



Apache Hadoop 3.x State of The Union and Upgrade Guidance

Wei-Chiu Chuang
@Cloudera,
Apache Hadoop PMC

Wangda Tan
Sr. Manager, Compute Platform
@Cloudera,
Apache Hadoop PMC



Agenda

- ❑ Hadoop Community Updates & Overview
- ❑ Updates from YARN, Submarine, HDFS, Ozone
- ❑ Upcoming releases
- ❑ Upgrade guidance



Community Updates





is hadoop dead



Is Hadoop Dead? How Kubernetes and Cloud-Native Could ...

<https://www.bmc.com> › [blogs](#) › [hadoop-cloud-native-kubernetes](#) ▼

Is Hadoop Officially Dead? - Datanami

<https://www.datanami.com>

Is Hadoop

<https://medium.com>

Fact or Fake?

[The Death of Hadoop? | Transforming Data with Intelligence](#)

<https://tdwi.org> › [articles](#) › [2019/02/26](#) › [ta-all-death-of-hadoop](#) ▼

[Why is Hadoop dying? | Packt Hub](#)

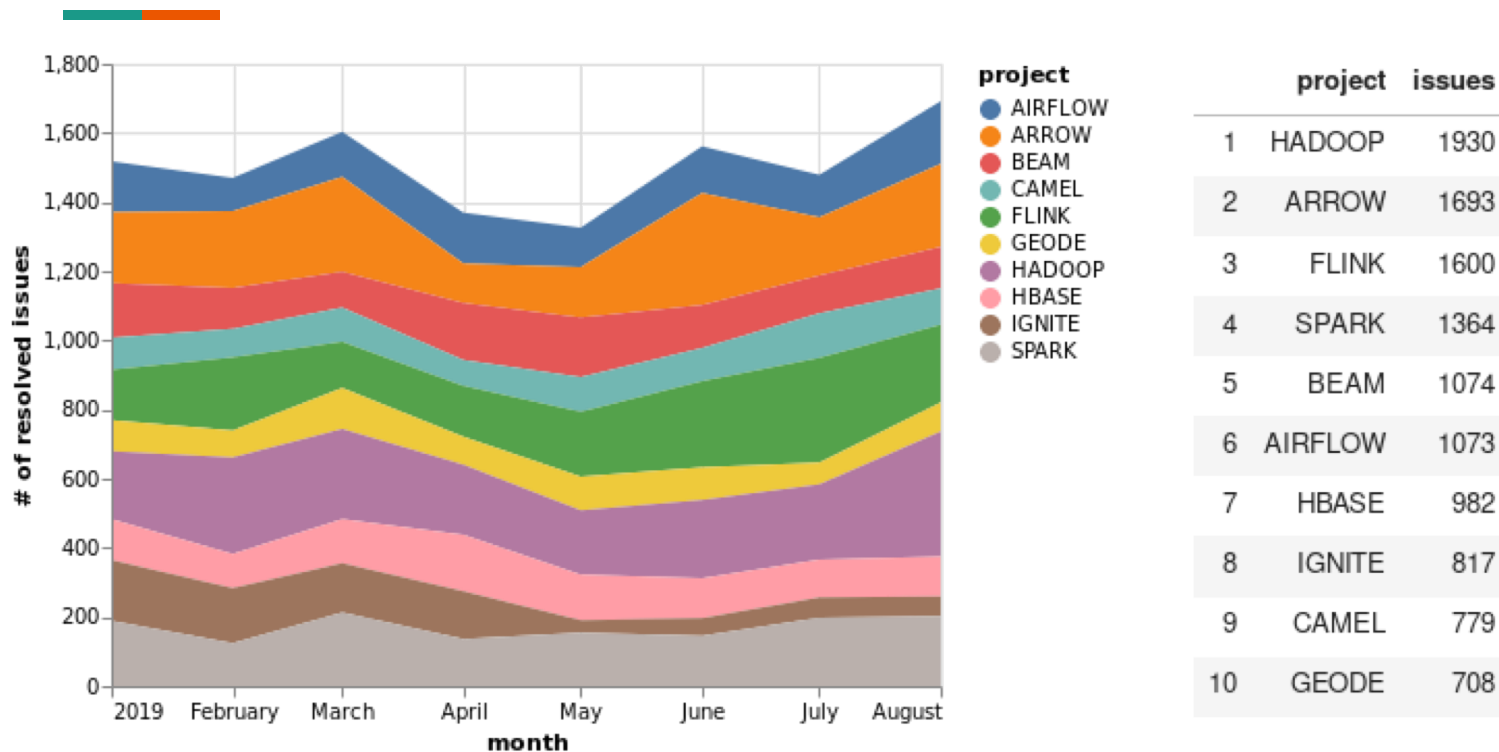
<https://hub.packtpub.com> › [why-is-hadoop-dying](#) ▼

