Prof. Dr. Stefan Tai
Marco Peise

Your name: _____

Enterprise Computing (WS 2016/17)

Exercise 3 (3 Portfoliopunkte)

Info:

- The solution to this exercise must be handed in by Wednesday, Nov 23th 2016, 12AM.
- Any written solution must be in an accessible PDF. Any source code in a separate ZIP File. All Files are uploaded in the Information System for Instructors and Students in a single ZIP File (https://isis.tu-berlin.de/course/view.php?id=8586).
- Please write your name on the solution sheet.

Task 1 – CAP Theorem (20+20 = 40%)

a) Shortly explain the CAP theorem by example of the Domain Name System (DNS).

b) Shortly describe two dimensions of data consistency from both a data-centric and a client-centric perspective.

Task 2 – Dynamo (10+10+20+10+10 = 60%)

a) Which queries does Dynamo support and for which type of data storage is Dynamo optimized?

b) Pessimistic replication (as implemented in Dynamo, for example) is used to offer high availability and low latency. True or false?

c) Dynamo uses vector clocks to determine the total order of write operations. Given the vector clocks in the table below with conflicting versions on servers A, B, and C. Please state whether or not the conflict can be reconciled automatically (yes/no) and how the vector clock must look like after a conflict resolution.

Vector clocks before conflict resolution	Can be reconciled automatically (yes/no)	Vector clocks after conflict resolution
D1 ([A,1]) D2 ([B,1]) D3 ([C,1])		
D1 ([A,1] [B,2]) D2 ([A,2] [B,2])		
D1 ([A,1] [B,4] [C,13]) D2 ([A,2] [B,3] [C,15]) D3 ([A,2] [B,4] [C,15]) D4 [A,1] [B,3] [C,14])		
D1 ([A,1] [B,2] [C,1]) D2 ([A,2] [B,2])		

d) Shortly explain the trade-off between consistency, read latency, and write latency in Dynamo and how a Dynamo-based application could be tuned either towards fast reads or towards fast writes using the (N,R,W) configuration.

e) What is the minimum cluster size, i.e., number of servers, of a Dynamo configuration (N=9, R=1, W=9) and why? For this minimum cluster size: how many data records are stored on each node after 1 million data records have been inserted into the Dynamo cluster by a client program?