

# mm-ADT

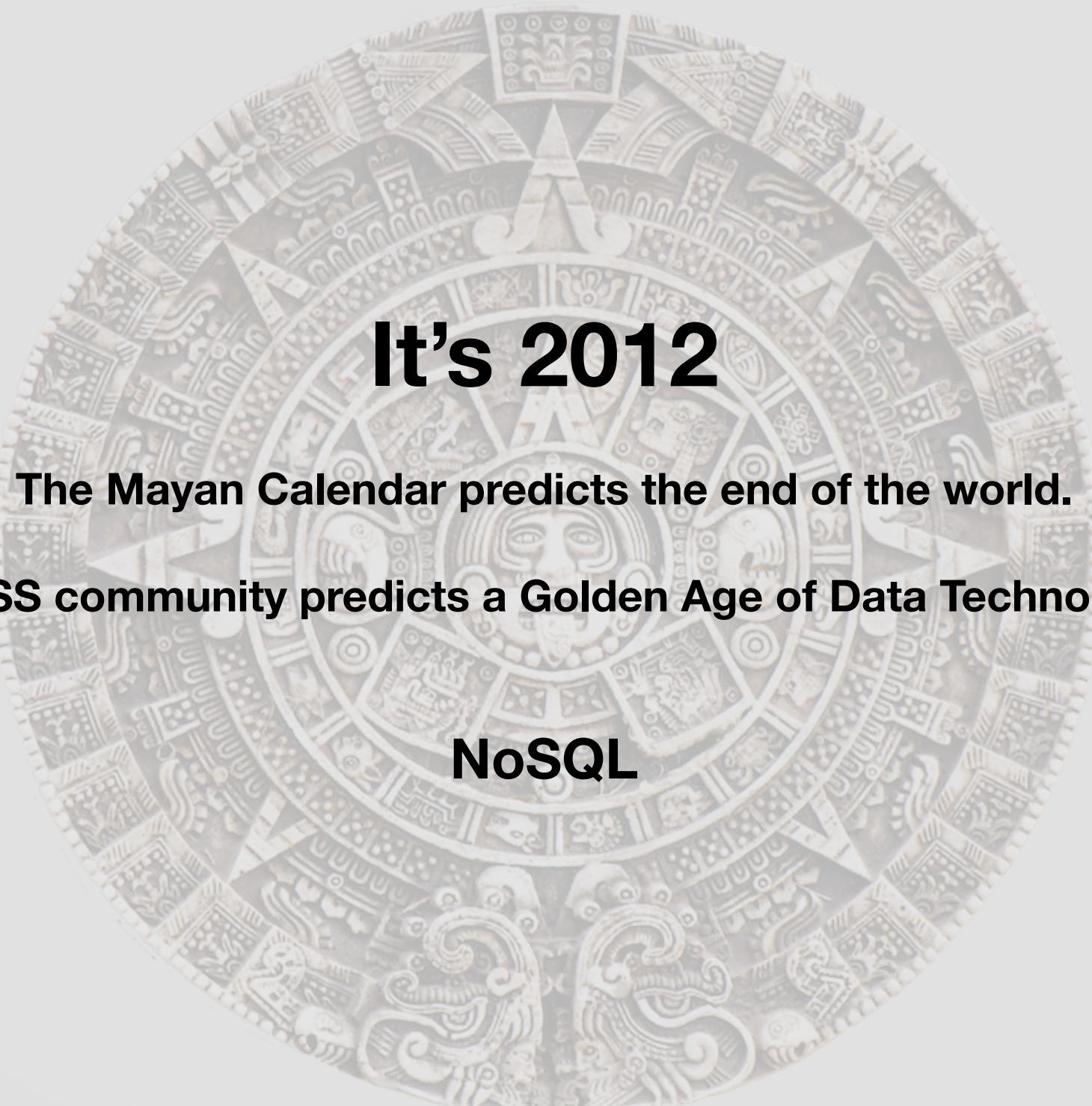
## A Virtual Machine/An Economic Machine



Dr. Marko A. Rodriguez



DataDays  
Austin TX, January 2020



# **It's 2012**

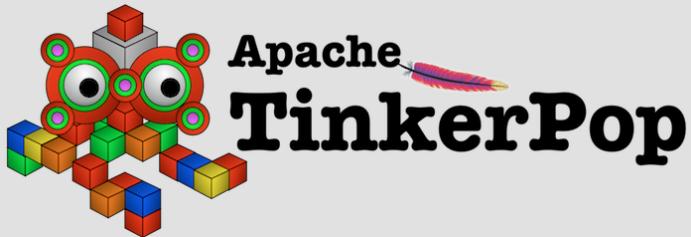
**The Mayan Calendar predicts the end of the world.**

**The OSS community predicts a Golden Age of Data Technologies.**

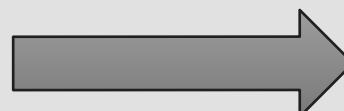
**NoSQL**



Document Database



Graph Database



Wide Column Store



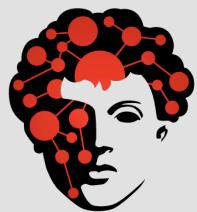
OLAP Processor



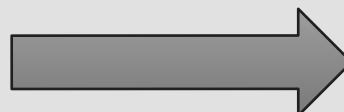
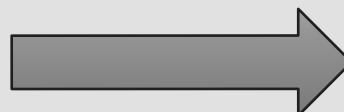
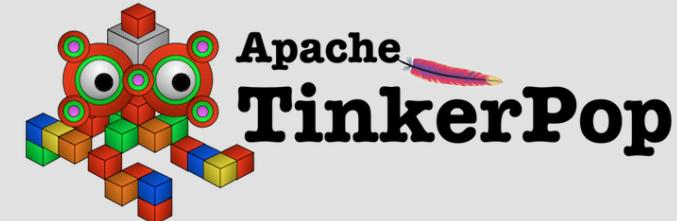
Index Store

Open Source Software  
**costs** money and time for  
research and development.

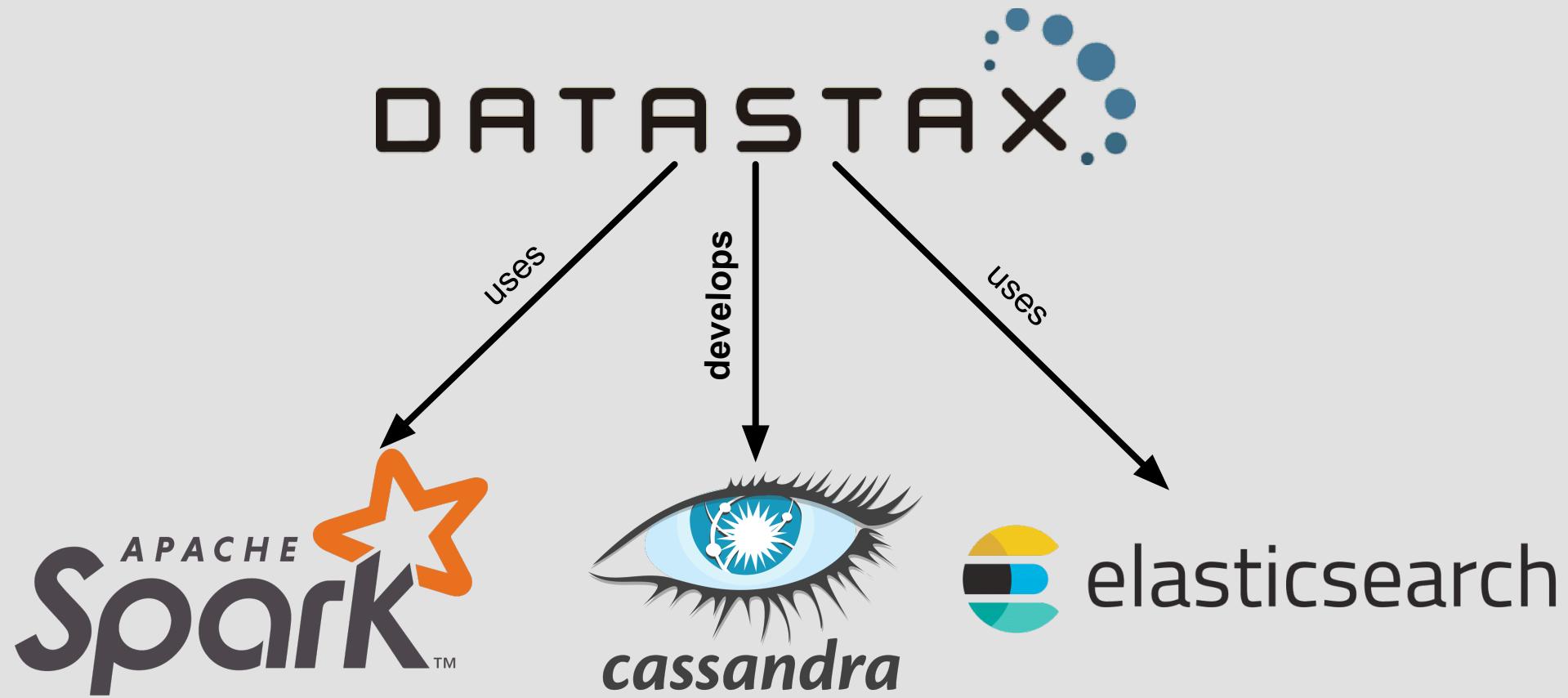
In order to provide the world  
free open source software many of our  
young innovators created companies to fund  
the development of their technology.