Is Hadoop dead and is it time to move to Spark? - Quora

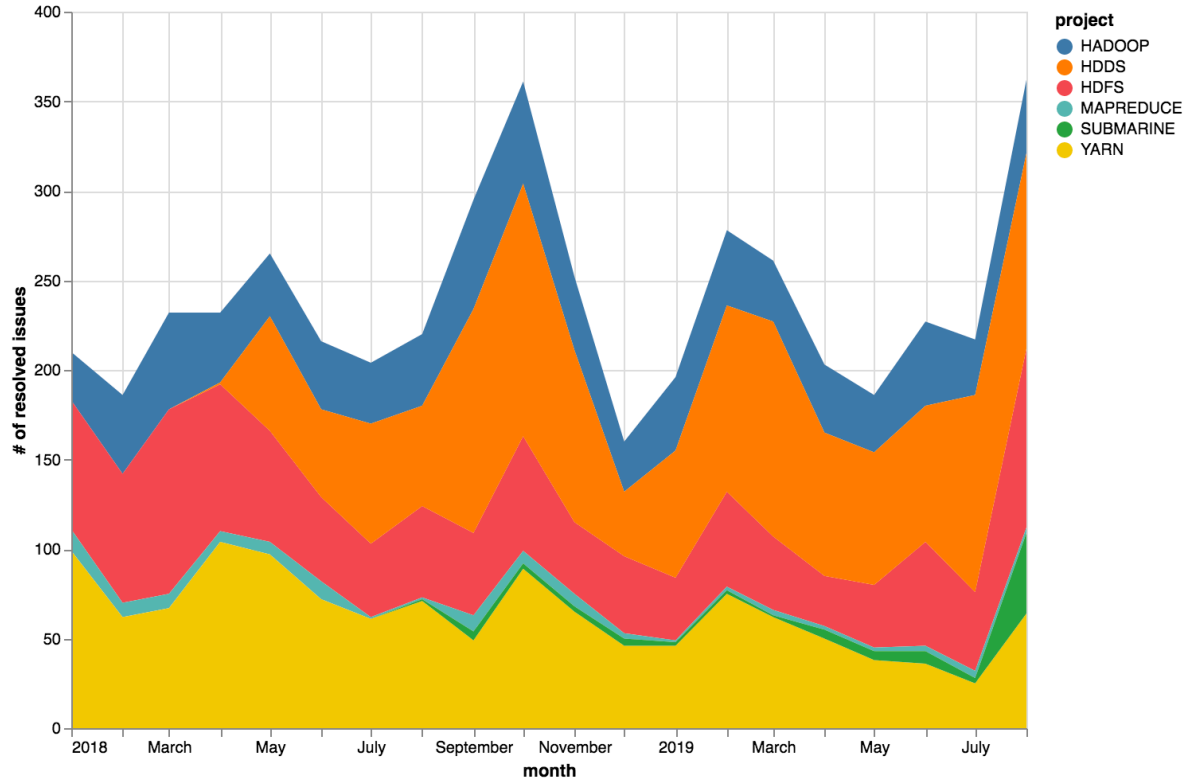
<https://www.quora.com> › [Is-Hadoop-dead-and-is-it-time-to-move-to-Spark](#) ▼

Jun 24, 2015 - ATA I think this is like asking: is Linux **dead** and should we move to Docker? or

