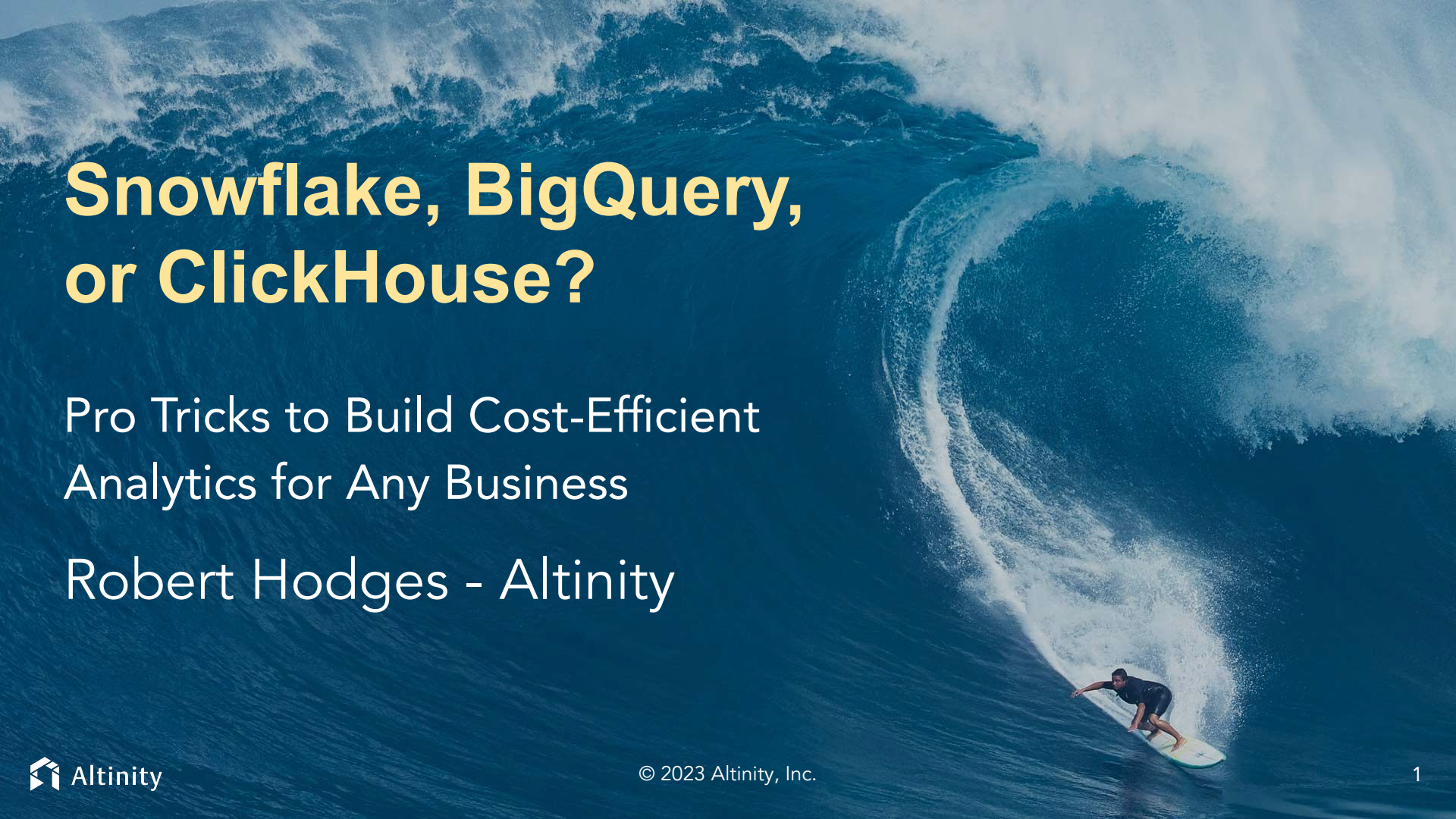


Snowflake, BigQuery, or ClickHouse?

Pro Tricks to Build Cost-Efficient Analytics for Any Business

Robert Hodges - Altinity

A brief message from our sponsor…

Robert Hodges

Database geek with 30+ years on DBMS. Kubernaut since 2018. Day job: Altinity CEO

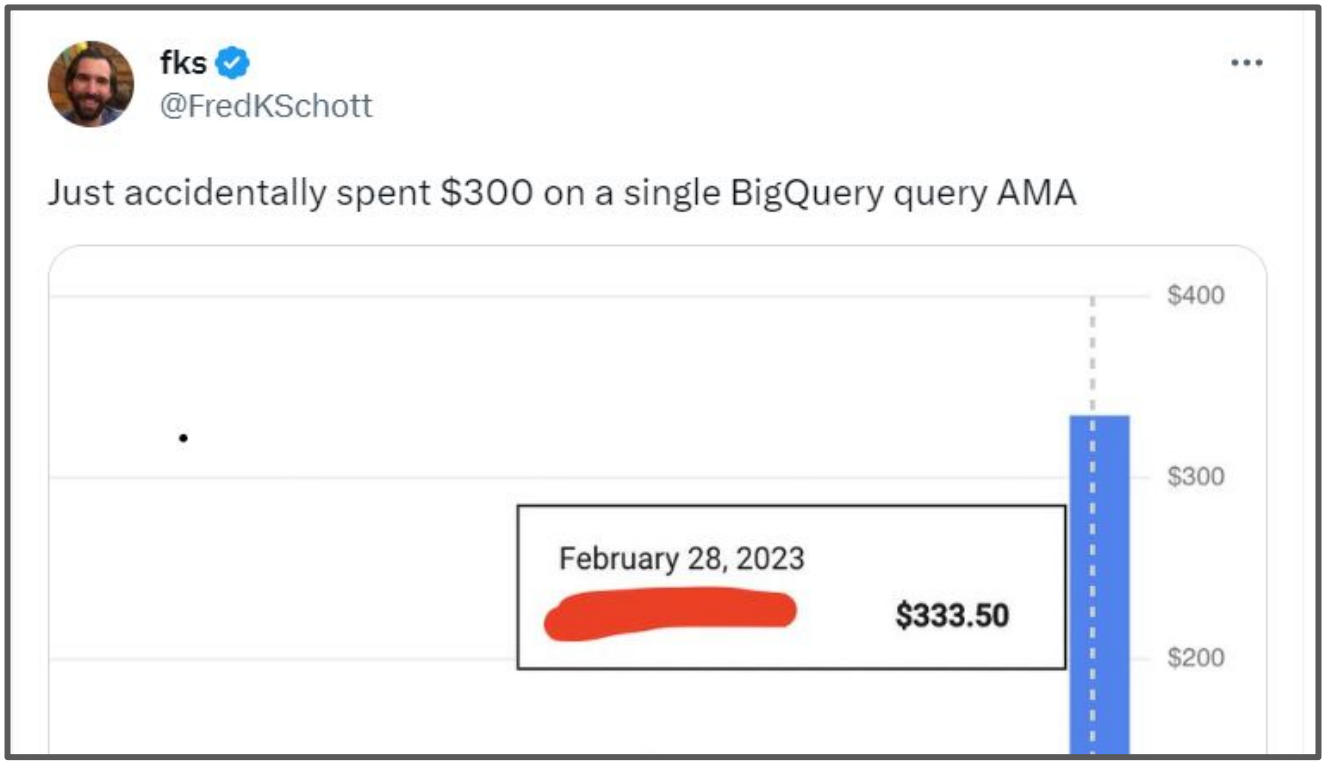
Altinity Engineering

Database geeks with centuries of experience in DBMS and applications

ClickHouse software and services: Altinity.Cloud and Altinity Stable Builds
Authors of Altinity Kubernetes Operator for ClickHouse

Introduction to analytic DBMS cost models

One way users learn about cloud cost-efficiency



Which leads to important intellectual questions

What's going on down there?

