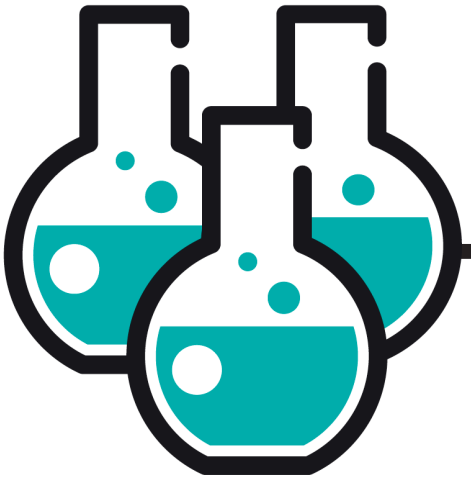


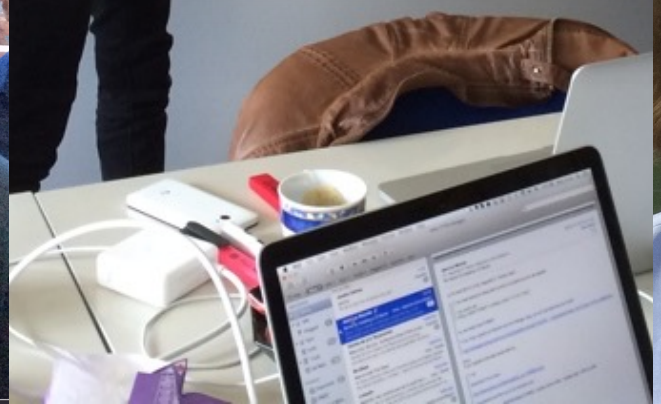
AntidoteDB

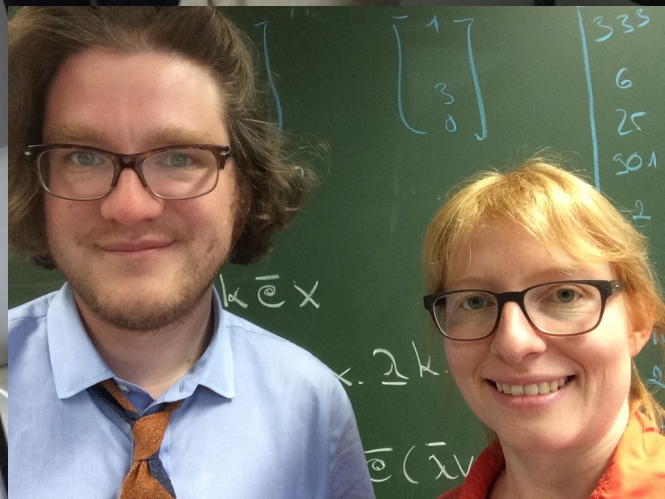
RainbowFS Workshop, March 28 2022

Annette Bieniusa, TU Kaiserslautern

The past

This is how it started…





The initial commit

Initial commit

13 README.md

floppy: A Riak Core Application

Application Structure

This is a blank riak core application. To get started, you'll want to edit the following files:

- `src/riak_floppy_vnode.erl`
  - Implementation of the riak_core_vnode behaviour
- `src/floppy.erl`
  - Public API for interacting with your vnode

Rename floppystore to antidote.

cmeiklejohn committed on 15 Nov 2014

https://github.com/AntidoteDB/antidote/commit/17ed1f622db2ca6ada6acc286eaaa692c766c17c

