Apache Cassandra™ in a Nutshell

Sponsored by DataStax

NoSQL Distributed Database

1 Installation = 1 NODE
- ✔ Capacity = ~ 2–4TB
- ✔ Throughput = LOTS Tx/sec/core

DataCenter (DC) | Ring

Communication:
- ✔ Gossiping
- ✔ No Master (peer-to-peer)
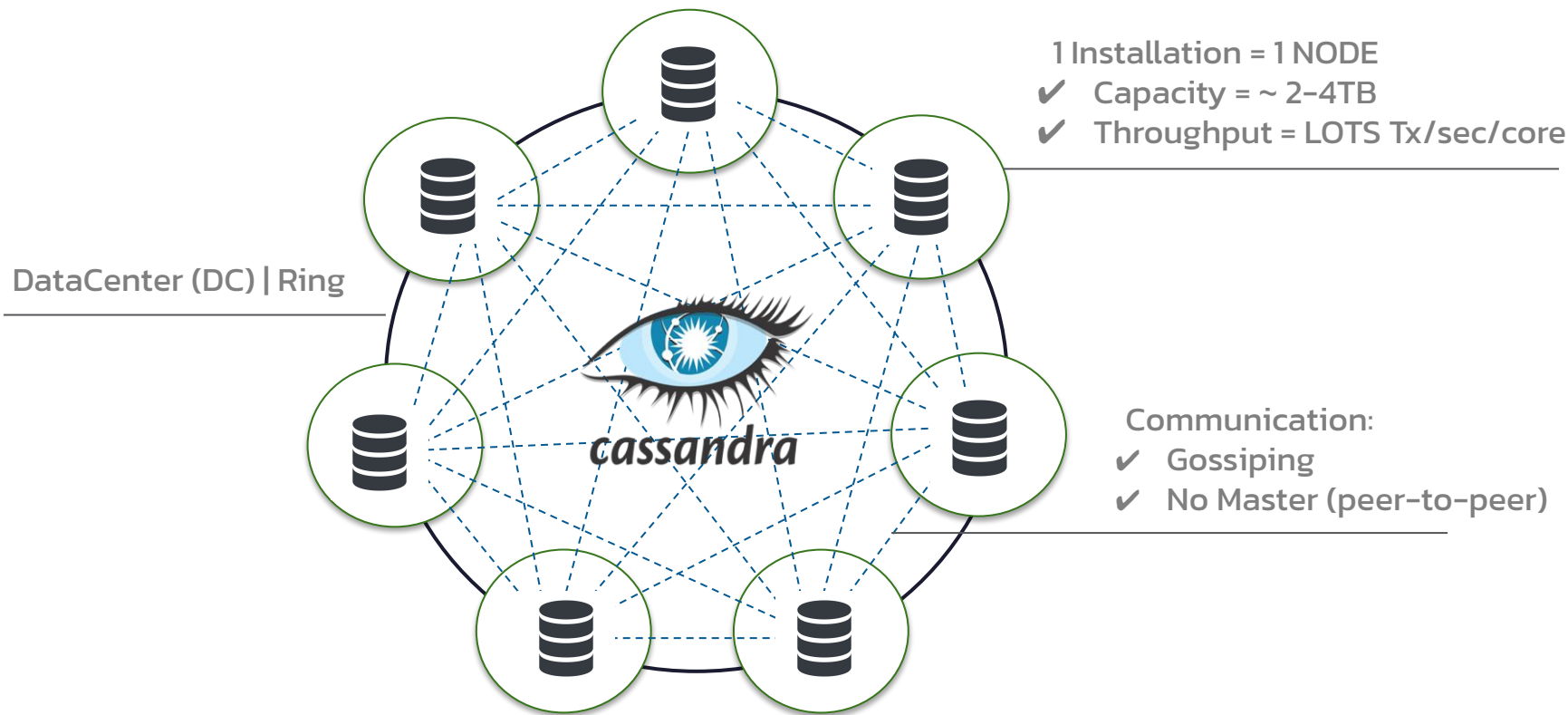
cassandra