Let's start by understanding how cloud businesses work

snowflake	31 Jan 2023	Margin
Total Revenue	$ 2,065,659.00	
Cost of Revenue	$ 717,540.00	34.74%
Gross Profit	$ 1,348,119.00	65.26%
Operating Expense	$ 2,190,386.00	
Operating Income	$ (842,267.00)	-40.77%

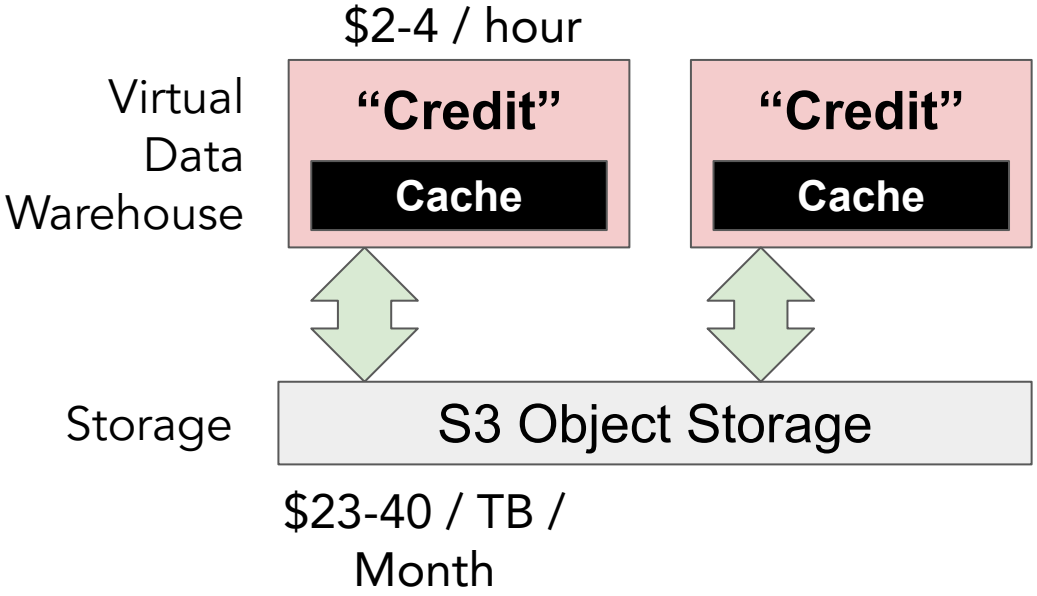
Cloud costs hide here!

Minimum markup on cloud costs:

