Raft Consensus Algorithm

What is Raft?

Raft is a consensus algorithm that is designed to be easy to understand. It's equivalent to Paxos in fault-tolerance and performance. The difference is that it's decomposed into relatively independent subproblems, and it cleanly addresses all major pieces needed for practical systems. We hope Raft will make consensus available to a wider audience, and that this wider audience will be able to develop a variety of higher quality consensus-based systems than are available today.

Hold on—what is consensus?

Consensus is a fundamental problem in fault-tolerant distributed systems. Consensus involves multiple servers agreeing on values. Once they reach a decision on a value, that decision is final. Typical consensus algorithms make progress when any majority of their servers is available; for example, a cluster of 5 servers can continue to operate even if 2 servers fail. If more servers fail, they stop making progress (but will never return an incorrect result).

Consensus typically arises in the context of replicated state machines, a general approach to building fault-tolerant systems. Each server has a state machine and a log. The state machine is the component that we want to make fault-tolerant, such as a hash table. It will appear to clients that they are interacting with a single, reliable state machine, even if a minority of the servers in the cluster fail. Each state machine takes as input commands from its log. In our hash table example, the log would include commands like *set x to 3*. A consensus algorithm is used to agree on the commands in the servers' logs. The consensus algorithm must ensure that if any state machine applies *set x to 3* as the n^{th} command, no other state machine will ever apply a different n^{th} command. As a result, each state machine processes the same series of commands and thus produces the same series of results and arrives at the same series of states.

Raft Visualization

Here's a Raft cluster running in your browser. You can interact with it to see Raft in action. Five servers are shown on the left, and their logs are shown on the right. We hope to create a screencast soon to explain what's going on. This visualization (RaftScope) is still pretty rough around the edges; pull requests would be very welcome.

The Secret Lives of Data is a different visualization of Raft. It's more guided and less interactive, so it may be a gentler starting point.

Publications

This is "the Raft paper", which describes Raft in detail: In Search of an Understandable Consensus Algorithm (Extended Version) by Diego Ongaro and John Ousterhout. A slightly shorter version of this paper received a Best Paper Award at the 2014 USENIX Annual Technical Conference.

Diego Ongaro's Ph.D. dissertation expands on the content of the paper in much more detail, and it includes a simpler cluster membership change algorithm.

More Raft-related papers:

- Doug Woos, James R. Wilcox, Steve Anton, Zachary Tatlock, Michael D. Ernst, and Thomas Anderson.
  Planning for Change in a Formal Verification of the Raft Consensus Protocol.
  Certified Programs and Proofs (CPP), January 2016.

- James R. Wilcox, Doug Woos, Pavel Panchekha, Zachary Tatlock, Xi Wang, Michael D. Ernst, and Thomas Anderson.
  Verdi: A Framework for Implementing and Verifying Distributed Systems.
  Programming Language Design and Implementation (PLDI), June 2015.

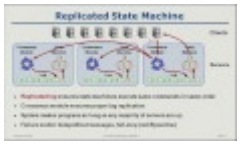
- Hugues Evrard and Frédéric Lang.
  Automatic Distributed Code Generation from Formal Models of Asynchronous Concurrent Processes.
  Parallel, Distributed, and Network-Based Processing (PDP), March 2015.

- Heidi Howard, Malte Schwarzkopf, Anil Madhavapeddy, and Jon Crowcroft.
  Raft Refloated: Do We Have Consensus?.
  SIGOPS Operating Systems Review, January 2015.

- Heidi Howard.
  ARC: Analysis of Raft Consensus.
  University of Cambridge, Computer Laboratory, UCAM-CL-TR-857, July 2014.

Talks

These talks serve as good introductions to Raft:

- 

  Video
- Video
- Video
- Video
- Video
- Video
- Video

Courses teaching Raft

This is a list of courses that include lectures or programming assignments on Raft. This might be useful for other instructors and for online learners looking for materials. If you know of additional courses, please submit a pull request or an issue to update it.

- IIT Bombay, CS 733: Cloud Computing, Sriram Srinivasan. Includes Raft programming assignment in Go (assignments). (Spring 2014, Spring 2015, ...)

Where can I ask questions?

The best place to ask questions about Raft and its implementations is the raft-dev Google group. Some of the implementations also have their own mailing lists; check their READMEs.

Where can I get Raft?

There are many implementations of Raft available in various stages of development. This table lists the implementations we know about with source code available. The most popular and/or recently updated implementations are towards the top. This information will inevitably get out of date; please submit a pull request or an issue to update it.