AURELIUS



# A Collaborative Ecosystem of Shared Technologies

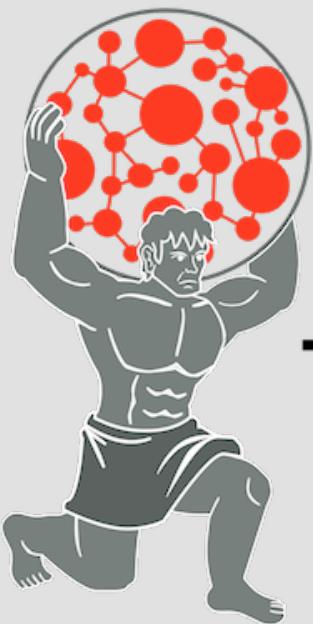


# A Collaborative Ecosystem of Shared Technologies

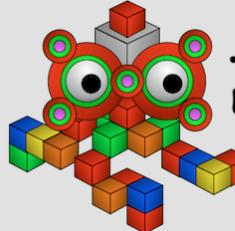


AURELIUS

uses



develops



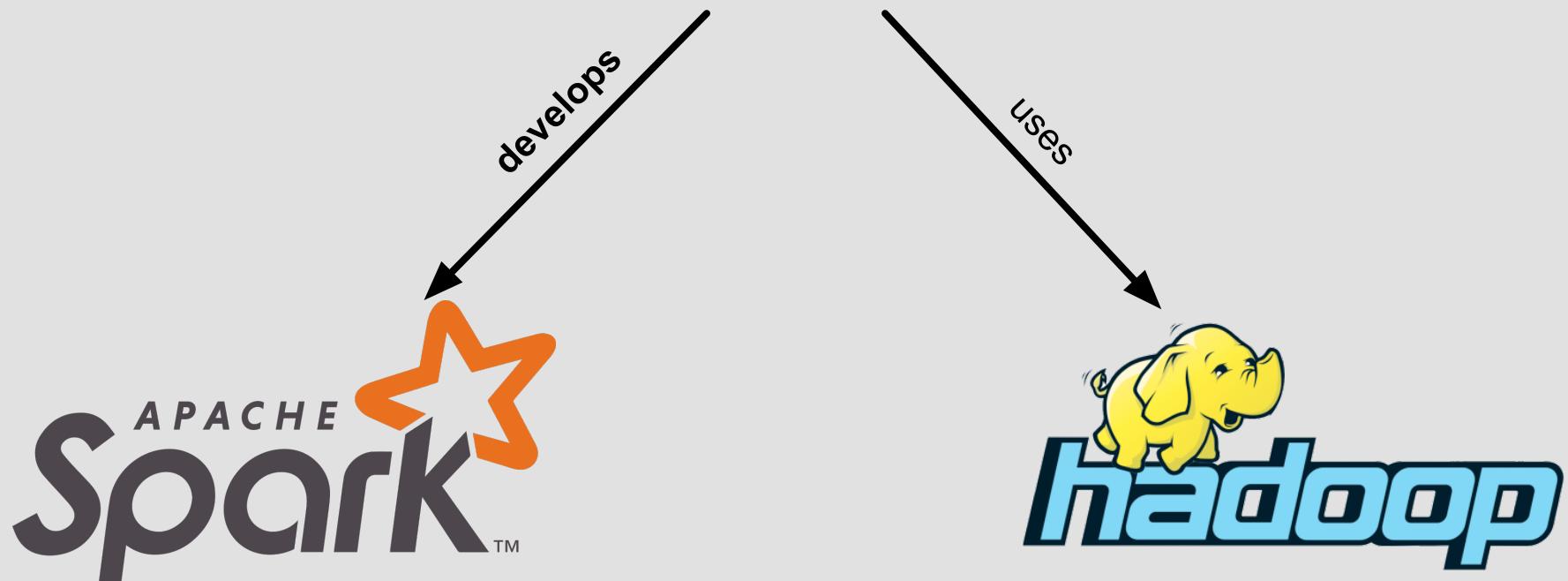
TinkerPop

develops

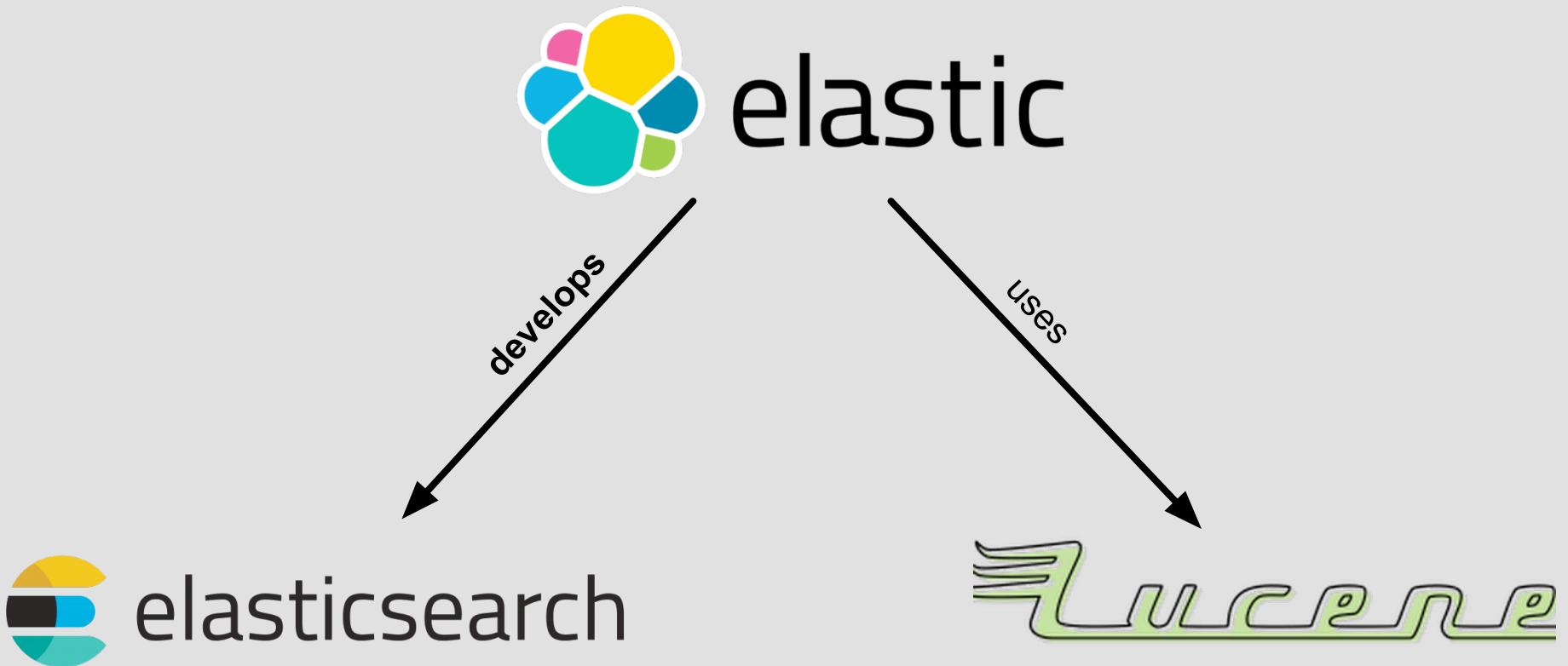
TITAN

# A Collaborative Ecosystem of Shared Technologies

 databricks



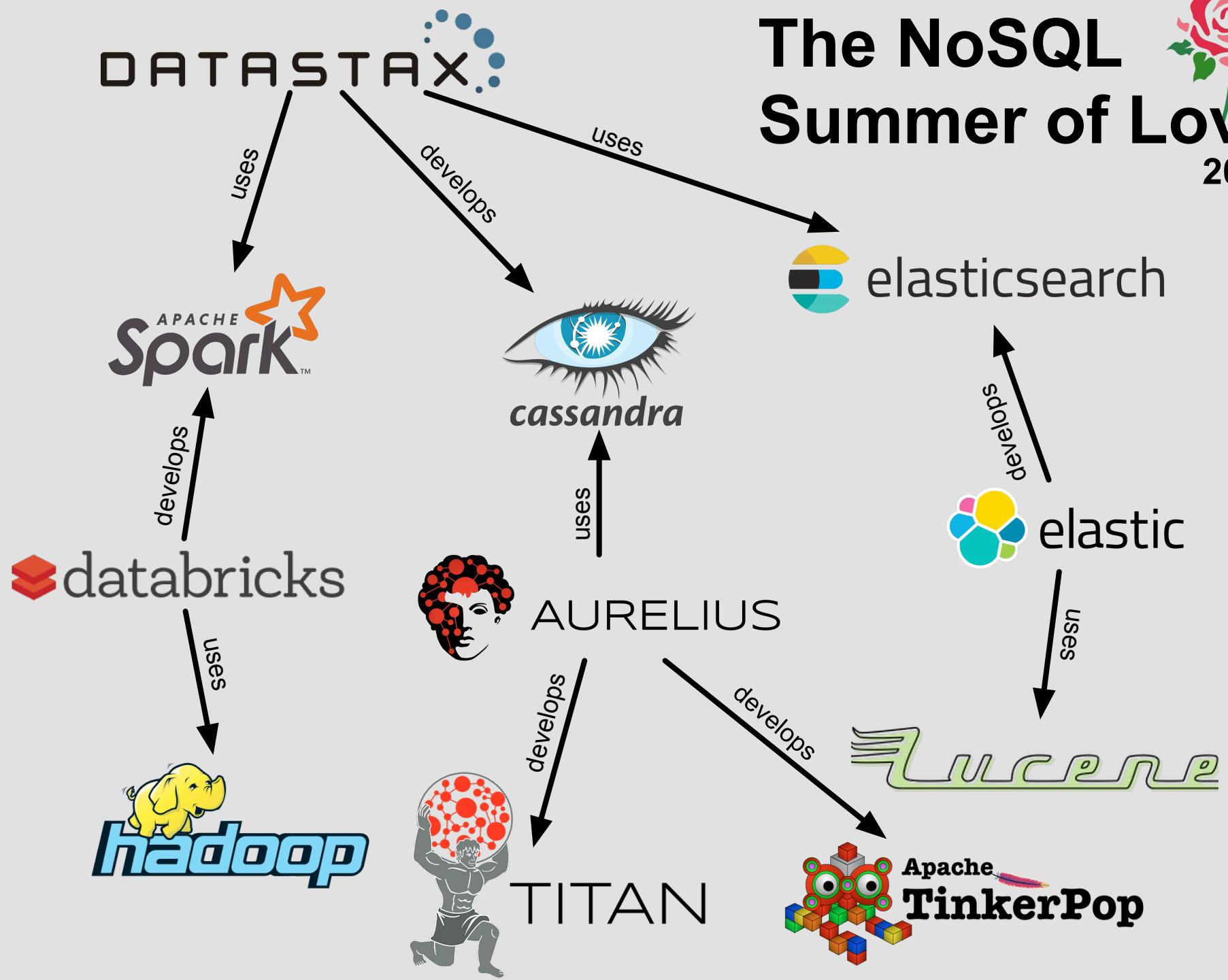
# A Collaborative Ecosystem of Shared Technologies





# The NoSQL Summer of Love

2012





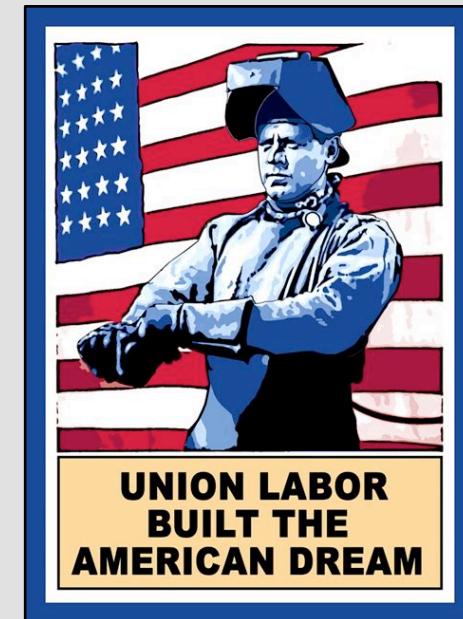
This generation of developers built the modern world's software data infrastructure.

I compare us to the bridge, dam, and interstate builders of the 1930's generation.

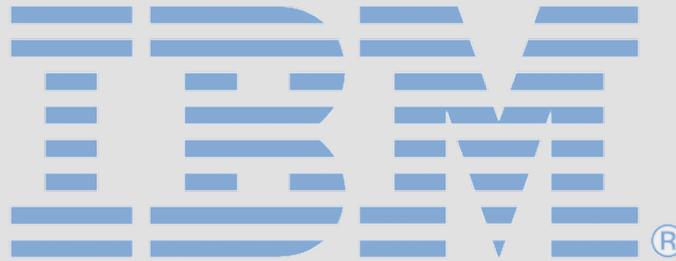
**We did something very good for this world.**



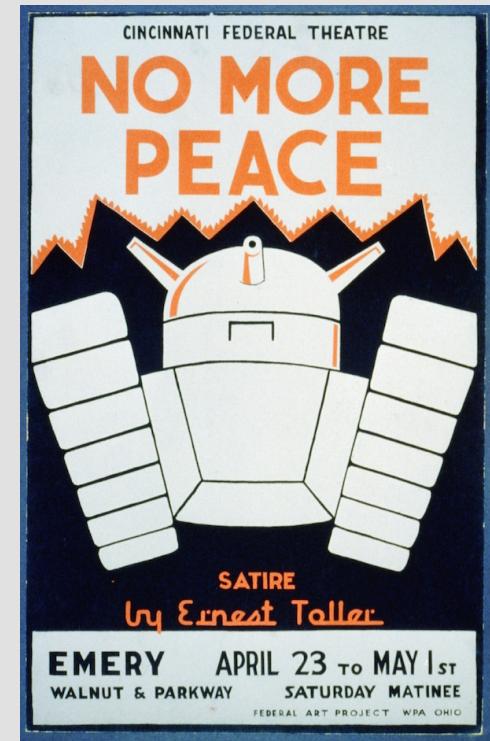
By making our software freely available, many institutions that would otherwise not pay for cutting-edge software have been able to advance their domain.



# And then came the parasites...



# Microsoft





## The Good

**Amazon's use of Apache TinkerPop for their internal operations is a perfect use of open source software.**

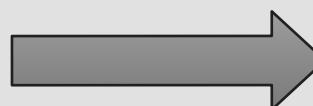
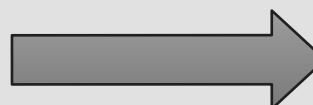
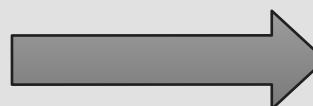
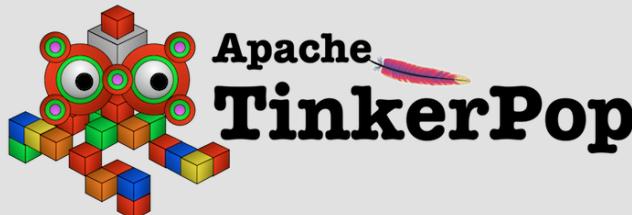
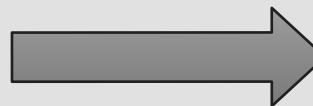
**Amazon** gets free software for their shipping logistics.

**Apache TinkerPop** is tested at extreme scale on a world-class problem.

**DataStax**, who were the primary developers at the time, is able to point (potential) customers to a success story.



# The Bad



## Amazon DocumentDB (with MongoDB compatibility)

Fast, scalable, highly available MongoDB-compatible database service

[Get started with Amazon DocumentDB](#)

## Amazon Neptune

Fast, reliable graph database built for the cloud

[Get started with Amazon Neptune](#)

## Amazon Managed Apache Cassandra Service

A scalable, highly available, and managed Apache Cassandra-compatible database service

[Get started with Amazon Managed Apache Cassandra Service \(preview\)](#)

## Amazon EMR

Easily Run and Scale Apache Spark, Hadoop, HBase, Presto, Hive, and other Big Data Frameworks

[Get started with Amazon EMR](#)

## Amazon Elasticsearch Service

Fully managed, scalable, and secure Elasticsearch service

[Get started](#)

# The Ugly

# Amazon “Acquires” OSS Companies for Free



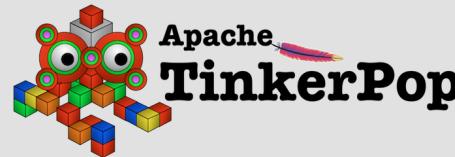
**Amazon DocumentDB (with MongoDB compatibility)**

Fast, scalable, highly available MongoDB-compatible database service

[Get started with Amazon DocumentDB](#)



AURELIUS



**Amazon Neptune**

Fast, reliable graph database built for the cloud

[Get started with Amazon Neptune](#)



**Amazon Managed Apache Cassandra Service**

A scalable, highly available, and managed Apache Cassandra-compatible database service

[Get started with Amazon Managed Apache Cassandra Service \(preview\)](#)



**Amazon EMR**

Easily Run and Scale Apache Spark, Hadoop, HBase, Presto, Hive, and other Big Data Frameworks

[Get started with Amazon EMR](#)



**Amazon Elasticsearch Service**

Fully managed, scalable, and secure Elasticsearch service

[Get started](#)

“Does Amazon contribute to these projects (benefitting users and the creators)?”  
Let’s see what the statistics say...