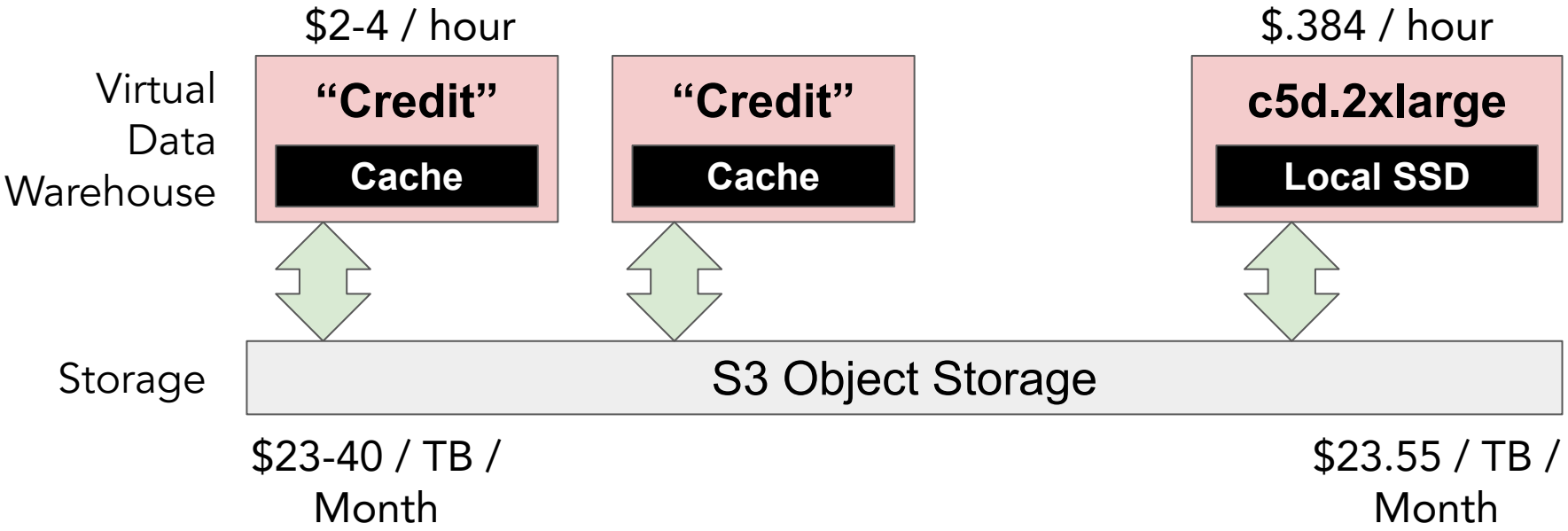
3x

Numbers in Thousands Source: finance.yahoo.com

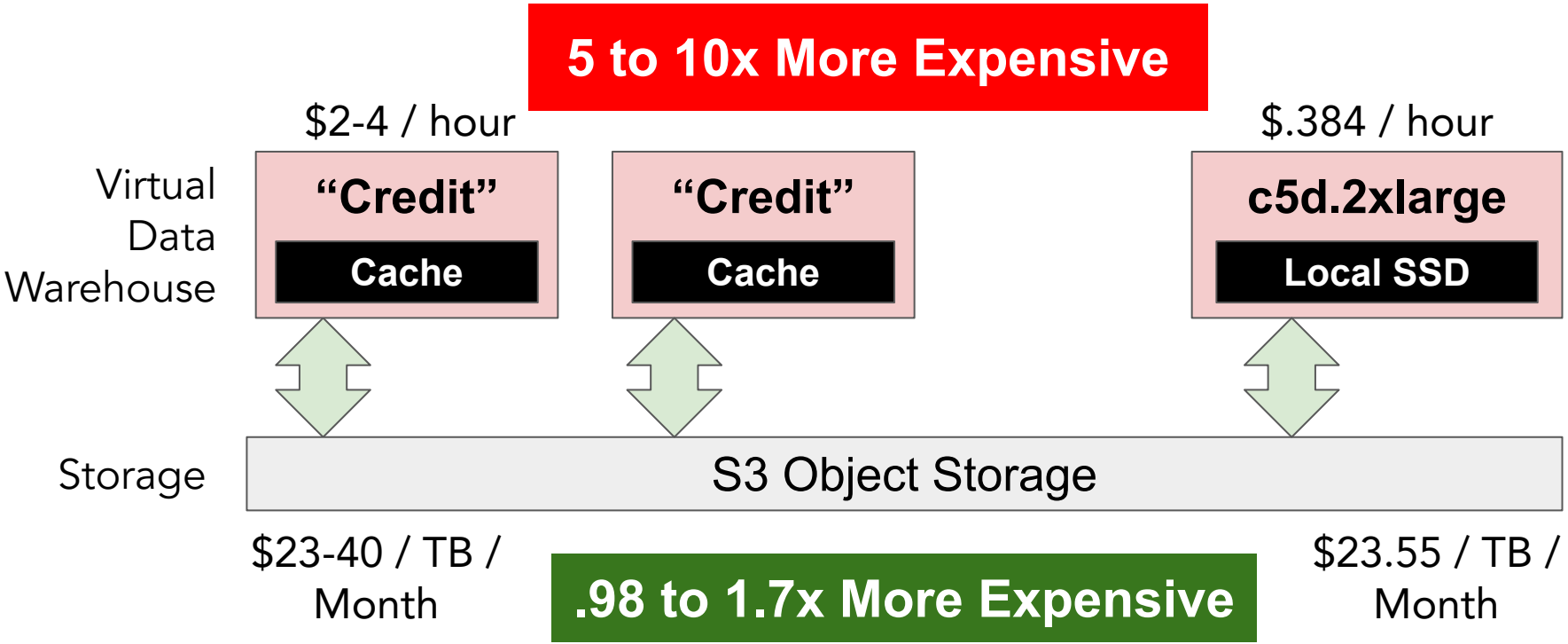
Snowflake Virtual Data WareHouse Model



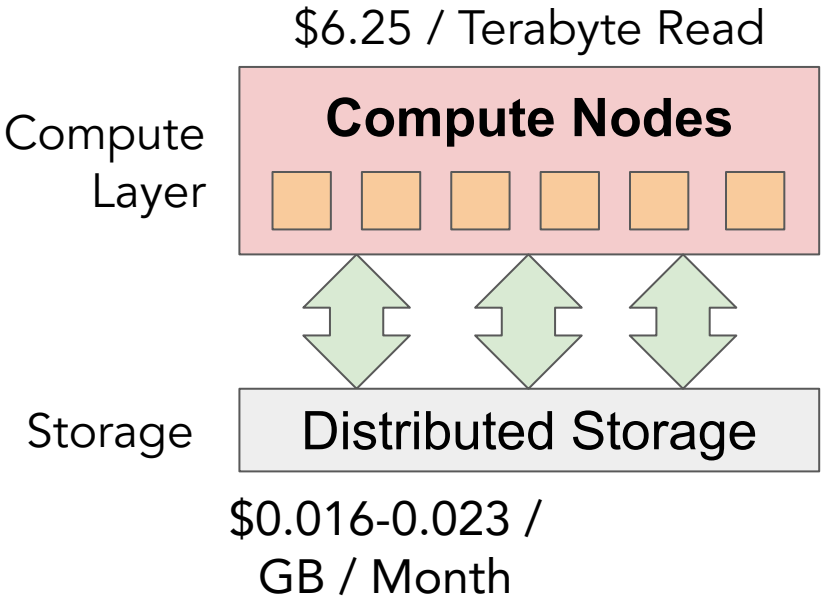
Snowflake Virtual Data WareHouse Model



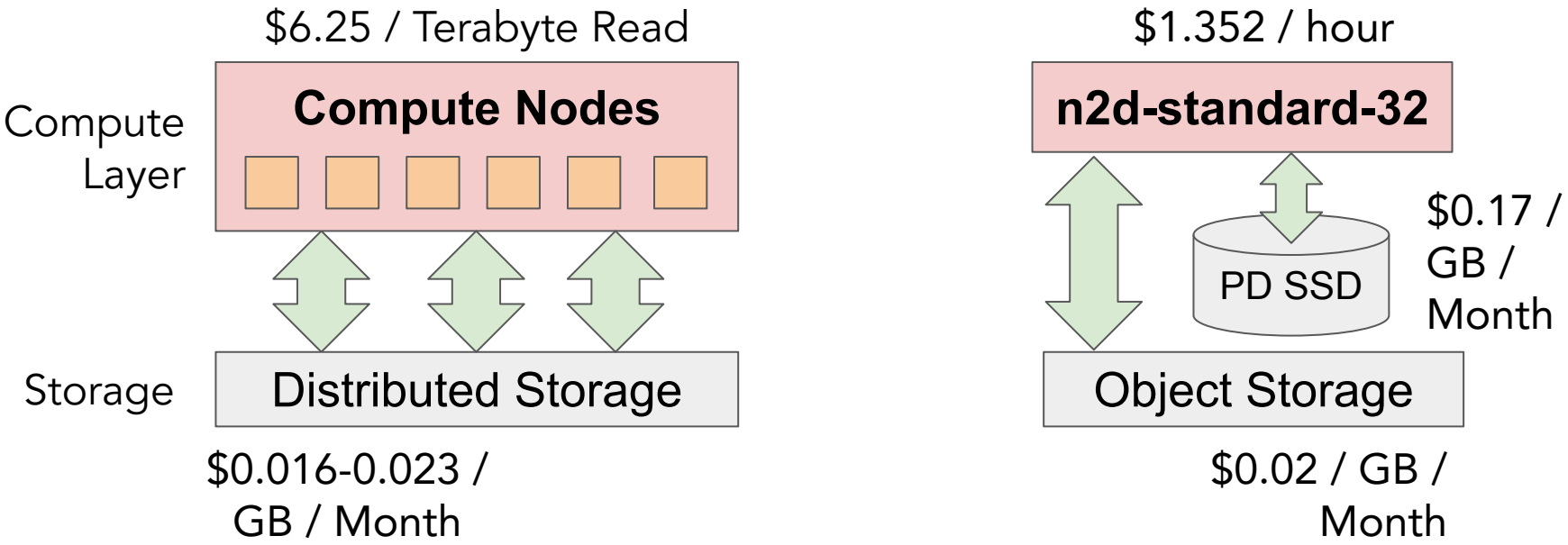
Snowflake Virtual Data WareHouse Model



BigQuery Serverless Query Model

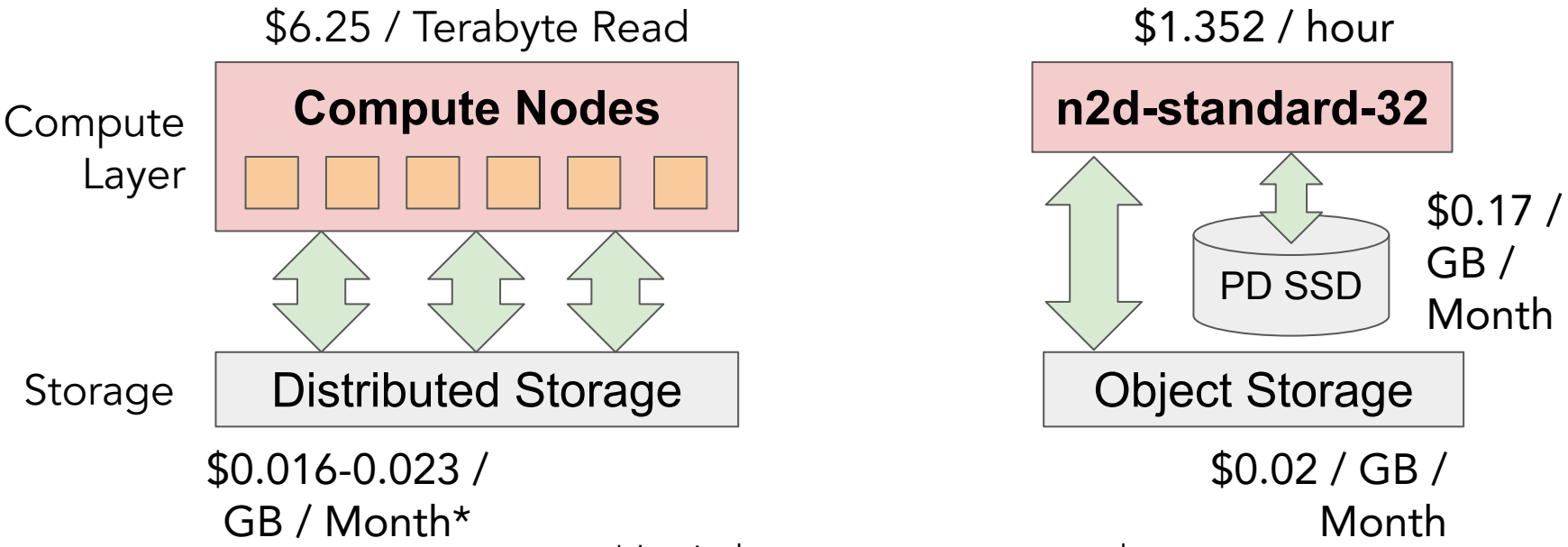


BigQuery Serverless Query Model



BigQuery Serverless Query Model

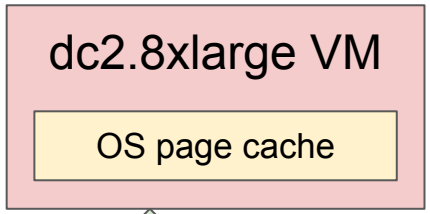
10x cheaper to
10x more expensive



