Exercise 1 (multiple choice)

=========================== Instructions

**#Excercise 1**: NoSQL Data Structures

Describe the pros & cons of four different NoSQL Data Structure families:

- key-value store

- column store/column family

- document store

- graph database

Sometimes NoSQL databases in the same family can be quite different. What are some differences between Couchbase and MongoDB.

**#Exercise 2**: Quorum and Dynamo Inspired Systems

[Resource] (http://www.allthingsdistributed.com/files/amazon-dynamo-sosp2007.pdf)

###Instructions

1. Name at least three systems that implement quorum protocols.

1. Define the following:

\* ‘W’

\* ‘R’

\* ‘N’

\* ‘Q’

1. Why is ‘N’ generally chosen to be an odd

integer?

1. What condition relating ‘W’, ‘R’, and ‘N’ must be satisfied to "yield a quorum like system"?

1. In the paper "Probabilistically Bounded Staleness,"

Berkeley researchers derive an analytic framework for the probability of reading a "stale" version of an object in a Dynamo-like system that implements quorum.

Using [this tool] (http://pbs.cs.berkeley.edu/\#demo) (lambda=0.1 for all latencies, tolerable staleness=1 version, 15,000 iterations/point), answer the following questions:

1. With what probability are you reading "fresh" data for n=3, w=2, r=2?

1. Does it depend on time? If so, why? If so, why not?

1. Compare the scenarios for (w,r,n) = (2,1,3) and (1,2,3).

1. Write down and explain the differences (if any) for the time dependence of P(consistent).

1. Is the (2,1,3) state symmetric with (1,2,3)?

1. Compare both P(consistent) and the median and 99.9% latencies.

1. Provide an intuitive explanation for your results.

1. Do either of these states favor consistency or availability? If so, why?

1. Perform a similar comparison for the (3,1,3) and (1,3,3) states. Do either of these states favor consistency or availability? If so, why?

1. In your opinion, assuming an n=3

system, what do you think is a reasonable choice for write heavy, read heavy, and read\~=write workloads?

**#Exercise 3:** Partitioning Strategies

What is the benefit of Consistent Hashing versus other approaches (like Range Partitioning)?

**#Exercise 4: Cloudant**

\*\*NOTE:\*\* Ensure that you replace "username" and "password" in the URLs in this section with your Cloudant username and password

1. Sign up for a free Cloudant account [here] (https://cloudant.com/sign-up/).

2. Create a DB

curl -X PUT -H 'Content-Type: application/json' https://username:password@username.cloudant.com/crud/

1. Run an insert using curl:

curl -d '{"season": "summer", "weather": "usually warm and sunny"}' -X POST https://username:password@username.cloudant.com/crud/ -H "Content-Type:application/json"

1. Ensure your new document is in the database:

curl https://username:password@username.cloudant.com/crud/\_all\_docs

Include the output from these requests in your homework submission.

\\*\*To turn in:\*\*

Upload one document with your responses to all four exercises.