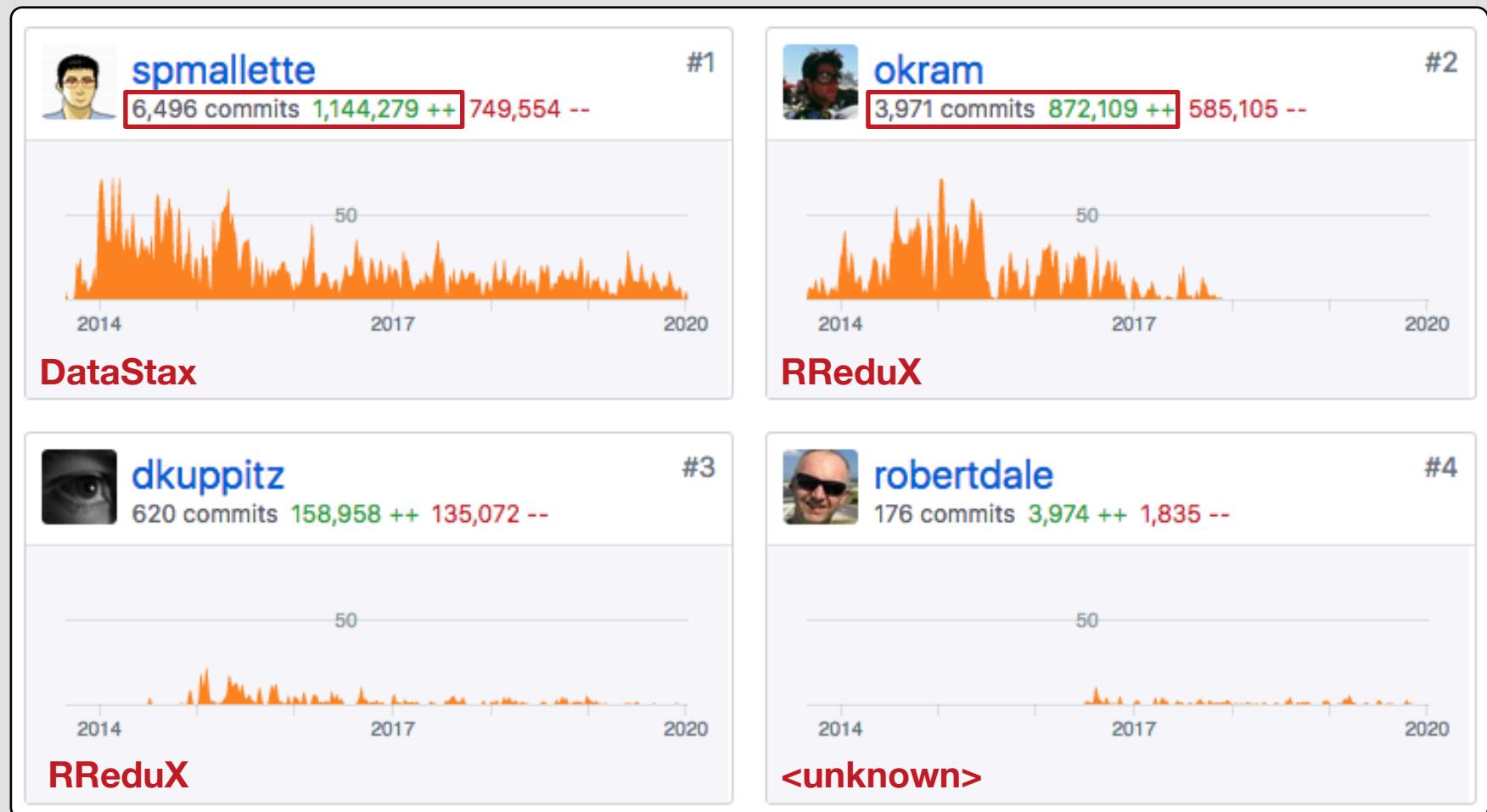
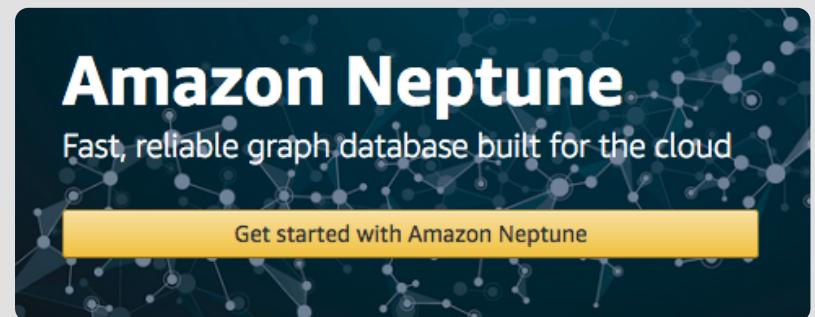
## Amazon DocumentDB (with MongoDB compatibility)

Fast, scalable, highly available MongoDB-compatible database service

[Get started with Amazon DocumentDB](#)



The top contributors do not work for Amazon



The top contributors do not work for Amazon



# Cassandra



## Amazon Managed Apache Cassandra Service

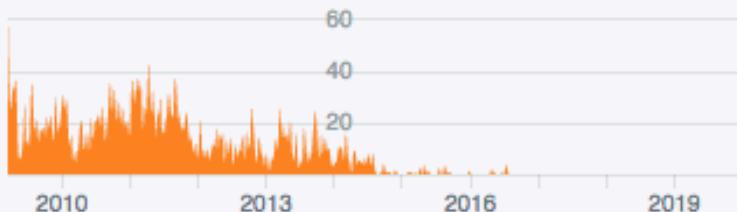
A scalable, highly available, and managed Apache Cassandra-compatible database service

Get started with Amazon Managed Apache Cassandra Service  
(preview)



jbellis

4,526 commits 547,981 ++ 459,184 --



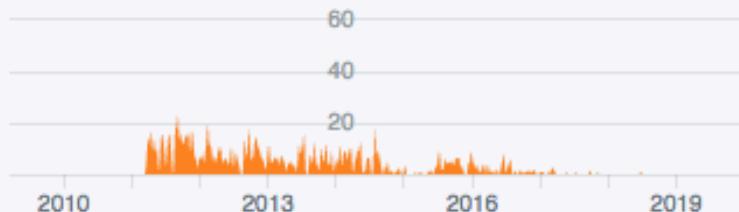
DataStax

#1



pcmanus

1,617 commits 254,428 ++ 294,714 --



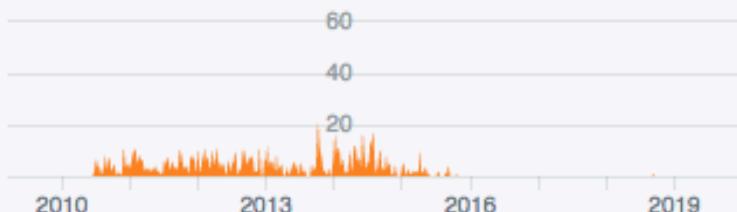
DataStax

#2



driftx

1,202 commits 83,533 ++ 56,204 --



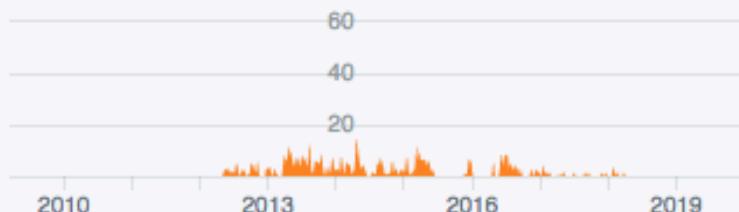
DataStax

#3



mebigfatguy

734 commits 18,152 ++ 16,863 --



Oracle

#4

The top contributors do not work for Amazon



## Amazon EMR

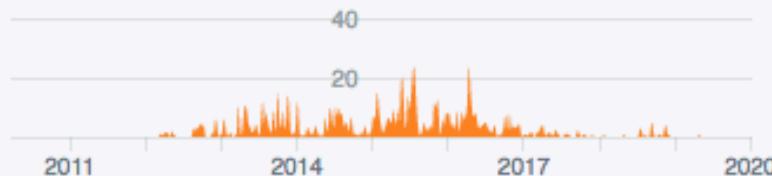
Easily Run and Scale Apache Spark,  
Hadoop, HBase, Presto, Hive, and other Big  
Data Frameworks

[Get started with Amazon EMR](#)



