



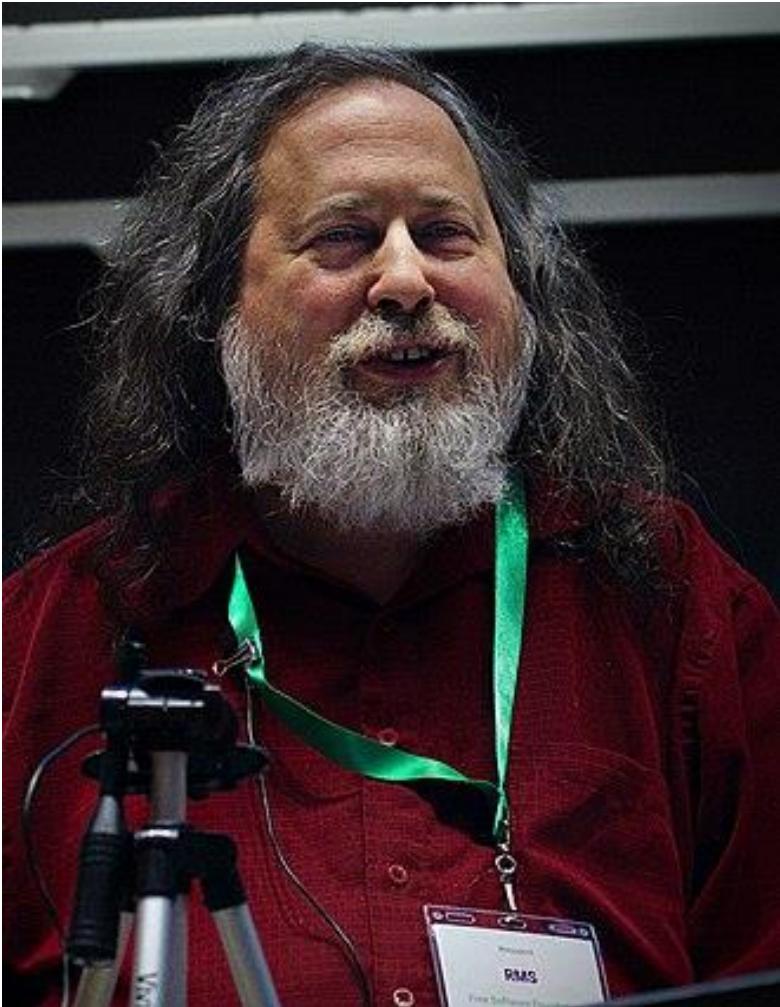
# Open Source Demystified!

*Exploring the real potential...*



**open**  
Source

**Chit-Chat**



## Richard Matthew Stallman

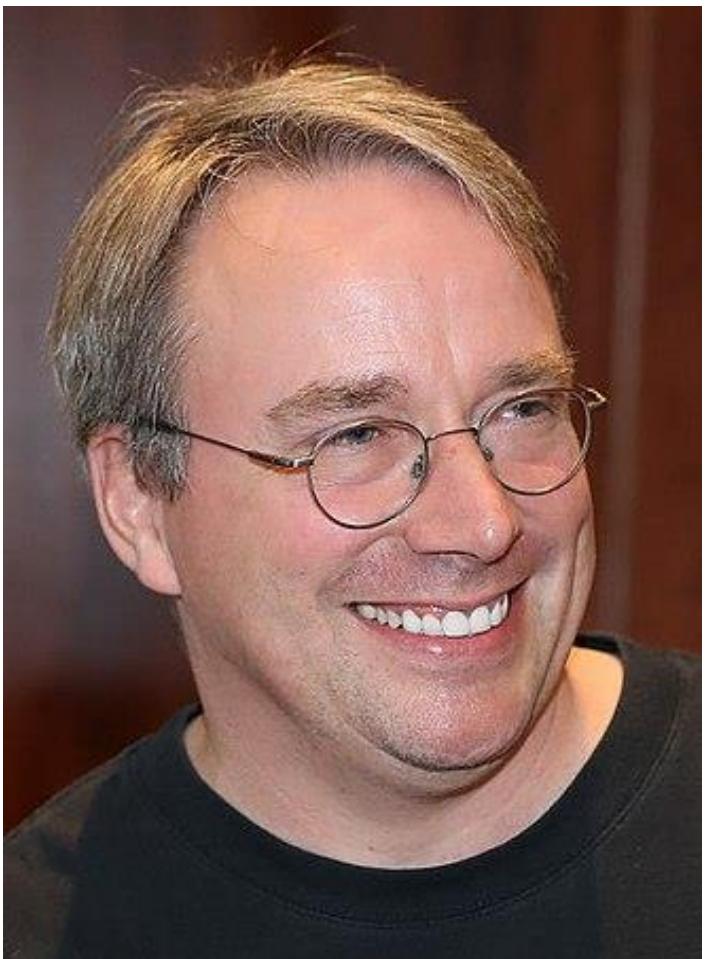
Started the GNU Project

Founded of Free Software Foundation (FSF) in October 1985

Developed GNU Compiler Collection and GNU Emacs

Wrote the GNU General Public License. (GPL)

[https://en.wikipedia.org/wiki/Richard\\_Stallman](https://en.wikipedia.org/wiki/Richard_Stallman)



## Linus Benedict Torvalds

Creator of Linux Kernel

Created the distributed version control system Git

From: torvalds@klaava.Helsinki.FI (Linus Benedict Torvalds)  
Newsgroups: comp.os.minix  
Subject: What would you like to see most in minix?  
Summary: small poll for my new operating system  
Message-ID: <1991Aug25.205708.9541@klaava.Helsinki.FI>  
Date: 25 Aug 91 20:57:08 GMT  
Organization: University of Helsinki

Hello everybody out there using minix -

I'm doing a (free) operating system (just a hobby, won't be big and professional like gnu) for 386(486) AT clones. This has been brewing since april, and is starting to get ready. I'd like any feedback on things people like/dislike in minix, as my OS resembles it somewhat (same physical layout of the file-system (due to practical reasons) among other things).

I've currently ported bash(1.08) and gcc(1.40), and things seem to work. This implies that I'll get something practical within a few months, and I'd like to know what features most people would want. Any suggestions are welcome, but I won't promise I'll implement them :-)

Linus (torvalds@kruuna.helsinki.fi)

PS. Yes - it's free of any minix code, and it has a multi-threaded fs. It is NOT portable (uses 386 task switching etc), and it probably never will support anything other than AT-harddisks, as that's all I have :-(.

[https://en.wikipedia.org/wiki/Linus\\_Torvalds](https://en.wikipedia.org/wiki/Linus_Torvalds)



open  
Source

What?



# MOVIE TIME!



Video Link : <https://www.youtube.com/watch?v=SpeDK1TPbew>

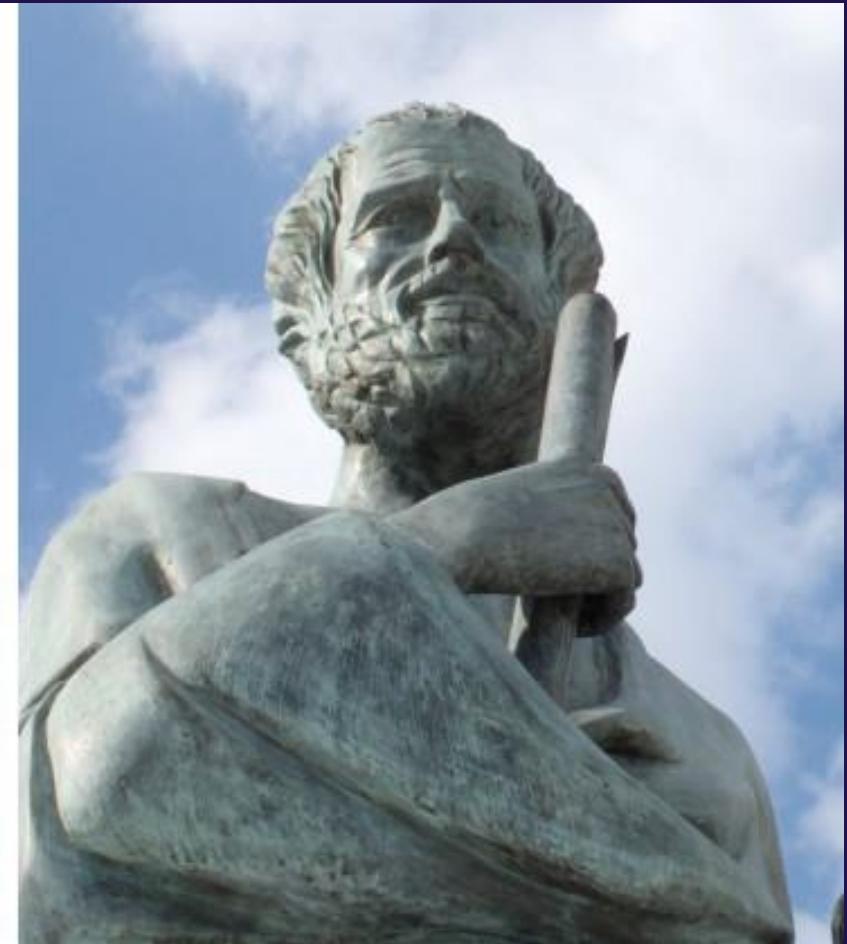
“F Prime has enabled a lot of goals we’ve had at JPL to design a truly reusable multi-mission flight architecture with the added bonus of the open-source collaboration and visibility afforded by the Mars Helicopter project,” Canham said. “It’s kind of an open-source victory, because we’re flying an open-source operating system and an open-source flight software framework, and flying commercial parts that you can buy off the shelf, if you wanted to do this yourself someday.” (The helicopter carries a combination of custom-made and off-the-shelf components – many from the world of cell phone technology – including its two cameras.)

# COMMUNITY

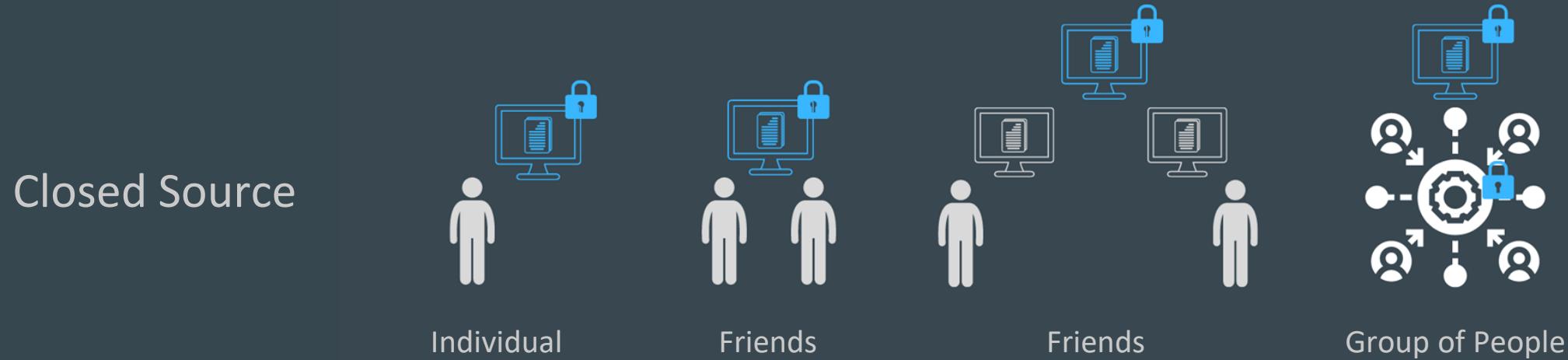


# Open Source Community ~~ Society

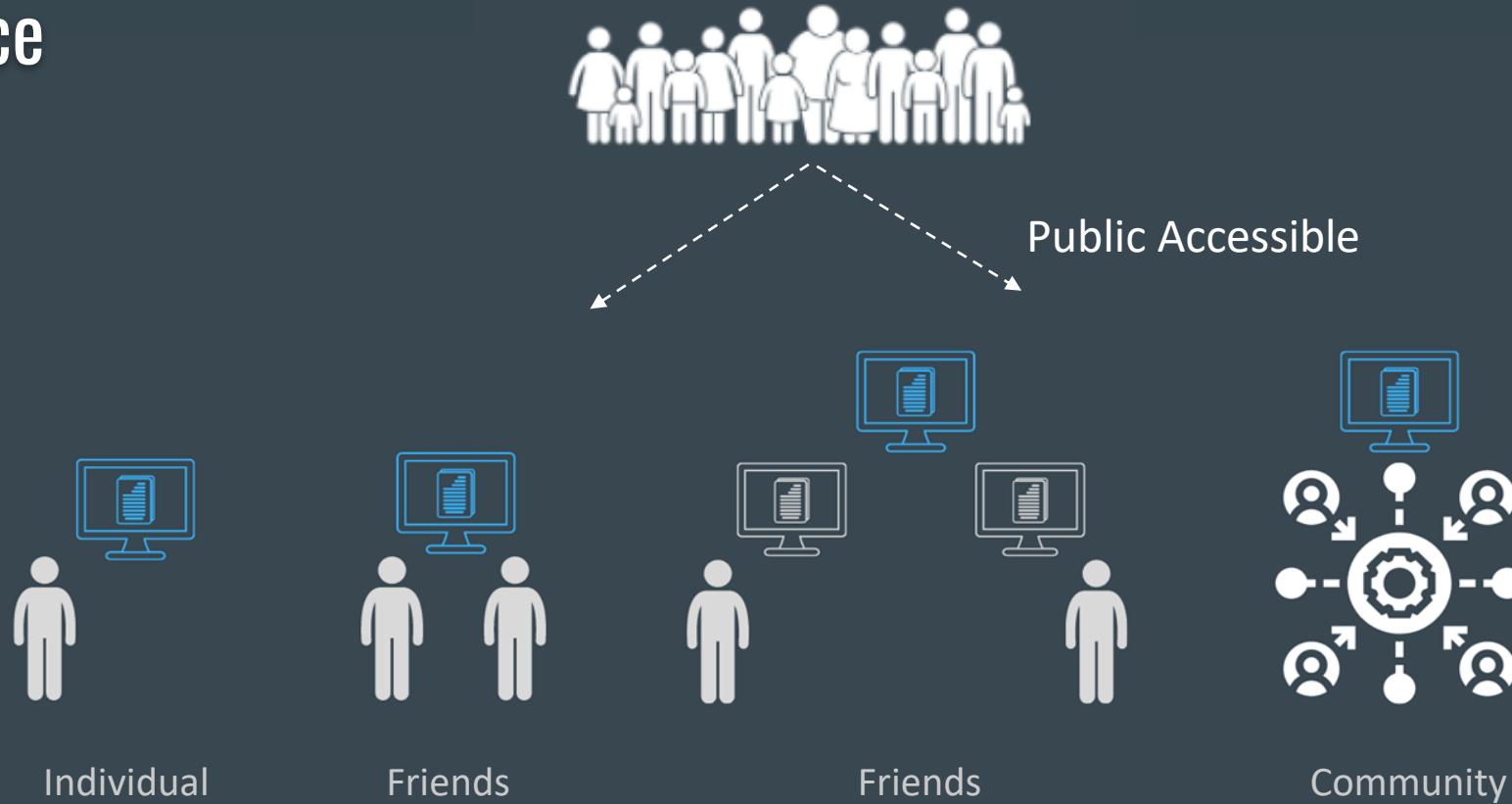
- Integrity
- Openness
- Sharing
- Collaboration
- Communication



# How do we do software development ?



# Open Source



Source Code is Open

Open Source



Need **NOT** be Free!

Free Software



Run, Edit, Contribute & Share

Need **NOT** be Free!

Freeware



Free to use (conditions apply!)

Mostly **NOT** Open Source!

# Open Source : Essentials



Common Place for Source Code

Examples : git, github, gitee, gitlab, bitbucket, ...



Developers

Using different programming languages  
(go, c, c++, java, scripts...)