~~NoSQL~~ Specialized, Purpose-built <u>Distributed</u> <u>Database</u>

1 Installation = 1 NODE
- ✔ Capacity = ~ 2–4TB
- ✔ Throughput = LOTS Tx/sec/core

DataCenter (DC) | Ring

Communication:
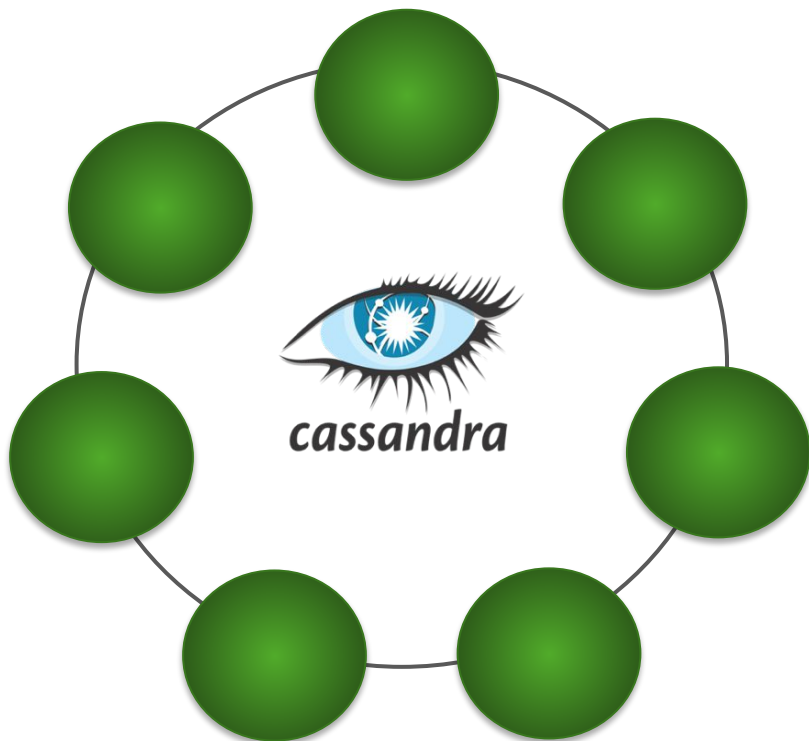- ✔ Gossiping
- ✔ No Master (peer–to–peer)
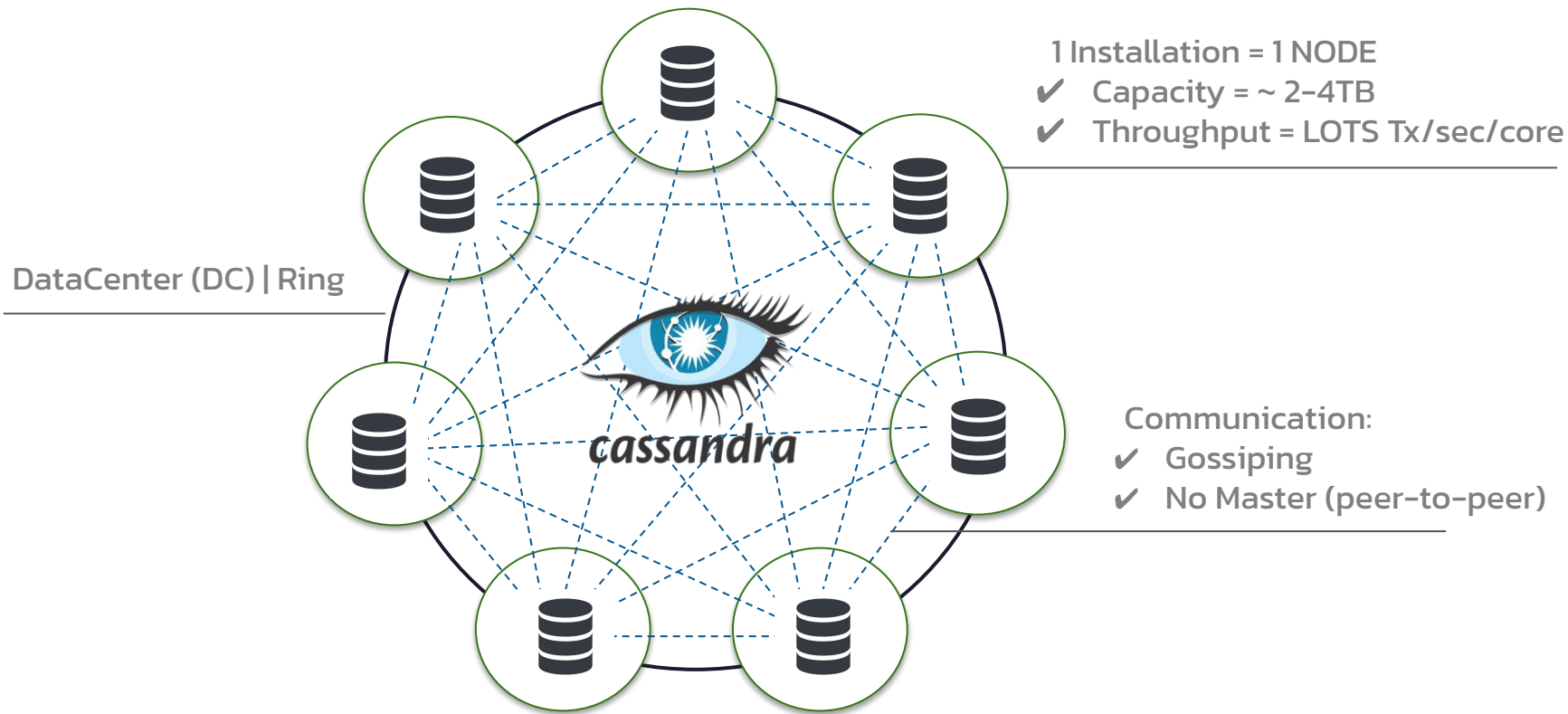
Apache Cassandra in a nutshell