* Logical storage = uncompressed

AWS RedShift “Buy the Box” Model

Compute



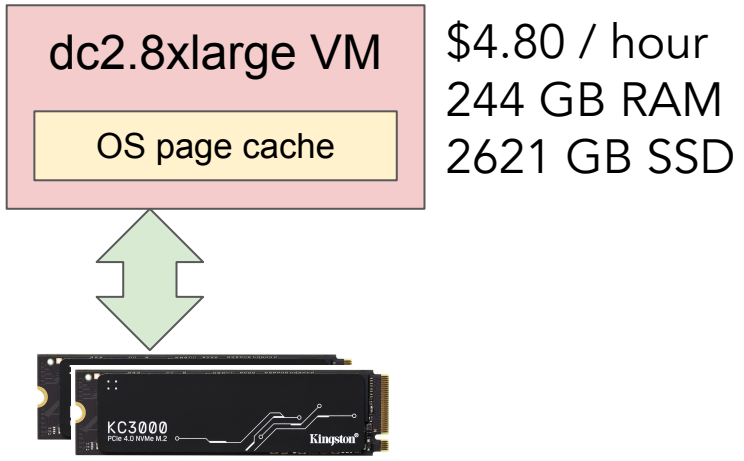
$4.80 / hour
244 GB RAM
2621 GB SSD

Attached Block Storage

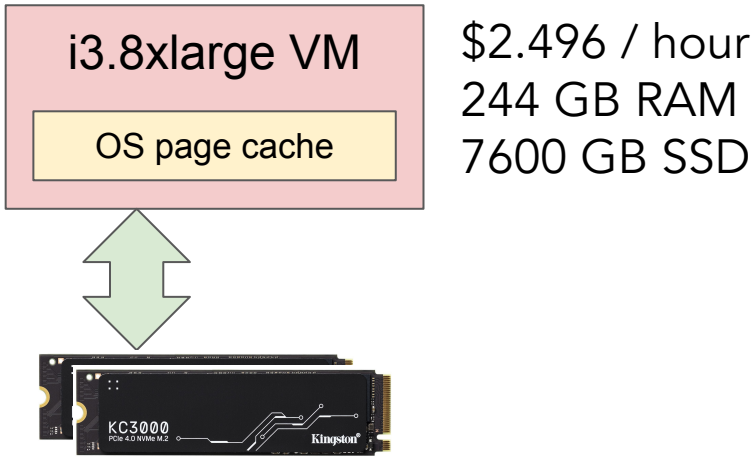
AWS RedShift “Buy the Box” Model

Compute

Attached Block Storage



$4.80 / hour
244 GB RAM
2621 GB SSD

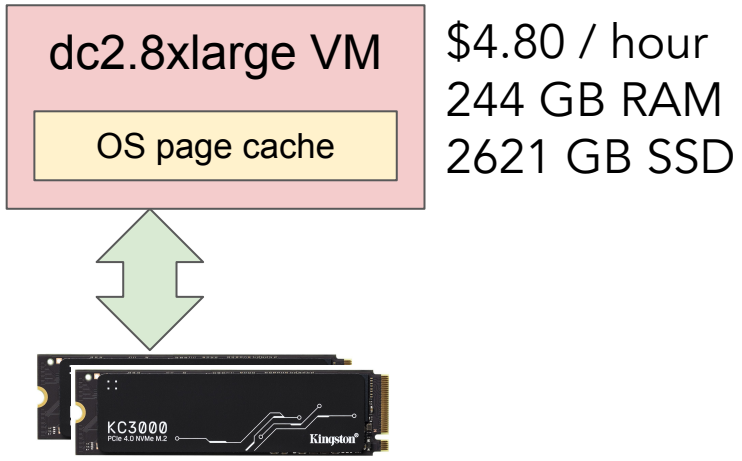


$2.496 / hour
244 GB RAM
7600 GB SSD

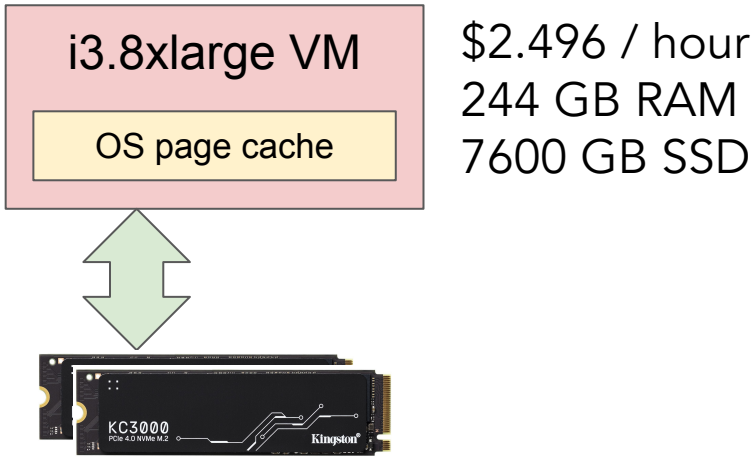
AWS RedShift “Buy the Box” Model

Compute

Attached Block Storage



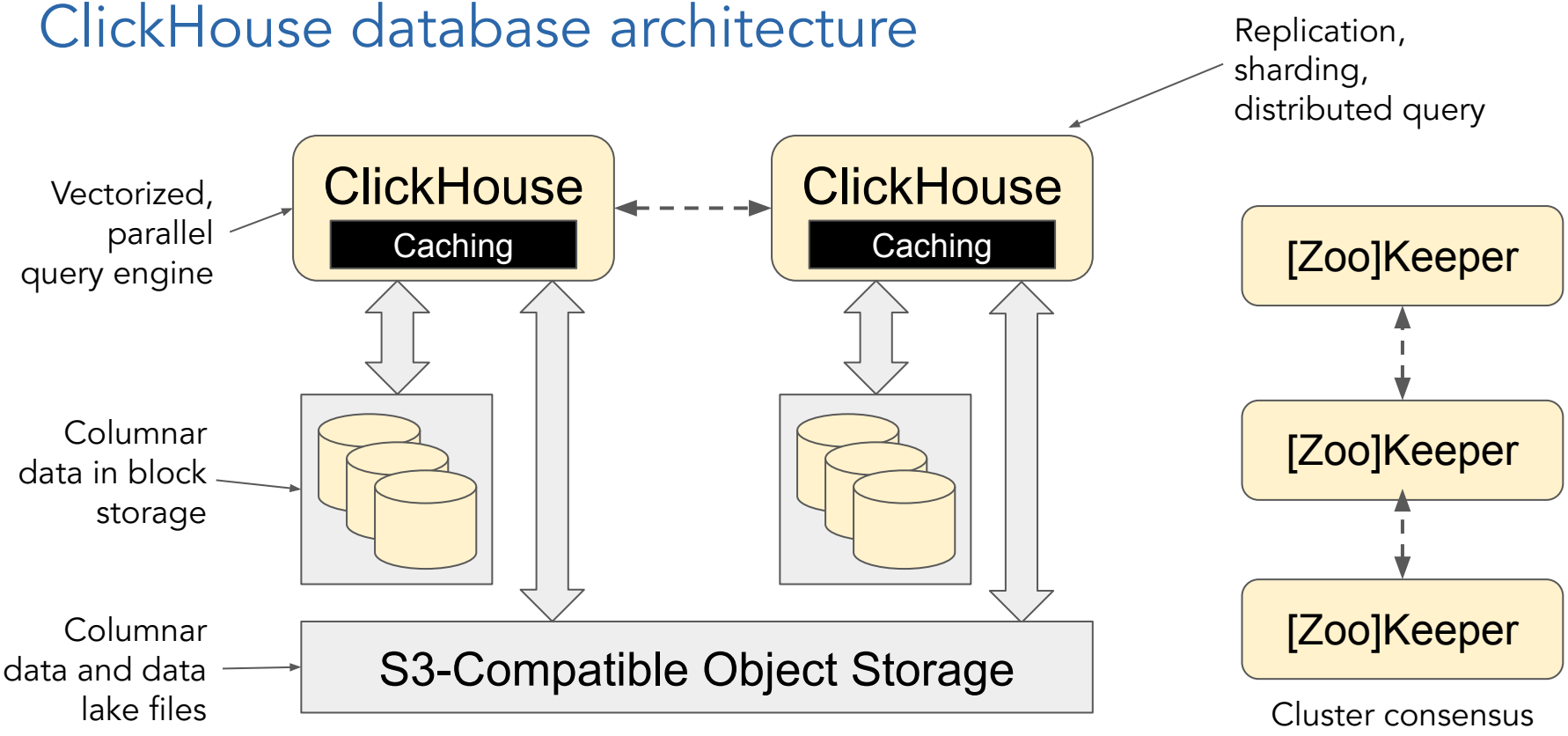
