

Using RethinkDB in Production for SageMathCloud

William Stein

University of Washington

SageMath, Inc.

<https://cloud.sagemath.com/>

November 17, 2015

What is Sage?

SageMath

- SageMath: big open source math software I started in 2004

SageMathCloud (SMC)

- **Launched:** 2013
- **Real-time editing like Google Docs:** Latex, IPython/Jupyter notebooks, Sage, Terminals, Teaching, etc.
- **Tech Stack:** RethinkDB, Linux, React.js, Node.js, SageMath/Python, CodeMirror, CoffeeScript
- **Users:** 4000+ daily active; nearly 1000 simultaneous
- **Production:** anger when it doesn't work – “my homework is gone!”
- **Open source:** 100% open source, GPL 3, etc.

Hi From Sage Days 70



Switched from **Cassandra** to RethinkDB this summer.

SMC Uses RethinkDB Heavily...

- **Setup:**
 - 6 Google compute engine nodes (quad-core n1-standard-4)
 - About 23 tables storing about 5 million documents
 - Replication factor 3, sharding of 3
 - Storage in persistent (network-mounted) SSD
- 5K-10K simultaneous changefeeds.

Operations

- **Backups:** periodic dump of most tables to json on a compressed filesystem, snapshot via bup (=git+more), rsync to google cloud storage and encrypted off-site USB drives.

RethinkDB team **amazing** at addressing all issues I encountered.

Show how SMC uses RethinkDB

- 1 Change name and see change in another browser.
- 2 Show changing project title and that appearing in another browser.
- 3 Draw a 3d plot in a sage worksheet
- 4 Open a Jupyter notebook – demo sync and history
- 5 No REST/API calls; instead, set entries in a table, back-end sees it, makes table change, all parts of all front-ends simultaneously see that (do a demo of project restart).

SMC Demo: Change username

The image shows a sequence of browser screenshots demonstrating a username change on SageMathCloud. The top-left screenshot shows the RethinkDB website. The middle-left screenshot shows the SageMathCloud account settings page with the first name 'William A.' and last name 'Stein'. The top-right screenshot shows the same page with the first name changed to 'William Arthur'. The bottom-right screenshot shows a code editor with a JavaScript snippet:

```
{ 'score' }.limit(3).changes()
```

 and a table of top player scores:

TOP PLAYER SCORES	
1.	alex: 64 points
2.	karl: 63 points

SMC Demo: Change Project Title

The screenshot displays two browser windows. The left window shows the RethinkDB website, and the right window shows the SageMathCloud project management interface. The right window is focused on a project titled 'SD70'. The project details form shows the title 'SD70' and the description 'Sage Days 70 in Berkeley -- IPython/Jupyter/Sage/SageMathCloud'. Below the form, a list of projects is visible, including 'mushroom_fox' and 'SD70'. The 'SD70' project is highlighted, showing its status as 'Running' and a list of users who have interacted with it.

RethinkDB

Projects

Showing projects

mushroom_fox now

SD70 less than a minute ago Sage Days 70 Workshop in Berkeley -- IPython/Jupyter/Sage/SageMathCloud

minute ago), John Jeng (12 hours ago), Harald Schilly (13 hours ago), William Arthur Stein

Running

Running

SMC Demo: 3D Plot

The image shows a SageMathCloud interface with two browser windows. The left window displays code for calculating the prime factorization of 2016 and generating a 3D plot. The right window shows the execution of the code, resulting in a 3D plot of a purple fractal structure within a 3D coordinate system.

```
1  
2 factor(2016)  
3 2^5 * 3^2 * 7  
4  
5 v = [(0,0,0)]  
6 for i in range(1000):  
7     v.append([a+random()-.5 for a in v[-1:]])  
8     line3d(v, color='purple', thickness=2)  
9
```

The 3D plot shows a complex, fractal-like structure rendered in purple, plotted within a 3D coordinate system. The axes are labeled with values: 11.02, 24.62, 11.43, and 24.41. The structure is composed of many small, interconnected segments, creating a dense, branching pattern.

SMC Demo: Jupyter Notebook

The image displays a Jupyter Notebook interface on SageMathCloud. The browser address bar shows the URL: `https://cloud.sagemath.com/projects/4a5f0542-5873-4eed-a85c-a18c706e8bcd/files/tmp/2015-11-11-rethink.ipynb`. The notebook title is "2015-11-11-rethink".