1. Big Data Ready
2. Linear Scalability
3. Read / Write Performance
4. Highest Availability
5. Self-Healing and Automation
6. Geographical Distribution
7. Platform Agnostic
8. Vendor Independent
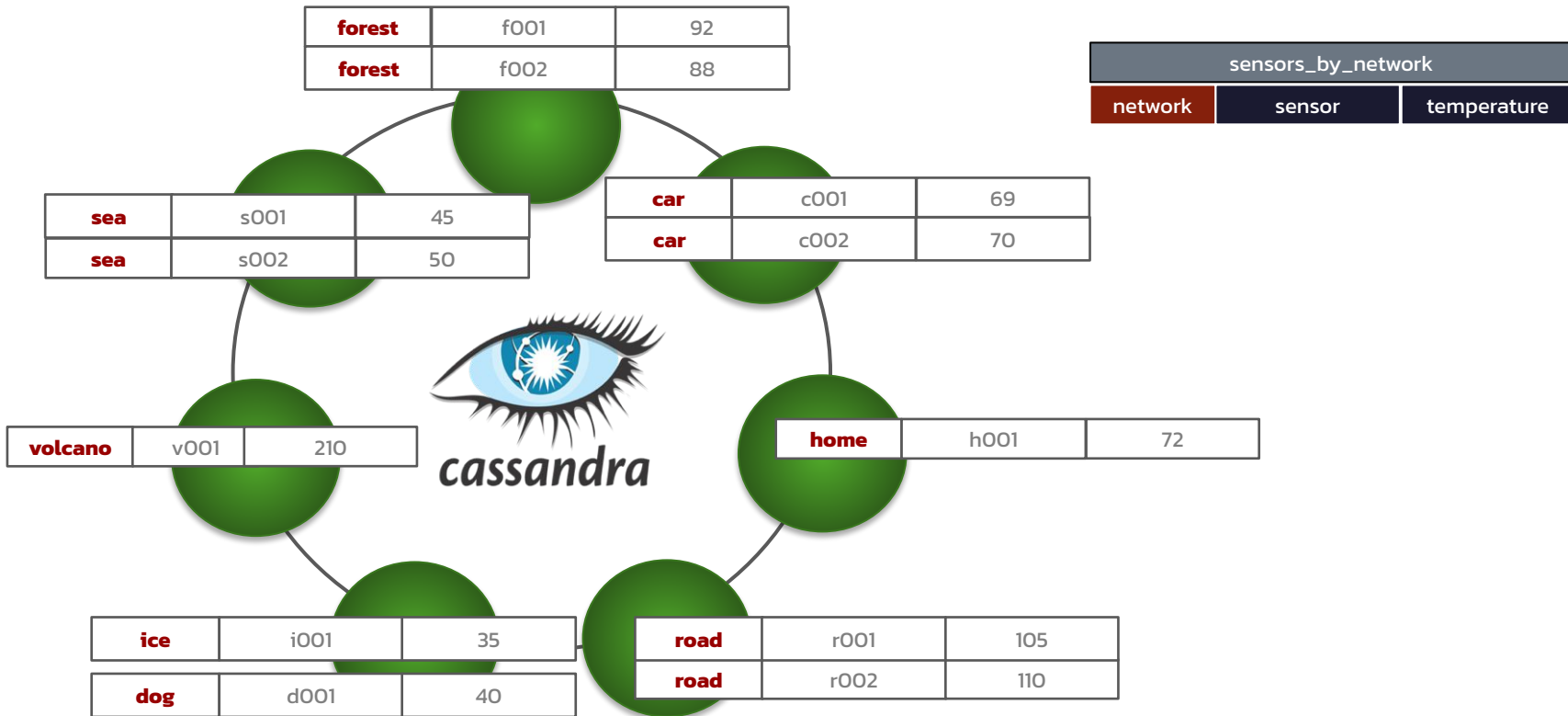
1. Data is Distributed