$4.80 / hour
244 GB RAM
2621 GB SSD



$2.496 / hour
244 GB RAM
7600 GB SSD

Redshift is 92% more costly with 66% less storage

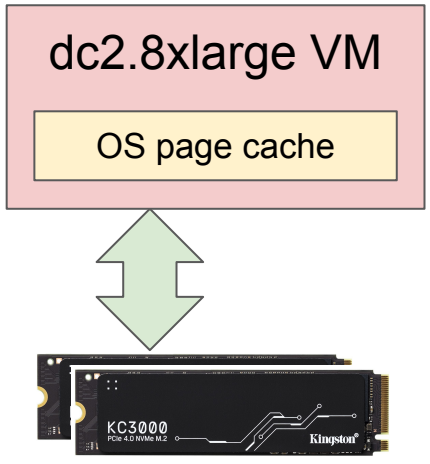
ClickHouse database architecture



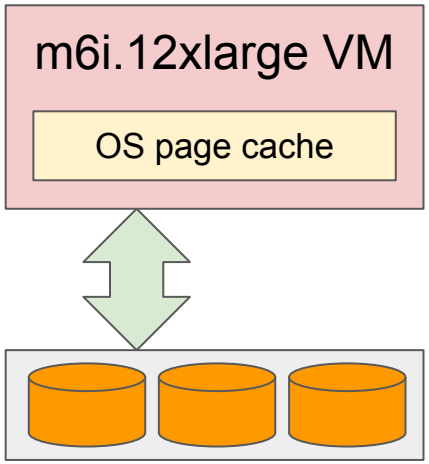
Better comparison: Modernized "Buy-the-Box"

Compute

Attached Block Storage



$4.80 / hour
244 GB RAM
32 vCPUs
2621 GB SSD



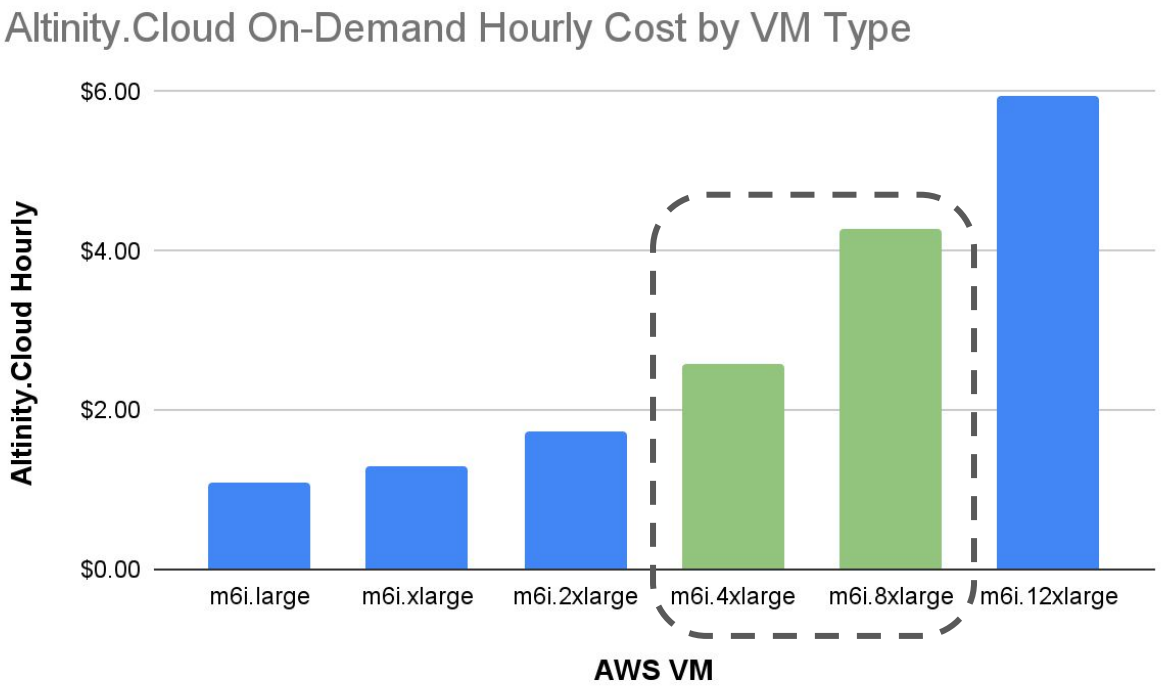
$2.64 / hour
196 GB RAM
48 vCPUs
2621 GB EBS

EBS gp3
1000 MB/sec

ALTINITY
Altinity.Cloud

Separate storage and compute lets you optimize cost!