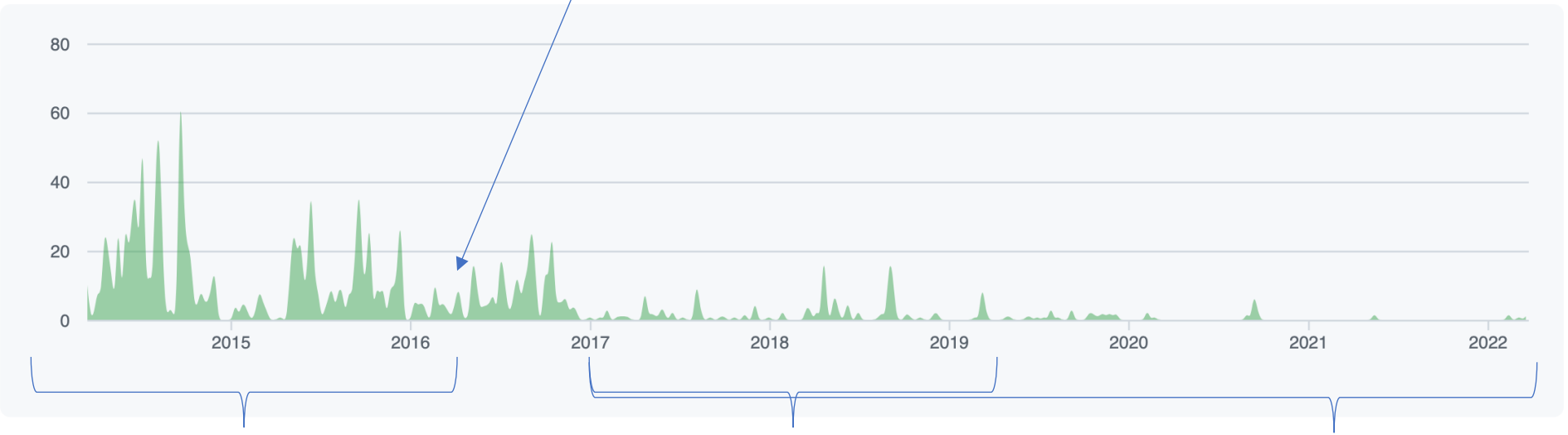
What happened

2016 IEEE 36th International Conference on Distributed Computing Systems

Cure: Strong semantics meets high availability and low latency

Deepthi Devaki Akkoorath*, Alejandro Z. Tomsic†, Manuel Bravo‡, Zhongmiao Li‡,
Tyler Crain†, Annette Bieniusa*, Nuno Preguiça§, Marc Shapiro†

*University of Kaiserslautern, †Inria & LIP6-UPMC-Sorbonne Universités
‡Université Catholique de Louvain, §NOVA LINCS