sensors_by_network		
network	sensor	temperature
forest	f001	92
forest	f002	88
volcano	v001	210
sea	s001	45
sea	s002	50
home	h001	72
car	c001	69
car	c002	70
dog	d001	40
road	r001	105
road	r002	110
ice	i001	35

Partition Key

Primary Key

1. Data is Distributed

forest	f001	92
forest	f002	88

sensors_by_network		
network	sensor	temperature

sea	s001	45
sea	s002	50

car	c001	69
car	c002	70

volcano	v001	210

home	h001	72

ice	i001	35
dog	d001	40

road	r001	105
road	r002	110

1. Data is Distributed

sensors_by_network		
network	sensor	temperature

forest

f001	92
f002	88

car

c001	69
c002	70

sea

s001	45
s002	50

volcano

v001	210

home

h001	72

ice

i001	35

dog

d001	40

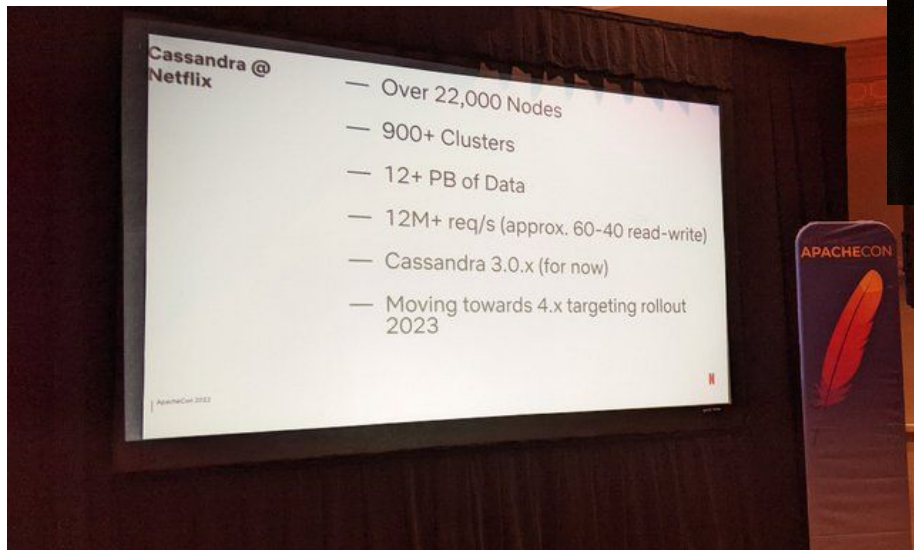**road**

r001	105
r002	110
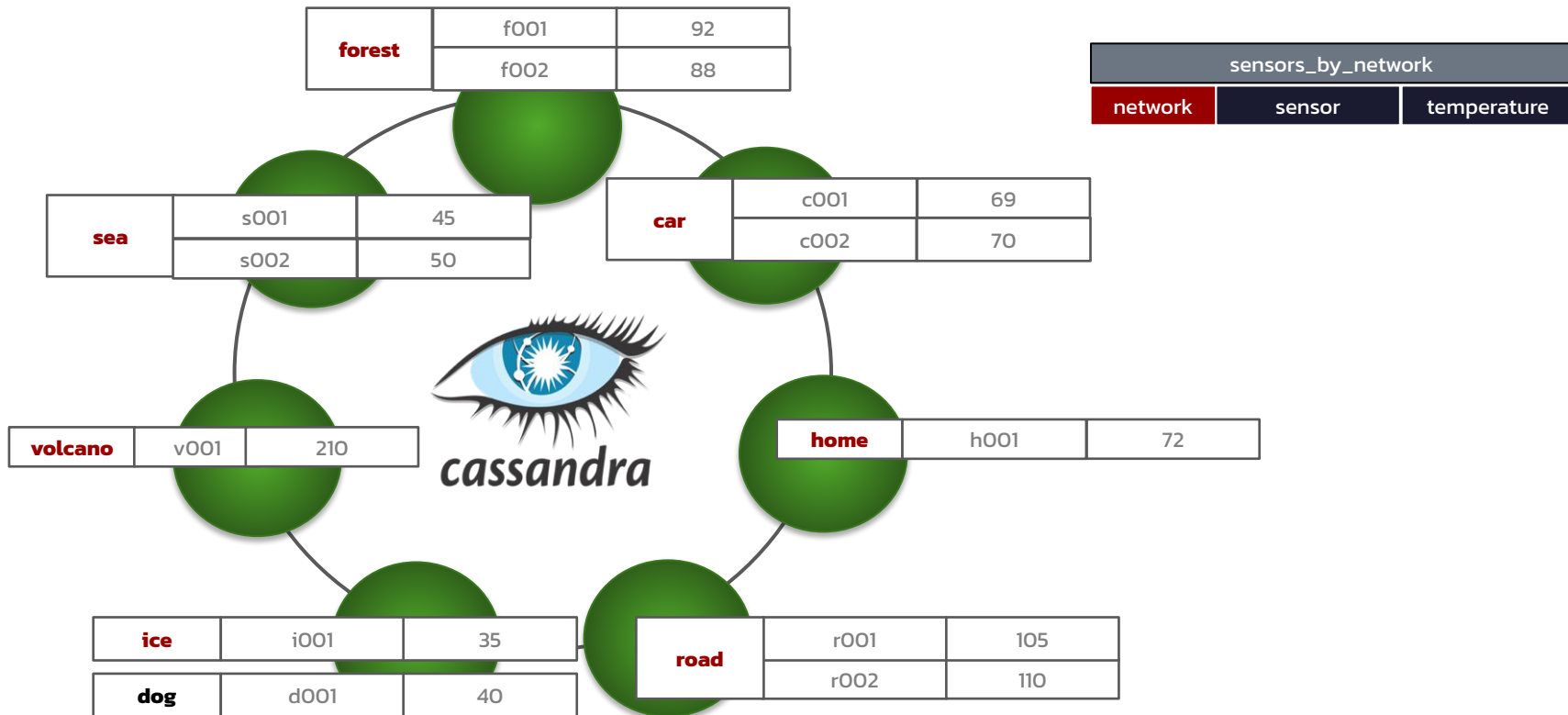
DEMO ?

2. Linear Scalability

- For volume or velocity, there are no limitations

- **Linear** - No overhead on new nodes, scales with your needs*

Client Writes/s by node count – Replication Factor = 3

1200000

1000000

800000

600000

400000

200000

0

0   50   100   150   200   250   300   350

174373

366828

537172

1099837

NETFLIX
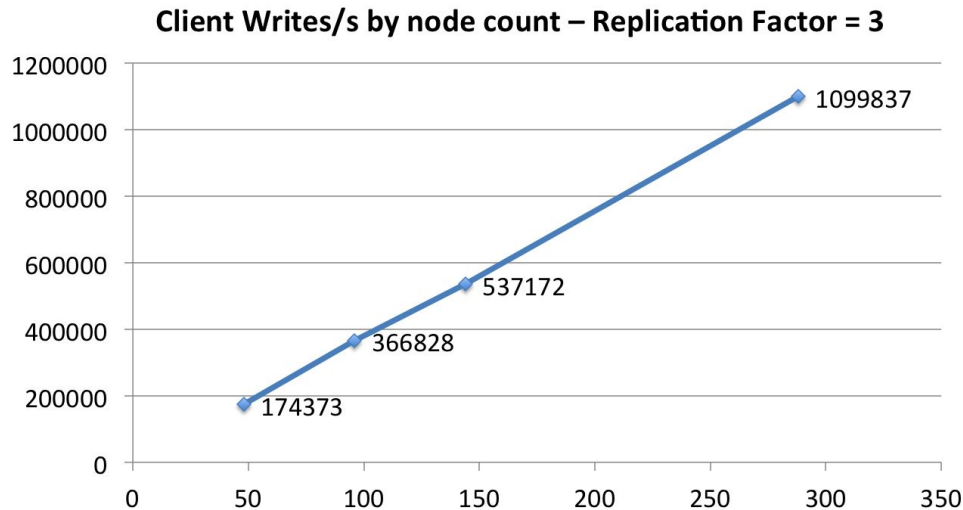
2. Linear Scalability

And many others...