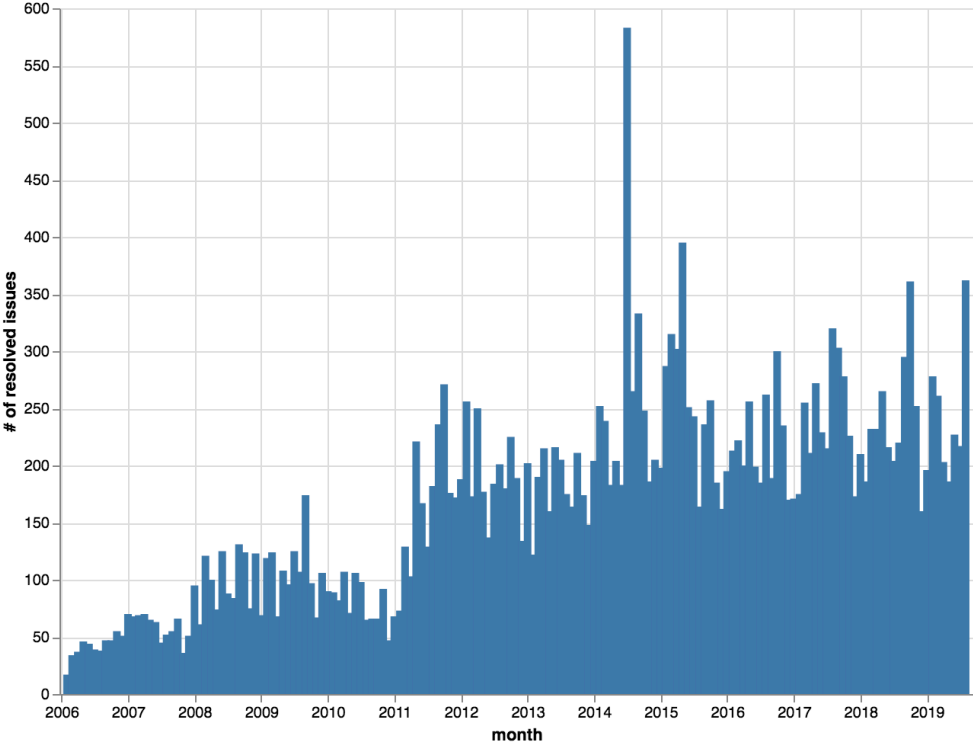
Resolved Issues by Top 10 ASF Projects



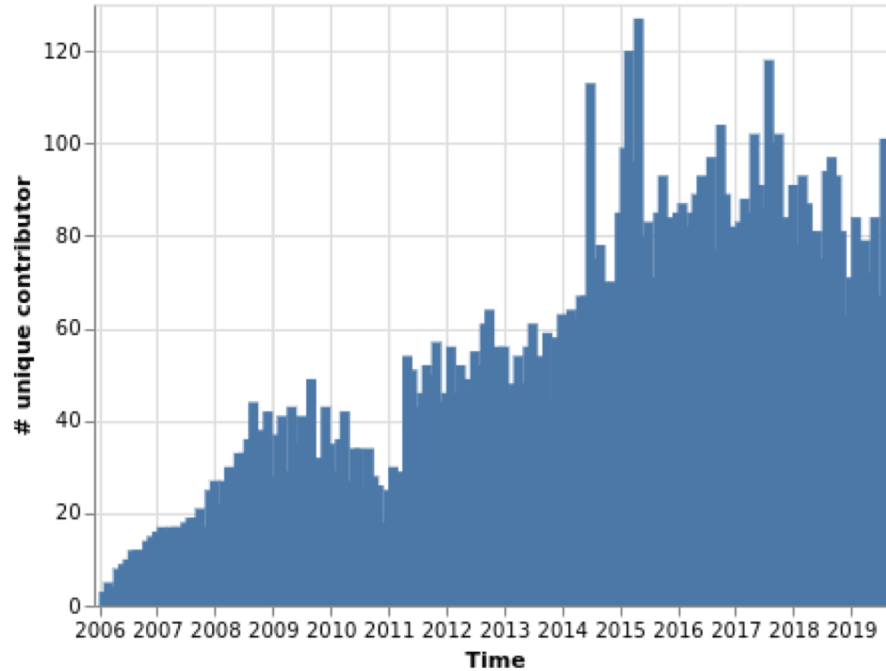
Resolved Issues within Hadoop by Subproject (Monthly)



Resolved issue in Hadoop (Monthly)



Number of Unique #Contributors of Hadoop (Monthly)



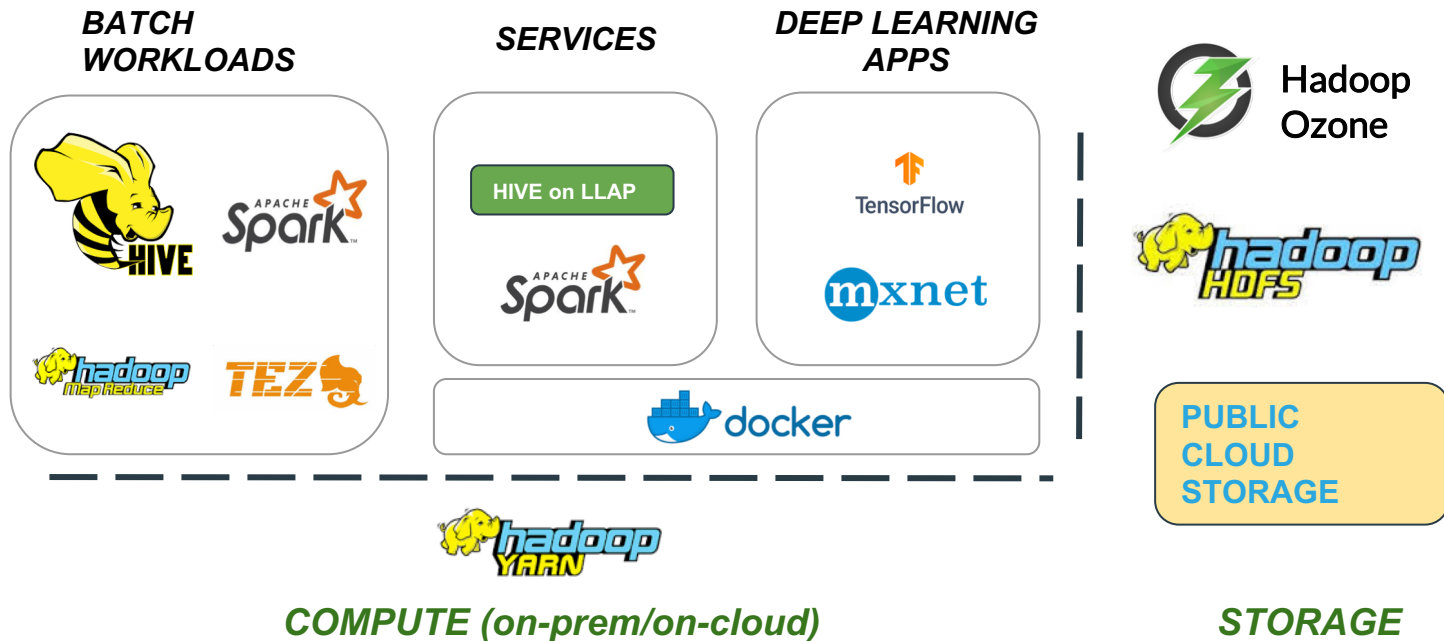
(All pictures credits to Marton Elek)