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
etcd/raft	Blake Mizerany, Xiang Li and Yicheng Qin	Go	Apache 2.0	Yes	Yes	Yes
TiKV	Jay, ngaut, siddontang, tiancaiamao.	Rust	Apache2	Yes	Yes	Yes
RethinkDB/clustering		C++	AGPL	Yes	Yes	Yes
SOFAJRaft	Boyan, Jiachun.	Java	Apache2.0	Yes	Yes	Yes
dragonboat	Lei Ni	Go	Apache2	Yes	Yes	Yes
hashicorp/raft	Armon Dadgar (hashicorp)	Go	MPL-2.0	Yes	Yes	Yes
LogCabin	Diego Ongaro (Stanford)	C++	ISC	Yes	Yes	Yes

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
braft	Zhangyi Chen, Yao Wang	C++	Apache 2.0	Yes	Yes	Yes
hazelcast-raft	Mehmet Dogan, Ensar Basri Kahveci	Java	Apache 2.0	Yes	Yes	Yes
willemt/raft	Willem-Hendrik Thiart	C	BSD	Yes	Yes	Yes
go-raft	Ben Johnson (Sky) and Xiang Li (CMU, CoreOS)	Go	MIT	Yes	Partial?	Yes
Kudu	David Alves, Todd Lipcon, Mike Percy	C++	Apache2	Yes	Yes	Yes
canonical/raft	Free Ekanayaka (Canonical)	C	Apache2.0	Yes	Yes	Yes
Ra	Team RabbitMQ	erlang	ASL2 / MPL1.1	Yes	Yes	Yes (w modifi
OpenDaylight	Moiz Raja, Kamal Rameshan, Robert Varga (Cisco), Tom Pantelis (Brocade)	Java	Eclipse	Yes	No	Yes
raftos	Alexander Zhebrak	Python	MIT	Yes	No	No

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
Zatt	Simon Accascina	Python	AGPL	Yes	Yes	Yes
bakwc/PySyncObj	Filipp Ozinov	Python	MIT	Yes	Yes	Yes
simpleRaft	Sean Reed	Python	MIT			
Qihoo360/raft	ZongzhiChen, AnAnZhaoKangWang	C++	GPL3	Yes	No	No
jgroups-raft	Bela Ban	Java	Apache2	Yes	Yes	Yes
.NEXT Raft	Roman Sakno	C#	MIT	Yes	Yes	No but enable third-p is sup
peterbourgon/raft	Peter Bourgon (SoundCloud)	Go	Simplified BSD	Yes	Yes	No
ckite	Pablo Medina	Scala	Apache2	Yes	Yes	Yes
akka-raft	Konrad Malawski	Scala	Apache2	Yes	Yes	Yes

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
liferaft	Arnout Kazemier	Javascript	MIT			
kanaka/raft.js	Joel Martin	Javascript	MPL-2.0	Yes	Yes	No
hoverbear/raft	Andrew Hobden, Dan Burkert	Rust	MIT	Yes		
Juno (BFT variant of Raft)	Brian Schroeder, Libby Kent, Stuart Popejoy, Will Martino.	Haskell	BSD			
jraft	Andy Chen	Java	Apache2	Yes	Yes	Yes
rafter	Andrew Stone (Basho)	Erlang	Apache2			
zraft_lib	Gunin Alexander	Erlang	Apache2	yes	yes	yes

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
verdi-raft	James Wilcox, Doug Woos, Pavel Panchekha, Zach Tatlock, Xi Wang, Mike Ernst, and Tom Anderson (University of Washington)	Coq	BSD	Yes	No	No
copycat	Jordan Halterman	Java	Apache2	Yes	Yes	Yes
Ratis		Java	Apache2			
ocaml-raft	Heidi Howard (Cambridge)	OCaml	MIT	Yes	No	No
skiff	Pedro Teixeira	Javascript	ISC	Yes	Yes	Yes
Permazen/RaftKVDatabase	Archie Cobbs	Java	Apache2	Yes	Yes	Yes
harryw/raft	Harry Wilkinson	Ruby	MIT			
kontiki	Nicolas Trangez	Haskell	BSD	Some	No	No

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
srned/Prez	Sureshkumar Nedunchezhian	C	BSD	Yes	No	No
floss	Alexander Flatter	Ruby	MIT			
py-raft	Toby Burress	Python	public domain	Lacking persistence	Yes	No
raft-clj	John Weaver	Clojure	Eclipse			
ScaleCube Raft Leader Election	ScaleCube	Java	Apache 2.0	Yes	No	No
whitewater	Adam Midvidy, Anh Mai, Karoun Kasraie, Sanketh Katta (Berkeley)	Bloom	MIT	Some correctness issues	No	No
allengeorge/libraft	Allen George	Java	BSD	Yes	No	No
Raft-php	Waqee Khalid	PHP	MIT	Yes	No	No
xraft	XnnYygn	Java	MIT	Yes	Yes	Yes

