Clojure & DynamoDB

2013-07-08

Stephen Johnston

@stevejjohnston

DynamoDB Overview

- In 2007, Amazon published "Dynamo: Amazon's Highly Available Key-value store"

- The technology behind Amazon S3.

- Several open-source implementations released based on this paper: Riak, Project Voldemort

- In 2012, Amazon offers "DynamoDB", a nosql database as a service

- Tunable read/write performance. (Provisioned capacity.)

Data Model

* DynamoDB allows users to setup one or more tables

* Each table has a primary key: Either a 'hash' or a 'hash + range'

* The remaining columns are schema-less

  * Each row in the table can have a varying number of columns

* Columns can be either string, numeric, or binary

Quick note on hashing

* The hash key determines where the data is written.

* We know that Dynamodb scales out horizontally.

* Thus, it is important that hashes have a nice even distribution.

Features

* Replicates data across 3 availability zones (data centers within the same region)

* Single digit millisecond response for get/put requests

* No administration

* Scales throughput / storage

* Local secondary indexes

* Transactions

Pricing

* Free Tier

  * 5 writes per second, 10 reads per second, 100MB data.

* After free tier, storage is 25¢ per GB per month

* Writes are $0.0065 per hour for each 10 units of write capacity

* Reads are $0.0065 per hour for each 50 units of read capacity

* Pay as you go

Pricing example

- 500 reads per second, 500 writes per second, 50GB of data

  - 500 read * $0.0065/50read * 24 * 365 / 12 = $47

  - 500 writes * $0.0065/10writes * 24 * 365 / 12 = $237

  - 50 * 25¢ = $12.50

  - Grand total per month: $296.50

Limits

* Reads and writes are capped at your provisioned capacity

* Your provisioned capacity can be increased as often as you like, but can only be decreased 4 times a day

* Can only increase capacity 100% at a time

* Tables have no practical limit on number of rows or bytes

* 256 tables per region

* An item (or row) has a max size of 64kb

* Batch reads: 100 rows at a time, or 1MB of data.

* Batch writes: 25 rows at a time, or 1MB of data.

DynamoDB and Clojure

* Weavejester's library: rotary

* ptaoussanis' library: faraday

* Clojure maps are a nice fit for schemaless items in DynamoDB

* { :user-id 100 :first-name "Bob" :last-name "Smith" }

Sample use case: 3 Billion phone calls a day

- Needs 35,000 writes per second

- Need to be able to reverse lookup a phone number

- Need to be able to see all the calls for a phone number, sorted by calltime (or sorted by another column, like duration or number-dialed.)

- Estimated Price?

    - Reserved capacity price: ~ $8,000 per month

Table setup

- my-table (reverse lookup table)

  - Primary hash key: 'phonenum' as string

  - Other columns: 'name' as string

  - Could grow the table to include additional columns as needed (address, account #, etc)

- phone-metadata

  - Hash key: 'phonenum' as string

  - Range key: 'calltime' as number

  - Other columns: duration, latitude, longitude, numberdialed

Demo ptaoussanis/faraday