Memcached

Catherine dewerd

What is it?

-Distributed memory object caching system
-Used to speed up dynamic web applications by alleviating database load
-Is a large hash table with keys up to 250 bytes, and values up to 1 MB that can span multiple web servers

So why is it useful?

When running a memcached server, you can take bits of code that take a while to run and put them in RAM, so when the database driven website needs access to the data it can get it faster, but all the data is disposable

Some History

-Originally developed for LiveJournal in 2003, and written in Perl by Brad Fitzpatrick
-Rewritten in C by Anatoly Vorobey
-Now used by YouTube, LiveJournal, Twitter, Wikipedia, Craigslist

Distributed cache … ?

- Can span many servers so it can grow easily

- Other examples include Aerospike, Ehchache, GigaSpaces, Hazelcast, Riak, Redis, SafePeak

Why memcached, not Redis?

- It’s a generic so it runs on most unix - like OS’s and on Windows, easier to learn, and more stable and portable.

- Best in situations where either the # of requests is high or the cost of generating a specific output is high

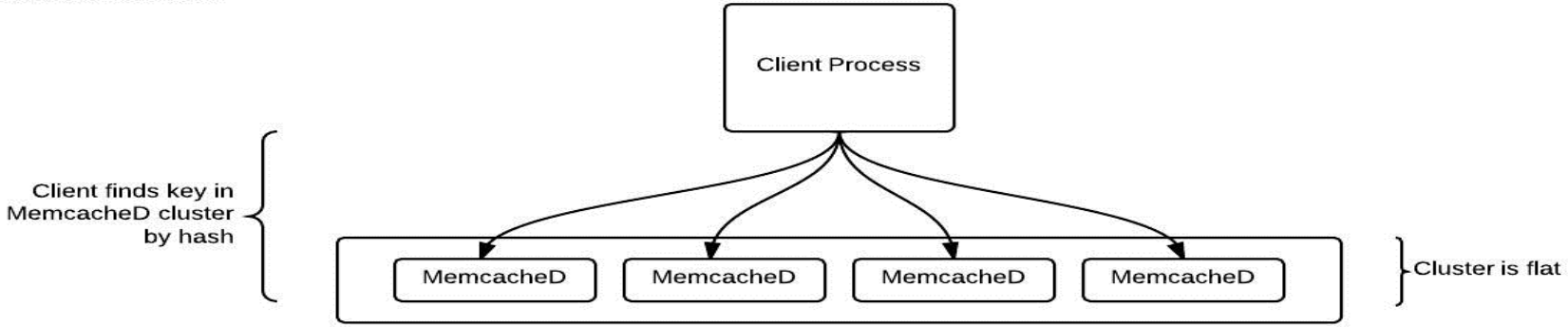
Pros

-Low complexity
-Easy configuration
-Atomic increment and decrement
-Simple to cluster
-Withstands a member dying
-Most languages have a memcached library