# Open Source : Typical Components

Governance



- Operation Structure
- Strategy & Budget
- Roadmap

Engineering



- Design and Development
- Project Management
- Source Version Control

Ecosystem



- Developers
- Vendors
- Users

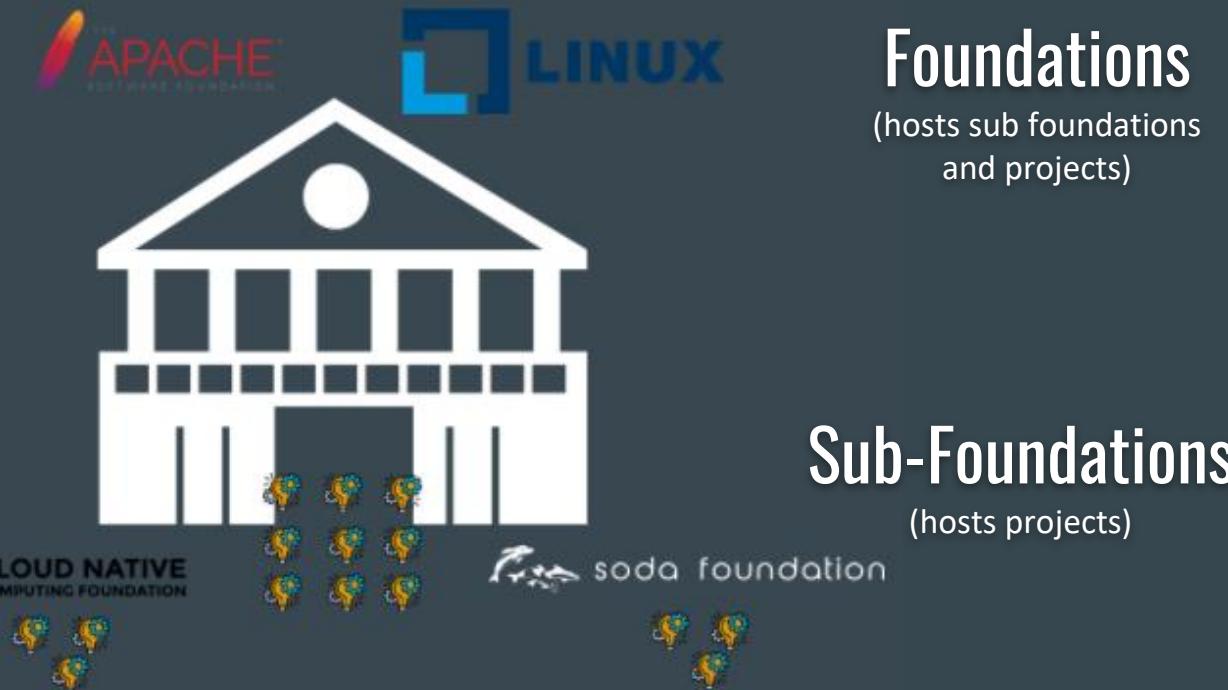
**..for a project to run smoothly and to be sustainable!**

# Project, Foundation and Sub Foundation

Projects



kubernetes



Foundations  
(hosts sub foundations  
and projects)

Sub-Foundations  
(hosts projects)

Why?



open  
Source

# There is an open source project for **that!**

**94M**

developers are on  
GitHub

**90%+**

of Fortune 100  
companies use  
GitHub

**90%**

of companies use  
open source\*

**413M**

open source  
contributions in  
2022

**80%** of organizations have increased the use of open source software over the last year.

**95%** of Organizations think that enterprise open source software is critical for their enterprise infrastructure.

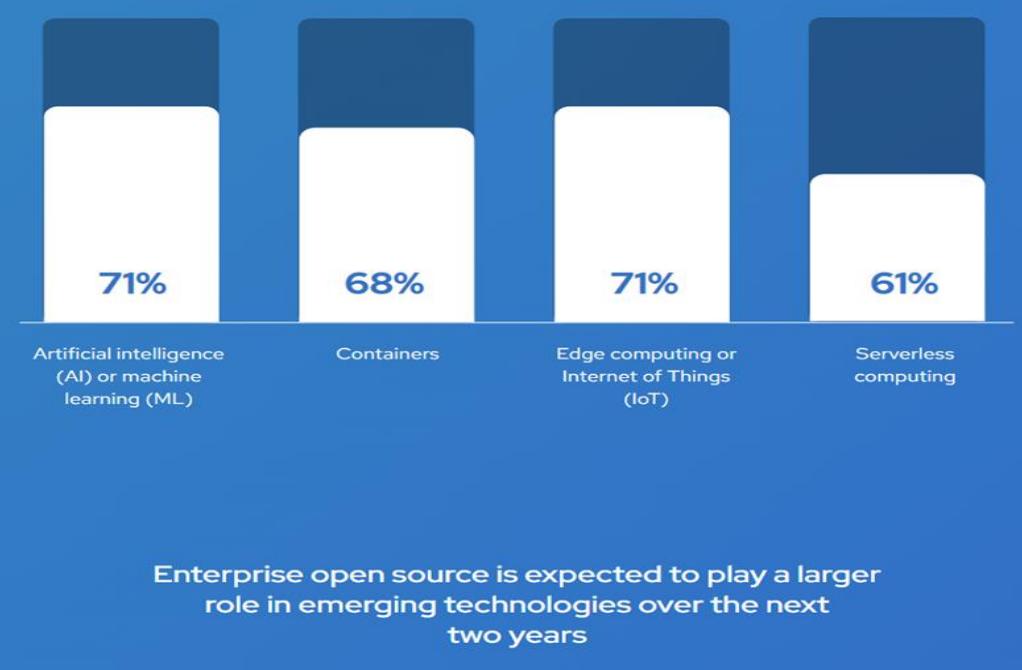
**82%** of IT leaders are more likely to select a vendor who contributes to the open source community!



# Open Source Collaboration makes your business more competent!

## The #1 reason

organizations use open source software is to have access to **innovations and the latest technologies.**



### Top reasons why enterprise open source vendors are preferred

- 1 They are familiar with open source processes 49%
- 2 They help sustain healthy open source communities 49%
- 3 They can influence the development of features that we need 48%
- 4 They are going to be more effective if I face technical challenges 46%

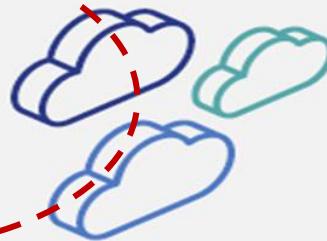
Enterprise open source solutions can help **boost 5G adoption** across industries like manufacturing, automotive, and transportation, not just telecommunications.

## Top 12 trends in data and storage

In 2022, the growth in data was 3 times higher than in 2021.



56% of end users deploy open source multi-cloud management in their production environments.



43% of end users demand the freedom to leverage multiple storage vendors.



Data security is the greatest challenge facing container deployments.



Public clouds run more than 40% of end-user organization workloads.



Primary data storage, complete data protection, and disaster recovery represent the top 3 use cases for cloud storage services.



Information security and data privacy are the leading reasons to use a private cloud solution.



The biggest challenge facing multi-cloud solutions is the security and protection of data.



Cloud technologies represent the most significant area of data and storage technology investment over the next three years.



AI-driven hybrid data management is considered the most critical area for data management and analytics over the next 2-4 years.

Data quality, governance, and security are top priorities when selecting metadata management solutions.

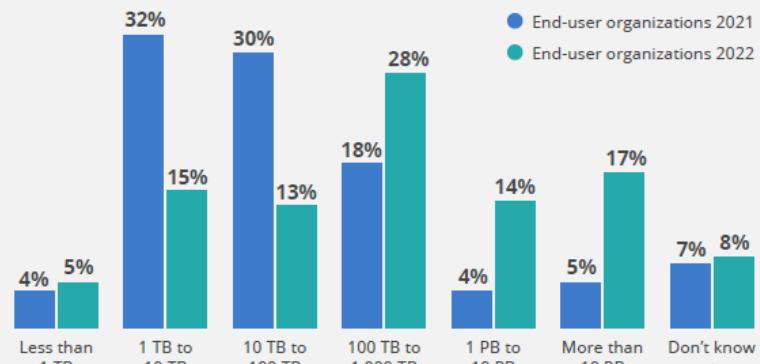


Cloud storage monitoring is the greatest challenge facing data and storage observability.



## AMOUNT OF DATA GROWTH BY CATEGORY IN 2021 AND 2022

How much is the approximate data growth per year for your organization? (comparison of 2021 and 2022 results)

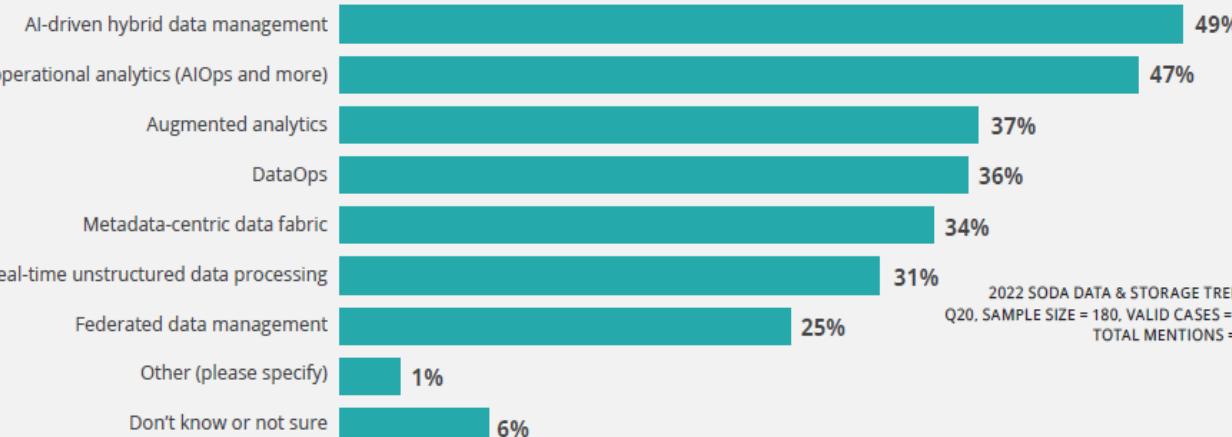


*Data is growing!*

*Cloud Technology is the top investment area in next 3 years*

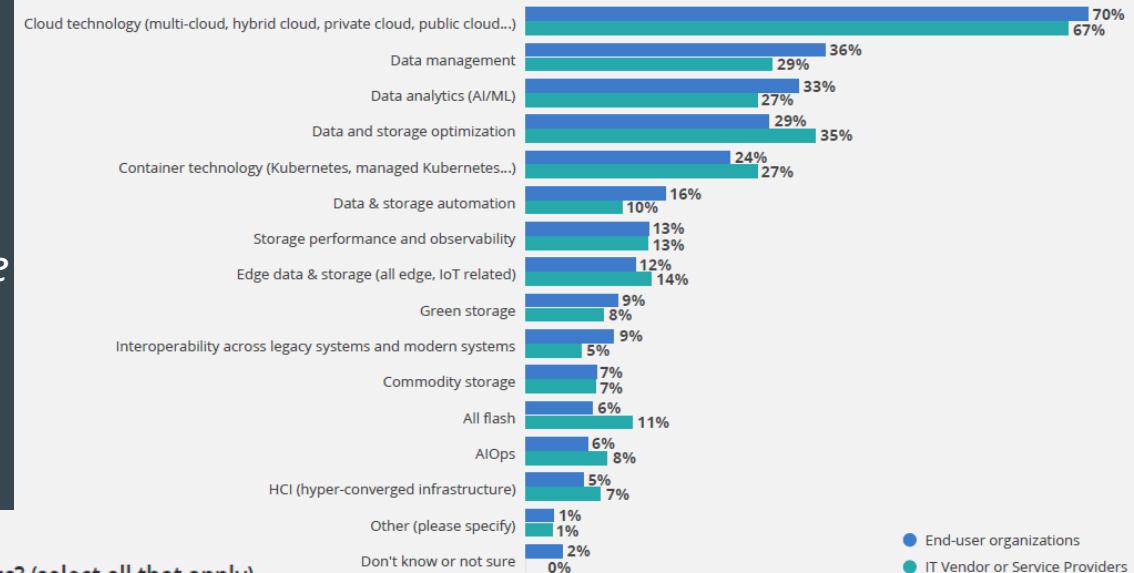
## KEY DATA MANAGEMENT AND ANALYTIC AREAS OVER THE NEXT TWO TO FOUR YEARS

Which are the key capabilities you think are critical in the next 2-4 years for data management and data analytics? (select all that apply)



## LEADING DATA STORAGE INVESTMENT AREAS OVER THE NEXT THREE YEARS

What are your organization's top 3 data and storage technology investment or deployment areas for the next 3 years? (segmented by end user and vendor/IT service provider)

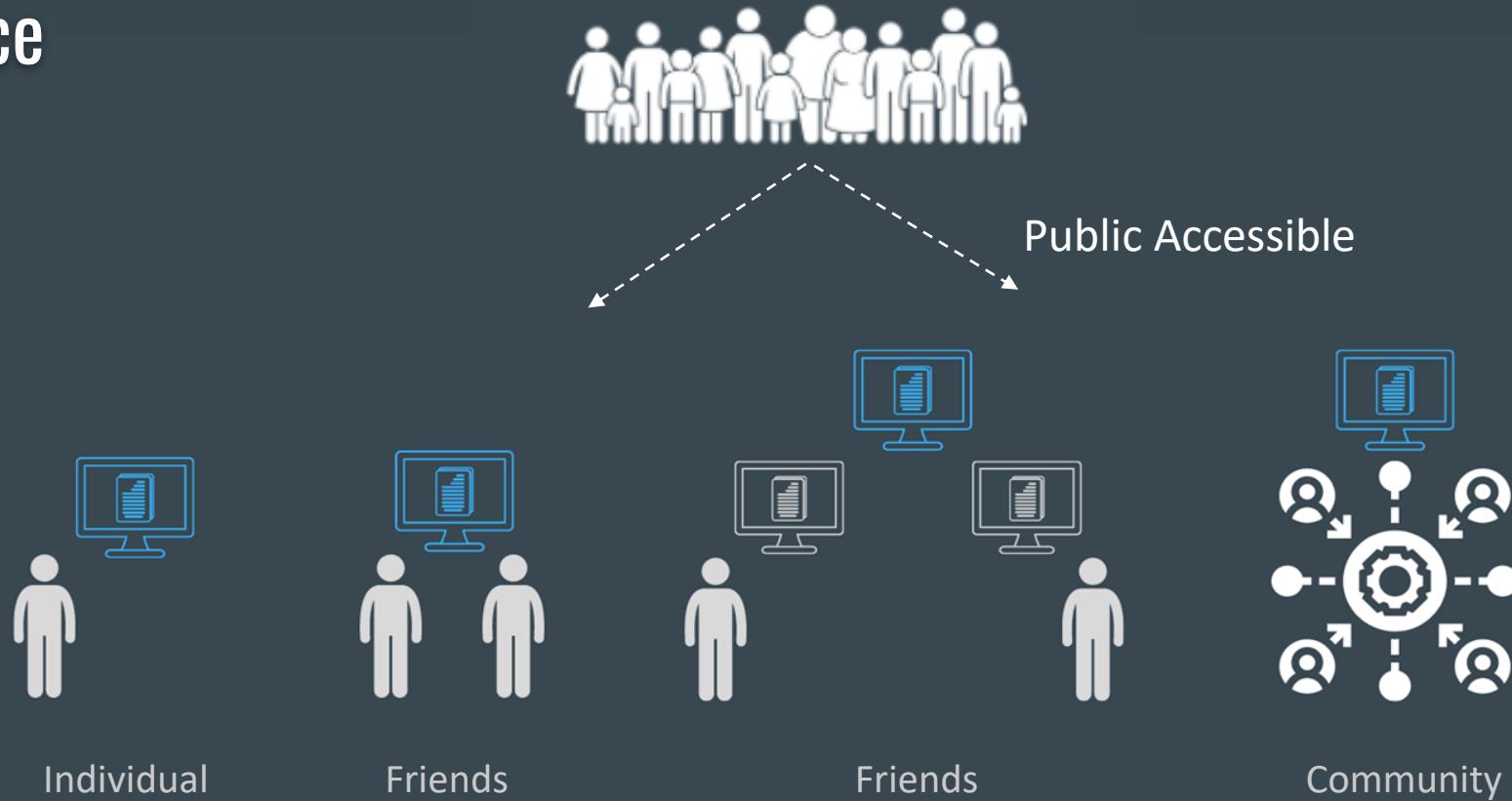


● End-user organizations  
● IT Vendor or Service Providers

*AI Driven Hybrid Data Management and Metadata management are the top trends for the next 2-4 years*

*Open Source adoption in Cloud and Kubernetes is increasing!*

# Open Source



**Experience Power of Collaboration and Open Culture!**

# Collaborative & Open Culture



Collaborative Goals Setting  
Collaborative Development  
Collaborative Research  
Collaborative Learning

# Practices to ensure effective collaboration!



## Open Communication

- What over Who (Community Calls, Slack, github discussions ...)