Effect of storage & compute on ClickHouse prices



Quick comparison of models

Altinity.Cloud Modern "Buy-the-Box"	Snowflake Virtual Data Warehouse	BigQuery On-Demand Query
Cheap compute	Expensive compute	Expensive for queries that scan a lot of data
More expensive storage	**Cheap storage**	**Cheap storage***
Storage / compute separation?	VDW plan type? Is it 24x7?	Arrangement of data? How many queries?
Customer facing analytics	In-house BI, limited customer facing analytics	In-house data exploration

* Only if it's compressed

Getting the best deal from cloud services

What Snowflake does well

- ✔ General purpose
- ✔ Serverless operation
- ✔ Handles large numbers of tenants with completely different applications
- ✔ Standards-compliant SQL
  - Complete implementation with ACID transactions
  - Sophisticated query optimizer
  - Efficient columnar storage with self-tuning partitioning and compression
  - Big table joins
- ✔ Excellent integration with tools

What Snowflake does not do

- ✘ Keep data in customer cloud account
- ✘ Minimize costs, especially for 24x7 analytics
- ✘ Deliver stable real-time response
- ✘ Handle SaaS tenant-facing analytics
- ✘ Avoid vendor lock-in

How can you get a better price on cloud analytics?

- Look for decoupled storage / compute
- Make sure you are charged for compressed storage
- Ask for discounts based on your monthly spend
- Look for price breaks that align with vendor's own discounts
  - E.g., Compute has high discounts
- Prepay / reserve capacity
- Buy on cloud marketplace and apply price to your commits

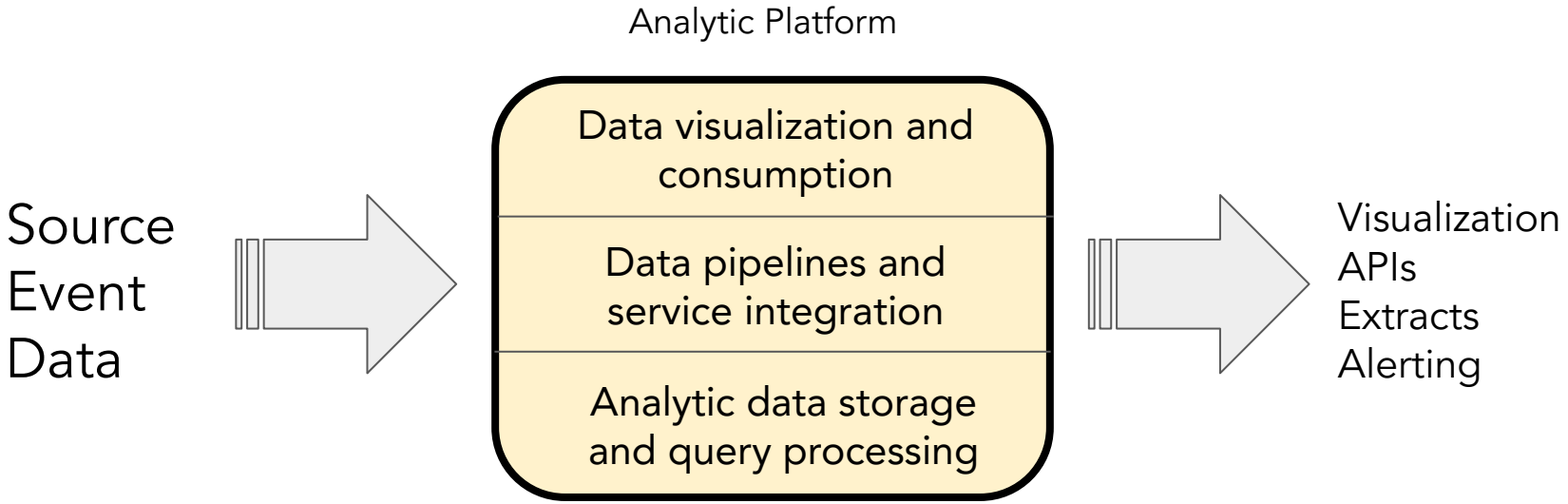
When is cloud analytic database pricing a good deal?

- Affordable
- Cost growth lower than your revenue growth
- Predictable maximum price
- No charge for unused resources
- Minimal extras (like transfer costs)
- Application changes don’t result in random price variations
- Vendor pricing is in line with their revenue

Getting a better deal by building with open source

Pick a specific problem

Deliver a GDPR-compliant replacement for Google Analytics



Reality check against Snowflake

Snowflake strengths

- ✘ General purpose
- ✔ Serverless operation
- ✘ Handle wide range of applications
- ✘ Standards-compliant SQL
- ✔ UI with SQL editing & management

Snowflake weaknesses

- ✔ Keep data in your own cloud account
- ✔ Minimize costs for 24x7 systems
- ✔ Deliver stable real-time response
- ✔ Handle SaaS user-facing analytics
- ✔ No vendor lock-in