SyncFree

LIGHTKONE
Lightweight computation for networks at the edge

RainbowFS

Contributors with > 1 commit

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

bieniusa
189 commits 40,191 ++ 11,503 --

mweberUKL
17 commits 2,028 ++ 209 --

albsch
75 commits 13,796 ++ 15,098 --

peterzeller
51 commits 5,066 ++ 3,786 --

FairPlayer4
5 commits 139 ++ 115 --

deepthidevaki
327 commits 16,662 ++ 16,109 --

red17electro
1 commit 0 ++ 16,636 --

lplit
3 commits 497 ++ 2 --

Eipifi
79 commits 4,598 ++ 4,492 --

cmeiklejohn basho
256 commits 8,130 ++ 19,920 --

aletomsic
161 commits 10,829 ++ 8,244 --

UCL Université catholique de Louvain

marsleezm
122 commits 7,659 ++ 3,430 --

tcrain
255 commits 15,307 ++ 8,261 --

itoumlilt
6 commits 154 ++ 349 --

angbrav
117 commits 6,099 ++ 3,272 --

UPMC SORBONNE UNIVERSITÉS

liaud
1 commit 483 ++ 17 --

gyounes
31 commits 9,435 ++ 9,176 --

ergl
2 commits 1,050 ++ 764 --

rogerpueyo
3 commits 4 ++ 4 --

balegas
41 commits 2,341 ++ 1,122 --

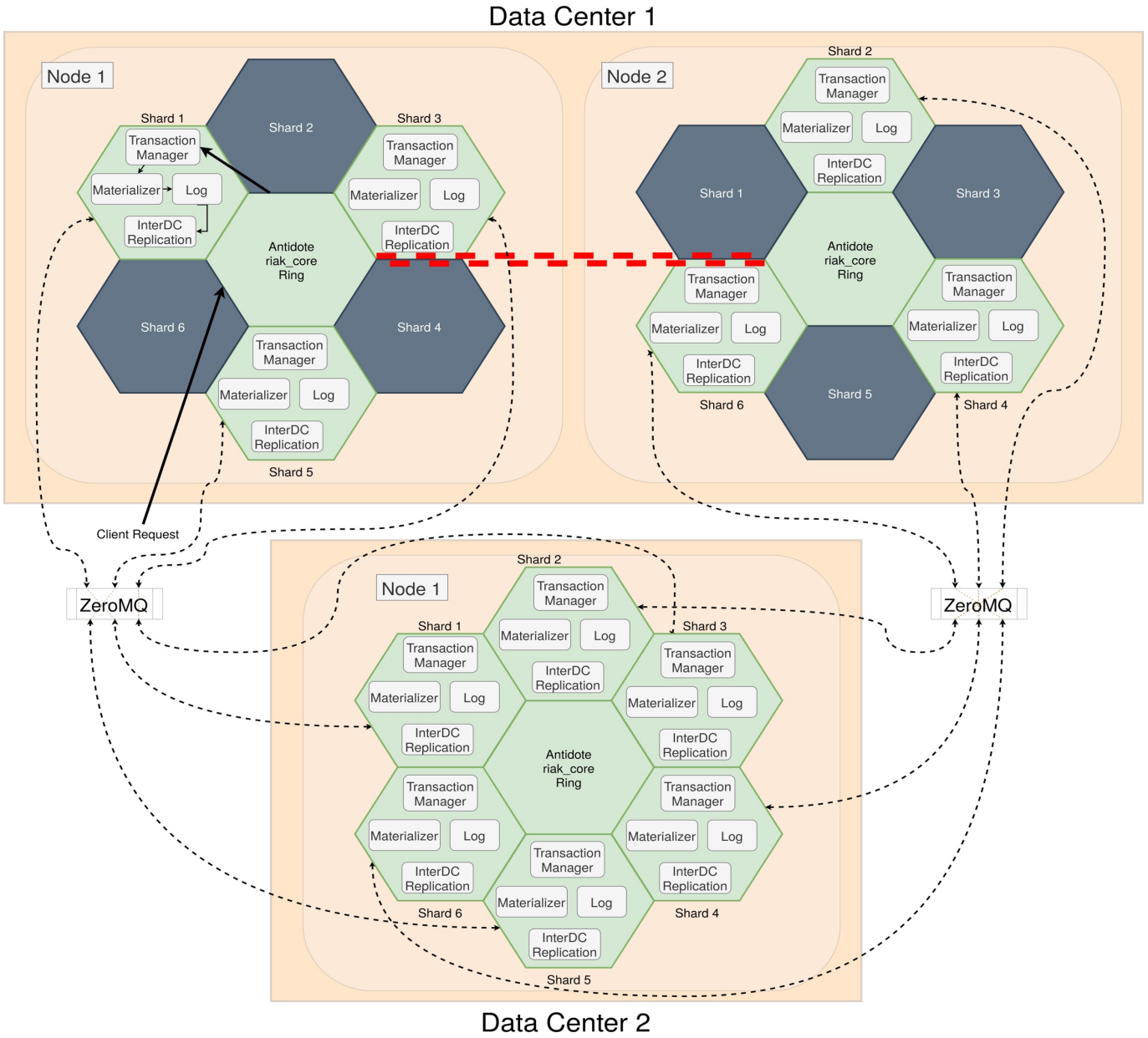
NOVALINCS LABORATORY FOR COMPUTER SCIENCE AND INFORMATICS

lythq
4 commits 705 ++ 139 --

Theses and Projects (small selection!)