Cassandra @ Netflix
- Over 22,000 Nodes
- 900+ Clusters
- 12+ PB of Data
- 12M+ req/s (approx. 60-40 read-write)
- Cassandra 3.0.x (for now)
- Moving towards 4.x targeting rollout 2023

Apache Cassandra at Apple Scale and Scope
- Over three hundred thousand instances
- Hundreds of petabytes of data
- Over two petabytes per cluster
- Millions of queries per second
- Thousands of clusters
- Thousands of applications

Y Hacker News new | past | comments | ask | show | jobs | submit

31. ▲ "Mother Will Be Pleased": How It Feels to Be Run over (1900) (publicdomainreview.org)
    41 points by samclemens 12 hours ago | hide | 13 comments

32. ▲ Cassandra at Apple: 1000s of Clusters, 300k Nodes, 100 PB (twitter.com/erickramirezau)
    258 points by mfiguiere 17 hours ago | hide | 156 comments

33. ▲ DNS based ad/abuse domain blocking (github.com/openwrt)
    14 points by behnamoh 7 hours ago | hide | discuss

34. ▲ Technical milestone reached: global earth system simulations with 1.2 km resoln (mpg.de)
    149 points by jokabrink 15 hours ago | hide | 73 comments

35. ▲ Why Disconnecting from Your Phone Is So Good for You – and How to Do It (time.com)
    67 points by ppjim 7 hours ago | hide | 30 comments

36. ▲ Happy 20th birthday Jira You suck so bad (rida.dev)
    70 points by julian_digital 2 hours ago | hide | 38 comments

Geographical Distribution

Cassandra's trademark is multi-datacenter deployments, granting you an exceptional capability for disaster tolerance while keeping your data close to your clients - worldwide.

All DCs are active (available for both writes and reads)!
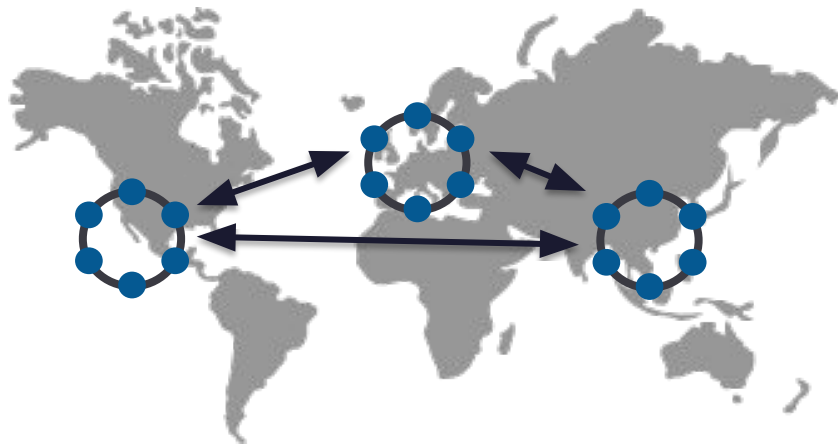
High Availability

Replication, Decentralisation, and Topology-Aware Placement Strategy take care of possible downtimes:

- Multiple Live Replicas
- No Single Point of Failure
- Network topology-aware data placement
- Client-side Smart Reconnection and Strong Retry Mechanism