Kubernetes enables a powerful alternative to proprietary cloud services

Enabled by Cloud Native Computing

Defined by Infrastructure as Code

The Modern Analytic Stack

Powered by Open Source

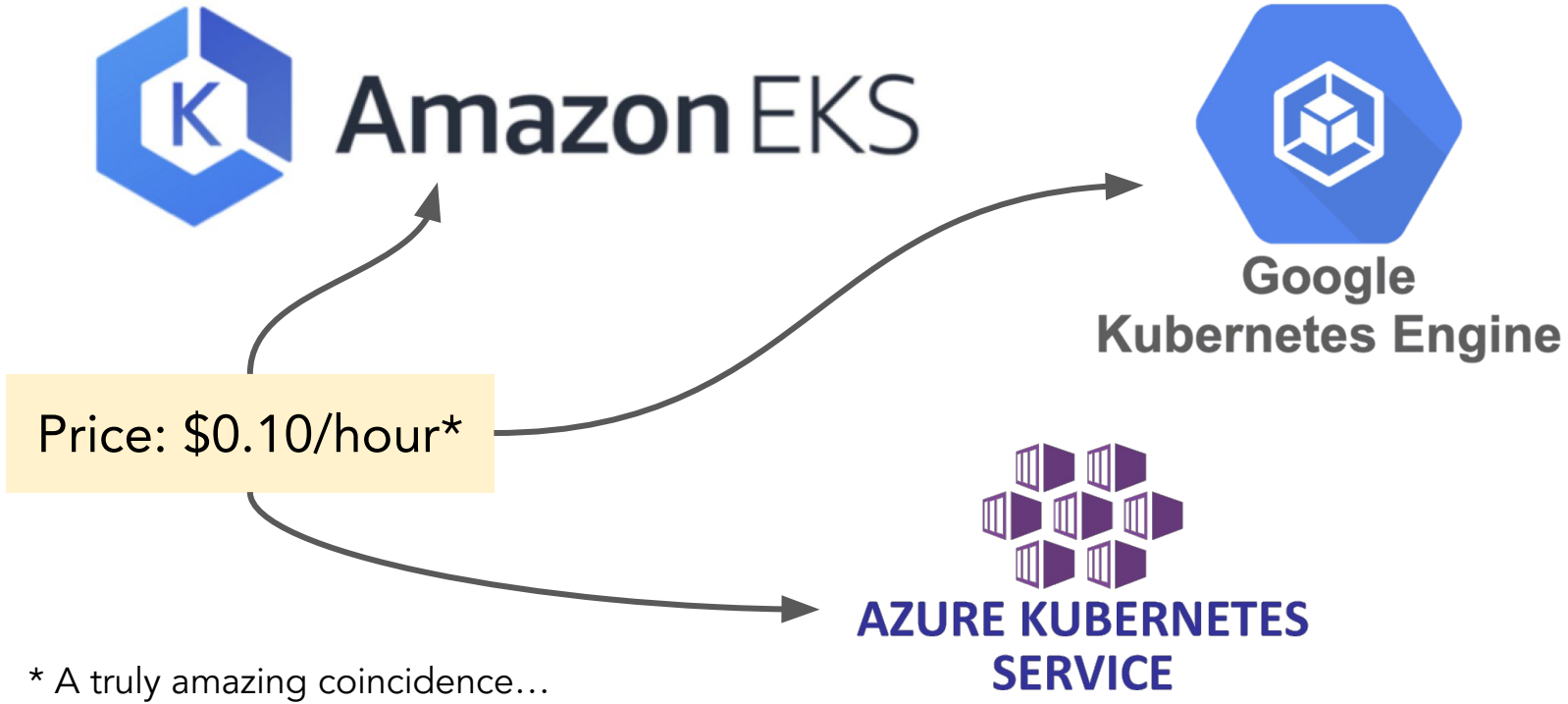
Deployed by GitOps

Operated on Kubernetes and Cloud

Modern analytic stacks are custom data platforms

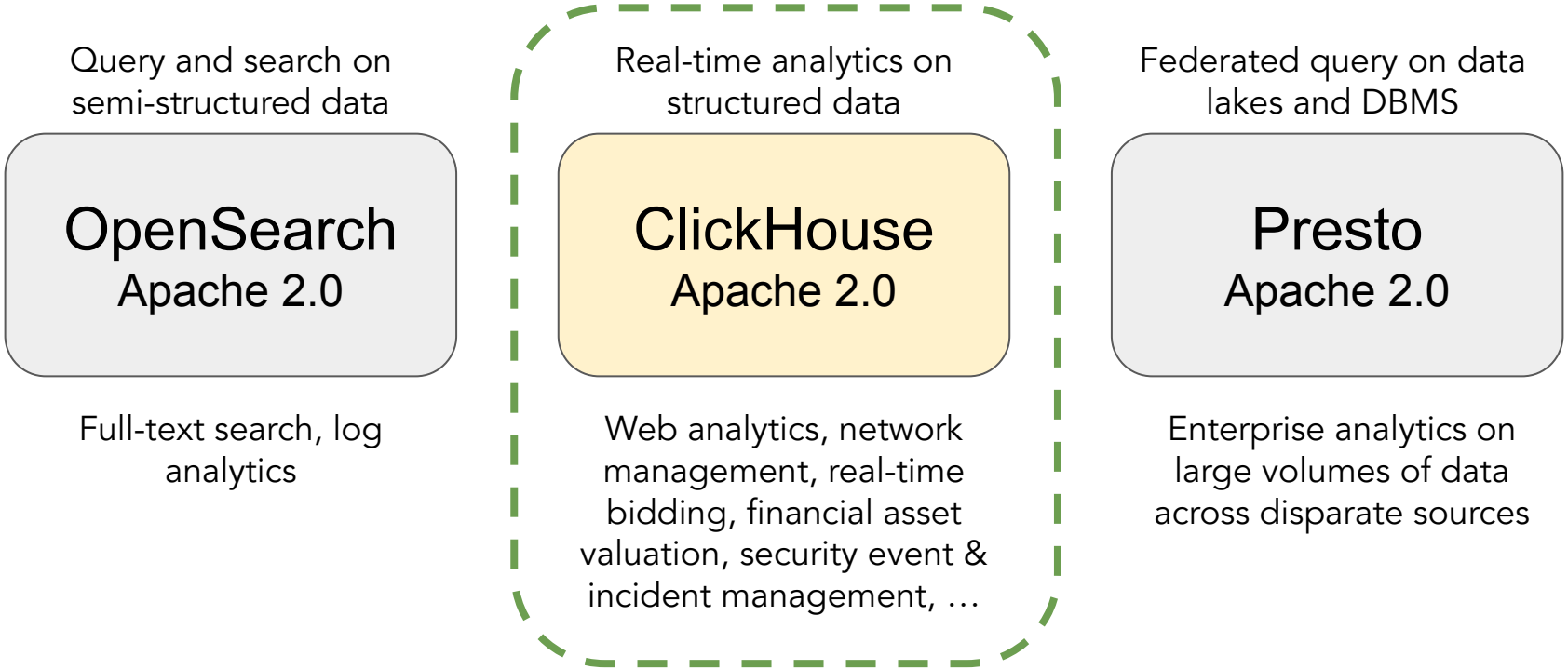
Consumption - BI Tools, Custom Apps, APIs
Orchestration - CDC, Event Processing, ETL/Reverse ETL
Storage - Database Engines
Management - Operators, Observability, Security
GitOps CI/CD Kubernetes 