## Value Development

- Why is it needed? Can't reuse? No to 'fancy pieces'! (github issue/PR handling)

## Efficient Tools

- What works for us/this? Add what we use! Can we reduce 'work' (mindmap, github, gitops)

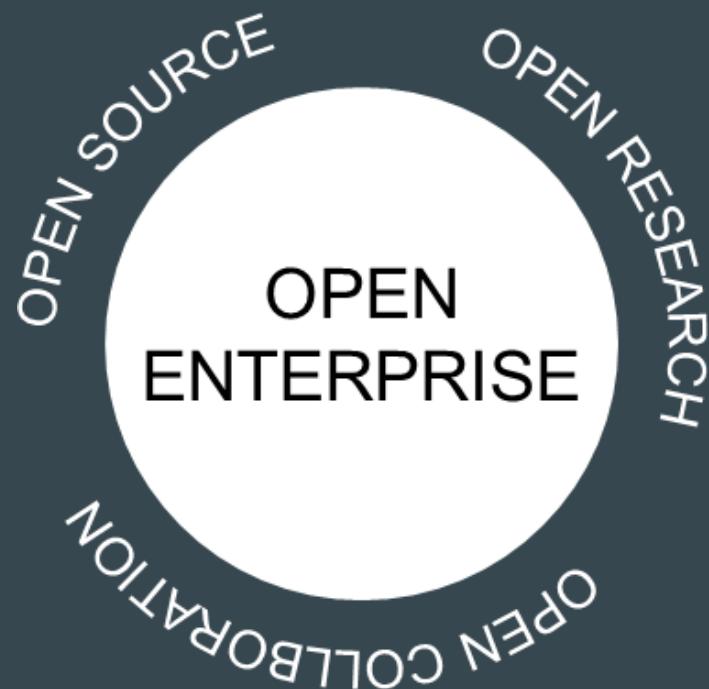
## Right Process

- fit for changing community? Transparent? Inclusive? Needed? (github projects, teams, PR review...)

## Self Learning

- Projects and People Self learn (project landscapes, meetups, events, outward view, technology..)

# Open X !



## Open Source

accelerate innovation and development with open source projects and ecosystems

## Open Research

expand insights and explore ideas with open research and case studies

## Open Collaboration

engage the community and build partnerships on a neutral platform

Why should I ?

Coding Skills

'Real' Projects

Technology

Communication

Confidence

Maturity

Visibility Unlimited!

Global Networking

Employability

Recognition

Leadership

Research & Innovation



**Transparent  
Democratic  
Value for All  
Sustainable**



open  
Source

How?

# Open Source Ecosystem

## Governance



Operation Structure  
Strategy & Budget  
Roadmap

*Hosts and Manages the open source projects*



### Foundations

Host Multiple Projects and Sub Foundations | Build Governance & Operations



### Sub Foundations

Host Multiple Projects | Build Governance & Operations



### Independent Project Teams

Host and Manages individual projects

## Engineering



Design and Development  
Project Management  
Source Version Control

*Technology and Open Source Software Development*

### Developers

Design and Development of Open source software

### Contributors

Contribute to non technical areas of the project

### Experts

Architecture and Technology Strategy Development

## Ecosystem



Collaborating Community  
Brand, Awareness, Deployments  
Requirements and Feedbacks

*Collaboration, Productization and Deployment*

### Users

Use the products and solutions | Gives requirements and feedbacks | Support projects /foundations

### Vendors

Build products and solutions | support the projects / foundations

### Supporters

Partnering companies and individuals supporting the project community ecosystem |

# How can I?

- Can Code?
- Can Test?
- Can Document?
- Can Software Design?
- Can do creative designs?
- Can evangelize?
- Can X..!

Want to  
grow

Ready to  
spare some  
time

Ready to  
spare some  
time  
consistently

# Quick Steps

- **Join an open-source project today!**
  - Do not “PLAN”! ☺
  - Do not “THINK TOO MUCH”! ☺
- **Engage and Contribute!**
  - Identify
  - Join the community
  - Listen
  - Start small
  - Contribute
  - **Sustain → MOST IMPORTANT!**
- **Be part of a community**
  - Local Meetup Groups
  - Project Slack
  - Events

# What we need?

**Programming Skills**  
Email ID, Github Id, Slack Id ☺  
**Technology Competency**

Time  
Open mind to learn & grow!



open  
Source

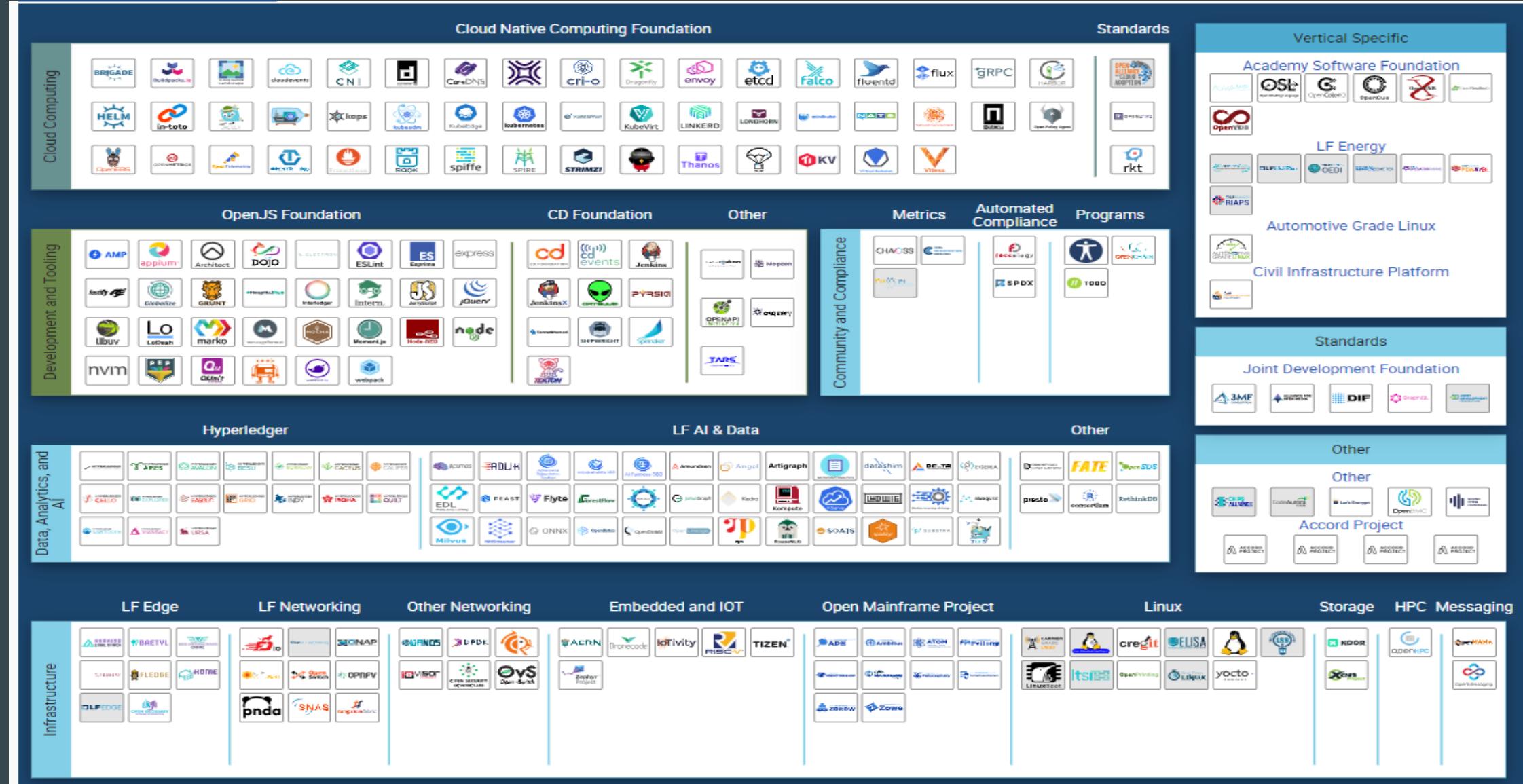
Growing Research, Growing Technology Landscapes...

More and More **domain-specific technology landscapes** of open source projects are being created.

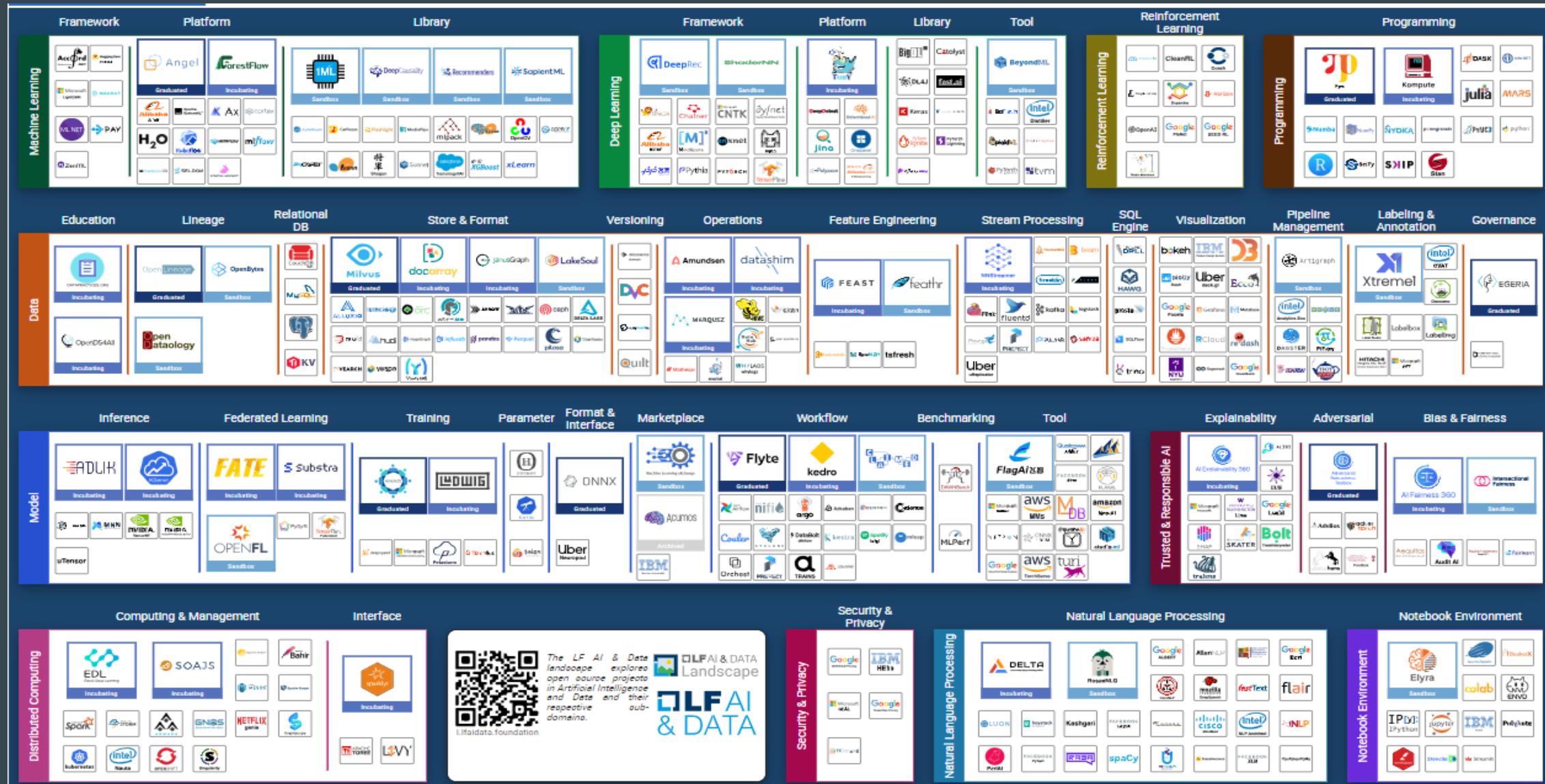
This **shows the scale** of open source project **research and development** in various technologies