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
Gondola	Patrick Chan, Wei-Cheng Pan	Java	New BSD	Yes	Yes	In prog
RaftCore	Guillermo Rodríguez	C#	MIT	Yes		
raft-kotlin	Sasha Chepurnoi	Kotlin	MIT	Yes	Yes	No
NRaft	Bert Willems (Premotion)	C#	MIT			
gaggle	Ben Ng	Javascript	MIT	Yes	Partial	No
barge	Dave Rusek	Java	Apache2	Yes	No	No
tetrapods/raft	Aaron Davidson	Java	Apache2	Yes		Yes
dupdob/RAFTiNG	Cyrille Dupuydauby	C#	Apache2			
noeleo/raft	Noel Moldvai, Rohit Turumella, Josh Muhlfelder, James Butkovic (Berkeley)	Bloom	Simplified BSD	Lacking persistence	No	No

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
dataleading/easyRaft	Shanliang Shen	Java	Apache2.0	Yes	No	In Pro
lite-raft	Luigi Tarenga	Shell	MIT	Yes	Yes	Yes
aioraft	lisael	Python	AGPL	Lacking persistence	Yes	No
rafute	Yuki Ito	Elixir	MIT	Yes	No	No
xingyif/raft	Yifan Xing	Python	Apache 2.0	Yes	Yes	Yes
chicm/CmRaft	Cheng Min Chi	Java	Apache2	Yes	No	No
archie/raft	Marcus Ljungblad	Scala		Some	No	No
Flotten	Henrik Feldt (Jayway)	F#	MIT	Some	No	No
Cornerstone	Andy Chen	C++	Apache2	Yes	Yes	Yes
Vesper	Nicola Manzini, Peng Ding.	Python	MIT	Yes	Yes	No

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
melee	Fredrick Galoso	Clojure	Eclipse			
C5 replicator	Ryan Rawson, Alex Newman, and Josh Greenberg	Java	Apache2	Yes	Yes	Yes
RSM	Olivier Paugam	Rust	MIT	Yes	Soon	Yes
Chillaxd	Yassine Lamgarchal	Python	Apache2	Yes	No	No
pontoon	Matt Reiferson	Go				
Riff	Aaron Pritzlaff	Scala	Apache2	Yes	Yes	Yes
IvanProdaiko94/raft-protocol-implementation	Ivan Prodaiko	Go	MIT	Yes	No	No
Raft-JVM	Tim Kellogg (Alteryx)	Java		No	No	No
dannycoates/raft-core	Danny Coates	Javascript	BSD			

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
cppa-raft	Zhang Yichao	C++	MIT	Partial	No	No
graft	Ben Mills and William Dix (Braintree)	Go		Partial		
huntlabs/hunt-raft	huntlabs	D	Apache2	Yes	Yes	Yes
meatcompute/raft-consensus	Mikaela Patella	Clojure	Eclipse			
RaftShop	Matteo Berti, Arnaldo Cesco , Salvatore Fiorilla	Jolie	LGPL	Yes		
mruby-flotte	Hendrik	mruby	Apache2	Yes	No	No
dinghy	Trevor Bernard (UserEvents)	Clojure	Apache2			
giraft	Patrick Van Stee	Ruby	MIT			
yora	Huy Le	Ruby	MIT	Yes	Yes	

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
r4j	Kaarel Kann	Java	Apache2	Yes	No	No
huckleberry	Jakob Sievers	Erlang				
zodiac-prime	Evan Phoenix (LivingSocial)	Ruby	MIT			
rodriguezvalencia/rafting	Sergio Rodriguez	Clojure	MIT	Partial	No	No
Raft4WS	Filipe Campos	Java	Apache2	Yes	No	No
fxsjy/Ins	Junyi Sun	C++	BSD	Yes	No	Yes
draft	Patrick Van Stee	Elixir				
rafterl	Eric Moritz	Erlang				
fsraft	Karl-Johan Nilsson	F#	public domain			

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
bspolley/raft	Alex Kaiser, Brennan Polley, Helen Weng (Berkeley)	Bloom		Some		
seaturtles	Lionel Barrow (Braintree)	Go				
drpicox/uoc-raft-2013p	David Rodenas	Java	GPL3			
scalaraft	Kim Je Min	Scala	Apache2			
pvilas/raft	Pere Vilas	Java				
chelan	Eric Jutrzenka	Scala		Yes	No	No
jalvaro/raft	Jordi Alvaro	Java				
cb372/raft	Chris Birchall	Scala				
jpathy/raft	Jiten Pathy	Go	WTFPL			

Name	Primary Authors	Language	License	Leader Election + Log Replication?	Membership Changes?	L Comp
viile/raft	viile	D	MIT	Yes	No	No
d-raft	Dengos Wei	C++	Apache2	Yes	No	No
raft-angular	Dushyant Sabharwal	Angular 4	MIT	Yes	Yes	No
raft-java	Wenwei Hu	Java	Apache2	Yes	Yes	Yes
Theseus	Eloquent Labs (Keenon Werling, Gabor Angeli, et al.)	Java	MIT	Yes	Yes	Yes