**rxin**

1,179 commits 207,022 ++ 229,472 --



DataBricks

#1



**mateiz**

830 commits 1,628,795 ++ 1,150,896 --



DataBricks

#2



**cloud-fan**

741 commits 111,841 ++ 87,572 --



DataBricks

#3



**pwendell**

654 commits 31,148 ++ 24,163 --



DataBricks

#4

The top contributors do not work for Amazon

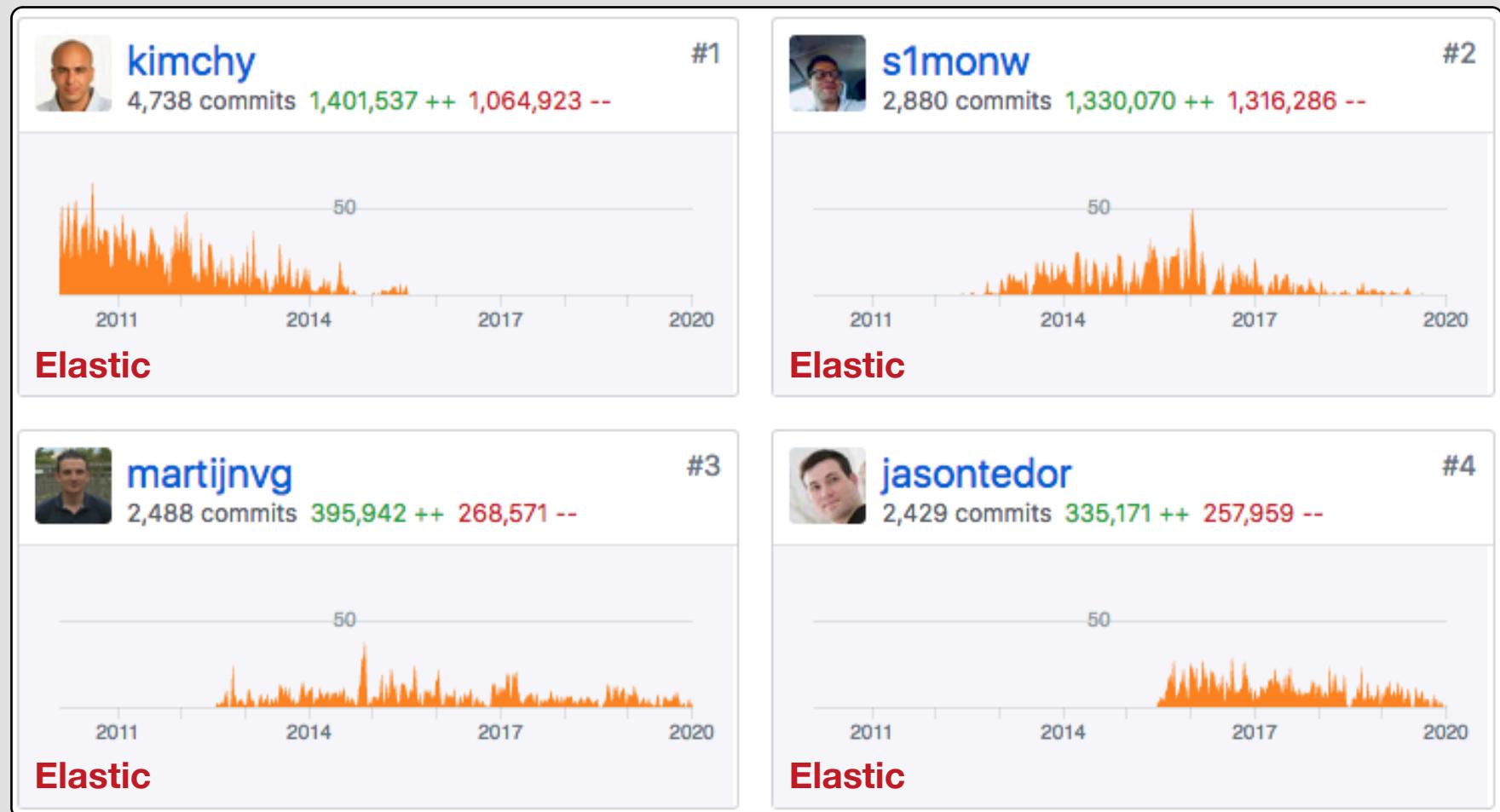


elasticsearch →

## Amazon Elasticsearch Service

Fully managed, scalable, and secure Elasticsearch service

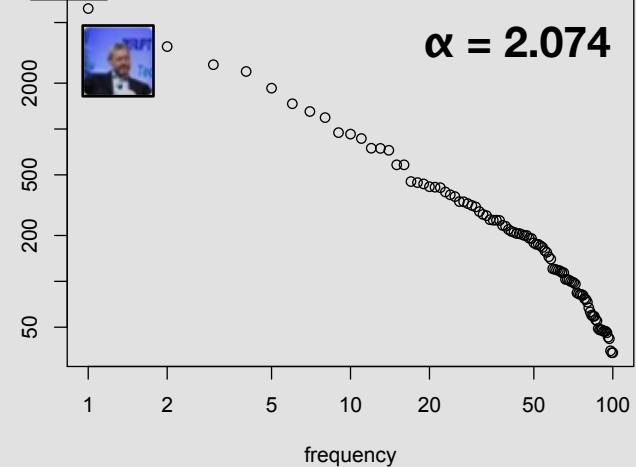
Get started



The top contributors do not work for Amazon

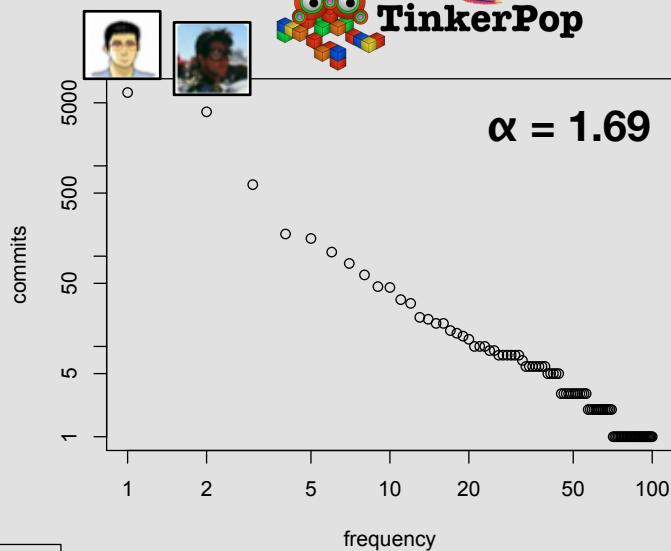


mongoDB®

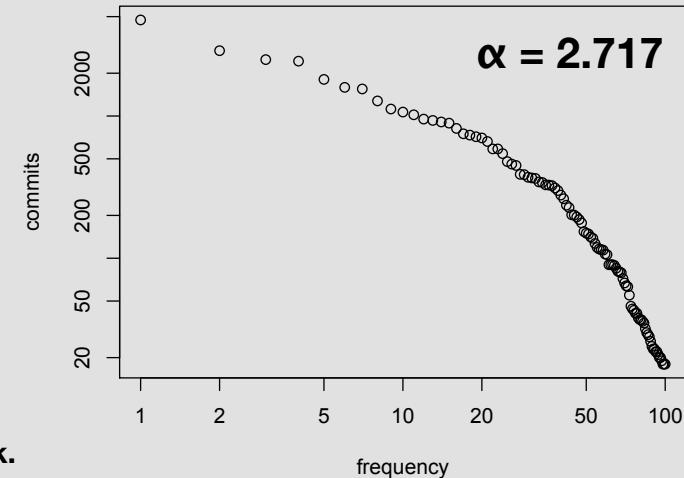
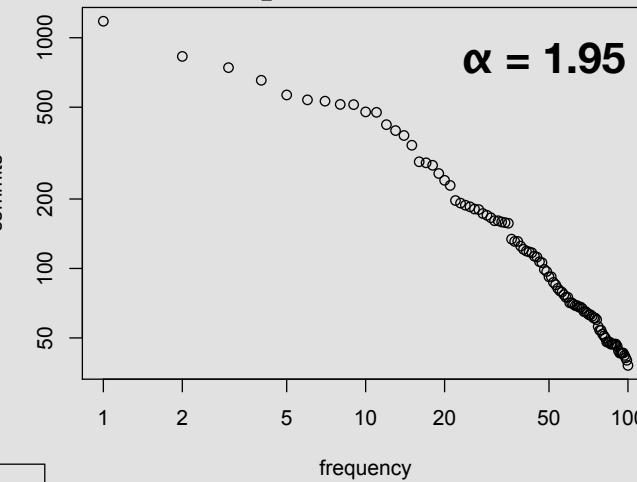
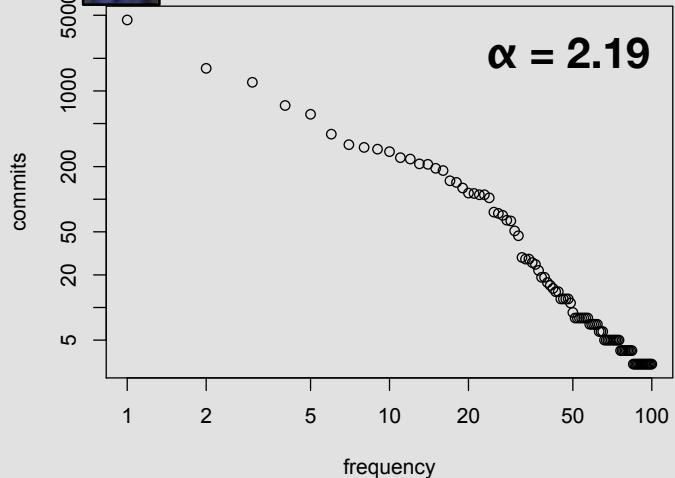


log/log-scale

APACHE  
Spark™



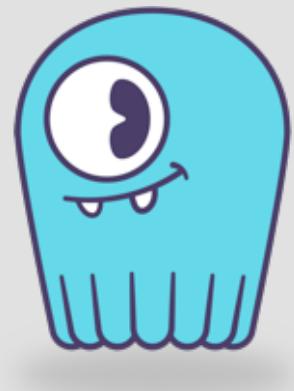
Cassandra



Smaller  $\alpha$  means  
fewer people do most of the work.

# SCYLLA.

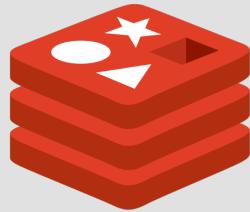
Yes, AWS **promises** to contribute back to Cassandra. The question is whether those contributions be more than self-serving ones ... there are **no incentives** to make real progress. To Amazon's credit, they already made a single major contribution in the form of the Dynamo paper, 12 years ago.



<https://www.scylladb.com/2019/12/04/managed-cassandra-on-aws-our-take/>



**Let's honor some other  
soldiers of the trade.**



# redis



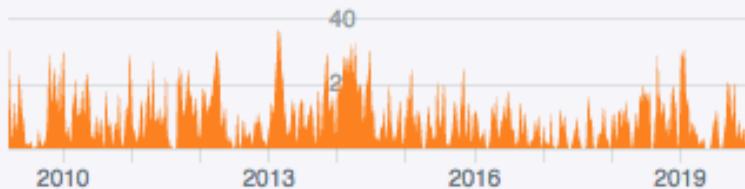
## Amazon ElastiCache for Redis

Redis compatible in-memory data store built for the cloud. Power real-time applications with sub-millisecond latency.

[Start with Amazon ElastiCache for Redis for free](#)

**antirez**

5,487 commits 482,042 ++ 266,027 --

**RedisLabs**

#1

**piatern**

496 commits 30,857 ++ 13,769 --

**RedisLabs**

#2

**mattsta**

151 commits 11,439 ++ 4,004 --

**CarrierDB**

#3

**soloestoy**

136 commits 1,419 ++ 591 --

**Alibaba**

#4

**The top contributors do not work for Amazon**



## Amazon EMR

Easily Run and Scale Apache Spark,  
Hadoop, HBase, Presto, Hive, and other Big  
Data Frameworks

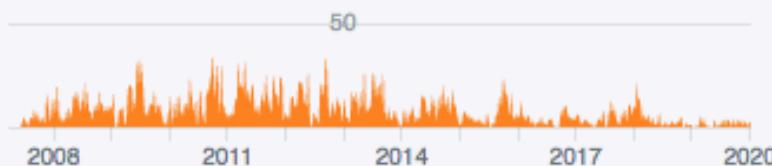
[Get started with Amazon EMR](#)



**saintstack**

5,305 commits 4,280,368 ++ 4,188,258 --

#1



Cloudera



**apurtell**

835 commits 194,362 ++ 92,678 --

#2



Salesforce



**tedyu**

795 commits 112,876 ++ 48,817 --

#3



<unknown>



**Apache9**

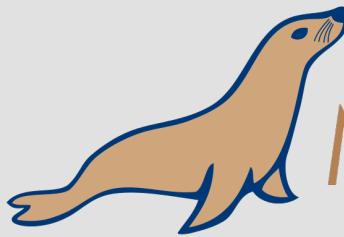
708 commits 237,054 ++ 208,183 --

#4



Xiaomi

The top contributors do not work for Amazon



# MariaDB



## Amazon Relational Database Service (RDS)

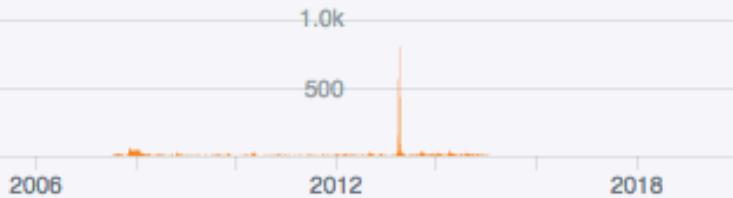
Set up, operate, and scale a relational database in the cloud with just a few clicks.

[Get started with Amazon RDS](#)



[prohaska](#)

4,612 commits 8,105,185 ++ 5,533,255 --



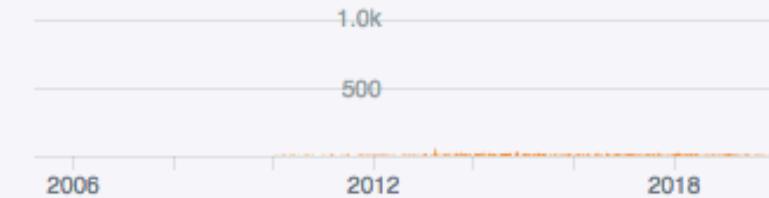
Percona

#1



[vuvova](#)

4,027 commits 8,973,386 ++ 6,548,651 --



MariaDB

#2



[fizzfaldt](#)

2,180 commits 279,052 ++ 182,962 --



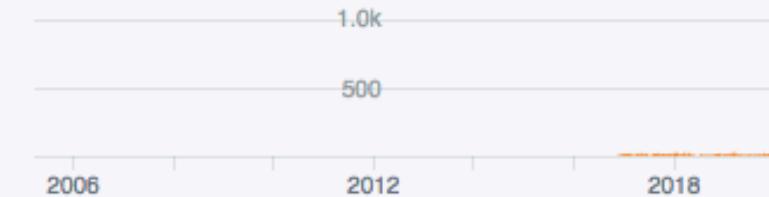
<unknown>

#3



[dr-m](#)