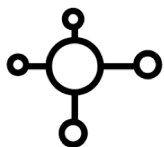
Hadoop 3.x Overview



Big Data/Long Running Services With Hadoop 3



Themes of Hadoop 3.x



Scalability



Containerization



Cost-efficiency



Cloud-native



Machine
Learning



YARN



Containerization

- ❑ Production-ready Docker container support on YARN.
 - ❑ Containerized Spark
 - ❑ Package/Dependency Isolation
- ❑ Interactive Docker Shell support (YARN-8762)
- ❑ OCI/squashfs (Like runc) container runtime.

Available since 3.1.0

Available since 3.3.0

Target 3.3.0

YARN in a cloud-native environment YARN-9548

Ongoing Effort

- ❑ Autoscaling
 - ❑ Scaling recommendations
- ❑ Smarter scheduling
 - ❑ **Bin-packing** Pack containers as opposed to spreading them around to downscale nodes better
 - ❑ Account for speculative nodes like spot instances
- ❑ Downscaling nodes
 - ❑ Improved Decommissioning
 - ❑ Consider shuffle/auxiliary services data

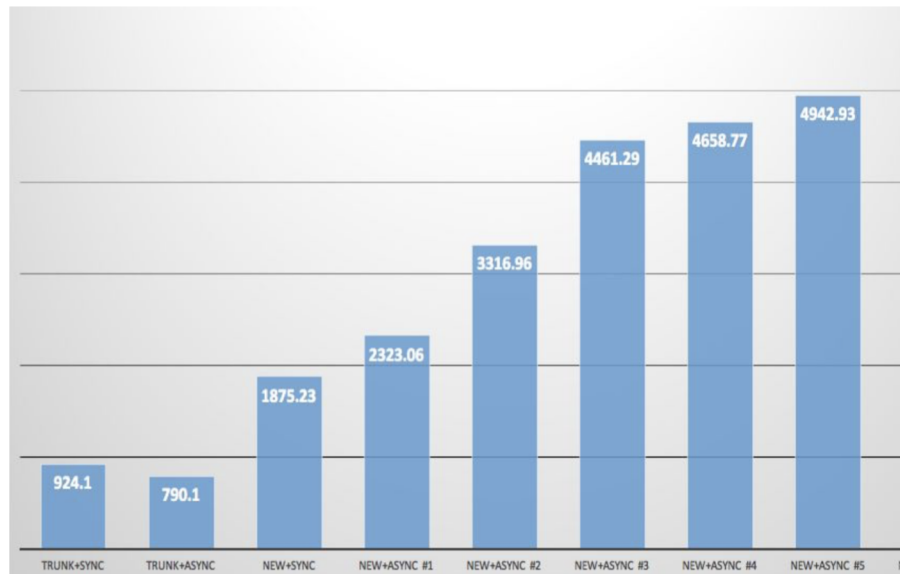
Global Scheduling Framework **YARN-5139**



Available since 3.0.0

Scheduler Capabilities enhancements

- Look at several nodes at one time.
- Fine grained locks.
- Multiple allocation threads.
- 5-10x allocation throughput gains.