More and More Open Source Projects in **Production**

# LF Overall Landscape



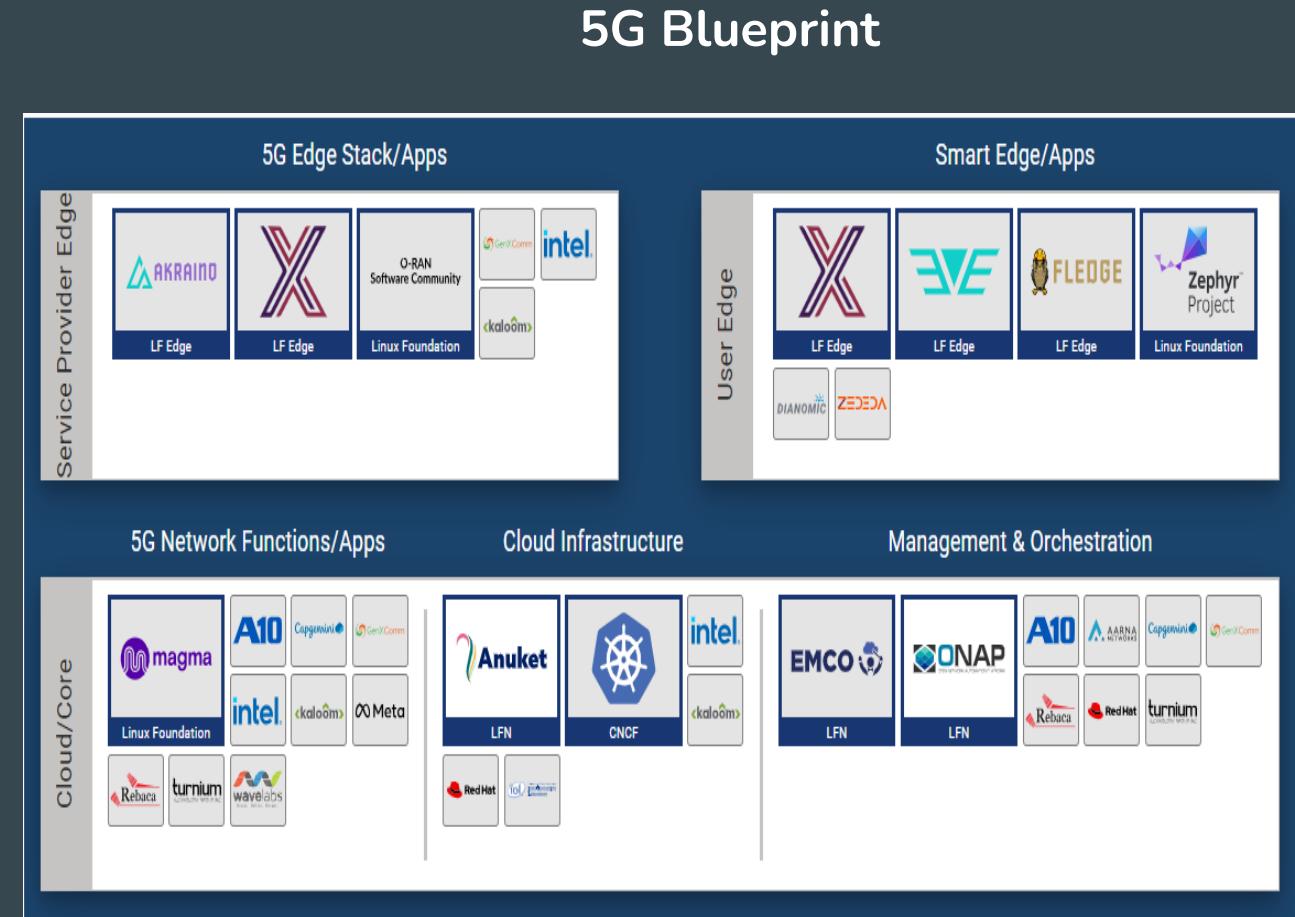
# LF AI & Data Project Landscape



*Gen AI is driving the AI & Data Projects growth*

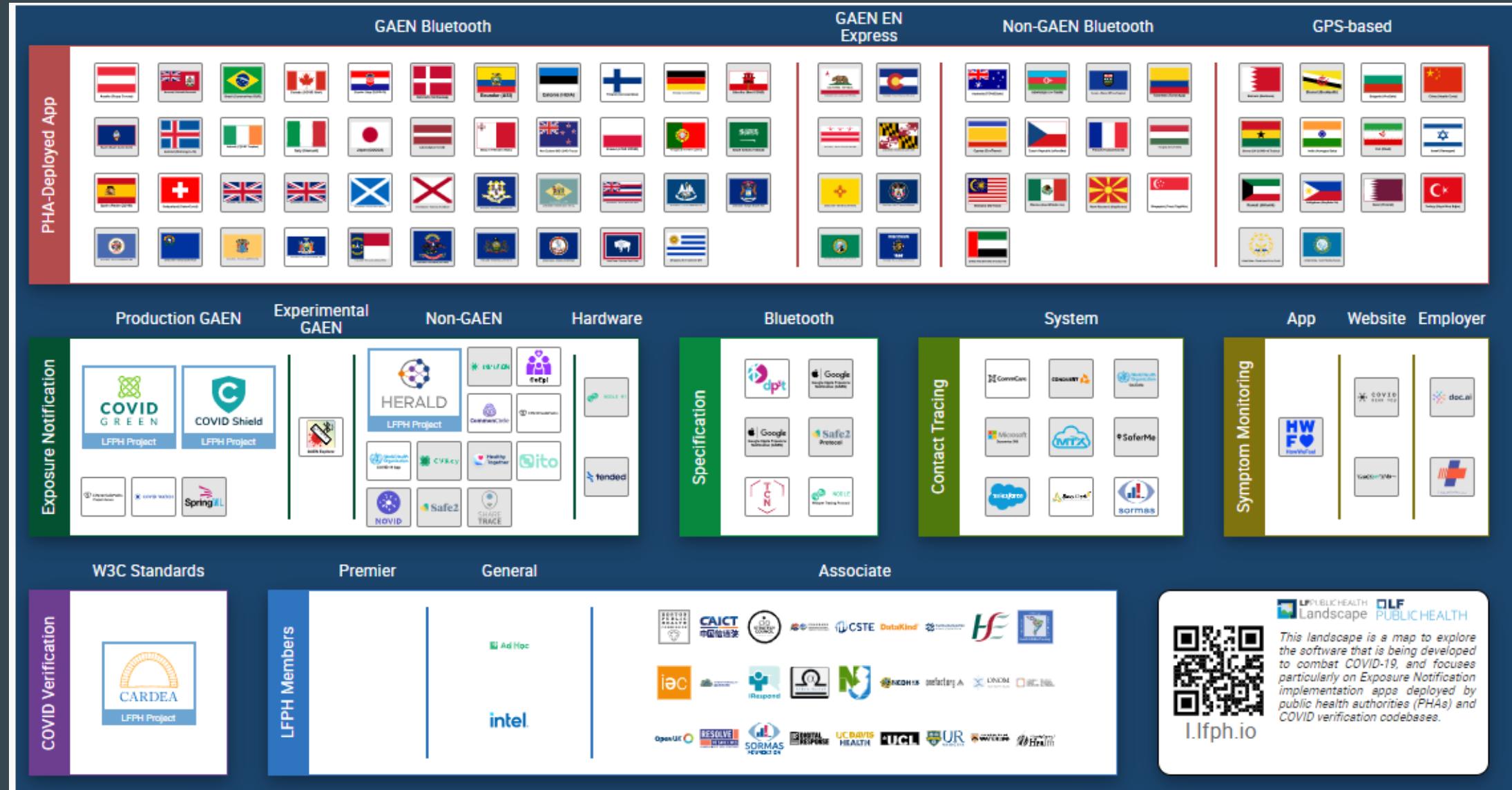
# LF Networking and 5G

## Networking



*Open Source Networking and 5G open source projects are getting higher traction!*

# LF Public Health

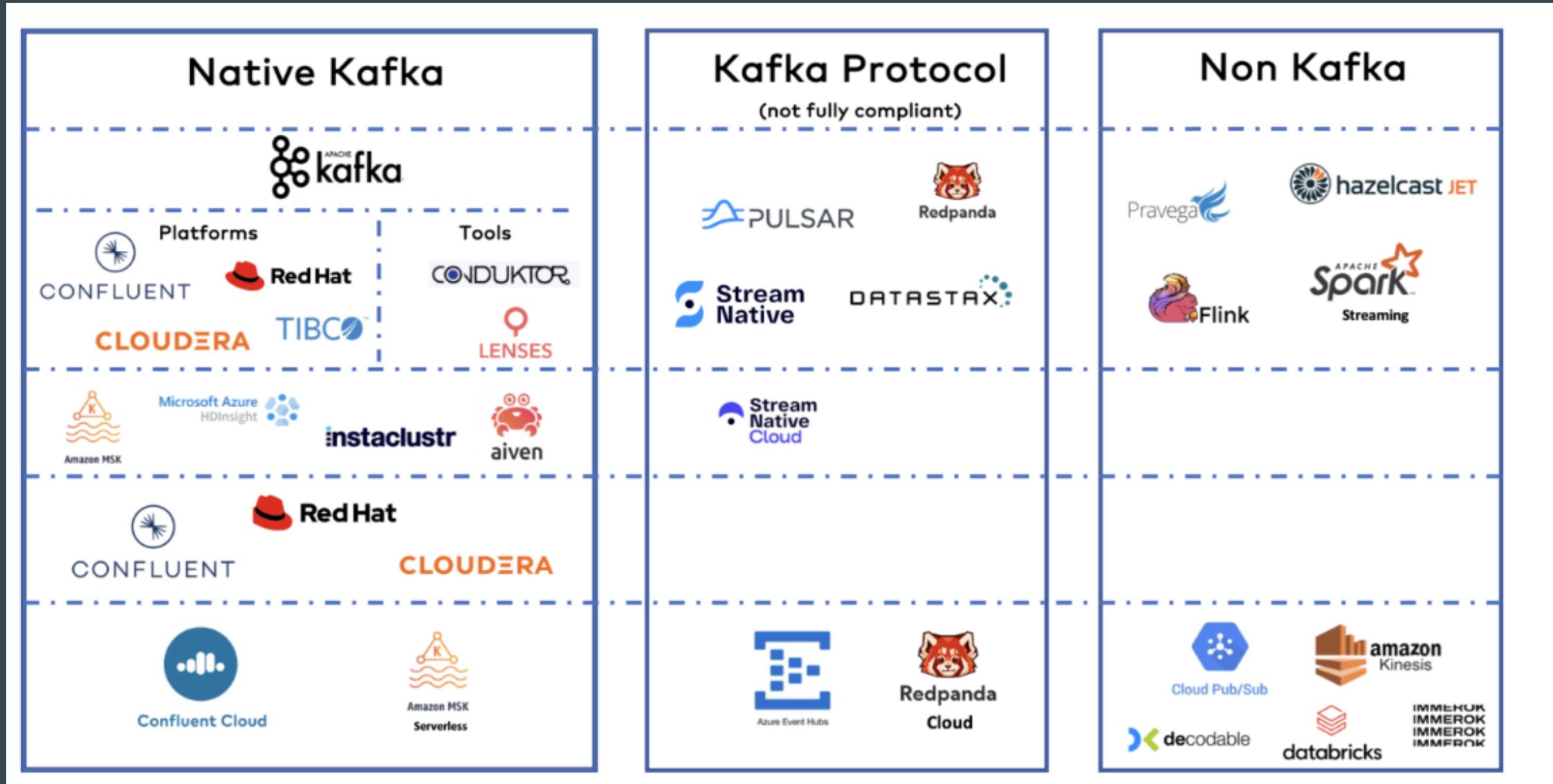


*LF and Industry collaboration on public health is becoming stronger...*

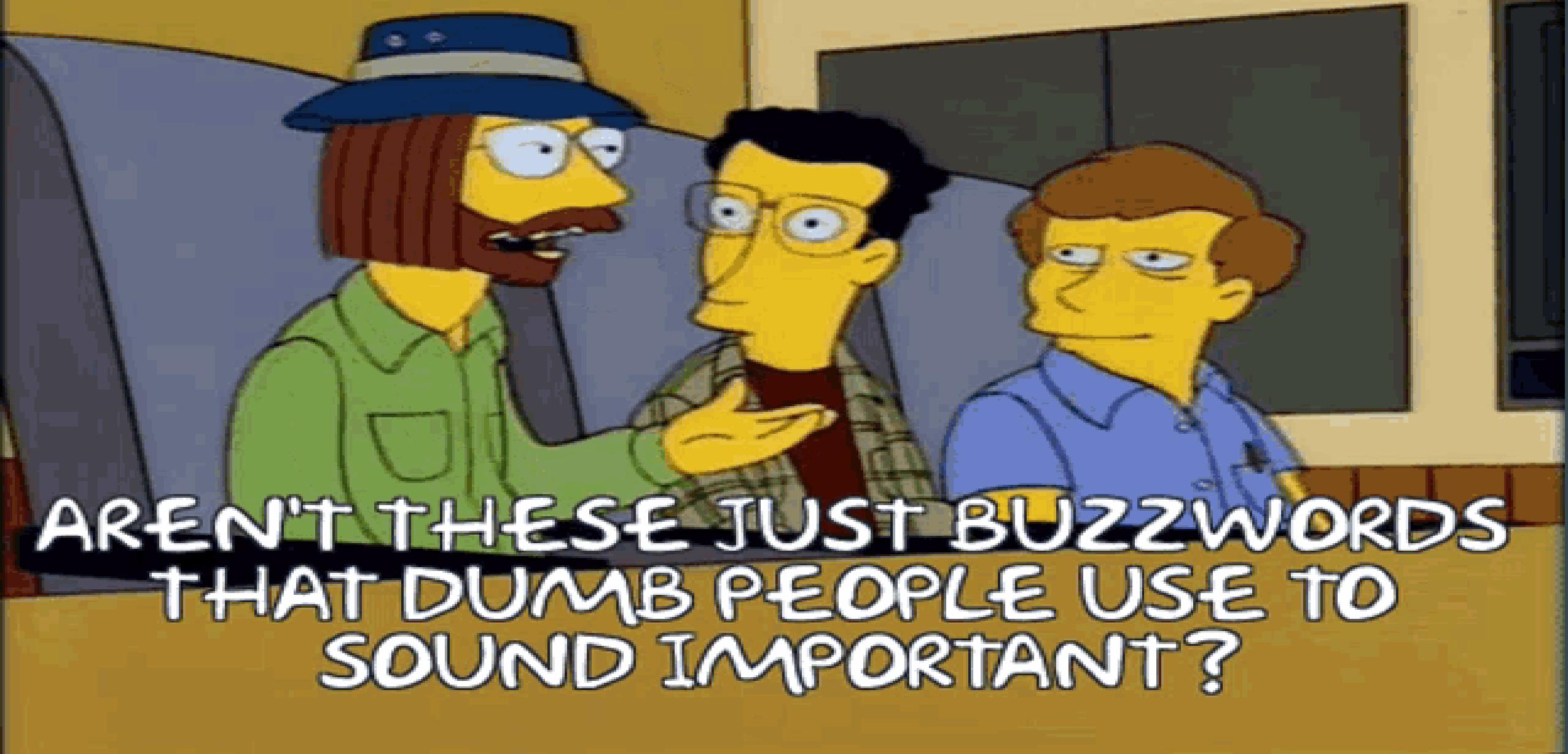
# Apache Foundation Projects across various technology domains

The screenshot shows the Apache Projects Directory page. At the top left is the Apache Software Foundation logo. To its right is the title "Projects Directory". Below the title is a navigation bar with links: Home, Committees, Projects (which is highlighted in blue), Releases, Statistics, Timelines, and a search bar labeled "Search...". On the left side, there is a sidebar titled "Project listings:" containing five items with icons: "By Name" (person icon), "By Committee" (two people icon), "By Category" (folders icon), "By Programming Language" (speech bubble icon), and "By Number of Committers" (numbers icon). Below the sidebar, a note states: "Please note that the information displayed here relies on the [DOAP files](#) which PMCs are encouraged to [provide](#). However DOAPs are not mandatory, and not all PMCs have provided a DOAP for all the projects they manage." A large section below is titled "Projects by category:" and includes a "TOC" link. At the bottom, there is a long list of project categories: big-data, build-management, c, c++, cloud, content, data-management-platform, data-visualization, database, distributed-sql-database, education, ftp, geospatial, go, graphics, groovy, hadoop, html, http, httpd-module, ide, identity-management, identity-provisioning, integration, iot, java, javaee, kerberos, library, mail, mobile, network-client, network-server, observability, osgi, php, python, regexp, retired, sdk, search, security, sql, templating, testing, virtual-machine, web-framework, xml.

# Apache Data Streaming Landscape 2023

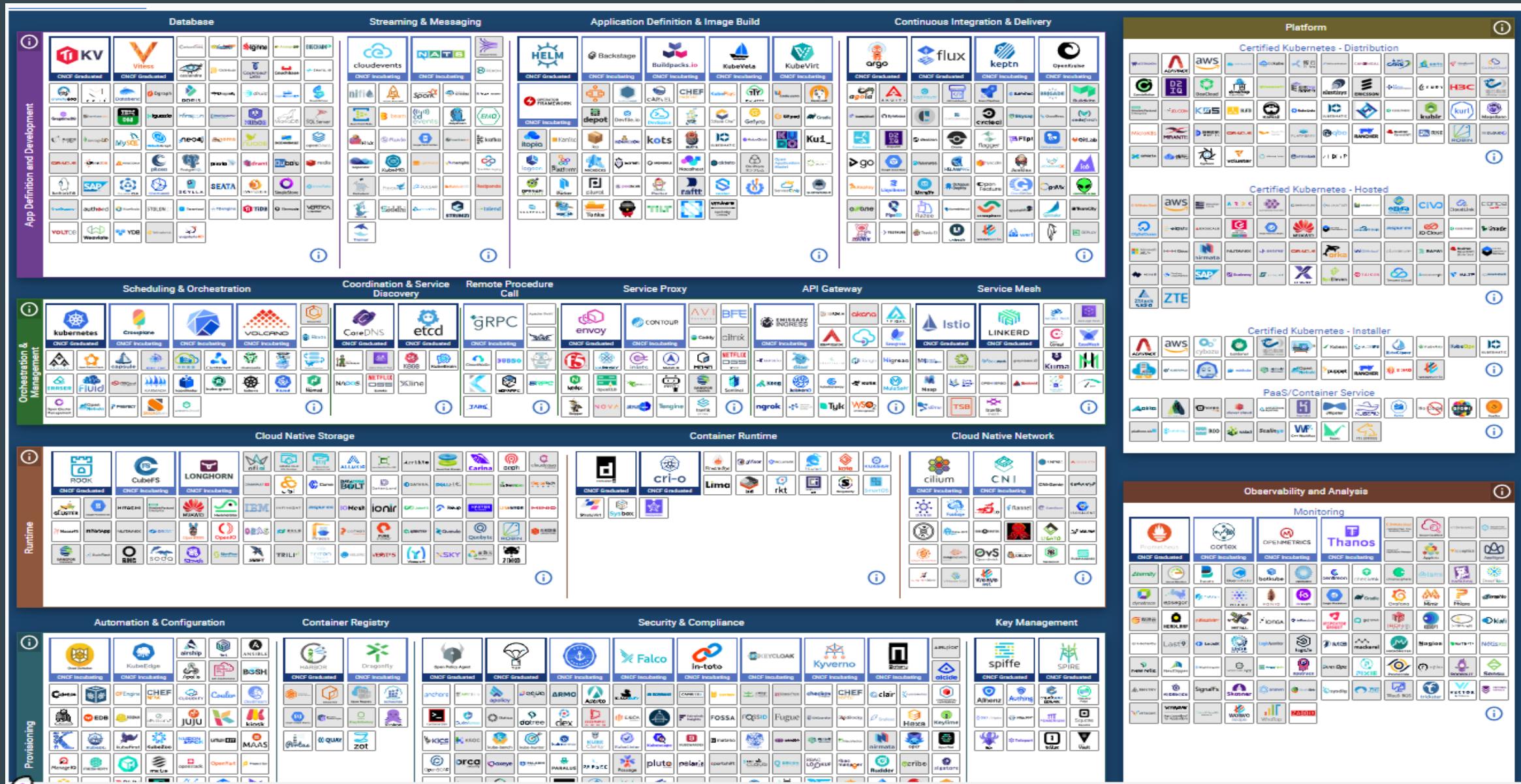


# Cloud, Cloud Native, Kubernetes...



AREN'T THESE JUST BUZZWORDS  
THAT DUMB PEOPLE USE TO  
SOUND IMPORTANT?