2,041 commits 189,273 ++ 590,054 --



MariaDB

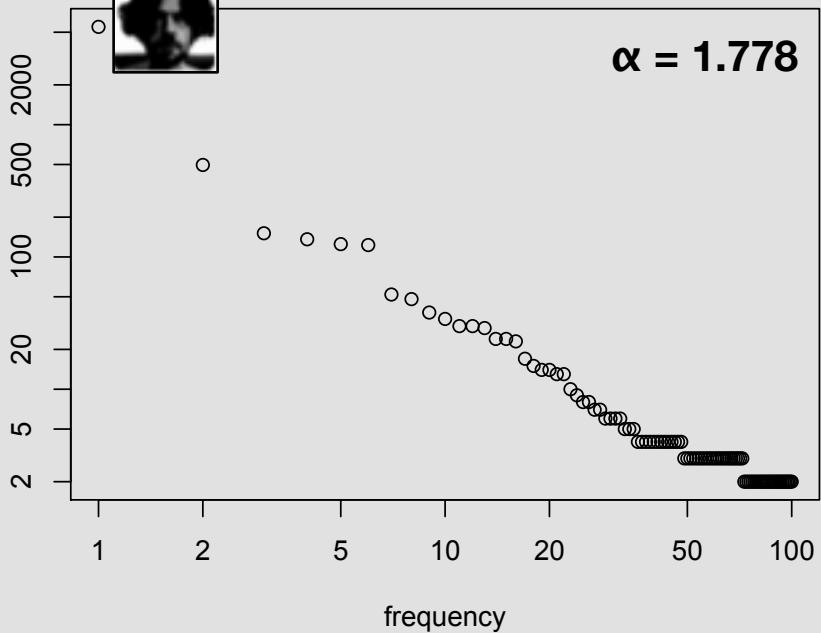
#4

The top contributors do not work for Amazon

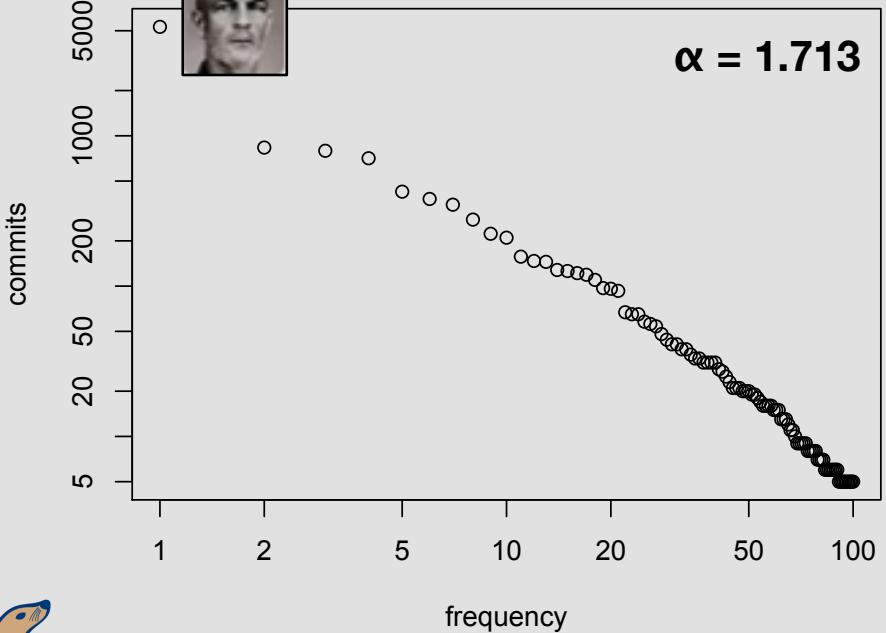


redis

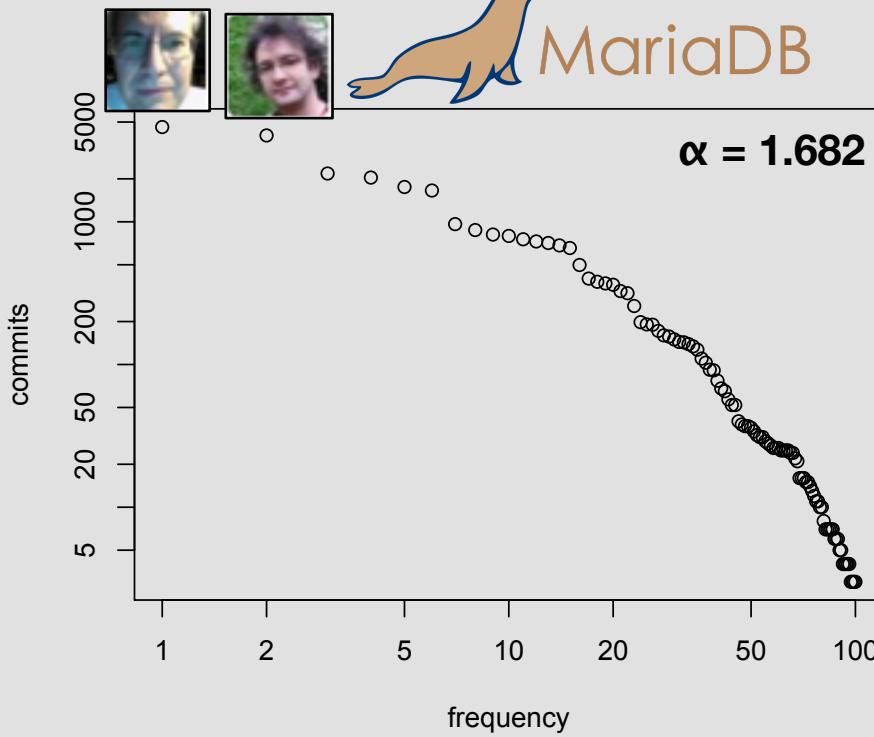
commits



APACHE  
HBASE



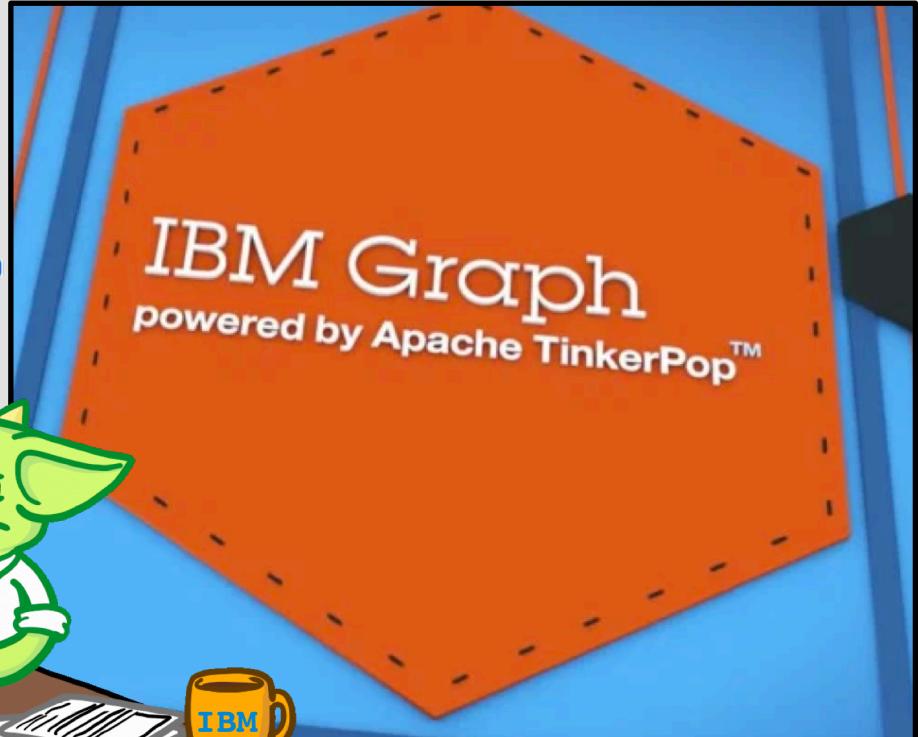
MariaDB





**It used to be...**





July 2016

@tachyeonz : IBM today took the wraps off a new cloud-based graph service based on open source [Titan and Apache TinkerPop](#) technologies. Called IBM Graph, the new service is suitable for production graph workloads of any size.

**Contribute nothing, take everything.**

Now it's the...



Software start-ups have a phrase for what Amazon is doing to them: 'strip-mining' them of their innovations.

## AWS guns for MongoDB with DocumentDB service offering full compatibility

AWS hawks database service compatible with version 3.6 of MongoDB



Amazon Web Services (AWS) has taken aim at MongoDB with the launch of DocumentDB, a database service that, it claims, is fully compatible with Register...

ENTERPRISE

## Amazon steps up its open-source game, and Elastic stock falls as a result

PUBLISHED TUE, MAR 12 2019 1:42 PM EDT



Jordan Novet  
@JORDANNOVET

The New York Times



DESIGN > OPEN SOURCE Expo hall at AWS re:Invent 2018

### Confluent Creates New 'Open Source' License to Stop Cloud Poaching

Expo hall at AWS re:Invent 2018

In his response, Gutmans argued that Amazon is not deliberately copying open-source software in order to profit from others' labor, but is instead just giving its customers what they want.

Mutually Exclusive?

# Predictions

Amazon will continue to release products based on Apache Software Foundation technology.

When Amazon has completed their exploitation and little competition exists they will use legal techniques to dismantle the Apache Software Foundation.

The Apache Software Foundation had \$1 million in their coffers for legal related issues in 2017.  
(prior to Amazon's 2018 \$1 million donation)

Open source developers will use less liberal licenses.  
Innovation in our industry will come to a standstill.

**DataDays will be subsumed by re:Invent.**

# Predictions

All popular open source technologies will be taken by Amazon.  
All popular closed source technologies will be cloned by Amazon.

Small agile data technology companies will struggle.

The majority of jobs in data technology  
systems engineering will be at Amazon.

Salaries in data technology will drop precipitously.

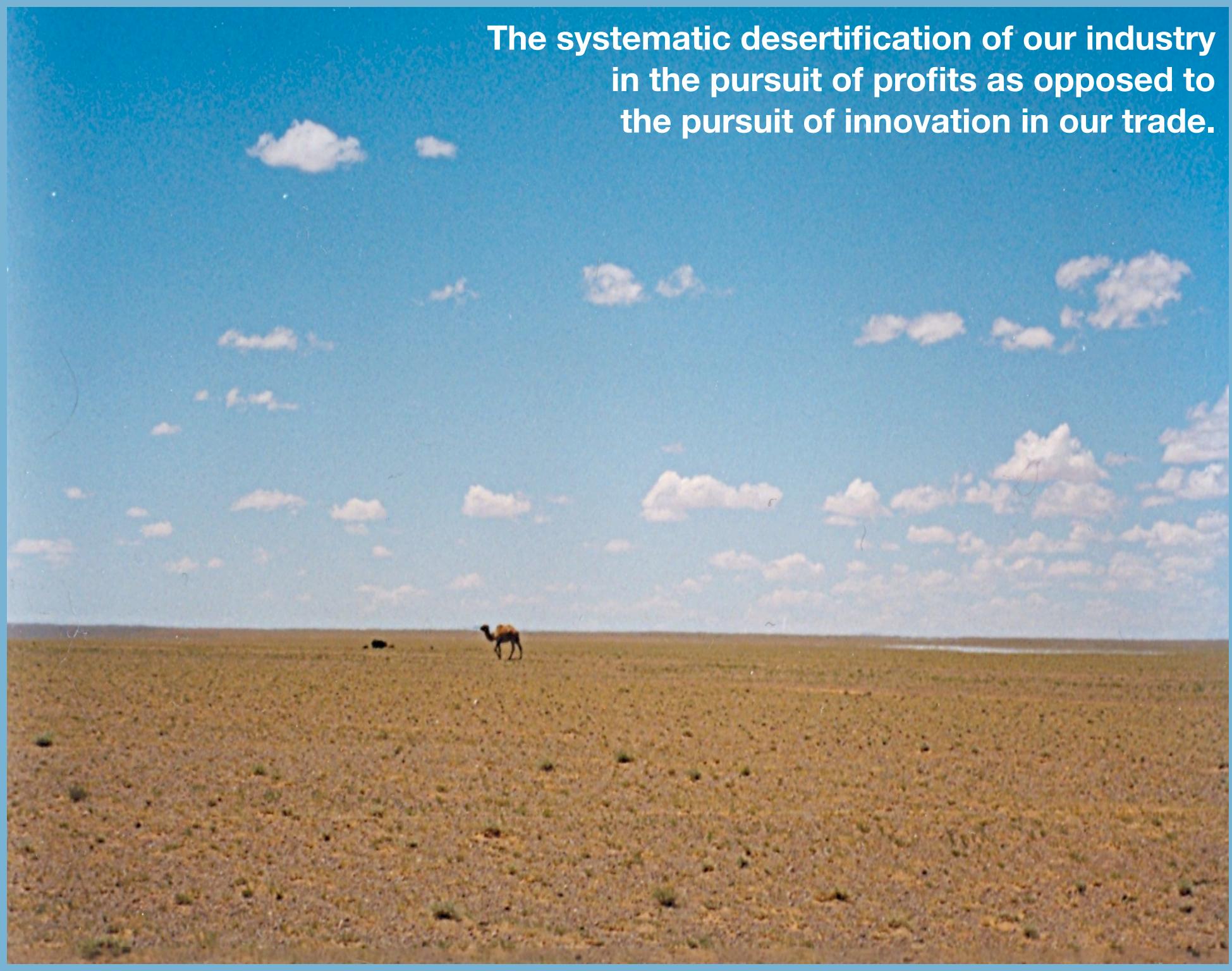
Amazon employees will try to create labor unions.

Amazon will squash any unionization efforts.

Their employees will be stuck in legal quagmires.

U.S. tax payers will be left to foot the bill on the most expensive  
anti-trust suit the U.S. government has ever endured.

**The systematic desertification of our industry  
in the pursuit of profits as opposed to  
the pursuit of innovation in our trade.**

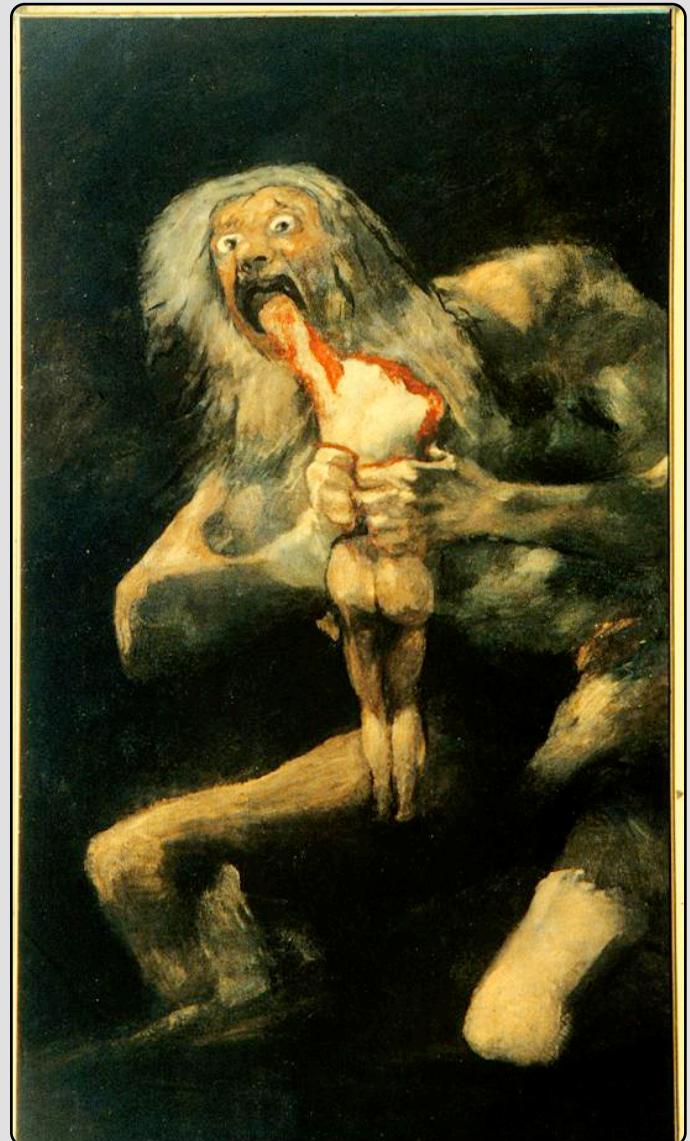


# A NoSQL Poem for Amazon/IBM/Microsoft

Oh father, love of mine,  
you have built such wonderful things  
with all your time.

In honor of your greatness,  
we followed behind and now of age,  
we give not what was mine.

May our software make your machines  
sing to the heavens for all time.