Other Enhancements

- Node Attributes: Tagging node with attribute and schedule containers based on that. (3.2.0)
- Placement Constraint: Affinity, Anti-Affinity, etc. (3.1.0)
- Dynamic Auto Queue Creation (Capacity Scheduler) (3.1.0)
- Scheduling Activity Troubleshooter. (3.3.0)

Total Outstanding Resource Requests: <memory:1720320, vCores:1260>

Priority ▲	AllocationRequestId	ResourceName	Capability	NumContainers	RelaxLocality	NodeLabelExpression	ExecutionType
20	-1	*	<memory:40960, vCores:30>	42	true		GUARANTEED

Showing 1 to 1 of 1 entries

Diagnostics in cache updated at 2019-05-20 21:13:44
 (Update via [Application Activities REST API](#))

Priority	AllocationRequestId	AllocationState	Diagnostics
20	-1	SKIPPED	1 node: insufficient resources=[vcores], 3 nodes: insufficient resources=[memory-mb, vcores]



Submarine



Machine Learning – Hadoop Submarine

- ❑ Started since Aug 2018.
- ❑ Benefit from Hadoop's feature like GPU/Docker on YARN support.
- ❑ Enables Infra engineers / data scientists to run deep learning apps
 - ❑ Tensorflow, Pytorch, MXNet.. on YARN/K8s
 - ❑ Supports Hadoop 2.7+.
 - ❑ LinkedIn TonY joined Submarine family



Machine Learning – Hadoop Submarine

- ❑ Lots of new stuff in upcoming releases (0.3.0).
 - ❑ Mini-submarine for easy trying Submarine from single node.
 - ❑ Brand-new Submarine web interface for end-to-end user Experiences.
 - ❑ Tensorflow/PyTorch on K8s.
- ❑ 15+ Contributors and community is fast growing..



Machine Learning – Hadoop Submarine Prod Use cases

NetEase:

- One of the largest online game/news/music provider in China.
- 245 GPU Cluster runs Submarine.
- One of the model built is music recommendation model which invoked 1B+/days.

LinkedIn:

- 250+ GPU machines
- 500+ TensorFlow trainings/day.
- Serves applications in recommendation systems and NLP.
- Collaboration on Submarine/TonY runtime and SDK development.

Ke.com:


- 50+ GPU machines (includes 19 multi-v100 GPU machines), based on Hadoop trunk (3.3.0).
- Serves applications like image/voice recognition, etc.


And many users are evaluating Submarine...

Machine Learning – Submarine new UI demo

New Submarine UI

127.0.0.1:8080/#/workbench/workspace/project
Project - Submarine Workbench

 **Submarine**

 liuxun

- 🏠 Home
- 📁 Workspace
- 🔧 Interpreter
- 📄 Job
- 📊 Data
- 📦 Model
- ⚙️ Manager
- 📈 仪表盘
- 📄 表单页
- 📄 列表页
- 📄 详情页
- 📄 结果页
- ⚠️ 异常页
- 👤 个人页
- ⚙️ 其他组件


Home / Workspace / Workspace / Project
🔍

Project

You can click on the project tree directory on the left and the project file list will be displayed on the right.

Project (2)
Release (4)
Training (8)
Team (3)
Shared (5)

+ Add New Project




Tensorflow-test1

python
test x
mnist x
+New Tag

Basic Operations on multi-GPU (notebook) (code). A simple example to introduce multi-GPU in TensorFlow. Basic Operations on multi-GPU (noteb...

☆ 10
👍 24
🗨 2

🔗 Edit
📄 Download
⚙️ Setting
⋮




Tensorflow-test1

python
test x
mnist x
+New Tag

Basic Operations on multi-GPU (notebook) (code).

☆ 10
👍 24
🗨 2

🔗 Edit
📄 Download
⚙️ Setting
⋮




Tensorflow-test1

python
test x
mnist x
+New Tag

Basic Operations on multi-GPU (notebook) (code). A simple example to introduce multi-GPU in TensorFlow. Train a Neural Network on multi-GPU ...

☆ 10
👍 24
🗨 2

🔗 Edit
📄 Download
⚙️ Setting
⋮



Tensorflow-test1

python
test x
mnist x
+New Tag

Basic Operations on multi-GPU (notebook) (code). A simple example to introduce multi-GPU in TensorFlow. Train a Neural Network on multi-GPU ...