And there

Your Record is here

And here

And here

And there
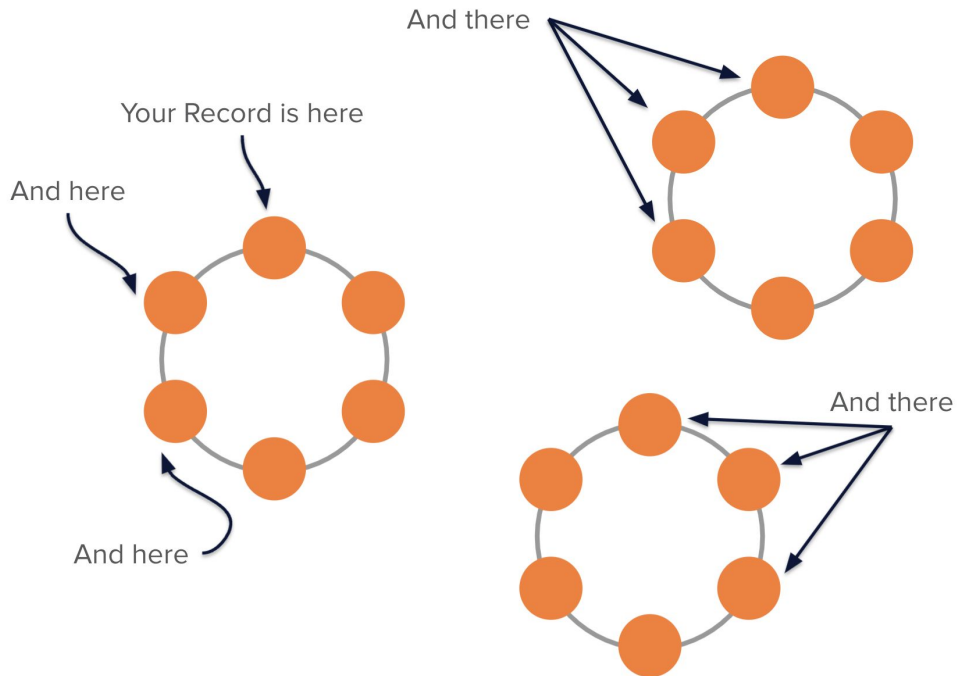
Platform Agnostic

Apache Cassandra is **not bound to any platform** or service provider, helping you build hybrid-cloud and multi-cloud solutions with ease.
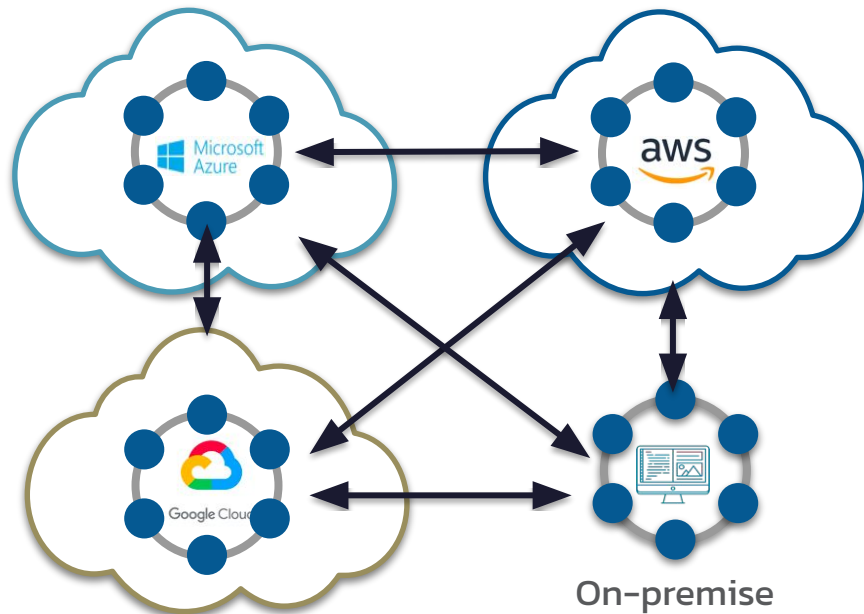
Vendor Independent

Cassandra doesn't belong to any of commercial vendors but controlled by a non-profit Open Source **Apache Software Foundation**, already familiar to you by *Hadoop*, *Spark*, *Kafka*, *Zookeeper*, *Maven* and many other projects.

Apache Cassandra
Built as a Technical Necessity

90% of Fortune 100 companies have adopted

High Availability
Always On
Every second of downtime translates into lost revenue

Linear Scalability
Hyper-Scalability
Millions of operations per day, hour, or second

Low Latency
Faster Pace
Every millisecond of latency has consequence

Global Distribution
Data Everywhere
On-premises, hybrid, multi-cloud, centralized, or edge

Thank you!