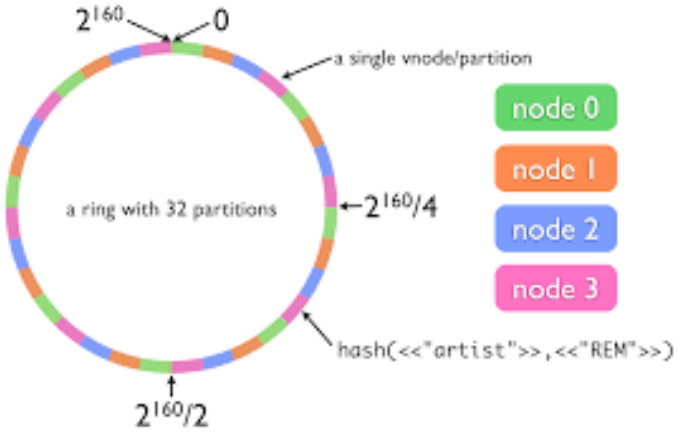
- Santiago Castineira: **Collaborative online web applications using Conflict-Free Replicated Data Types** (Master's Thesis, 2014)
- Gonçalo Thomas: **FMKe: a Real-World Benchmark for Key-Value Data Stores** (Master's thesis, 2017)
- Tim Dellmann: **Implementation of a calendar app on a weakly consistent data storage** (Bachelor's thesis, 2017)
- Gonçalo Cabrita: **Non-uniform replication for replicated objects** (Master's thesis, 2017)
- Luc François: **Big Sets for Antidote** (Master's thesis, 2017)
- Pedro Lopes: **Antidote SQL: SQL for Weakly Consistent Databases** (Master's thesis, 2018)
- Server Khalilov: **Offline caching in web applications for AntidoteDB** (Master's thesis, 2018)
- Ala Harirchi: **Minidote+: A Transactional CRDT Store for the Edge** (Master's thesis, 2020)
- Yannick Wagner: **Skalierbare Verteilung und Leistungsauswertung von Antidote mit Kubernetes** (Bachelor's thesis, 2020)
- Kevin Bartik: **Caching and Storage for distributed transactional CRDT databases** (Master's thesis, 2020)
- Pascal Grosch: **Adopting Random Slicing for Riak Core** (Master's thesis, 2021)
- Ayush Pandey: **Persisting the AntidoteDB Cache: Design and Implementation of a Cache for a CRDT Datastore** (Master's thesis, 2022)
- Ali Hussein Rezzae: **Range query implementation for distributed key-value stores** (Master's thesis, ongoing)
- ...

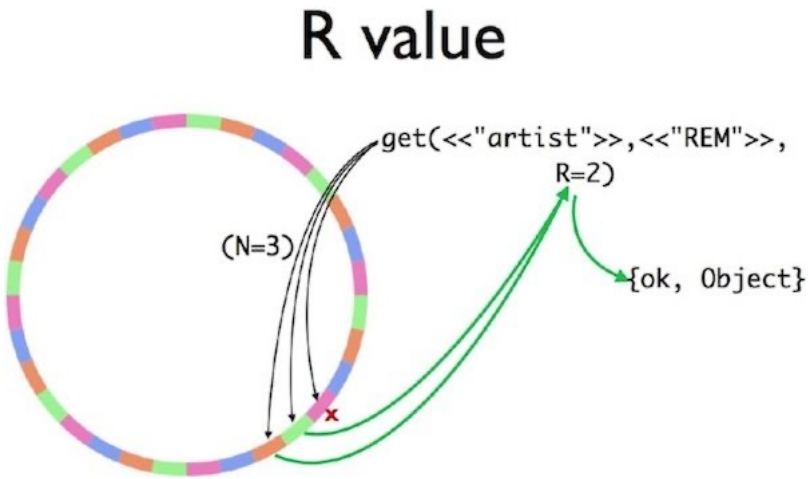
Architecture



riak_core_light

- Management of a ring overlay for distributed systems
- Trusted, fault-tolerant, and highly scaling system
- Contributions: Complex refactoring & minimization
- Open sourced in Summer 2019 → already industry adaptation
- Problem: Antidote misses intraDC replication
  - Cure protocol not designed for quorum-based replication