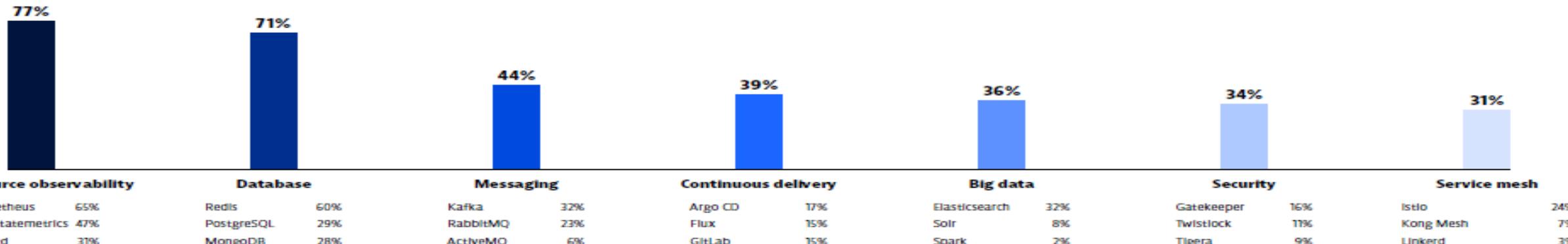
# LF Cloud Native Project Landscape



*Open Source project growth in Cloud Native and Kubernetes is exponential*

# Open source software drives a vibrant Kubernetes ecosystem

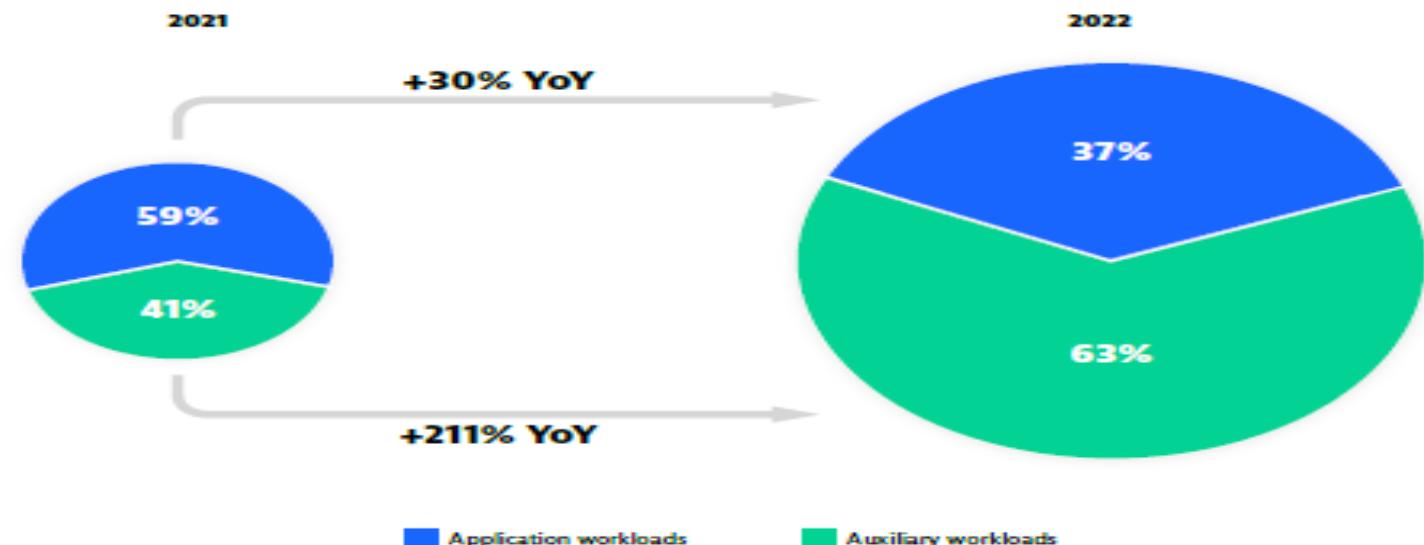
Focusing on non-application workloads, organizations use an increasing variety of technologies. These results reflect the need to enhance Kubernetes with better observability, security, and service-to-service communications. Similarly, other technologies enable specific use cases like CI/CD tools or databases. Across all categories in the Kubernetes survey, open source projects rank among the most frequently used solutions.



## Kubernetes is emerging as the “operating system” of the cloud

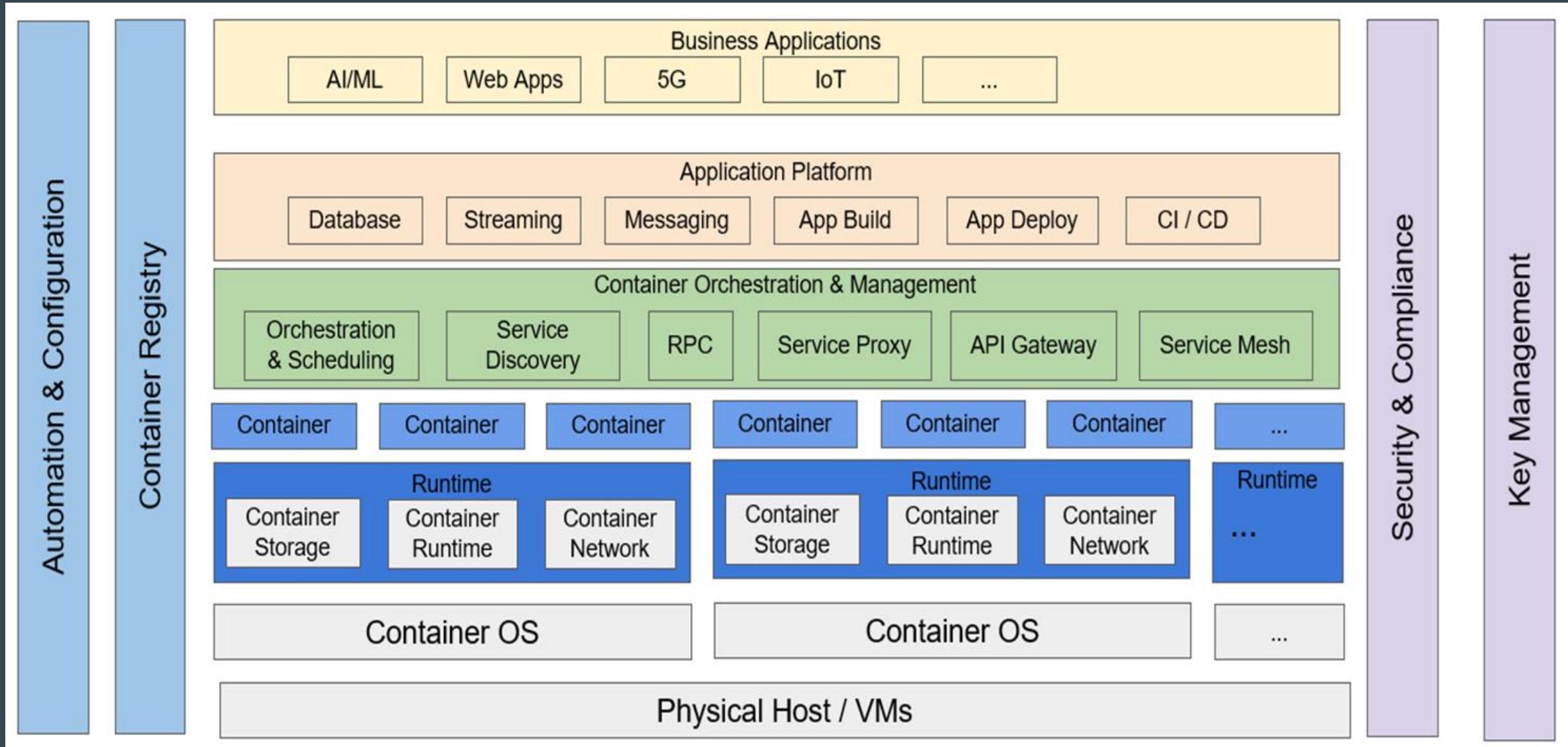
As the ideal orchestration platform for running cloud-native microservice applications, Kubernetes comes with the benefit of built-in deployment, scaling, and resiliency capabilities. In 2021, in a typical Kubernetes cluster, application workloads accounted for most of the pods (59%). By contrast, all non-application workloads, such as system and auxiliary workloads, played a relatively smaller part.

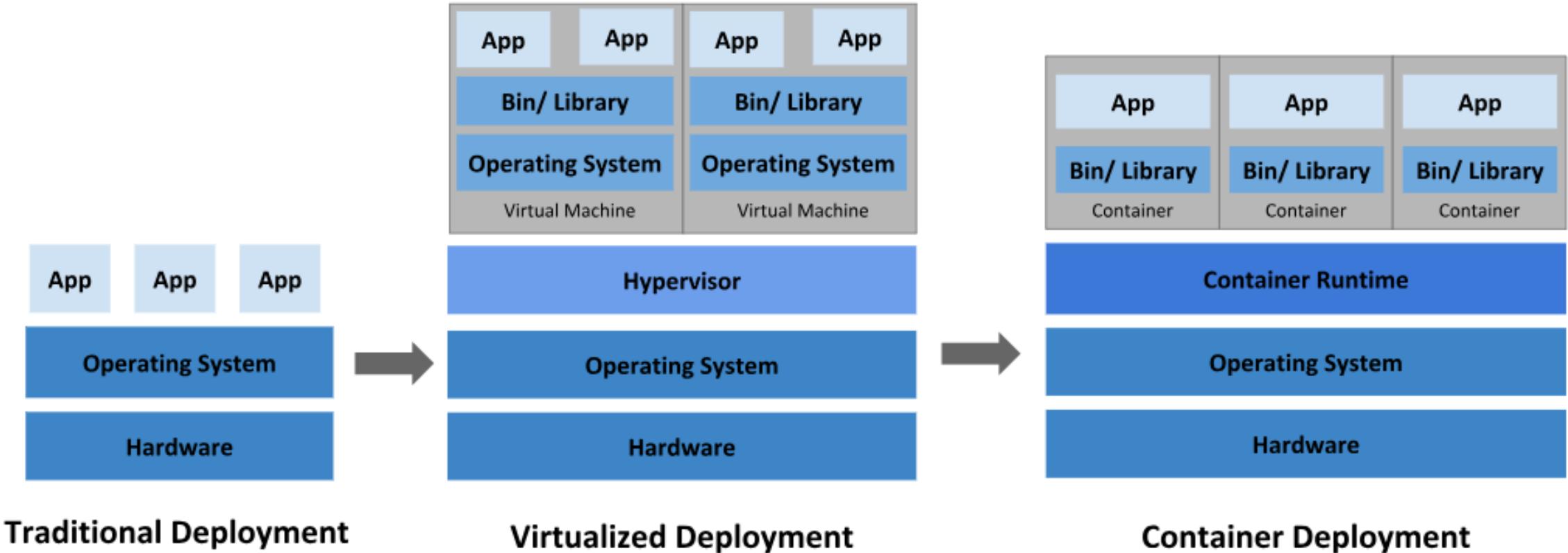
But in 2022, this picture reverses. As Kubernetes adoption has grown, auxiliary workloads now outnumber application workloads (63% vs. 37%). This switch reflects that organizations are implementing more advanced Kubernetes platform technologies such as security controls, service meshes, messaging systems, and observability tools. At the same time, organizations are using Kubernetes for a broader range of use cases, including build pipelines and scheduled utility workloads, among others. Kubernetes becomes the platform for running almost anything. As such, Kubernetes is emerging as the “operating system” of the cloud.



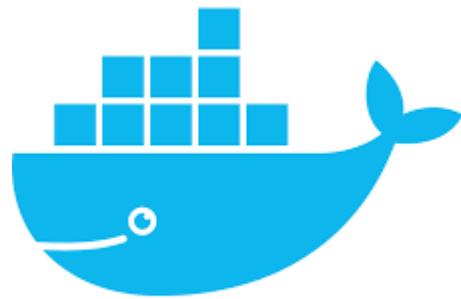
In 2021, application workloads dominated, whereas in 2022, auxiliary workloads were predominant, showing a broader range of use cases.