# The Virtual Machine

# Modern Data System Components

## Storage



## Processor



## Language





# Synthetic Data Systems

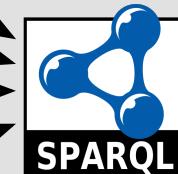
## Storage

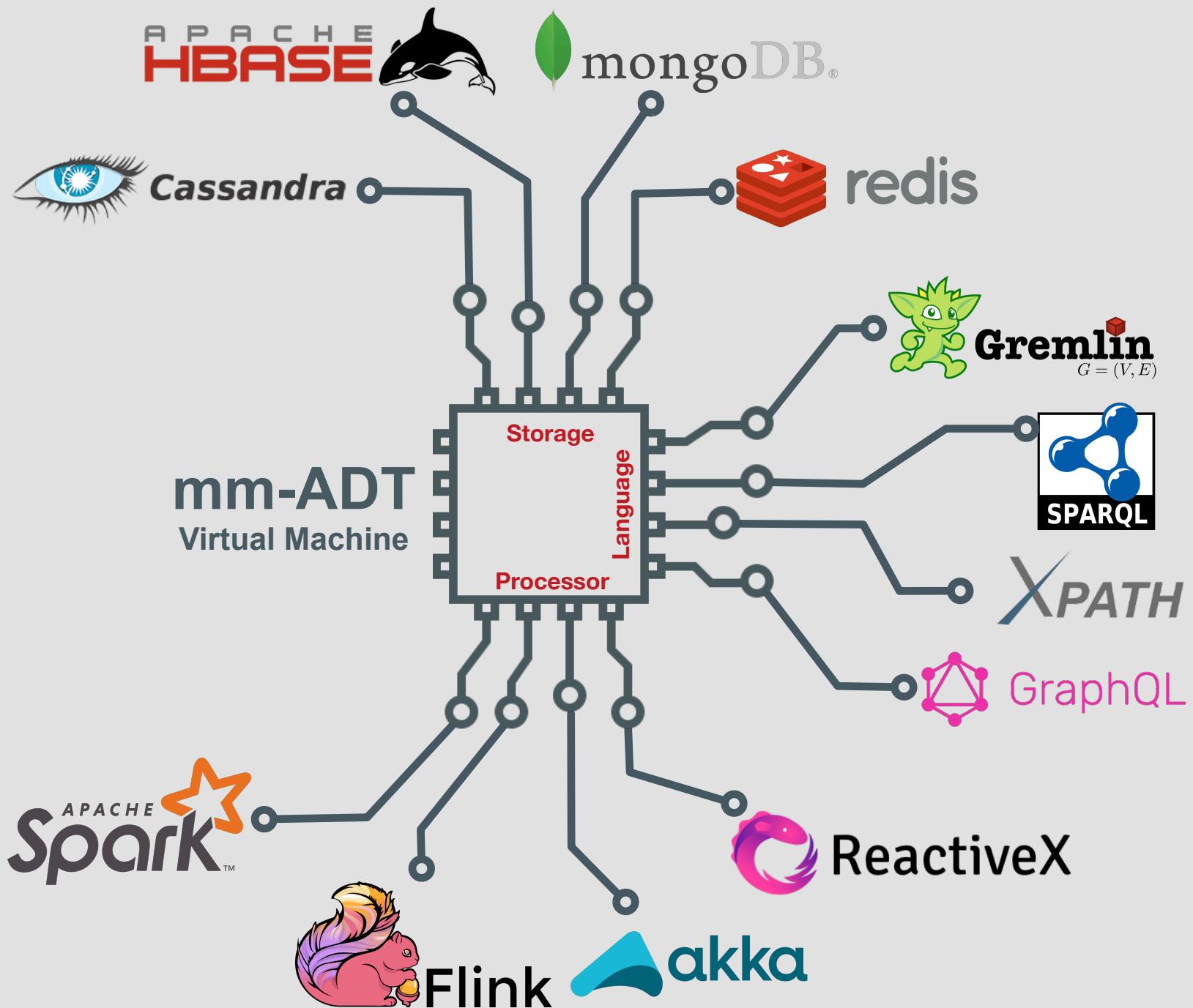


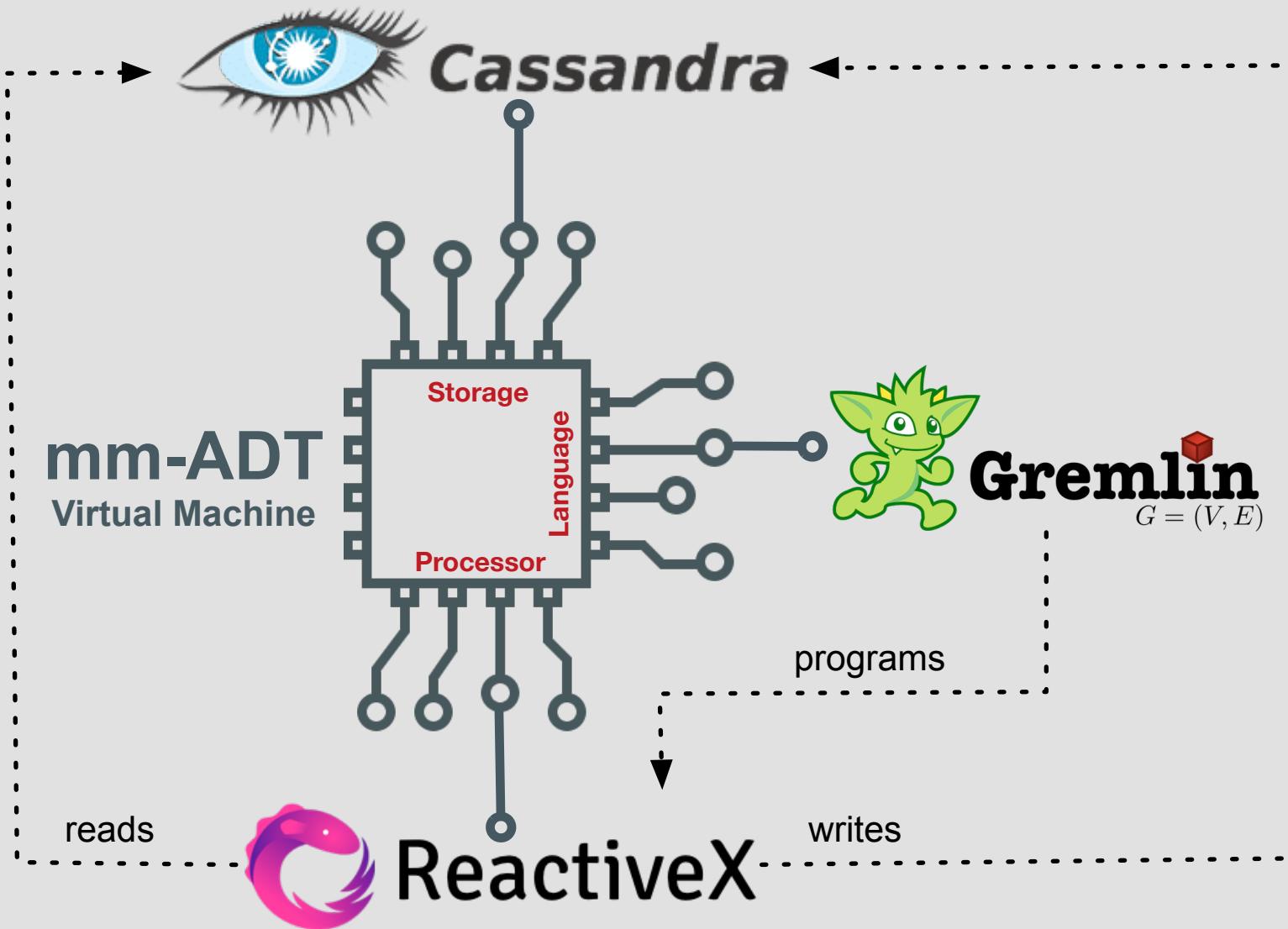
## Processor



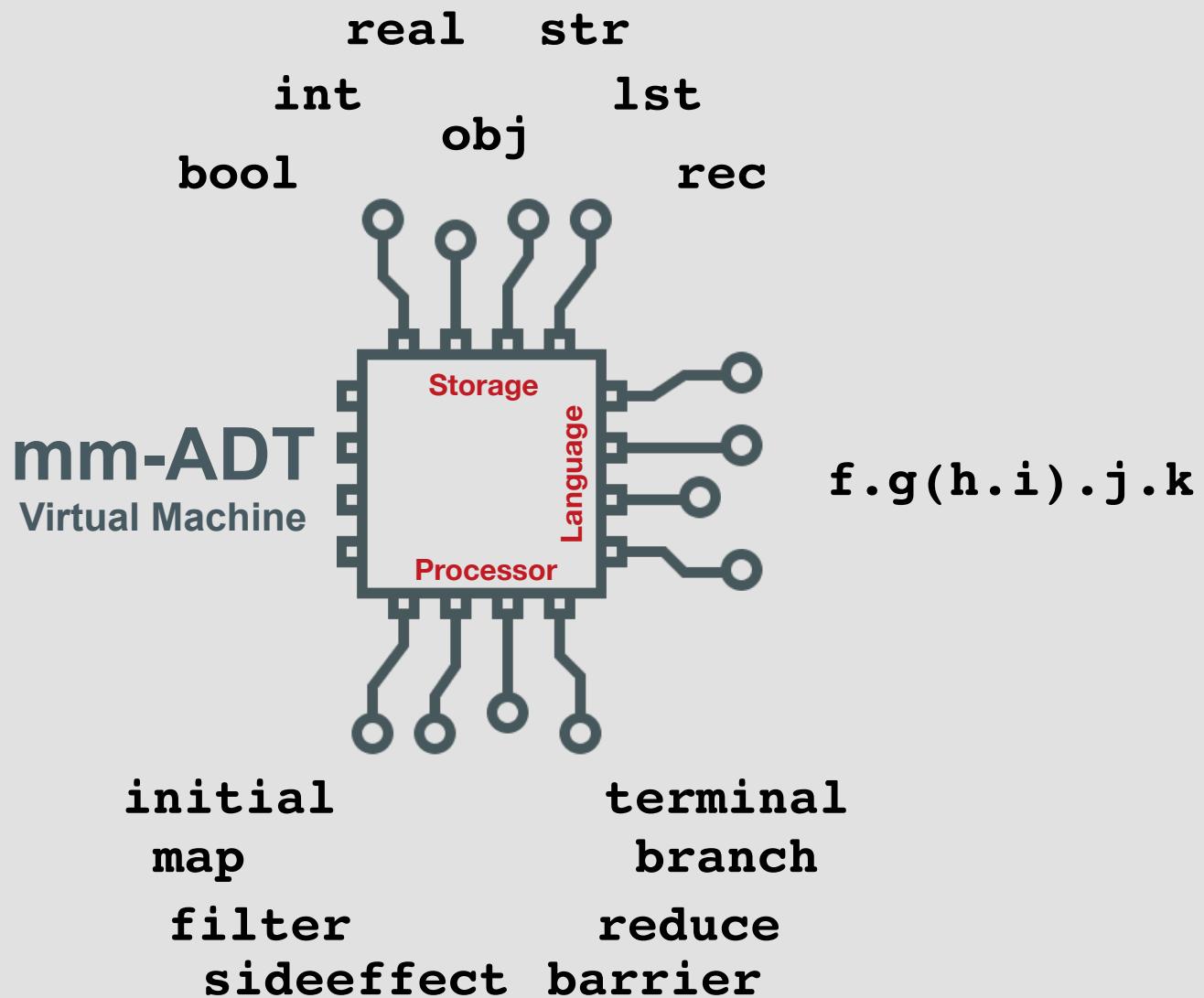
## Language







# Model Languages w/ Instructions

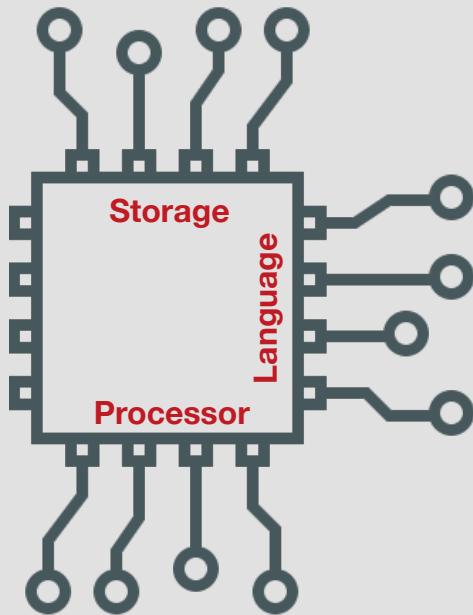


# Model Processors w/ Operators

## storage exposes graph model

```
graph -> [graph --> vertex{*}
           |vertex -> ['id' :int,
                         'inE' :edge{*},
                         'outE':edge{*}]
           |edge    -> ['label':str,
                         'outV' :vertex,
                         'inV'  :vertex]]
```

**mm-ADT**  
Virtual Machine



```
initial      [=graph]
filter       [is,[get,'id'][eq,1]]
map         [get,'outE']
filter       [is,[get,'label'][eq,'knows']]
map         [get,'inV']
map         [get,'id']
```

processor supports requisite operators

pattern matching  
"if lhs, then rhs"

[ a->b | c->d | e->f ]



`g.v(1).out('knows').id()`



```
int{*} <=[=graph]
[is,[get,'id'][eq,1]]
[get,'outE']
[is,[get,'label'][eq,'knows']]
[get,'inV']
[get,'id']
```

language compiles to the graph model

mm-ADT is about creating **models** in terms of other **models**.

The base model is always **mm**.

The ‘domain of discourse’ of the mm-ADT VM.

Commonly used models are “standardize” by mm-ADT.

**mm** : multi-model

**kv** : key/value

**rdb**: relational

**pg** : property graph

**wc** : wide column

**doc**: documents

**rdf**: resource description framework

Storage providers expose their system via these models.

Language providers write compilers to these models.

## A user's schema is a model...

```
social -> [ person -> [ 'ssn'      :int,  
                           'friends':person{*} ] ]
```

# A user's schema is a model...

```
social -> [ person -> [ 'ssn'      :int,  
                           'friends':person{*} ] ]
```

...that can be embedded in other models.