The notebook content includes the following code cells:

```
In [1]: from sage.all import factor
In [2]: factor(2016)
Out[2]: 2^5 * 3^2 * 7
In [5]: print range(100)
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99]
```

A revision control overlay is visible, showing a warning: "Warning: Any changes you make to revisions displayed below will". The overlay includes a slider and a timestamp "Revision 11:51:25".

How SageMathCloud uses Changefeeds

Motivation

- Make front-end development easier
- Simplify code connecting the front-end to back-end (one declaration instead of messages flying all over)

Inspiration

- Facebook's GraphQL – but simpler

Goal

- Have declarative client-side queries and database schema
- Instant notifications about changes.

Building a GraphQL-like API on RethinkDB and Node.js

(do not look at this)

Browser (or iOS/Android at some point) client query:

- JSON object that describes what result should look like; null's get filled in. `{table:{foo:bar, stuff:null}}` gets one record in table where `foo="bar"` and `{table:[{foo:bar, stuff:null}]}` gets them all.
- If `changes=true`, then any time RethinkDB table changes, client gets updates, and anytime client makes changes, they get pushed to back-end to RethinkDB.
- Tables can be "virtual", and not correspond to actual RethinkDB tables. e.g., different permissions, or involving multiple tables (so joins, technically; they also have a killfeed).
- Show `schema.coffee`.
- Text editing: describe algorithm based on the above, which isn't deployed yet.

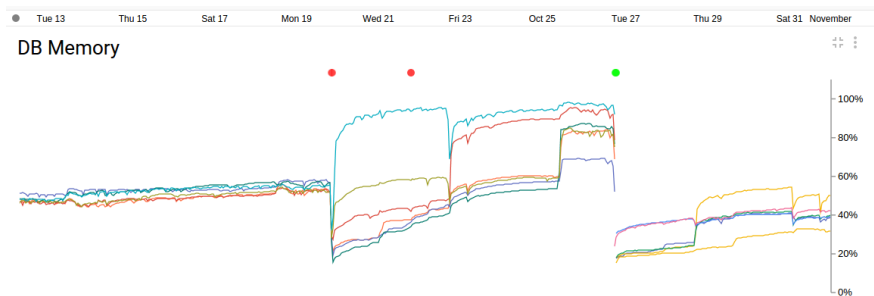
Example: Server Overload

About 3 week of data for November 2015 across 6 nodes. At one point (with 6 n1-highcpu-2s), we hit a threshold (with around 850 simultaneous users) and the backend collapsed.

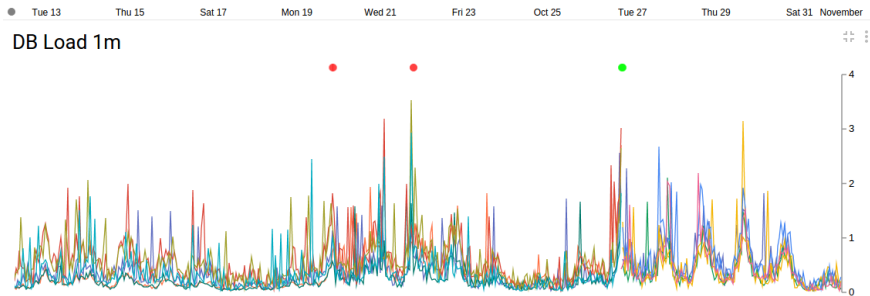
Solution

A new node had to be added (Tue 27th).

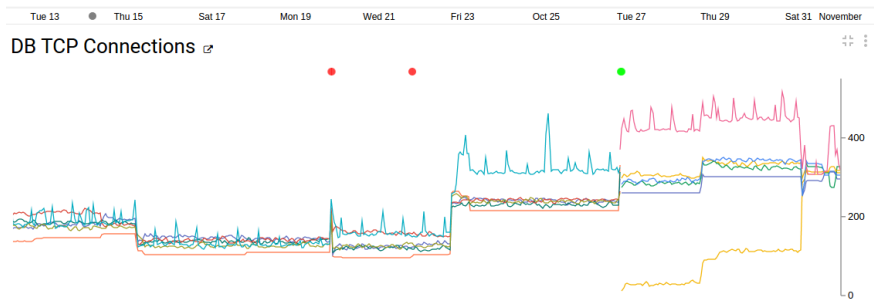
Memory usage across database nodes



CPU Load (1 min) across database nodes



TCP connections across database nodes



Thanks!

Sign up today!

<https://cloud.sagemath.com/>