☆ 10
👍 24
🗨 2

🔗 Edit
📄 Download
⚙️ Setting
⋮

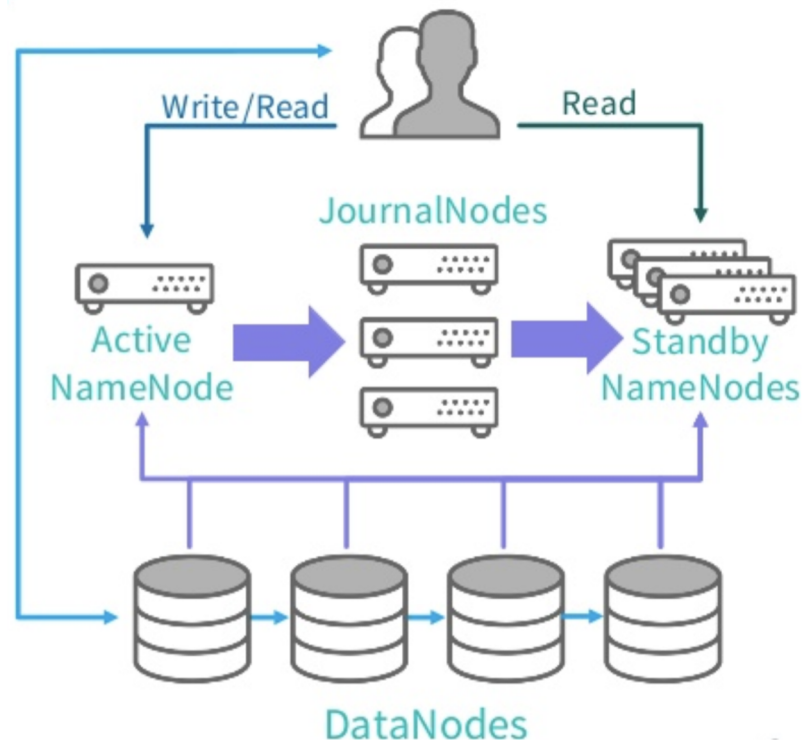
🌐 Submarine Site
 Copyright © 2018 submarine.apache.org



Storage

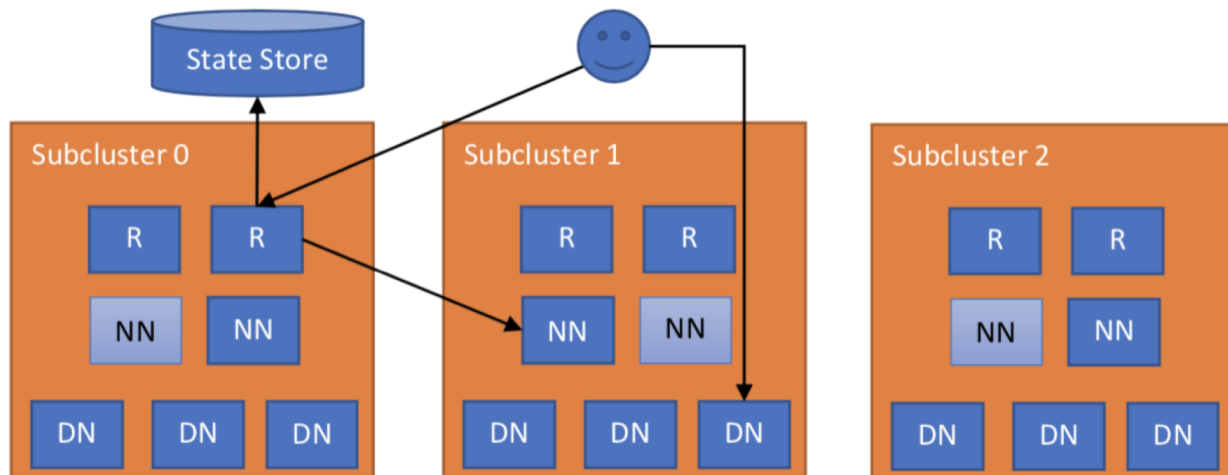
HDFS Updates - Consistent Read from Standby

- ❑ Offload reads to non-active NameNodes to improve overall file system performance.
- ❑ Consistency: if a client can report the last transaction ID seen by it, then a standby can allow a read if it has caught up to that transaction ID seen by the client.
- ❑ Used in production at Uber and LinkedIn.



HDFS Updates - Router Based Federation

- ❑ Router based Federation Supports Security.
- ❑ Lots of work on scalability and the ability to handle slower sub-clusters.
- ❑ We are seeing usage across the industry



And many more HDFS features

- Selective Wire Encryption
- Cost based Fair call queue
- Dynamometer
- Storage Policy Satisfier
- Support Non-volatile storage class memory in HDFS cache directives

Ongoing development

- RPC support for TLS
- KMSv2
- OpenTracing integration
- JDK11 support

Cloud Connector Updates - S3A/S3Guard

S3A File system supports Delegation Tokens.

- ❑ Full user + secret + encryption keys: simplest, but secrets do not leave your system.
- ❑ Generated session tokens + encryption keys: keeps the long lived secrets locally; life of non-renewable tokens limited

S3Guard is no longer considered experimental

- ❑ Maintain consistency through corner cases involving partial failure of rename/delete operations.
- ❑ Out of band support - detecting and adapting to other applications overwriting files.
- ❑ Tracking of etag and version Ids for stricter consistency when you want to defend against OOB changes.
- ❑ “authoritative mode” improves performance dramatically.