**social model => relational model => mm-ADT model**

```
social->[person -> [ 'ssn'      :int~x,  
                           'friends':person{*} ] ]
```

**social model => graph model => mm-ADT model**

```
social->[person -> [ 'ssn'      :int~x,  
                           'friends':person{*} ] ]
```

# A user's schema is a model...

```
social -> [ person -> [ 'ssn'      :int,
                           'friends':person{*} ] ]
```

**...that can be embedded in other models.**

**social model => relational model => mm-ADT model**

```
social->[person -> [ 'ssn'      :int~x,
                       'friends':person{*} <=(row{*} <=[=rdb][get,'people'][is,[get,'id']
                                                               [eq,[=rdb][get,'links']
                                                               [is,[get,'source'][eq,x]]
                                                               [get,'sink']]])]
          <=(row <=[=rdb][get,'people'][is,[get,'id'][eq,x]]))
```

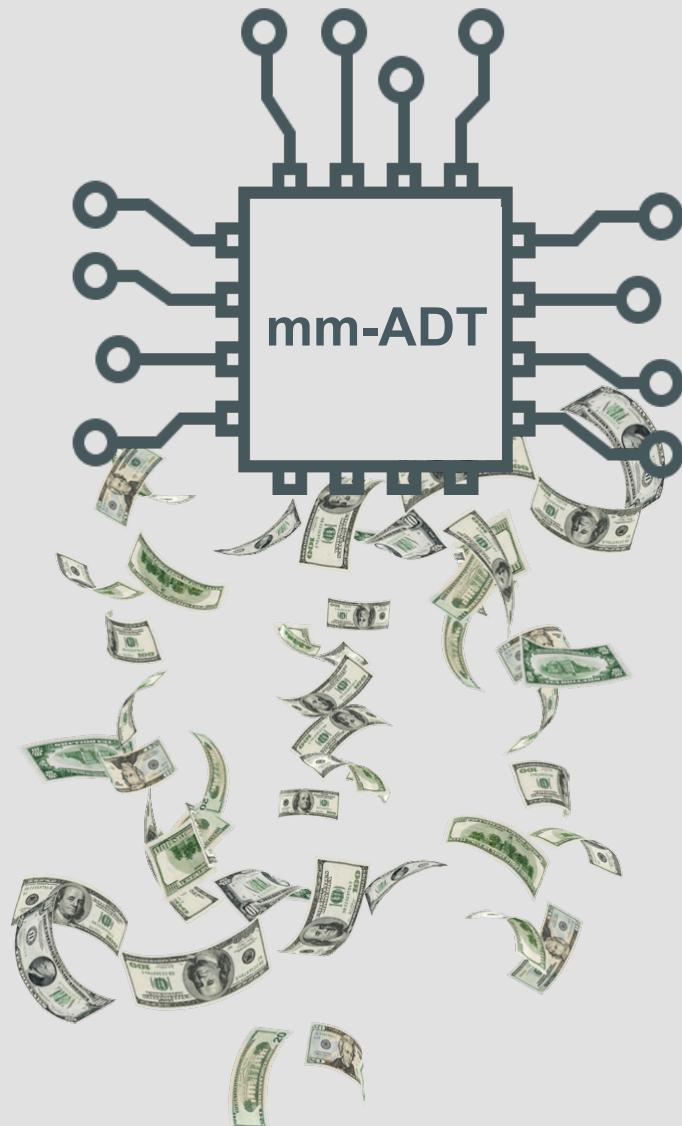
**social model => graph model => mm-ADT model**

```
social->[person -> [ 'ssn'      :int~x,
                       'friends':person{*} <=(vertex{*} <=[=graph][is,[get,'id'][eq,x]]
                                                               [get,'outE']
                                                               [is,[get,'label'][eq,'knows'])
                                                               [get,'inv']]])
          <=(vertex    <=[=graph][is,[get,'id'][eq,x]]))
```



# The Economic Machine

**The mm-ADT virtual machine**  
serves as a universal integrator for the technologies of our industry.

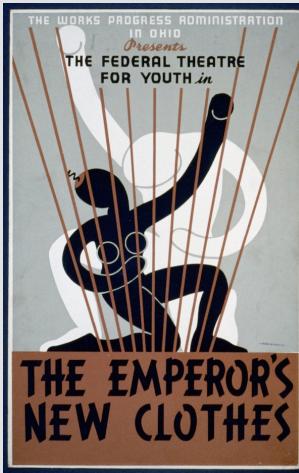


**The mm-ADT economic machine**  
serves as a universal integrator for the creators in our industry.



## Independent Open Source Developers Suffer Economically

Independent open source developers spend their time developing software that provides no financial compensation and, paradoxically, as the popularity of their project increases, success can lead to burnout and financial ruin.



## Entrepreneurial Open Source Developers Lack a Competitive Edge

Entrepreneurial open source developers may form a company around their open source project. Unfortunately, such companies are easily obviated in the market because the core product is freely licensed and competing firms can offer the same product without having to invest in the research and development effort.



## Corporate Open Source Developers Sacrifice Autonomy

Corporate open source developers are in the problematic situation where their work is dependent on their sponsoring organization who also controls the direction and pace of development as well as the interoperability and inclusion of any novel innovations within the project.

# The OSS Developer

...is an artist, not an employee.

...needs an agent, not a employer.

...seeks to mentor, not to manage.

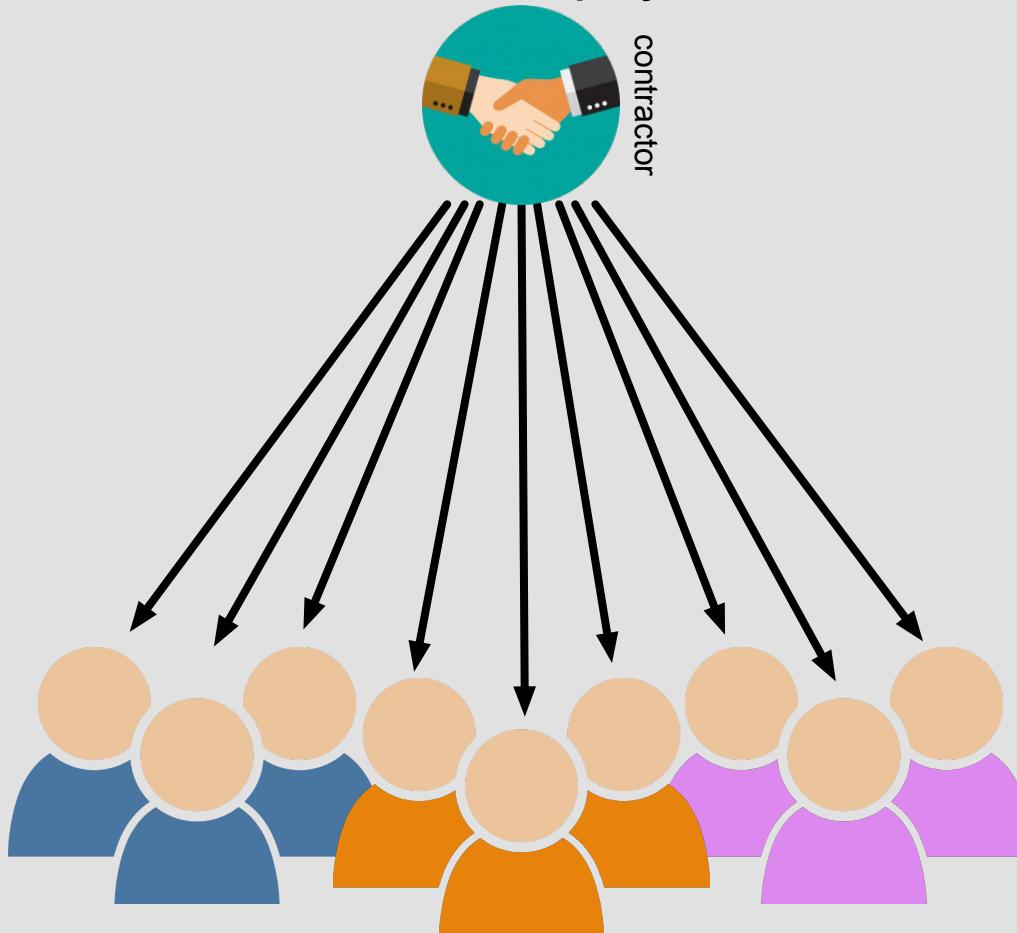
...should be patronized, not invested in.



...needs to be financially compensated for their work.

# A Royalty-Based Open Source Model

Think  
Etsy+Apache



Data Technology Developers

# A Royalty-Based Open Source Model

An mm-ADT component  
is created by an  
open source development team

**design:** Specify the component mm-ADT model.

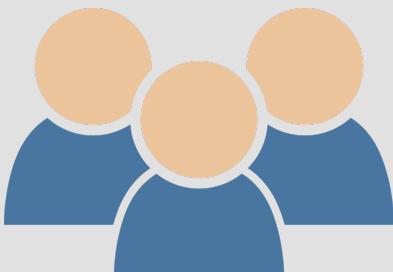
**develop:** Engineer the component binding.

**price:** Decide the cost for licensing the component.

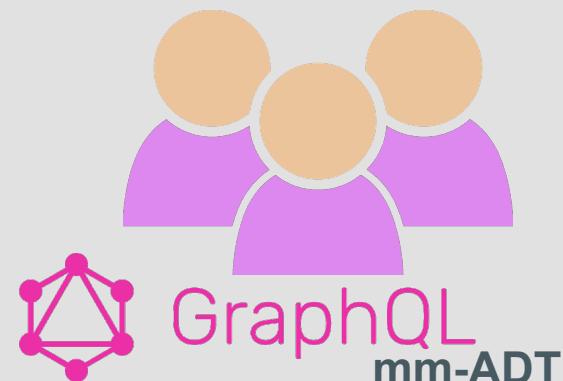
**promote:** Write blog posts, academic articles, tutorials.

**Maintain:** Fix bugs and update over mm-ADT VM versioning.

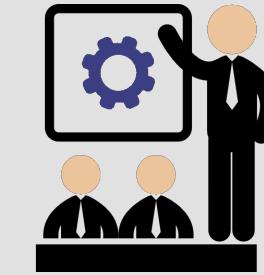
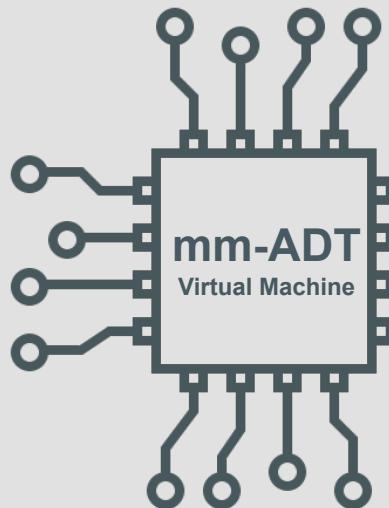
**support:** Mailing list Q&A, documentation, issue tracker.



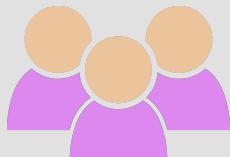
**Cassandra**  
mm-ADT



# A Royalty-Based Open Source Model



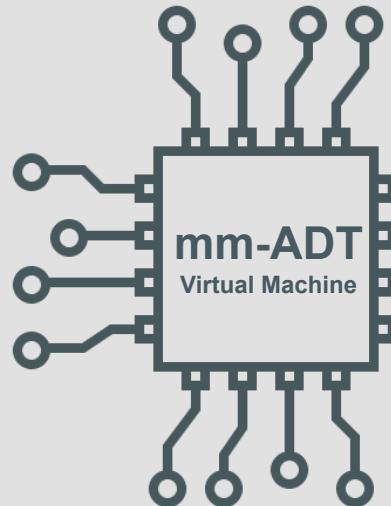
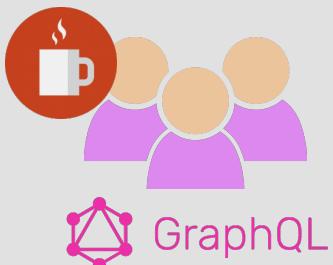
MicroCyberRobo Corp



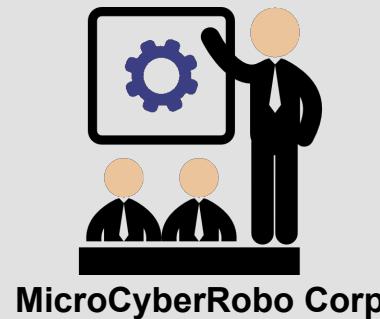
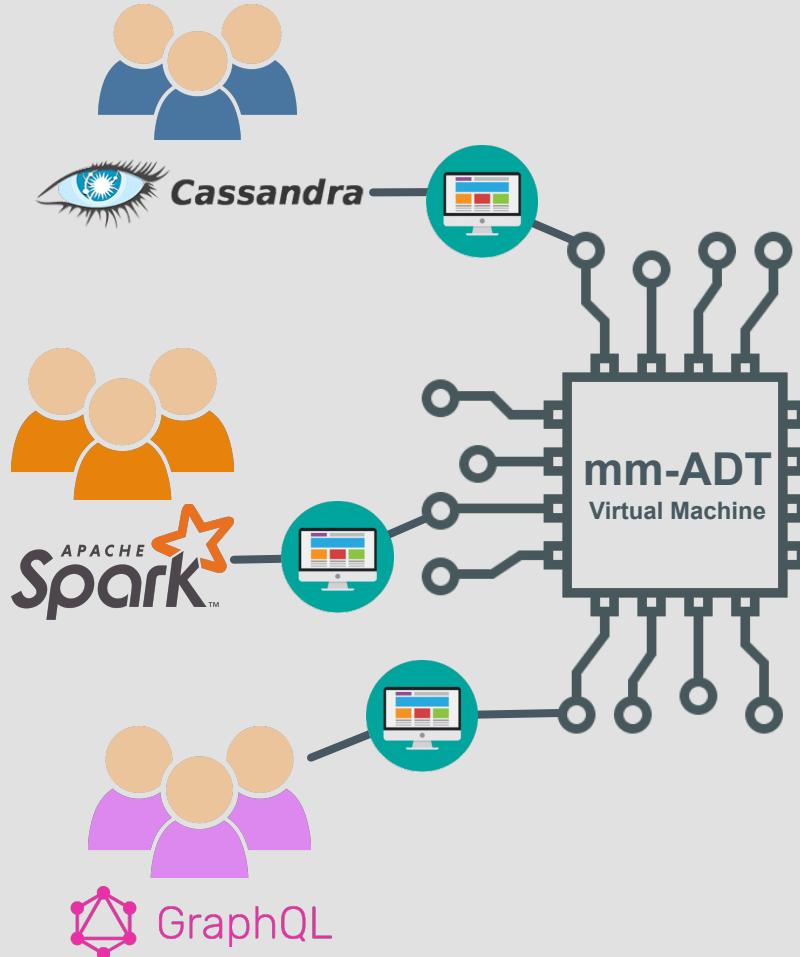
# A Royalty-Based Open Source Model