https://www.infoq.com/articles/dynamo-riak-random-slicing/

Scott L. Fritchie
@slfritchie

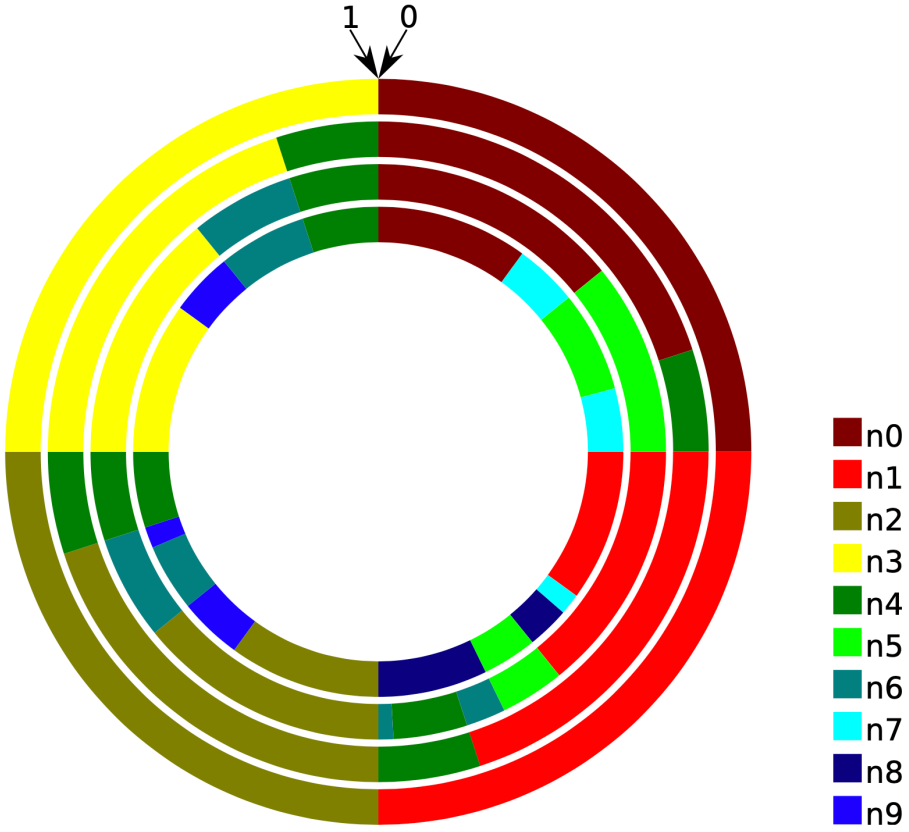
InfoQ has published my article comparing the Random Slicing consistent hashing algorithm with Riak Core's Dynamo-style hashing. What was great in 2008 isn't always great today. Find out why at

https://www.infoq.com/articles/dynamo-riak-random-slicing/

Assumptions and limits of Riak Core's implementation of the Dynamo-style consistent hashing algorithm