# Cloud Native Stack

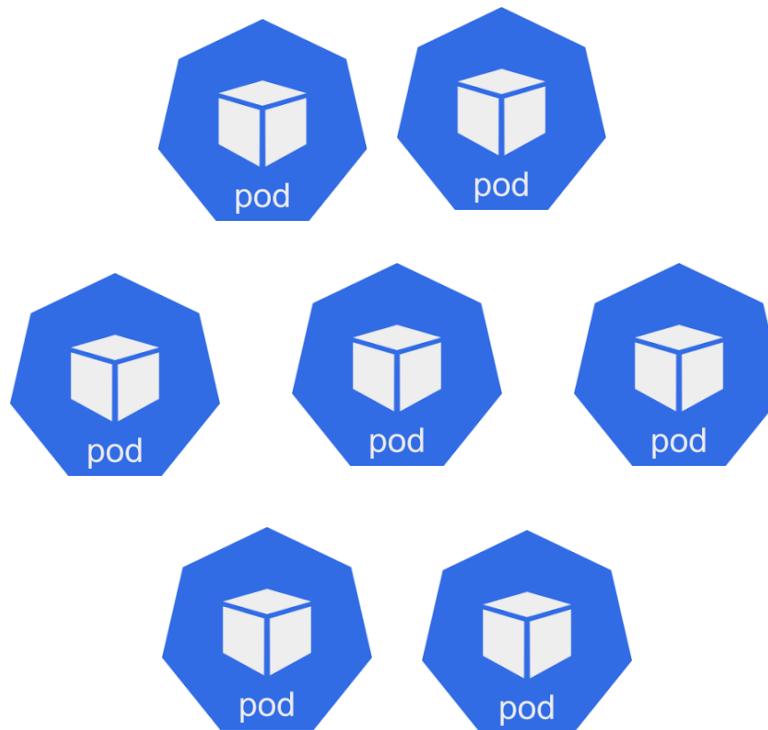




# Containers



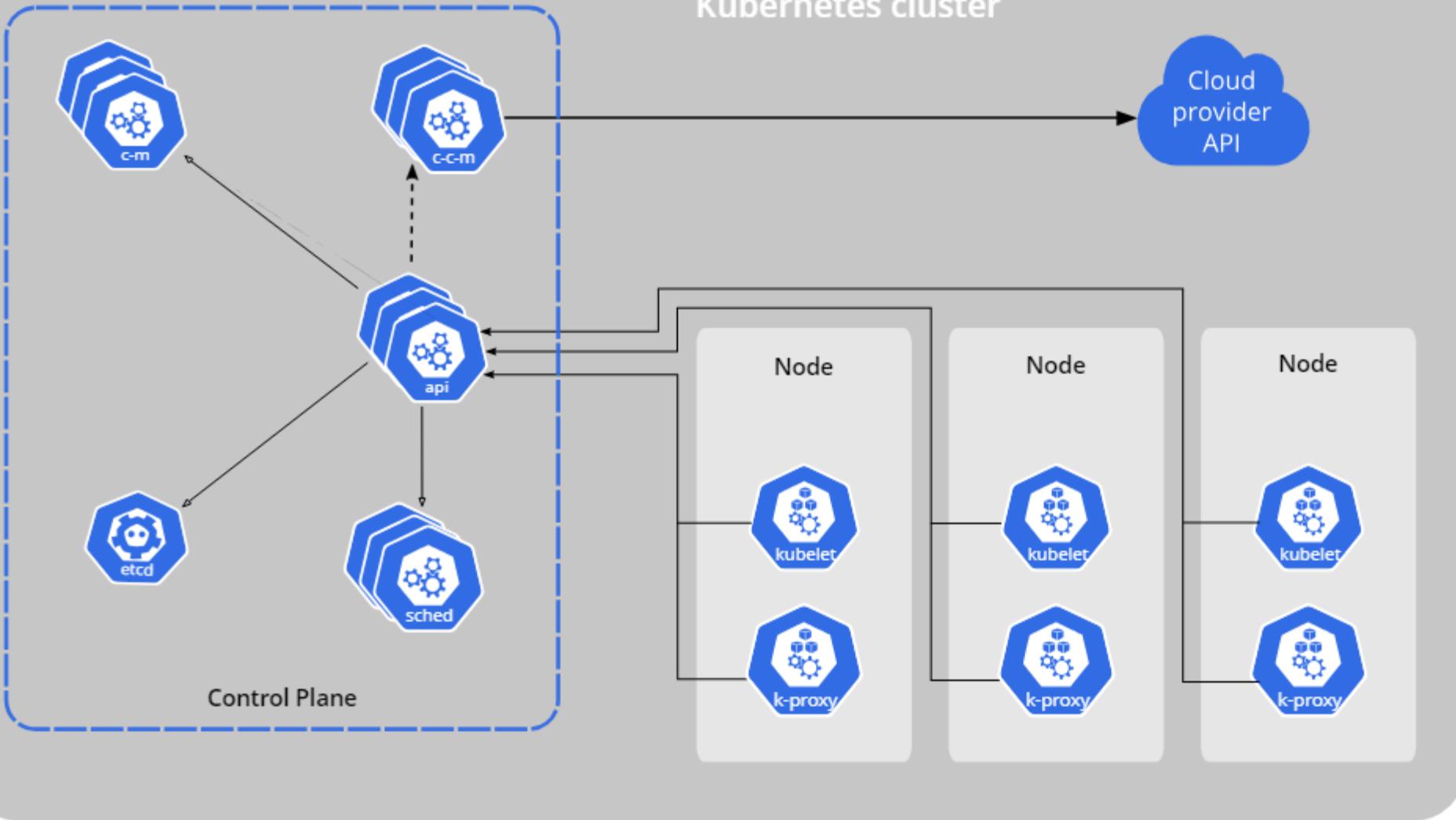
# Pods



# Kubernetes



## Kubernetes cluster



API server



Cloud controller manager  
(optional)



Controller manager



etcd  
(persistence store)



*kubelet*



*kube-proxy*

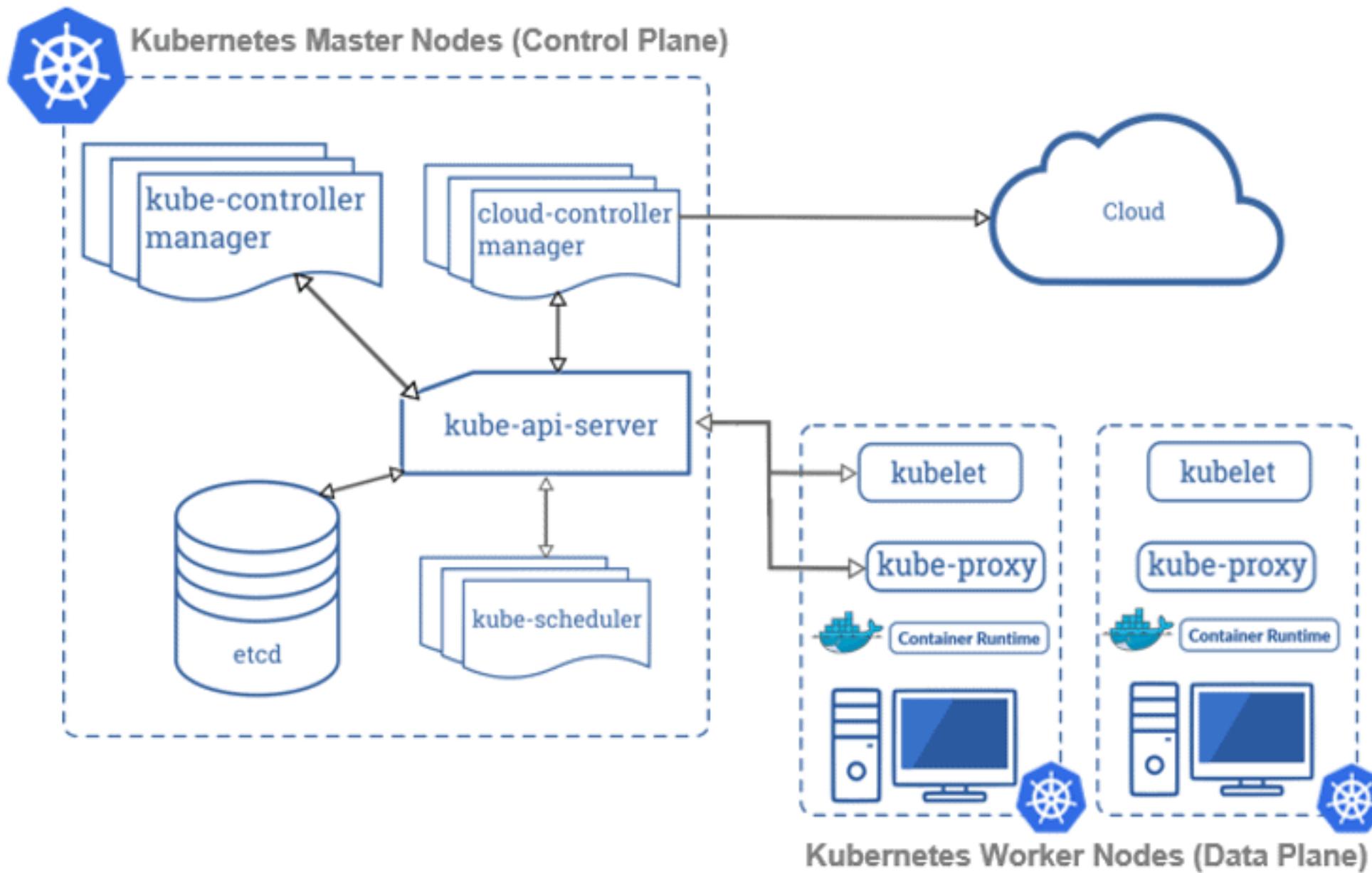


Scheduler



Control plane





# Opportunities - Kubernetes

## Security

- Zero trust arch
- Runtime security tools
- Policy-driven security management
- Pen testing
- Assessments
- AI/ML

## Observability

- Distributed Monitoring
- Async Alerts
- AI/ML
- Performance
- “Real” Observability

## Data Management

- Hybrid
- Heterogenous
- Distributed
- Unstructured
- Lake / Pool / House .. ☺
- AI/ML

# Knowing SODA Foundation Better!

*One Open Ecosystem for Data, Infinite Possibilities*

## SODA Introduction

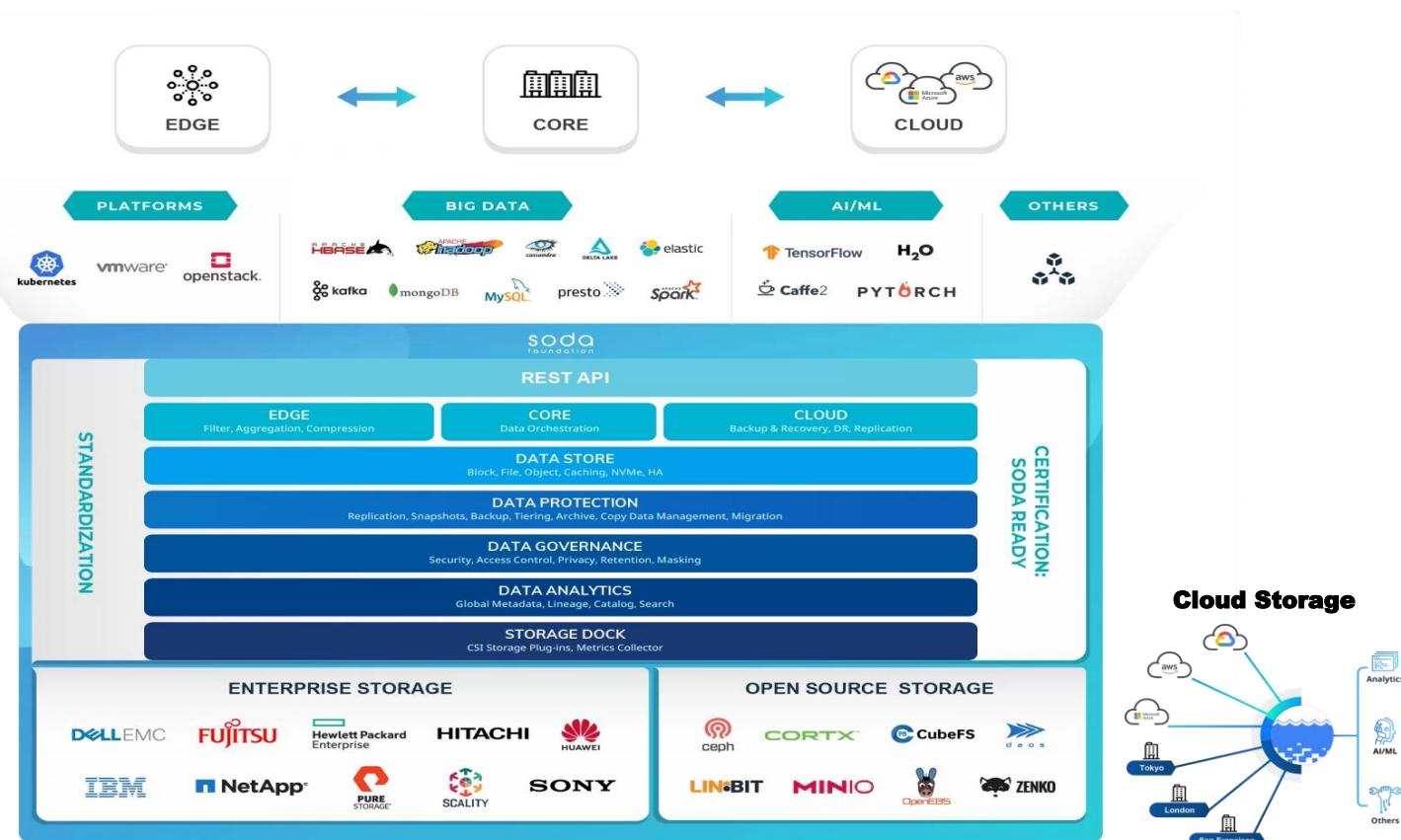


# What is SODA Foundation?

Sub Foundation Chartered under Linux Foundation | Launched on June 29, 2020

aims to foster an ecosystem of open source data management and storage software for data autonomy. SODA Foundation offers a neutral forum for cross-projects collaboration and integration and provides end users quality end-to-end solutions

- Open Source Projects
- Open Research
- Open Collaboration
- Open Standards



# SODA Members

## Premier Members



## General Members



## Supporters



## Associate Members



## Alliance Partner



Run by Industry Organizations

# SODA Help Organizations Thrive With Open Source

**Open Source:** accelerate innovation and development with open source projects and ecosystems

**Open Collaboration:** engage the community and build partnerships on a neutral platform

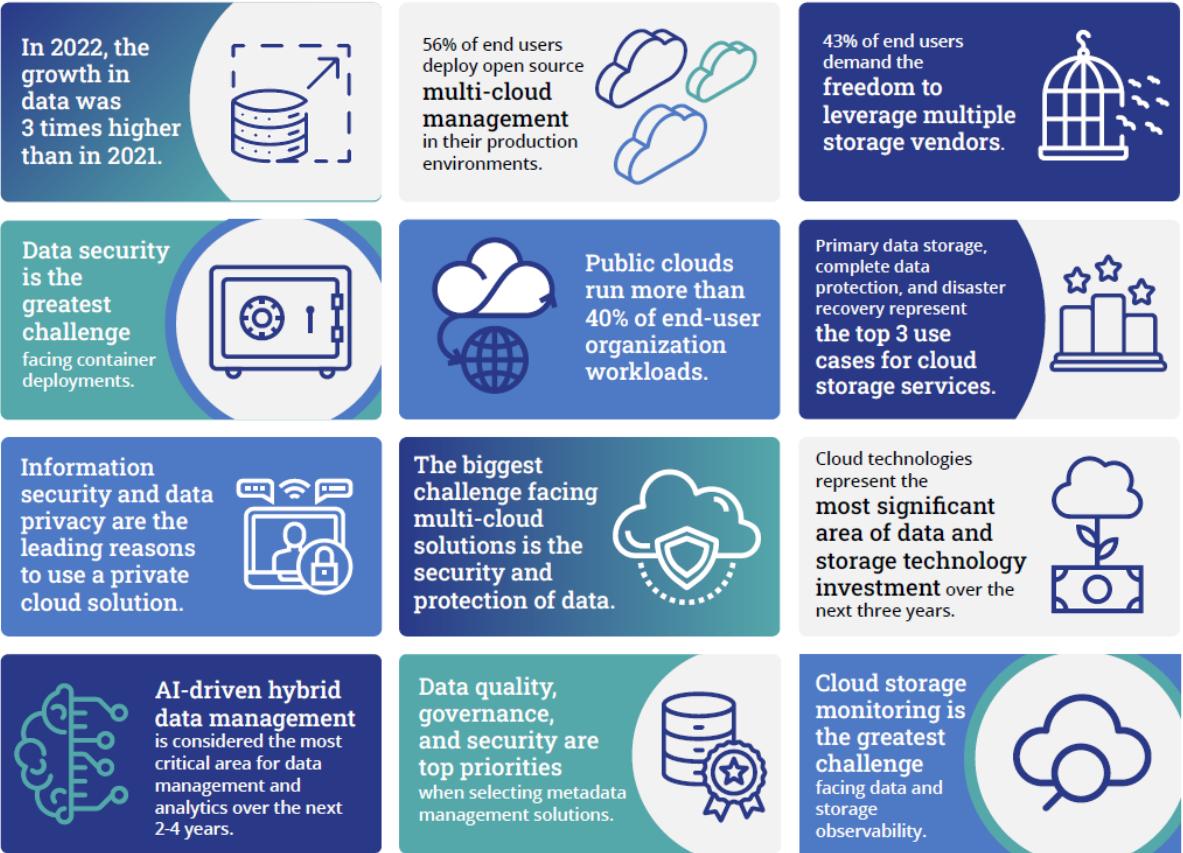
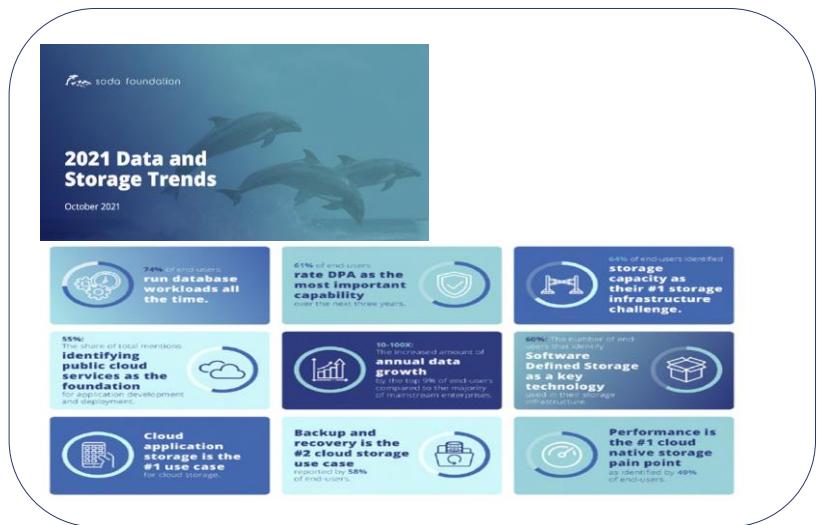
**Open Research:** expand insights and explore ideas with open research and case studies



