**Tech + Research Projects - Exposing Flaws in Cloud Data Storage**

Nistha Mitra, Najlaa Bouras, Brea Austin, Anjali Viramgama

Faculty Member- Daniel Abadi

Amazon, Microsoft, and Google all widely use cloud computing product. It enables them and the users to deploy and maintain application in their respective clouds. Each cloud has multiple databases. We are working with the ones that guarantee consistency and efficiency. Which means that any change in data done at one geographical location, would change it all over the world instantly. Our team is trying to find the loophole in that guarantee and the tradeoffs involved.

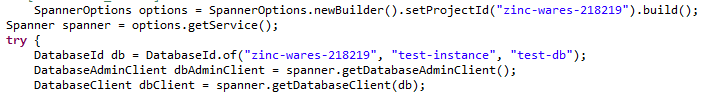
**What’s the problem?**

For Example, you are shopping online. Each item is saved in the database. When you want to buy it, you are sending a request to the database to get access to it. And the database returns an answer according to the availability or other factors. This is called a transaction.

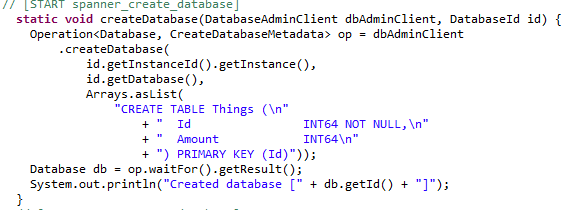
When one part of the dataset has a lot more transactions, it’s called a hotset. Example, the dataset of Christmas items on Amazon during the Christmas season. The no. of people trying to buy an item is a lot more than items available.  This is a high contention scenario.

We are testing the efficiency in terms of transaction per second (throughput) and latency which is the time taken to conduct the transaction. Our experiment involves sending transactions in varying contention scenario from different geographical standpoints.

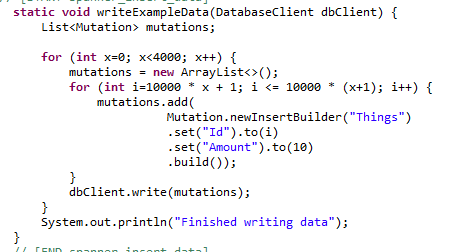
We are using Google Spanner for our experiment.



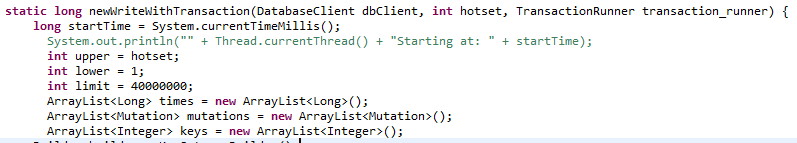
***Driver that sets up the different clients. The “setProjectId” may or may not be necessary.***



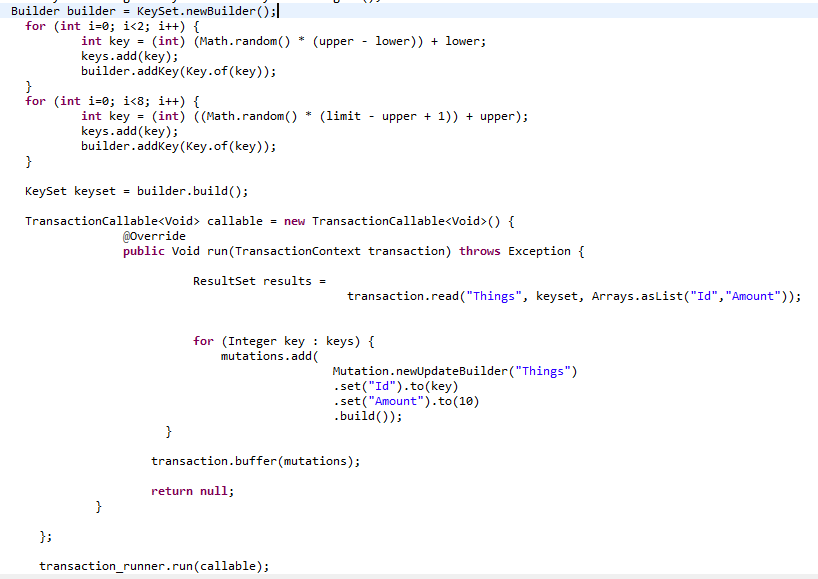
***Creates a database with two columns, Id and Amount***

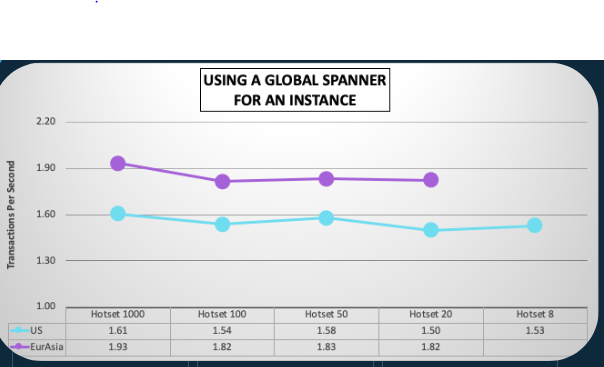


***Inserts 40,000,000 records into the “Things” database. Inserting this many record will take a few minutes.***



***Transaction method signature and variable definitions for the implementation below.***

***Sample read-write transaction that accesses 2 records from the hot set and 8 records from the cold set. The writes in this specific implementation of a read-write transaction do not depend on the results of the read.***

**Conclusion** 

We saw an decrease in throughput and increase in latency when we moved from a single region instance to multiple region instance. In general, it is not as consistent as guaranteed.