Second exam 2007/2008

August 20, 2008 13:30-15:30h, LA A107

Please pay attention to the following:

- **This exam has to be completed in 2 hours.**
- It is **NOT** allowed to use the book or any other material.
- You can answer in either Dutch or English.

Distribution of points:

- 10 points for showing up;
- Other questions as indicated.

1 Question 1 (10 points)

The book distinguishes between shared data and controlled duplication. Describe the similarities and differences between them from the perspective of information accuracy and availability.

2 Question 2 (15 points; 5 points for a, b and c each)

There are many acronyms in the world of AIS. Describe the relationship between each of the following pairs of acronyms:

a) **SOAP** (Simple Object Access Protocol) and **RPC** (Remote Procedure Call)

b) **WSDL** (Web Services Definition Language) and **RMI** (Remote Method Invocation)

c) **ORB** (Object Request Broker) and **CORBA** (Common Object Request Broker Architecture)

3 Question 3 (20 points; 10 for a, 10 for b)

Suppose we have a database server which is attached to a 1 Gb/sec (one gigabit) network, with a protocol overhead such that transmitting one **byte** takes 10 bits. The database server processes queries where the message carrying the query is on average 1kB (one kilo**byte**) and the message carrying the result is on average 10kB (ten kilo**byte**) of XML. Query processing (from the moment the request has been delivered to the database engine to the moment the result is ready for transmission) takes on average 0.3 μsec. As XML messages are

quite verbose, it is often a good idea to compress them, especially because the compression can be done almost completely in parallel with transmitting the compressed message (after a short setup delay, transmission of the beginning of the compressed message starts while the end is still being compressed). This would reduce the size of the message by a factor C (with C a real number between 0 and 1 exclusive), but would increase the query processing time with D μsec.

a) How long does it take to process a query, including communication but excluding compression?

b) Whether it is wise to compress depends on whether the increase in processing time (D) is lower than the decrease in the time it takes to transmit the result message. We are interested in the break-even point B: the value for D where the increase in processing time is equal to the decrease in message transmission time, given a compression factor C. Give a formula for the relation between B and C, draw its graph and explain whether your answer makes sense.

4 Question 4 (15 points)

Imagine we have two applications A and B which communicate using Message Queuing. A sends a request to B and B sends one response back for each request. Because processing the requests is computationally expensive, there are three servers running the application B. For reasons of resiliency, there are two other servers running application A. There are no databases involved in the applications.

Draw a picture with both applications, all servers and all queues needed in this setup. Mark all arrows with "request", "response" or "both".

5 Question 5 (15 points, 5 for a, b and c each)

For the migration of a current information system to a new information system two approaches are **chicken little** and **cold turkey**.

a) Describe both approaches.

b) What is/are the function(s) of gateways in both approaches? Do(es) the function(s) differ between gateways in the two approaches?

c) In which circumstances will cold turkey be the best choice?

6 Question 6 (15 points; 5 points each for a, b, c)

Subprime Lendings, Inc. is a provider of credit for everyone who's short on cash but wants to live like a millionaire. To get credit, a customer enters into a contract with *Subprime Lendings*, after which the customer can transfer money from *Subprime Lendings* to his/her bank account, whenever he or she wants, as often

as he or she wants, up to a certain limit. For instance, a customer with a credit limit of € 10,000.- can make ten withdrawals of € 1,000.- each.

There are two processes:

- The process in which a customer enters into a contract with *Subprime Lendings*, executed once per customer. If successful, *Subprime Lendings* records a withdrawal limit and sets the current debt at € 0.
- The process in which a customer actually withdraws money from his/her account with *Subprime Lendings*, executed one or several times per customer. Each time this process is executed, *Subprime Lendings* checks whether the sum of the current debt and the withdrawal amount is lower than the credit limit.

The IT department of *Subprime Lendings* maintains a portfolio of web services. In this exam, we focus on two of these services used in the second process:

- `Check_limit`, which takes as input an identifier that identifies the relevant contract and a withdrawal amount, and returns a 'true' when the current debt plus the withdrawal amount is lower than the limit and 'false' otherwise.
- `Withdraw`, which takes as input an identifier that identifies the relevant contract and the withdrawal amount. The service transfers the money and increases the current debt of the customer. The service does not return any value.

The software application that handles new credits uses these two web services. A requirement is that this happens according to the ACID-properties.

a) Of course, *Subprime Lendings* wants to avoid that a customer is able to withdraw more than the limit by starting two or more withdrawals at the same time. Explain this in terms of the ACID-properties.

b) Assume that there are three separate software applications: the application that handles new contracts (by using the two web services), a contract management system that provides the `Check_limit` service and a payment system that provides the `Withdraw` service. One can use a piece of middleware called a Distributed Transaction Coordinator (DTC) to ensure that the ACID properties hold. Explain how this works.

c) To increase resiliency, both the contract management system as well as the payment system are each replicated: there are two instances of each of them, running on separate servers. Explain how to do this using the two-database approach (the two instances of the contract management system each have their own database, and so do the two instances of the payment system), not the clustering approach (where the two instances of the contract management system share one mirrored contract database, as do the two instances of the payment system with their database). Motivate your answer.