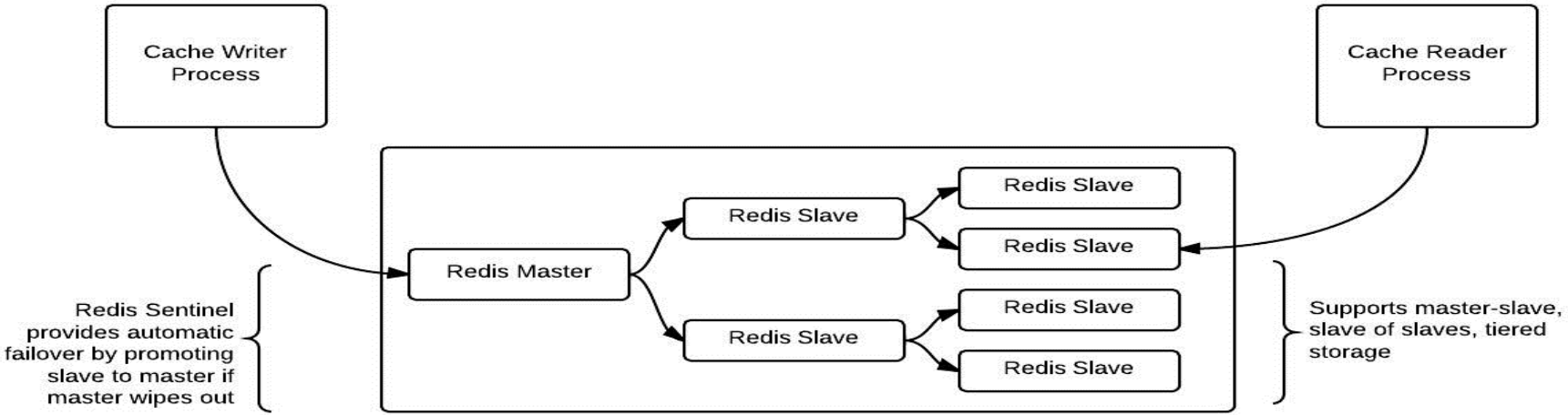
Cons

-Doesn’t do anything but in-memory key/value store
-Caches shared by client don’t scale across AWS zones
- Unbalanced clusters require a full system restart
-Adding a member to the pool requires reconfiguring and rebooting the client

MemcacheD Cluster



Redis Cluster



- Just maps Id’s to values; never needs to scan an entire table
- Simple protocol, no parsing SQL
- Does not require reliability of database, so it *can be run on the same machine as is running the application.

*this is a bad idea in general, just because you can doesn't mean you should

How is data disposable?

It’s cached. The data in the RAM is a copy of the stuff in the actual database, so if the cache goes away, no actual data is lost. The cache is there for ease of access.

Basic Example, working with PHP

First run an instance of memcached on each computer with some allocated RAM.

Then make an array of the IP addresses and port numbers of the machines

Example continued

Make a memcache object:

```
$memcache = new Memcache();
foreach($MEMCACHE_SERVERS as $server){
    $memcache->addServer ( $server );
}
```

Then for more efficiency

```
$huge_data_thing = $memcache->get("huge_data_thing");
if($huge_data_thing === false){
    $huge_data_thing = array();
    // if this thing takes a long time
    $sql = "SELECT * FROM hugetable WHERE timestamp > lastweek ORDER BY
timestamp ASC LIMIT 50000";
    $res = mysql_query($sql, $mysql_connection);
    while($rec = mysql_fetch_assoc($res)){
        $huge_data_fthing[] = $rec;
    }
    // put it in a cache for 10 minutes
    $memcache->set("huge_data_thing", $huge_data_thing, 0, 600);
}
```

Further efficiency

- Add things to the memcache that are used frequently (100+ times a second)
- Things that are big
- Lots of things (to take advantage of the distributed nature)

Memcache vs Memcached

-Both are clients of the memcached server, but memcached offers more features
-BUT: they serialize data differently (memcache interprets things as strings while memcached interprets things as integers), so switching between them is difficult.

Depends on libevent library

- provides asynchronous event notification
- Meant to replace the event loop in event driven network servers
- Provides a mechanism to execute a callback function when an event occurs on a file descriptor, a callback is signaled, or a timeout reached.

Commercially supported distributions

Couchbase Server
GigaSpaces Java based Memcached (fault tolerant)
Hazelcast Memcached (clustered, elastic, fault-tolerant, Java based)

Engine Yard, Jelastic, Heroku, Google App Engine, AppScale, Microsoft Azure, Amazon Web Services all offer memcached services

Has been implemented in

C
Java
Ruby
Go
PHP

Notes:

As of 10/9/2014, Redis has implemented built in data clustering, therefore making the two fairly equal. While Redis is still more complicated to learn, its built in clustering, and more powerful data types make the community consider it “better” though for simple data types memcached is still useful

Sources

http://www.memcached.org/

Memcached Wiki

http://php.net/manual/en/book.memcached.php

http://stackoverflow.com/questions/10765897/whats-the-memcached-server

Discussion on Memcached vs Redis

Stack Overflow Memcached vs Redis

Memcached Basics