ABFS: “Azure Datalake Gen 2” Connector

- ❑ A high performance cloud store & filesystem for Azure
- ❑ Added in Hadoop 3.2.0;
- ❑ Stabilization in trunk with all fixes backported to 3.2.1
- ❑ Has a similar extension point for Delegation Token plugins as S3A. (though implementing DTs is “left as an exercise”. Contributions welcome)

*Credit to Thomas Marquardt and Da Zhou @Microsoft for their work
—and welcome to the Hadoop Committer Team!*



Hadoop Common



On going Effort

- RPC support for TLS
- KMSv2
- OpenTracing integration
- JDK11 support



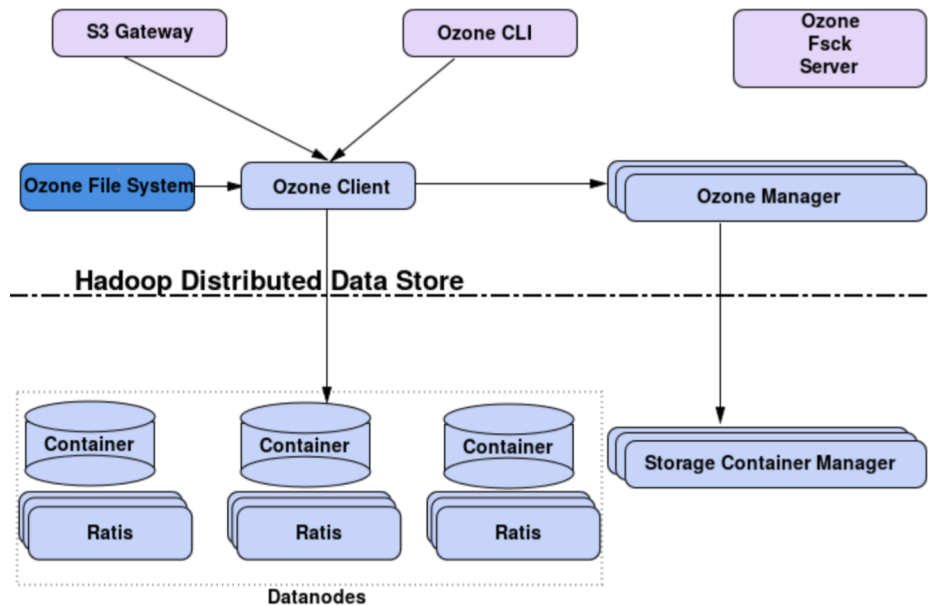
Ozone



Ozone



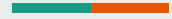
- ❑ Object Store made for Big Data workloads.
 - ❑ A long term successor of HDFS.
 - ❑ In-place upgrade from HDFS (roadmap)
- ❑ Contribution from Hortonworks/Cloudera/Tencent
- ...
- ❑ Tremendous progress over past year



Ozone Upcoming releases



- ❑ Three Alpha Releases so far.
 - ❑ 0.2: basic object store.
 - ❑ 0.3: s3 protocol.
 - ❑ 0.4: Security and Ranger support.
- ❑ 0.4.1 release (Native ACLs) coming out soon (December-ish).
- ❑ 0.5.0 will be the beta release.
 - ❑ Reliability and performance improvement.
 - ❑ HA



Releases



Release Plan - Core Hadoop

2018	2.6.5	2.7.5 - 2.7.7	2.8.3 - 2.8.5
	Stabilization, Maintenance, Bug fix	Stabilization, Maintenance, Bug fix	Stabilization, Maintenance, Bug fix
	2.9.0 - 2.9.2	3.0.0 - 3.0.3	3.1.0 - 3.1.1
	YARN Federation, Opportunistic Containers	EC, Global scheduling, multiple resource types, Timeline Service V2, RBF for HDFS	GPU/FPGA, Long Running Services, Placement Constraints, Docker on YARN GA
2019	3.1.2	3.2.0	2.10 (Planned)
	Stabilization, Maintenance, Bug fix	Node Attributes, Submarine, Storage Policy Satisfier, ABFS connector	<ul style="list-style-type: none"> <input type="checkbox"/> YARN resource types/GPU support (YARN-8200) <input type="checkbox"/> Selective wire encryption (HDFS-13541) <input type="checkbox"/> HDFS Rolling upgrades from 2.x to 3.x(HDFS-14509)
	3.1.3 (RC0, Target Sep 2019)	3.2.1 (Sep 2019, released).	3.3.0 (Planned)
	Stabilization, Maintenance, Bug fix	Stabilization, Maintenance, Bug fix (GA of 3.2)	<ul style="list-style-type: none"> <input type="checkbox"/> OCI/SquashFS <input type="checkbox"/> NEC Vector Engine <input type="checkbox"/> Consistent reads from Standby <input type="checkbox"/> NVMe for HDFS cache


Release Plan - Submarine

- Voted to become a separate Apache project
 - No longer part of Core Hadoop releases