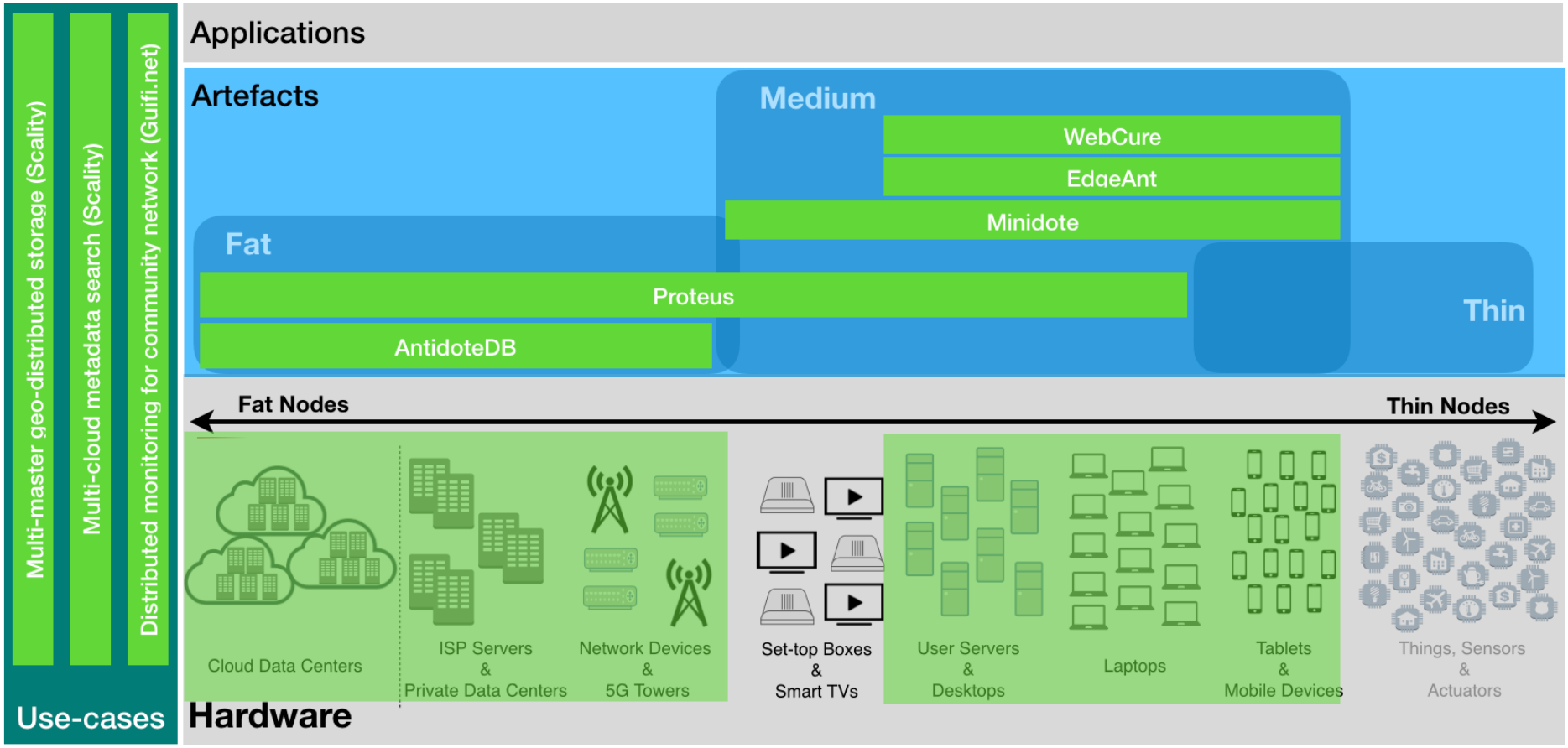
1. *The only string-to-integer hash algorithm is SHA-1.*
2. *The hash "rings" integer interval is the range 0 to 2^160-1.*
3. *The **number of partitions is fixed.***
4. *The number of partitions must be a power of 2.*
5. *The **size of each partition is fixed.***
6. *Historically, the "claim assignment" algorithm used to assign servers to intervals on the ring were buggy and naive and frequently created imbalanced workload across nodes.*
7. *Server capacity adjustment by "weighting" did not exist: **all servers were assumed equal.***
8. ***No support for segregating extremely "hot" keys**, for example, Twitter's "Justin Bieber" account.*
9. *No effective support for "rack-aware" or "fault domain aware**" replica placement policy"***

Random Slicing



Random Slicing example scaling from 4 to 5 to 7 to 10 nodes with homogeneous capacity

AntidoteDB Ecosystem in 2020



LIGHTKONE

Lightweight computation for networks at the edge

AntidoteDB Ecosystem

Minidote

- Mini-version of Antidote for mid-edge systems
- Reduced functionality (no transactions)
- Derived by requirements of Guifi Monitoring use case