# Open and Neutral Governance



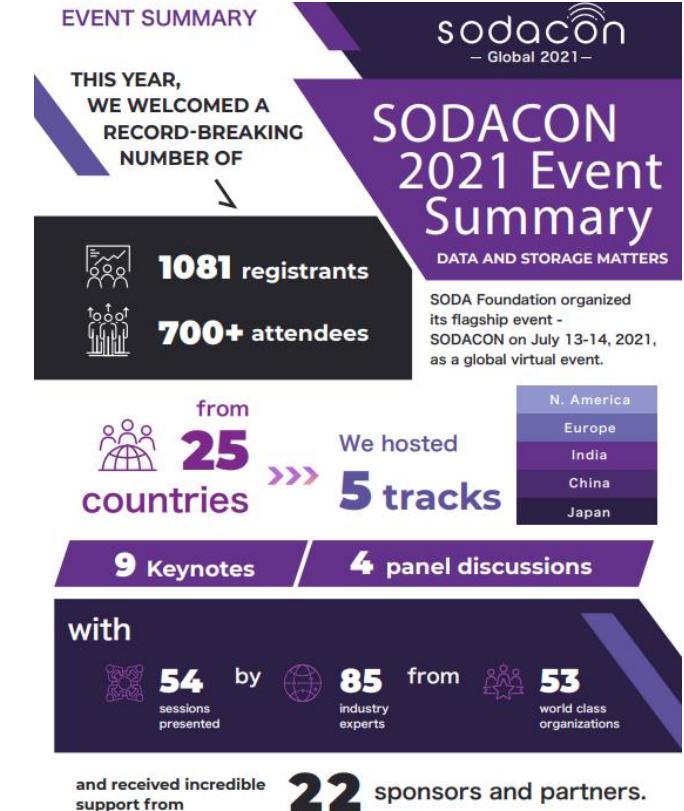
# Technology Trends and Research: SODA Data & Storage Surveys



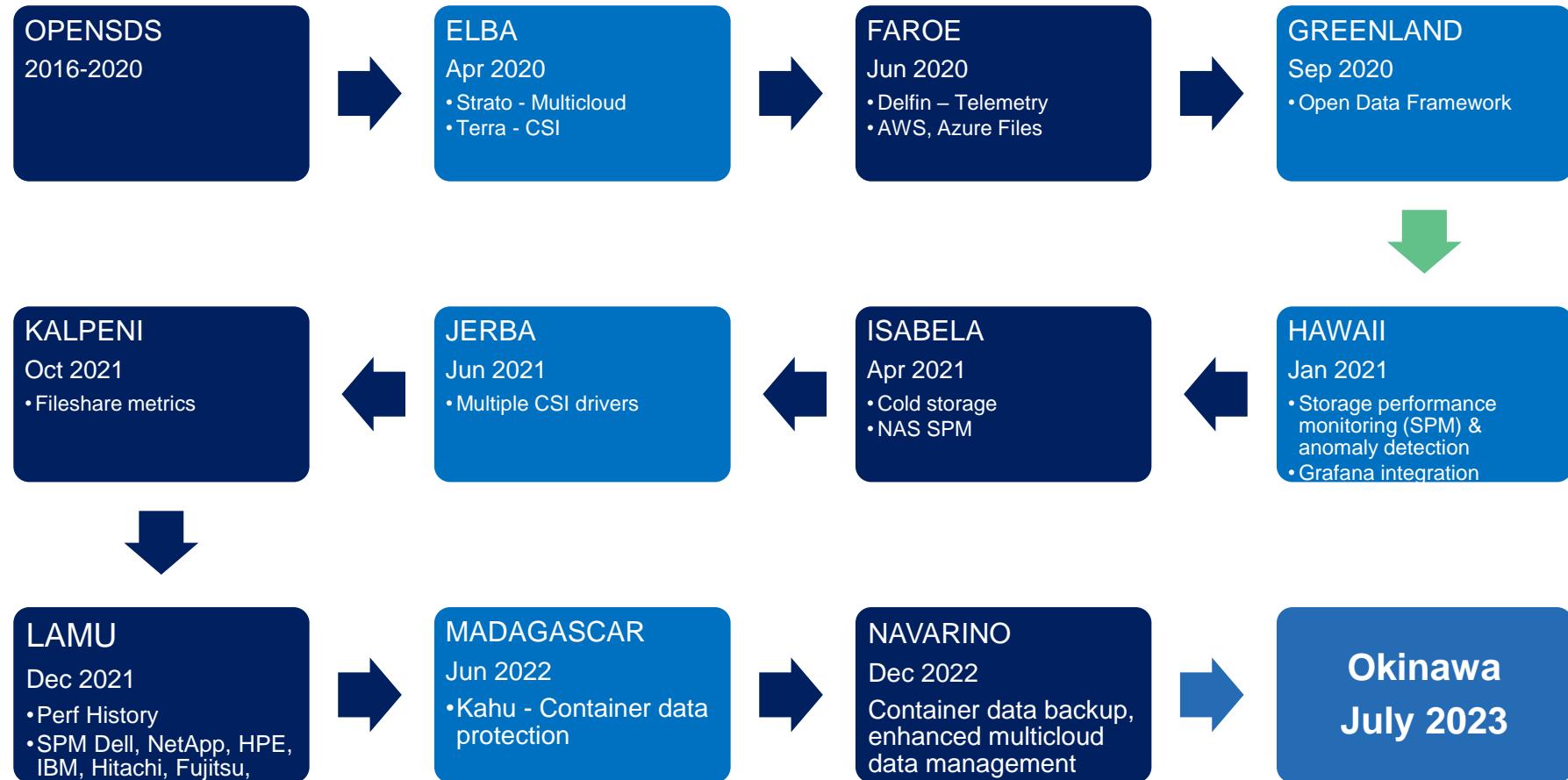
Research Collaboration and Surveys for Data & Storage Technology Direction

soda foundation

# Global Events: Technology and Development Collaboration



# Regular Project Releases: *For Data Framework and Solutions*



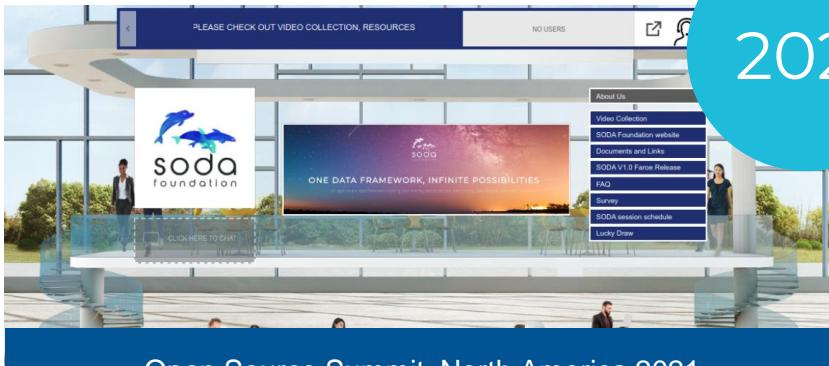
# Glimpse of SODA Events

2022

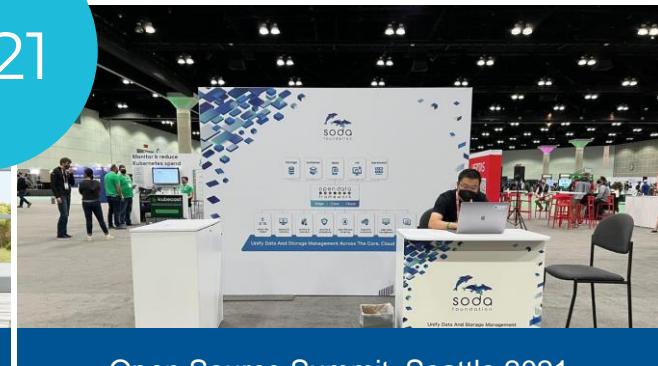


SODACON2022, Yokohama, 2022

2021



Open Source Summit, North America 2021



Open Source Summit, Seattle 2021

2020



Kubernetes Forum, Delhi 2020

2019



SODA China Community Launch, Beijing 2019



SODA Forum, Tokyo 2019

2017



Open Source India, Bengaluru 2019



Open Source Summit, Lyon 2019

2018



1<sup>st</sup> Japan Meetup, Tokyo 2018

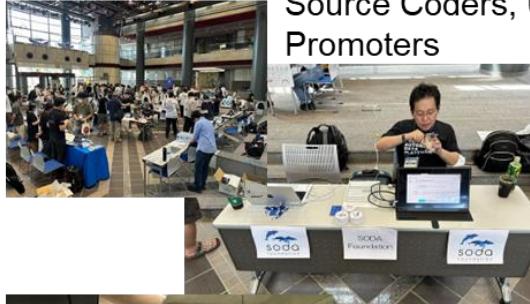


OpenSDS Mini Summit, Prague 2017

# 2023



## COSCUP 2023 Japan



## Conference for Open Source Coders, Users & Promoters



## SODA China Meetup 2023 Shanghai, China September 26, 2023

SODA China Meetup, organized by SODA Foundation and NextArch Foundation in collaboration with LF OSS Asia, convened over 400 online & in person attendees for an insightful exploration journey on unstructured data trends and learned about Crystal, a SODA project for unified metadata management.



## SODA Data Vision 2023 Bilbao, Spain September 18, 2023

Data Vision 2023, organized by SODA Foundation in collaboration with LF OSS Europe, convened over 400 online & in person attendees in Bilbao, Spain, on September 18. The event featured a full day of 4 keynotes, 10 sessions, and 2 panel discussions, bringing together experts, architects, developers, and end-users to explore recent innovations, future trends, use cases, and challenges in data and storage. Representatives from various organizations, including Kurago, SUSE, Veeam Software, AWS, Microsoft, and more, participated in discussions. The Students Challenge recognized top achievers, with Som Shanker Pandey securing a SUSE eLearning Silver subscription, and Utkarsh Umre, Vaishali Rawat, and Kelvin Parmar winning free access to a Kubernetes for Developer Course. The event fostered collaboration and knowledge exchange within the SODA community.



## DATA VISION by SODA September 18, 2023 Bilbao, Spain + Virtual #datavision23

## SODA at Open Source India (OSI) 2023 Bengaluru, India October 19-20, 2023

SODA Foundation participated as a Strategic Community Partner at OSI Days 2023 in Bengaluru on October 19 and 20. The event, organized by the EPI group in collaboration with India's Open Source Community and Industry, attracted over 4,700 attendees, with engagement across 100+ sessions, 100+ exhibitors, and 10+ tracks. On Day 2, Steven Tan, SODA Foundation Chair and SVP & CTO Cloud Solution at Futureus, delivered the opening keynote remotely, along with Sanil Kumar D, Founder & CEO of Cad Labs, who presented on the SODA Project. The event provided a platform for attendees to learn and explore innovations in unstructured data trends and learned about Crystal, a SODA project for unified metadata management. Additionally, SODA hosted an onsite booth over hot days, attracting over 600 attendees for information on SODA Projects & Ecosystem, an engaging them with interactive quiz events.





Sub Foundation



**CLOUD NATIVE**  
COMPUTING FOUNDATION

Cloud Native Projects :  
Kubernetes and more!  
<https://landscape.cncf.io/>

Sub Foundation



**soda**  
foundation

Data Management Projects

What we have...

# Growing Project Ecosystem

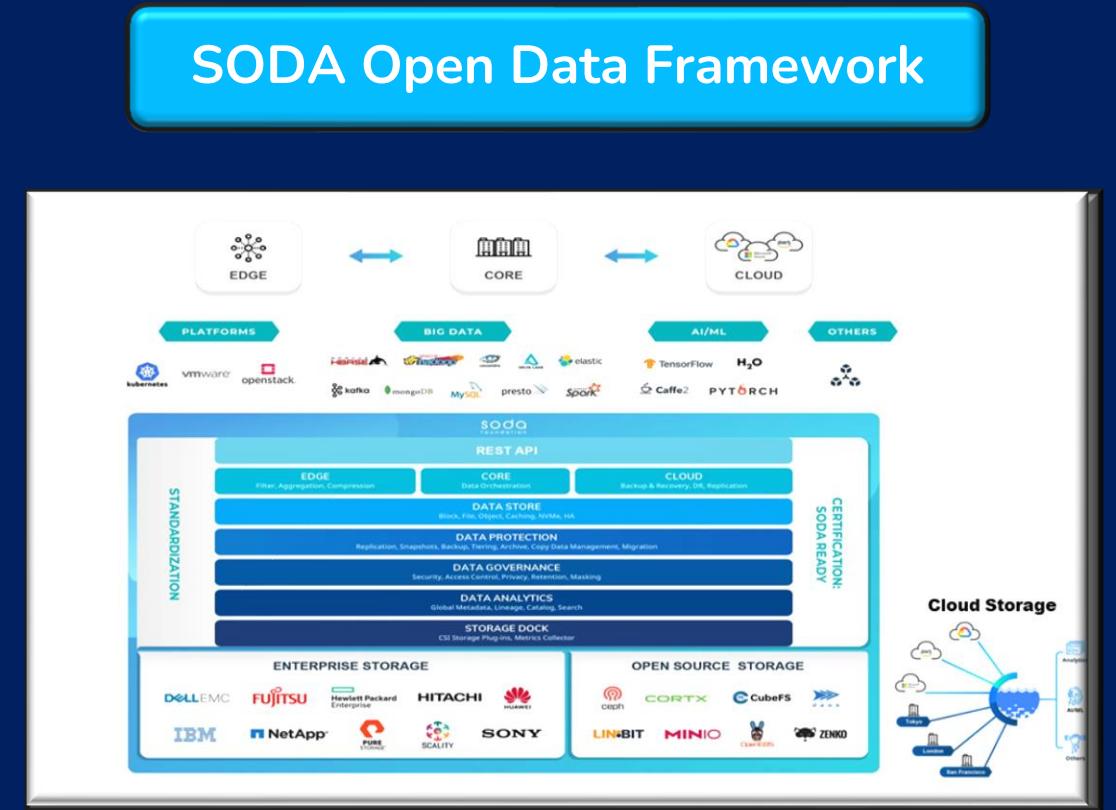
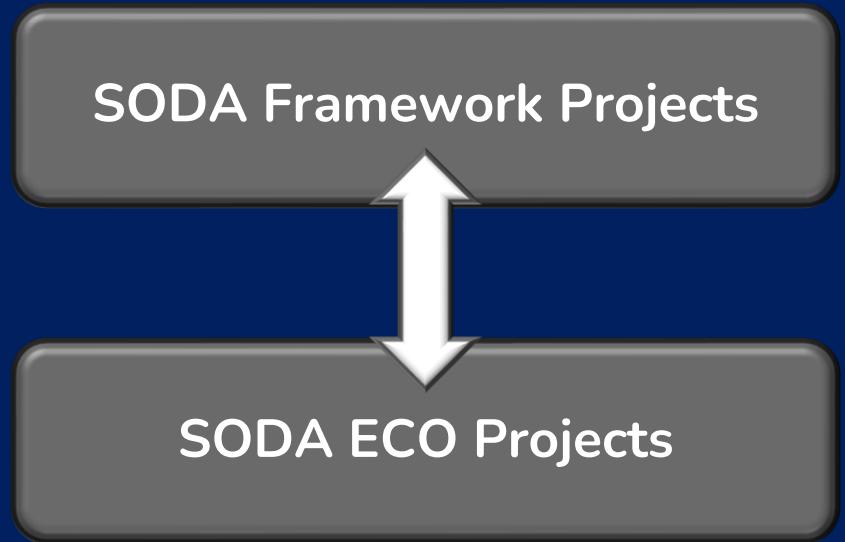


SODA Framework Projects

New SODA Framework Projects (2022)

SODA ECO Projects

Framework Projects for SDS Controller, Multicloud, Container Data Management and Data Lake |  
10 External Projects in Data & Storage joined as SODA ECO Projects to build collaborative solutions.