2018	0.1.0	
	<ul style="list-style-type: none"> <input type="checkbox"/> YARN <ul style="list-style-type: none"> <input type="checkbox"/> Distributed Tensorflow <input type="checkbox"/> MXNet 	
2019	0.2.0	0.3.0 (Planned)
	<ul style="list-style-type: none"> <input type="checkbox"/> Support for other runtimes <ul style="list-style-type: none"> <input type="checkbox"/> Pytorch <input type="checkbox"/> LinkedIn's TonY <input type="checkbox"/> Zeppelin Notebook support 	<ul style="list-style-type: none"> <input type="checkbox"/> Support K8s runtimes <input type="checkbox"/> Mini-submarine <input type="checkbox"/> Submarine-workbench <input type="checkbox"/> Submarine SDK

End of Life Policy

- ❑ EOL of Releases with no maintenance release in long term (1.5+ yrs)
- ❑ Security-only releases on EOL versions if requested.
- ❑ EOLed Versions
 - ❑ Hadoop 2.7.x (and lower)
 - ❑ Hadoop 3.0.x



Upgrades (Hadoop 2 -> Hadoop 3)

Express/Rolling Upgrades

Express Upgrades

- Stop the world Upgrades
- Cluster downtime
- Less stringent prerequisites
- Process
 - Upgrade masters and workers in one shot

Rolling Upgrades

- Preserve cluster operation
- Minimizes Service impact and downtime
- Can take longer to complete
- Process
 - Upgrades masters and workers in batches

Recommendation - Express or Rolling?

Major version upgrade

- Challenges and issues in supporting Rolling Upgrades

Technical challenges with rolling upgrade

- Lot of work done/WIP by Hadoop community to support upgrades without Downtime. Should be part of releases soon.
- Backward incompatible changes blocks rolling upgrade.

Recommended

- Express Upgrade** from Hadoop 2 to 3

Compatibility

Wire compatibility

- Preserves compatibility with Hadoop 2 clients
- Distcp/WebHDFS compatibility preserved

API compatibility

Not fully, but minimal impact.

- Dependency version bumps
- Removal of deprecated APIs and tools
- Shell script rewrite, rework of Hadoop tools/scripts.

Source & Target Versions

Upgrades Validated with

Hadoop 2 Base version	Hadoop 3 Base version
Apache Hadoop 2.8.x	Apache Hadoop 3.1.x

Why 2.8.x release?

- Most of production deployments are close to 2.8.x

What should users of 2.6.x and 2.7.x do?

- Do more validations before upgrading, we do see some users directly upgrade from 2.7.x to 3.x.

Upgrade Process/Details

Refer to our earlier talk for further details

[Migrating Hadoop cluster and workloads from Hadoop 2 to Hadoop 3](#)

Many successful use cases for Hadoop 3.x (New And Upgrade) in Production

Summary of upgrade

❑ Hadoop 3

Eagerly awaited release with lots of new features and optimizations !

- ❑ Lots of large clusters already on Hadoop 3 at enterprises
- ❑ **Express Upgrades** are recommended
- ❑ If you haven't upgraded yet, **NOW** is the best time!



Questions?



Rate today's session

Cyberconflict: A new era of war, sabotage, and fear

See passes & pricing

David Sanger (The New York Times)
9:55am-10:10am Wednesday, March 27, 2019
Location: Ballroom
Secondary topics: Security and Privacy

 Add to Your Schedule
 Add Comment or Question

Rate This Session

We're living in a new era of constant sabotage, misinformation, and fear, in which everyone is a target, and you're often the collateral damage in a growing conflict among states. From crippling infrastructure to sowing discord and doubt, cyber is now the weapon of choice for democracies, dictators, and terrorists.

David Sanger explains how the rise of cyberweapons has transformed geopolitics like nothing since the invention of the atomic bomb. Moving from the White House Situation Room to the dens of Chinese, Russian, North Korean, and Iranian hackers to the boardrooms of Silicon Valley, David reveals a world coming face-to-face with the perils of technological revolution—a conflict that the United States helped start when it began using cyberweapons against Iranian nuclear plants and North Korean missile launches. But now we find ourselves in a conflict we're uncertain how to control, as our adversaries exploit vulnerabilities in our hyperconnected nation and we struggle to figure out how to deter these complex, short-of-war attacks.

David Sanger
The New York Times



David E. Sanger is the national security correspondent for the *New York Times* as well as a national security and political contributor for CNN and a frequent guest on *CBS This Morning*, *Face the Nation*, and many PBS shows.

Session page on conference website

✓ Attending

Notes

Remove

Cyberconflict: A new era of war, sabotage, and fear

9:55 AM - 10:10 AM, Wed, Mar 27, 2019

Speakers



David Sanger
National Security Correspondent
The New York Times

Ballroom

Keynotes

David Sanger explains how the rise of cyberweapons has transformed geopolitics like nothing since the invention of the atomic bomb. From crippling infrastructure to sowing discord and doubt, cyber is now the weapon of choice for democracies, dictators, and terrorists.

SESSION EVALUATION

O'Reilly Events App