EdgeAnt

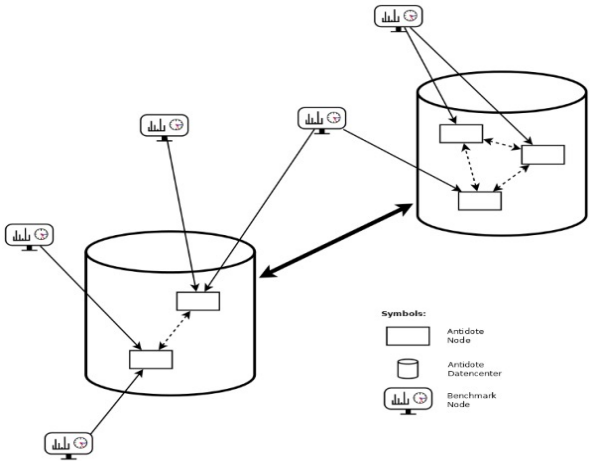
- (In-memory) cache located in Edge devices
- Uses Antidote as backend server storage for synchronization
- Partial replication
- Offline support

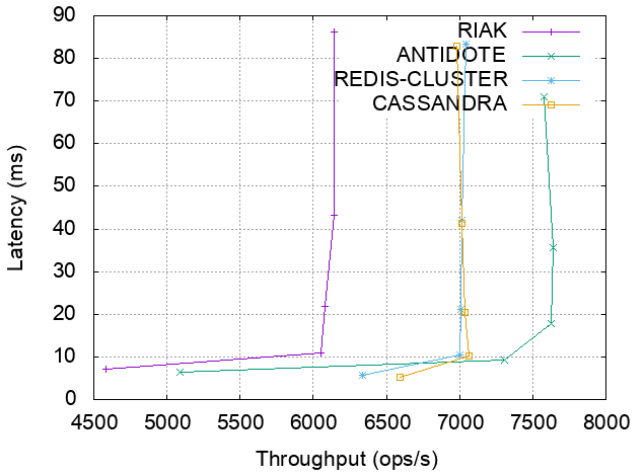
WebCure

- Active client-cache for web apps
- Demo app: Shared Calendar

Deployment and Benchmarking

- Public images on DockerHub
- Deployment with Kubernetes
- Setup to scale benchmarks
  - New Antidote Basho Bench branch



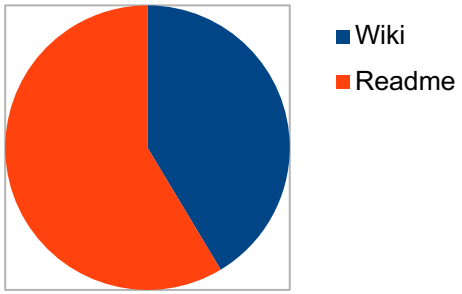


Monitoring

- Industry-level real time monitoring and alarm system
  - Over 30 custom metrics supported
  - Monitoring can (should) be deployed on remote systems

Support structures for developers

- Open sourced on Github:
  > 80 forks, ~~344~~ (2018) ~~401~~ (2019) ~~481~~(2020) 636 (2022) Stars
- Support via Slack channel (99 members)
- Weekly meetings (till 2021)
- Tutorials
- Over 5000 lines of documentation text



The future

What's next?

- Checkpointing and log truncation
  - ➢Talk by Ayush Pandey
- Storage layer
- Performance: Improving throughput
- Robustness
  - Backpressure and failure modes under high throughput
  - Refactoring of inter-DC communication
  - Intra-DC shard replication
- Programming model and API

Thank you!


AntidoteDB
A planet-scale, available, transactional database with strong semantics
Why Antidote?
MacBook Pro
Bières artisanales aux céré
ANTIDOTE
BIÈRE ARTISANALE
aux Céréales d'Auvergne
AMBRÉE