## SODA Framework Projects

SODA Framework projects focus on solving a common problem faced by end users  
Managed and maintained by SODA Foundation directly



### KAHU

Backup, recover and migrate K8s clusters data anywhere with no vendor lock-in

PROTECT



### STRATO

Move data across multicloud environment with a common S3 compatible interface

MOVE



### DELFIN

Observe, monitor and manage alerts for any storage anywhere

OBSERVE



### CRYSTAL

Unified metadata for unstructured data across on-premise and cloud storage

DISCOVER

<https://github.com/sodafoundation>

SODA ECO projects are partner open source projects join SODA project landscape to **build collaborative solutions for end to end data & storage use cases**



# SODA Framework Projects



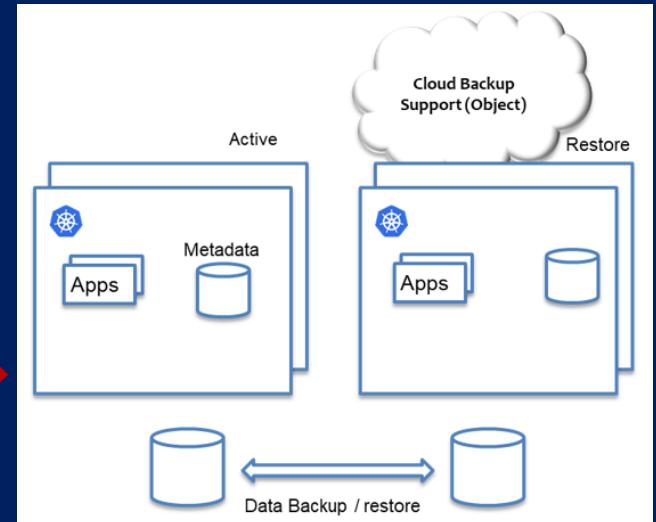
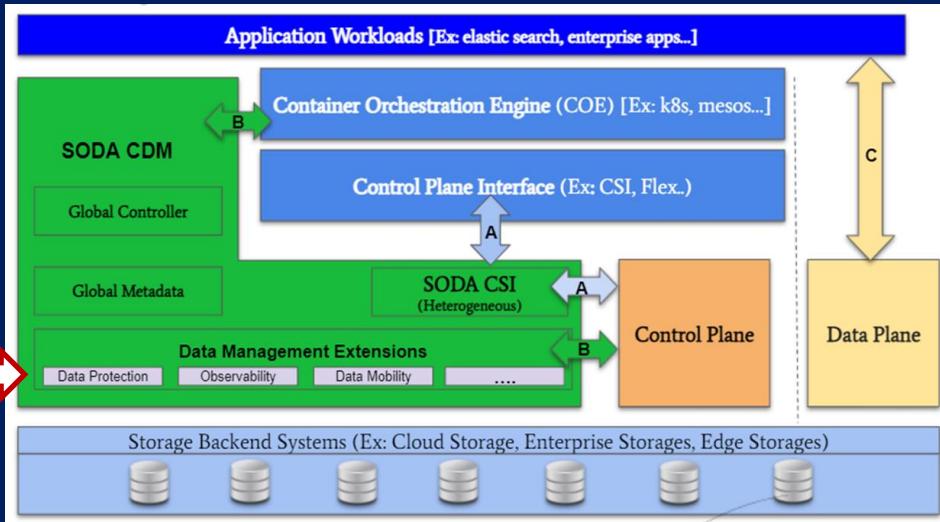
KAHU

Backup, recover and  
migrate K8s clusters data  
anywhere with no vendor  
lock-in

PROTECT

### Backup and Restore

- metadata and data
- snapshot support
- prehook / post hook support
- cross cluster backup and restore
- Backup of CSI provisioned and non CSI provisioned volumes



Container Data Management → Container Data Protection → Kahu → Backup / Restore → Kahu

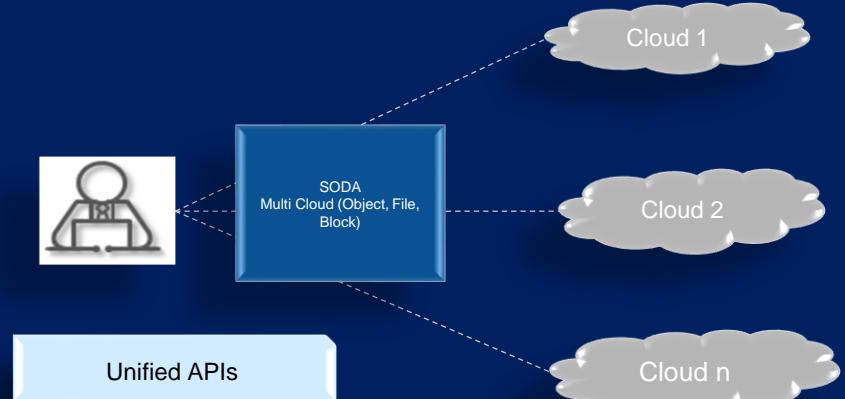
### Storage Provider Framework

- Pluggable volume/metadata backup providers
- Runtime co-existence of multiple backup providers

### Automation & Orchestration

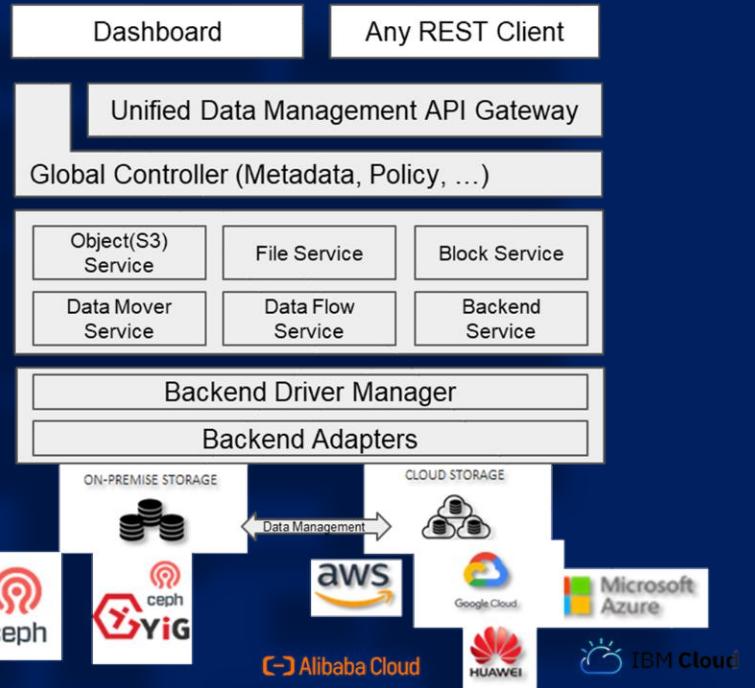
- Scheduled backup
- Policy / Event-driven backup
- Provider-independent service plan automation

  
**STRATO**  
 Move data across  
 multicloud environment  
 with a common S3  
 compatible interface



## MOVE

- Unified multicloud API
- S3 compatible
- Support all major cloud vendors
- Data migration
- Data Lifecycle management
- Hybrid data management for object storage (on-prem-cloud)
- File/Block/Object
- User Level Tiering Plan – Storage Service plan
- Smart Archival





## DELFIN

Observe, monitor and manage alerts for any storage anywhere

## OBSERVE

### Resource monitoring

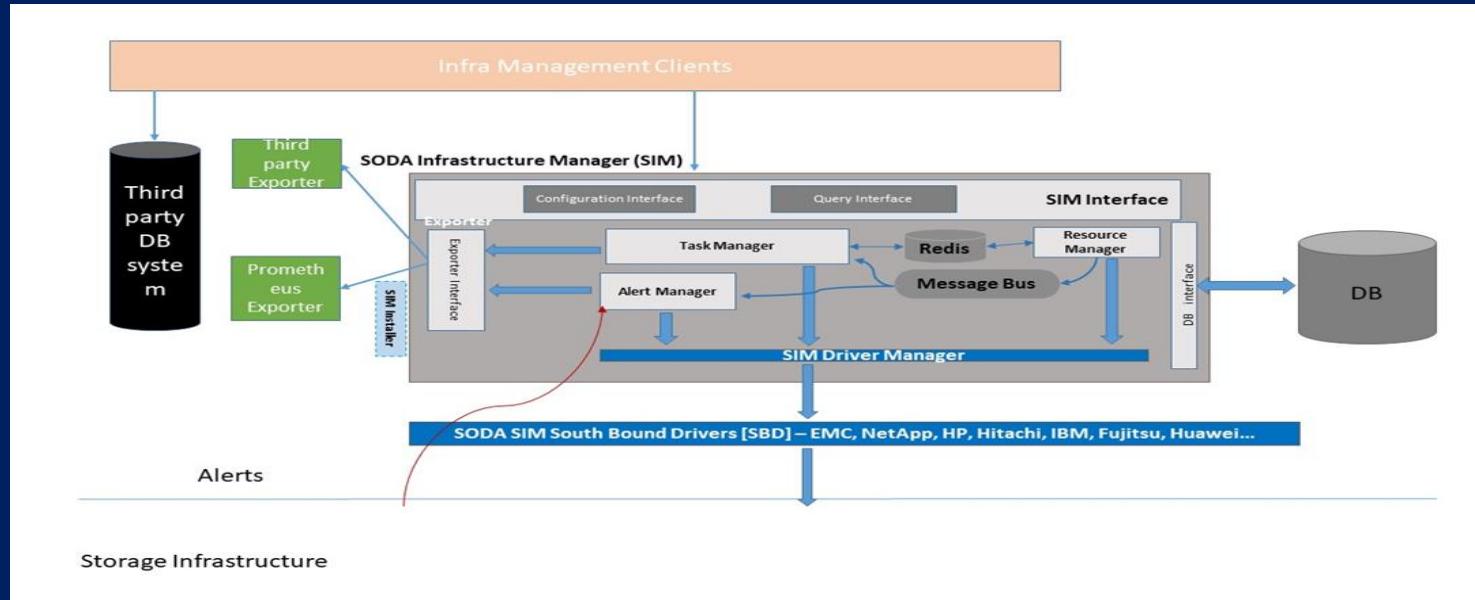
- Storage, Pool, Volume, Port, Network & Controllers
- Filesystem, Share, Quota, Qtree
- Mapping views, topology relations

### Performance monitoring

- Bandwidth, Throughput, IOPS, Response time, Capacity

### Alert Management

- Alert Notifications and Management



### Driver Manager

- Easy support for new storage devices
- Support different vendor management platforms

### Plugin Manager

- Easy support for new clients (export monitoring data to Prometheus, Grafana, Kafka or custom clients)

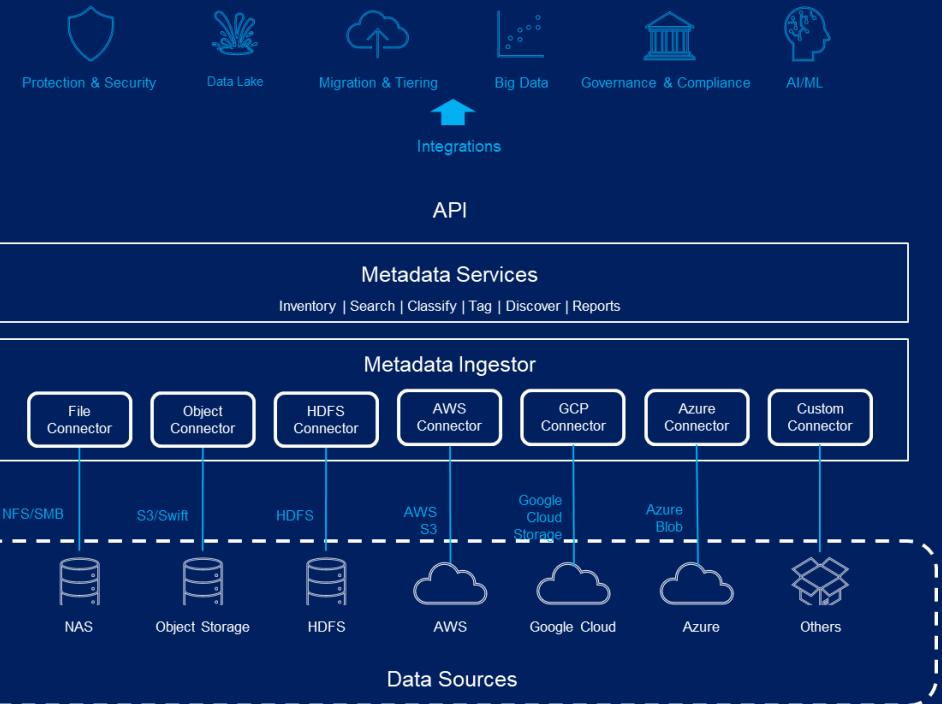
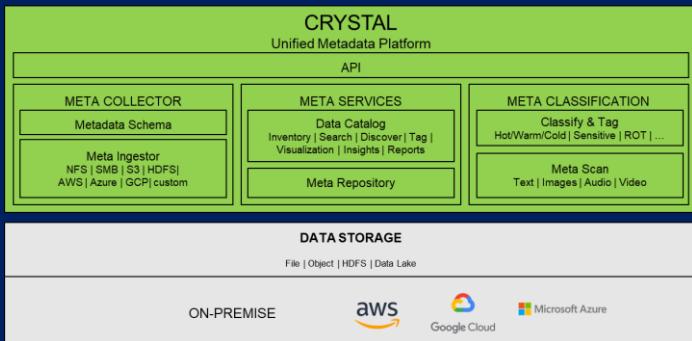


## CRYSTAL

Unified metadata for  
unstructured data across on-  
premise and cloud storage

### DISCOVER

- The project just started
- Added with some base code derived from **SODA Strato** (s3 metadata management)
- S3 metadata get, store and search
- Basic demo of apis



<https://github.com/sodafoundation/crystal>

What we plan...



### KAHU

Backup, recover and migrate K8s clusters data anywhere with no vendor lock-in

#### PROTECT

- Replication and Failover
- Backup – Restore Performance Benchmarking



### STRATO

Move data across multicloud environment with a common S3 compatible interface

#### MOVE

- Data Mover Performance Optimization
- Re-usable components and reorganization of the project inline with SODA Crystal



### DELFIN

Observe, monitor and manage alerts for any storage anywhere

#### OBSERVE

- Kubernetes Storage Monitoring
- Performance benchmarking



### CRYSTAL

Unified metadata for unstructured data across on-premise and cloud storage

#### DISCOVER

- Common metadata schema design
- Analysis and value positioning



soda foundation

# Smart and Secure Data Management

Making Cloud Native Data Management  
Smarter and Secured!

Getting Ready.....



## Cloud

- Hybrid Cloud Data Orchestration & Management
- Cloud Native Observability
- Metadata management for unstructured data

## AI

- AI Data exporters
- LLM-based analytics/prediction?
- LLM based assistant modules for data management
- Vector Databases
- Data Mesh

## Security

- Cloud Data Security
- Cloud Native Security hardening for data
- Security assessment



<http://bit.ly/soda-starter>

<https://www.sodafoundation.io/slack/>

**Let us build Next Gen Smarter Data First Solutions in Open Source!**

Some more links we discussed:

<https://www.meetup.com/sodafoundationindia/> - join soda India meetup

<https://www.sodafoundation.io/slack/> - join soda slack

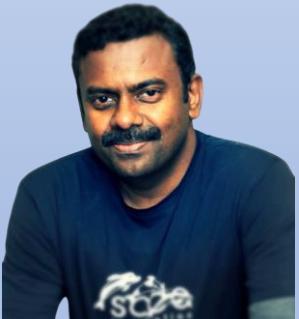
<https://github.com/> - to create your github id

<https://www.linkedin.com/company/sodafoundation> - soda foundation linkedin

So..

get best of  
Let us Open Source!

# Thank You!



## Sanil Kumar D.

Founder CEO Caze Labs | SODA Foundation TOC Co-chair |  
Industry Awarded Technologist | Patents & Research Papers | Speaker |  
Founder 123Life | PhD Scholar (Open Source & Observability) | ACM Eminent Speaker