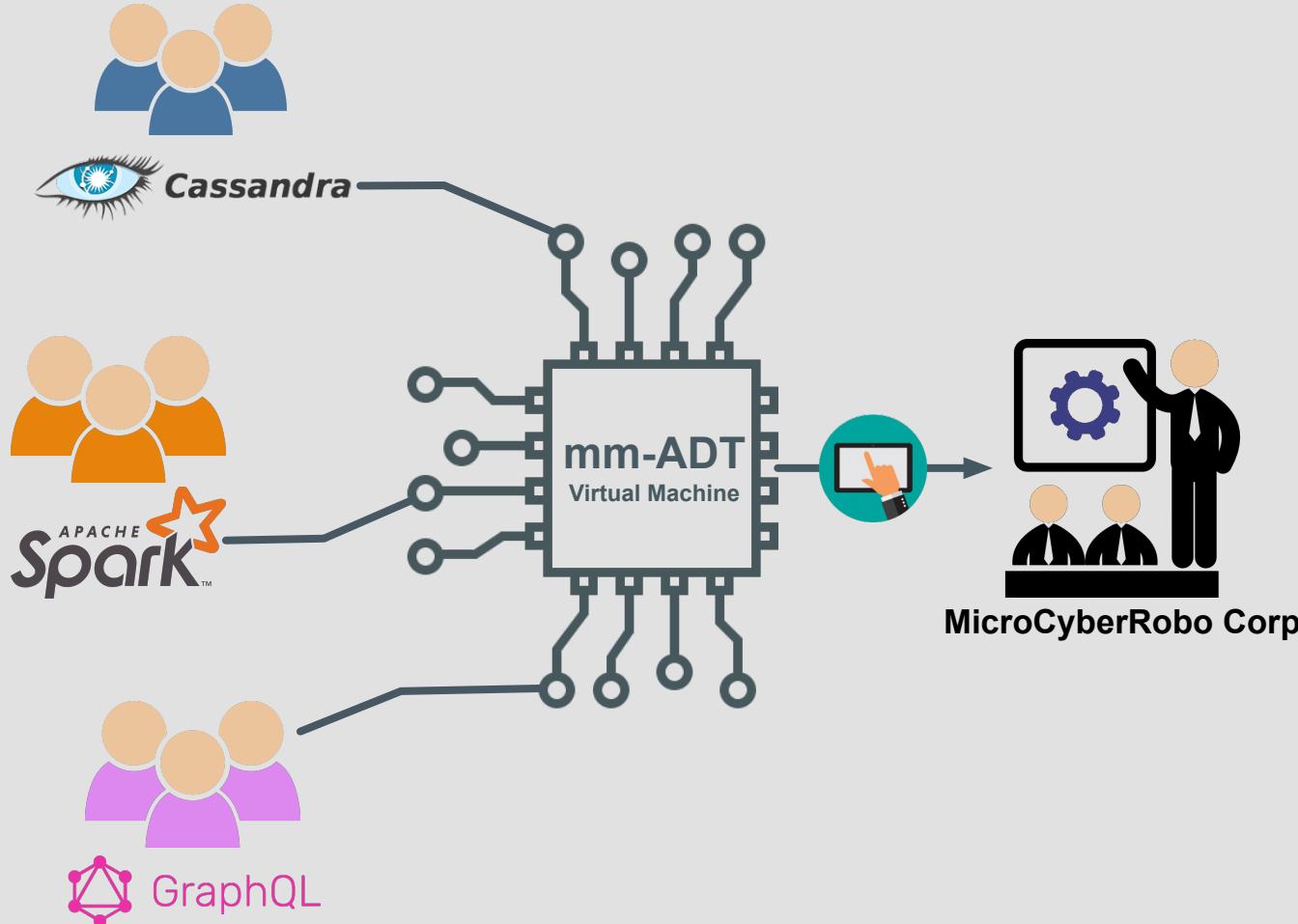
 **Cassandra**



# A Royalty-Based Open Source Model



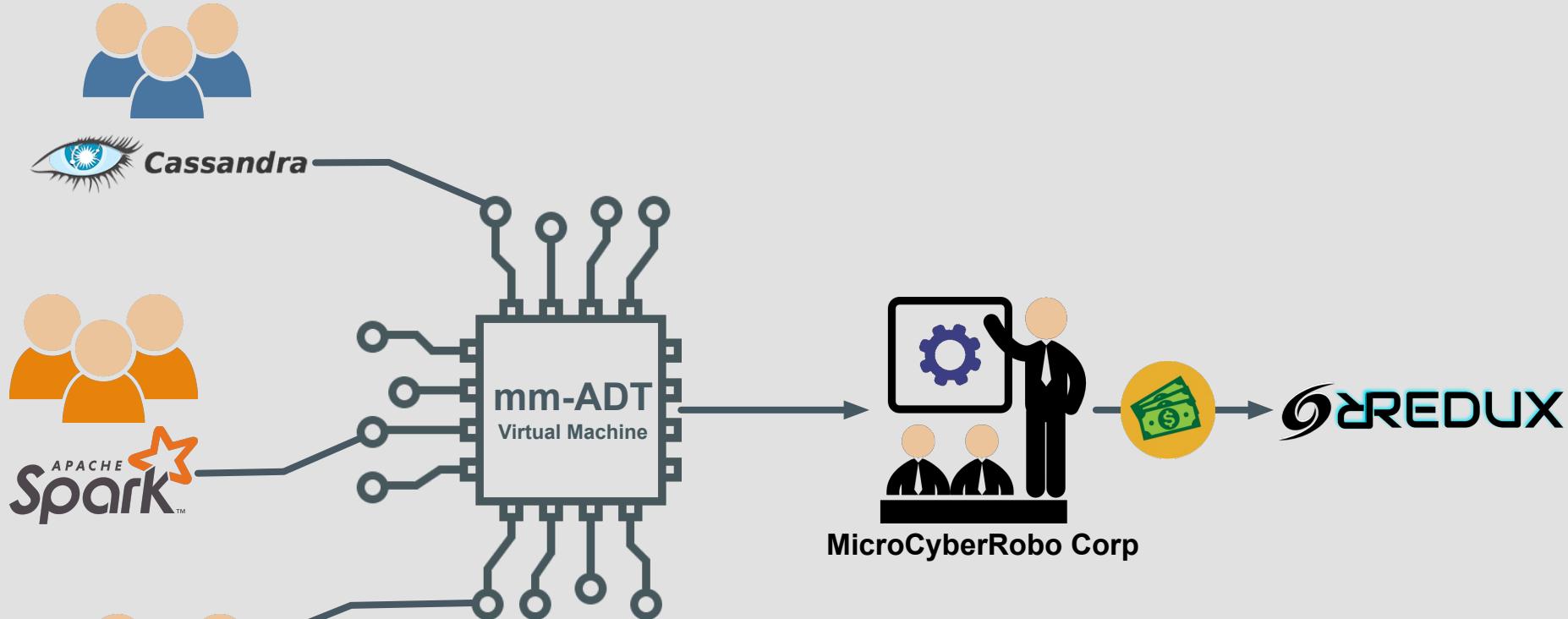
# A Royalty-Based Open Source Model



REDUX

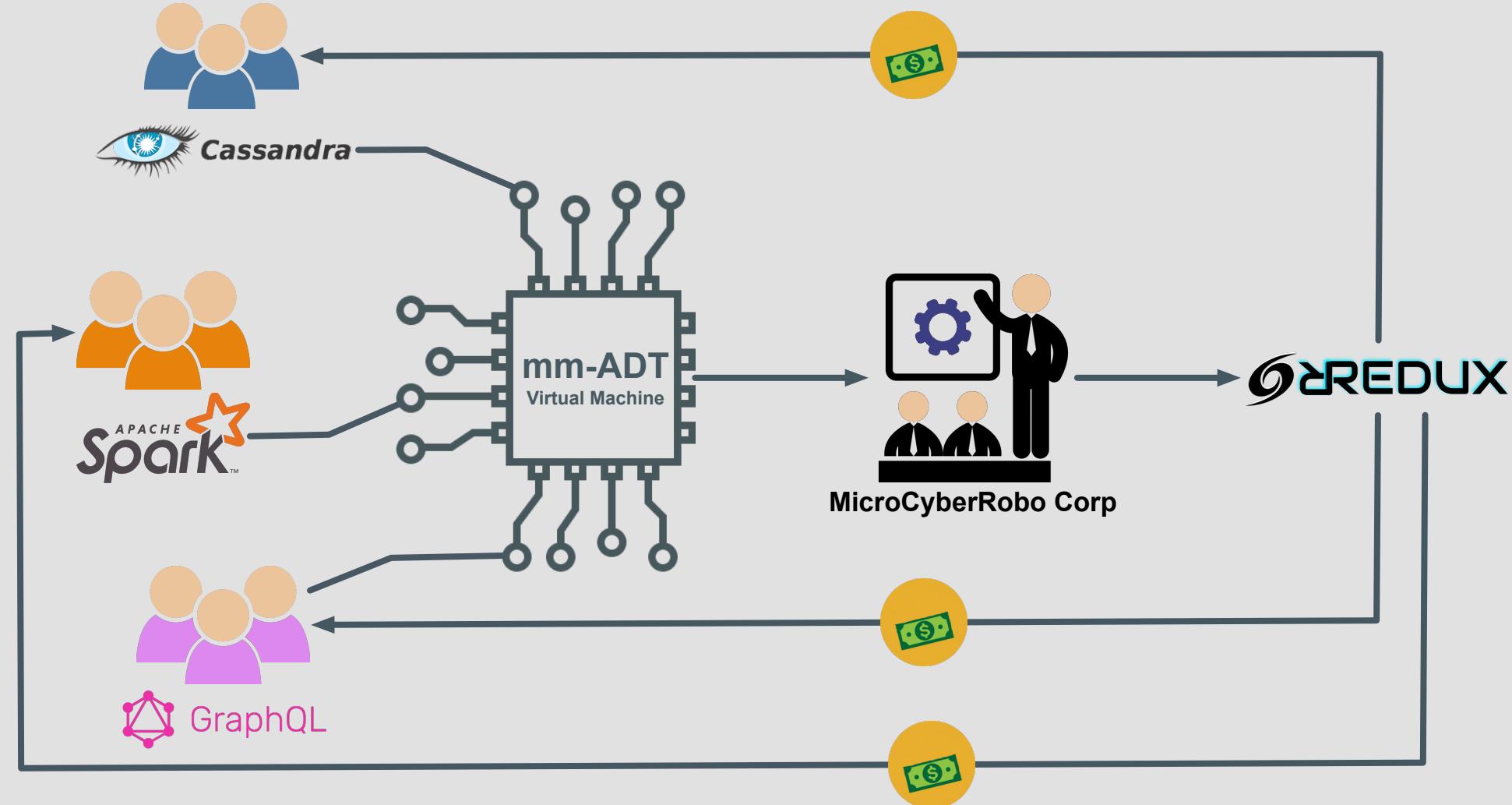
Compose

# A Royalty-Based Open Source Model



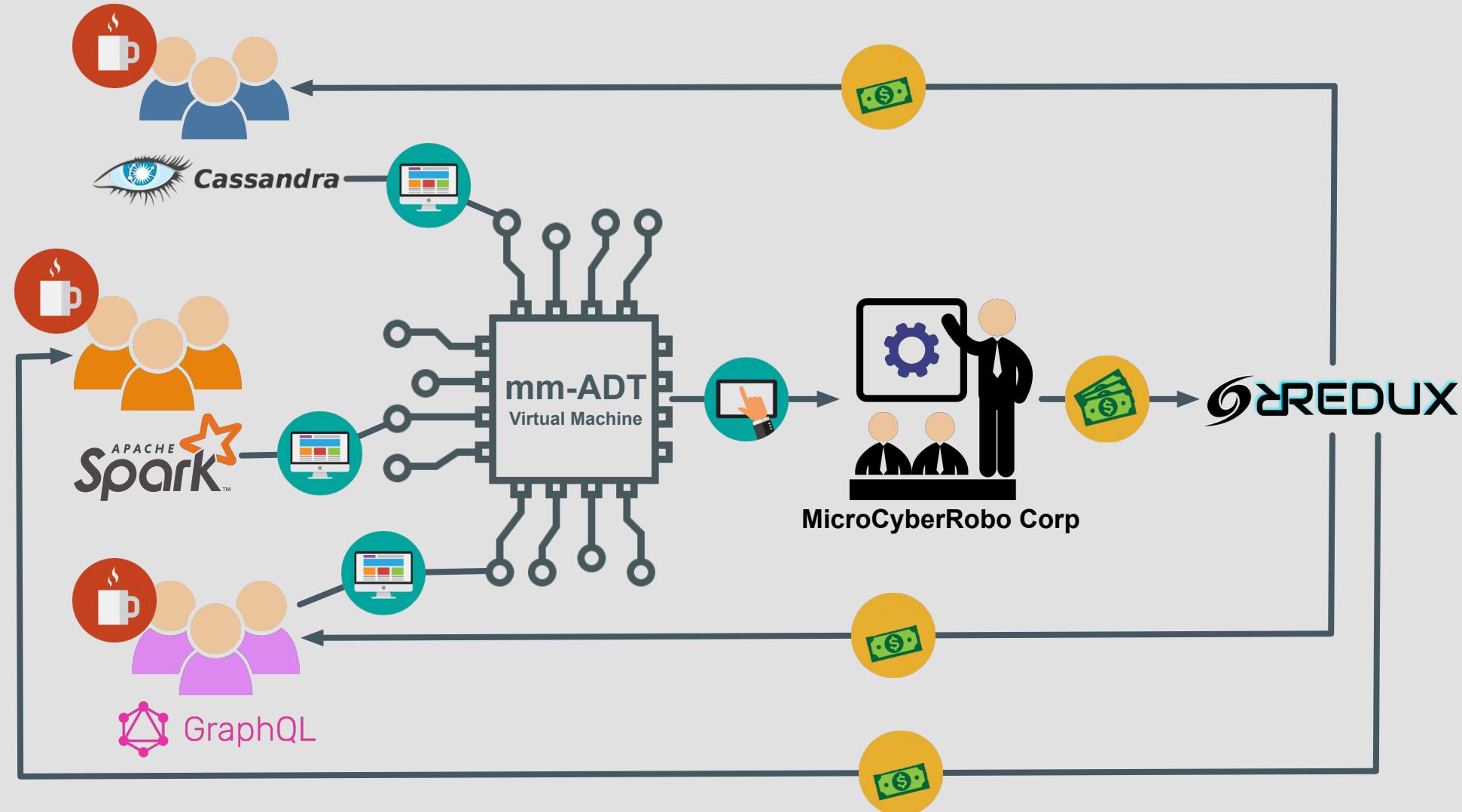
Purchase 

# A Royalty-Based Open Source Model



**Payout**

# A Royalty-Based Open Source Model



Design



Develop



Compose



Purchase



Payout





**Fin.**