Step 1: Choose a Kubernetes distribution

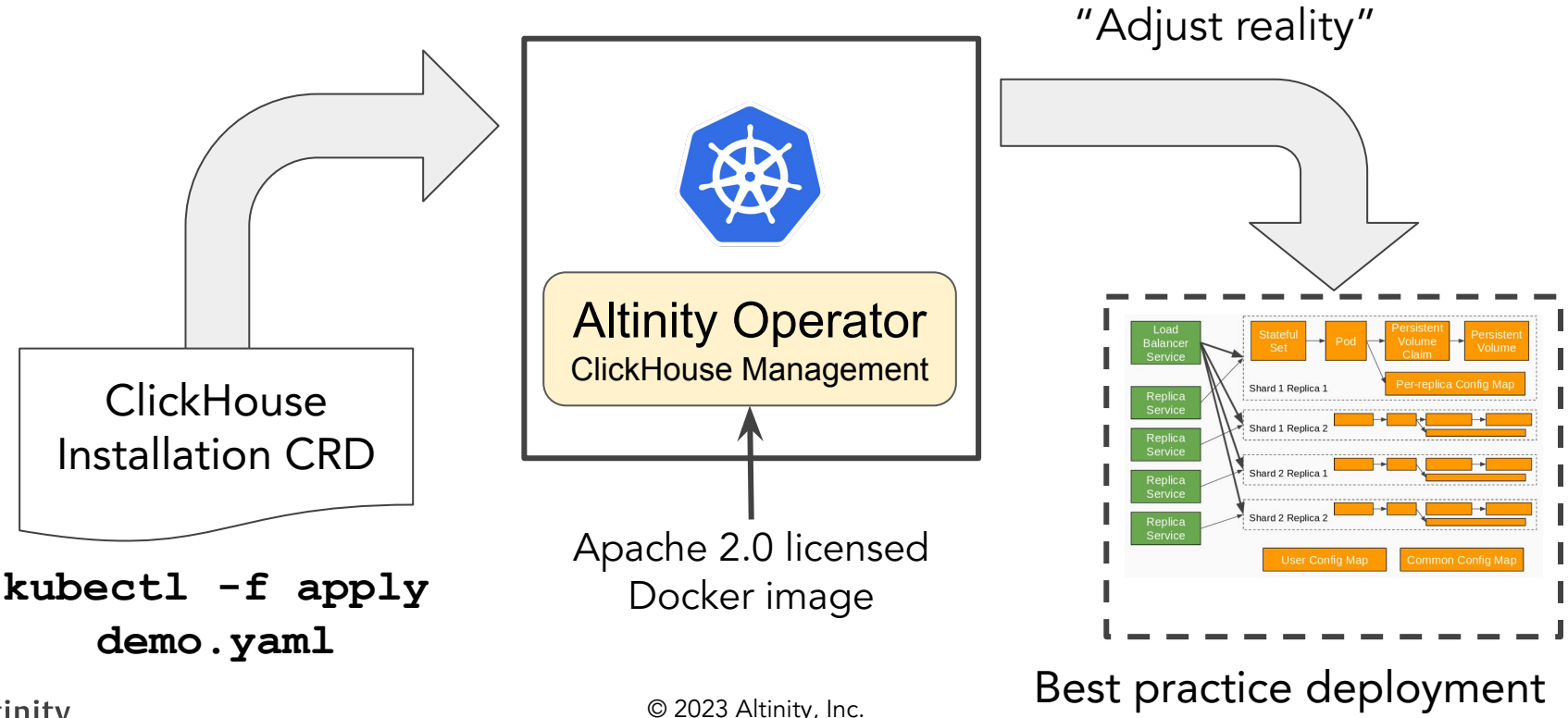


* A truly amazing coincidence…

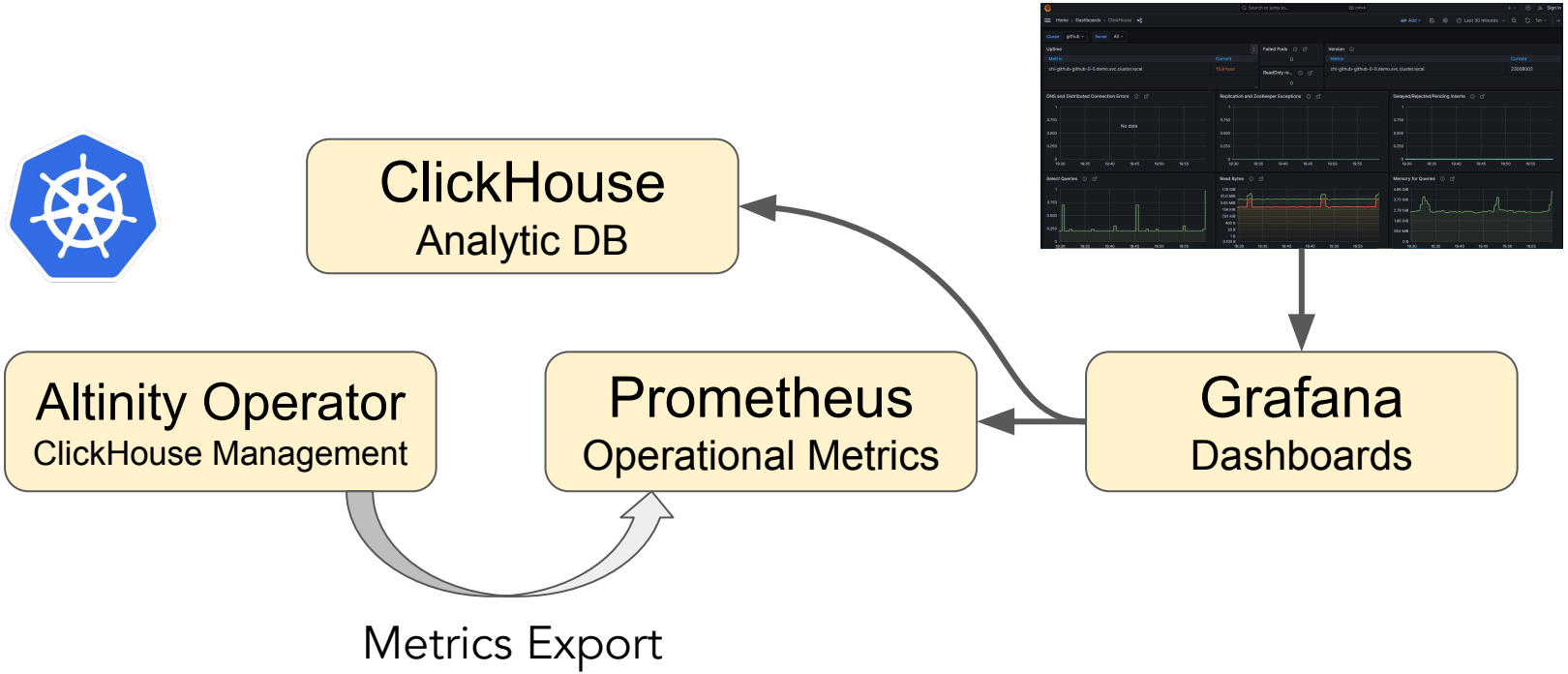
Step 2: pick an open source analytic database



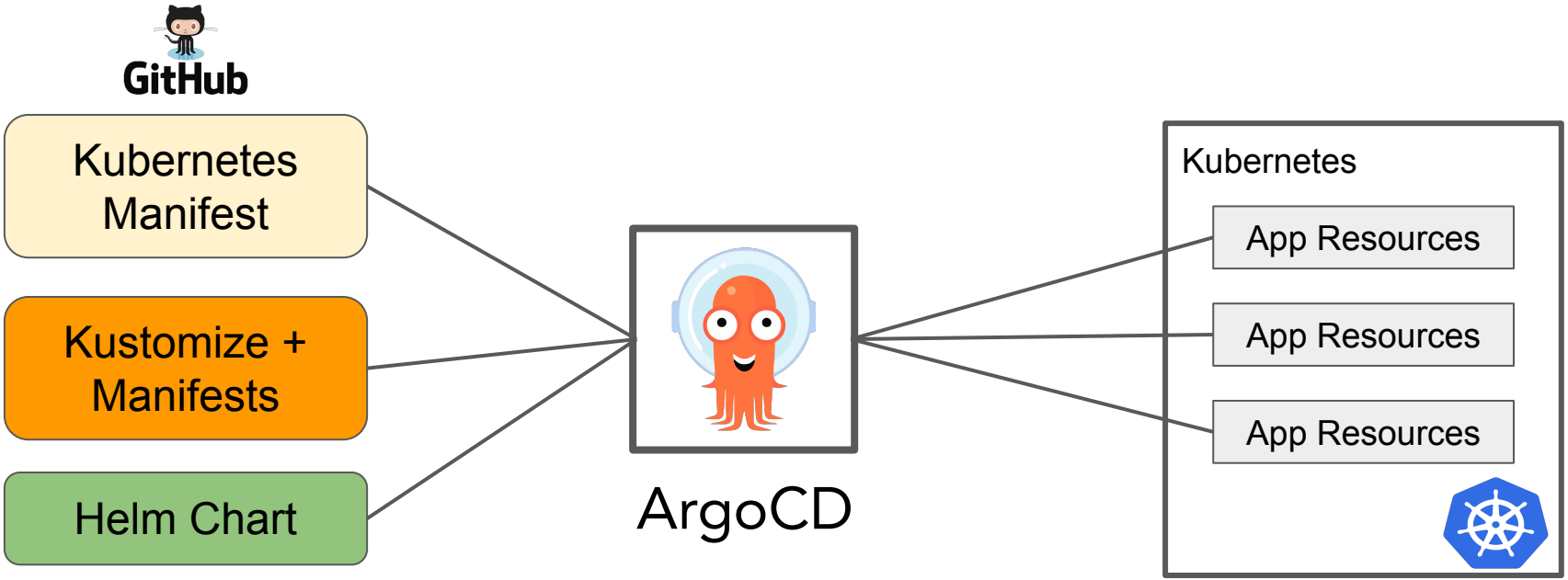
Step 3: Pick an operator to run the database



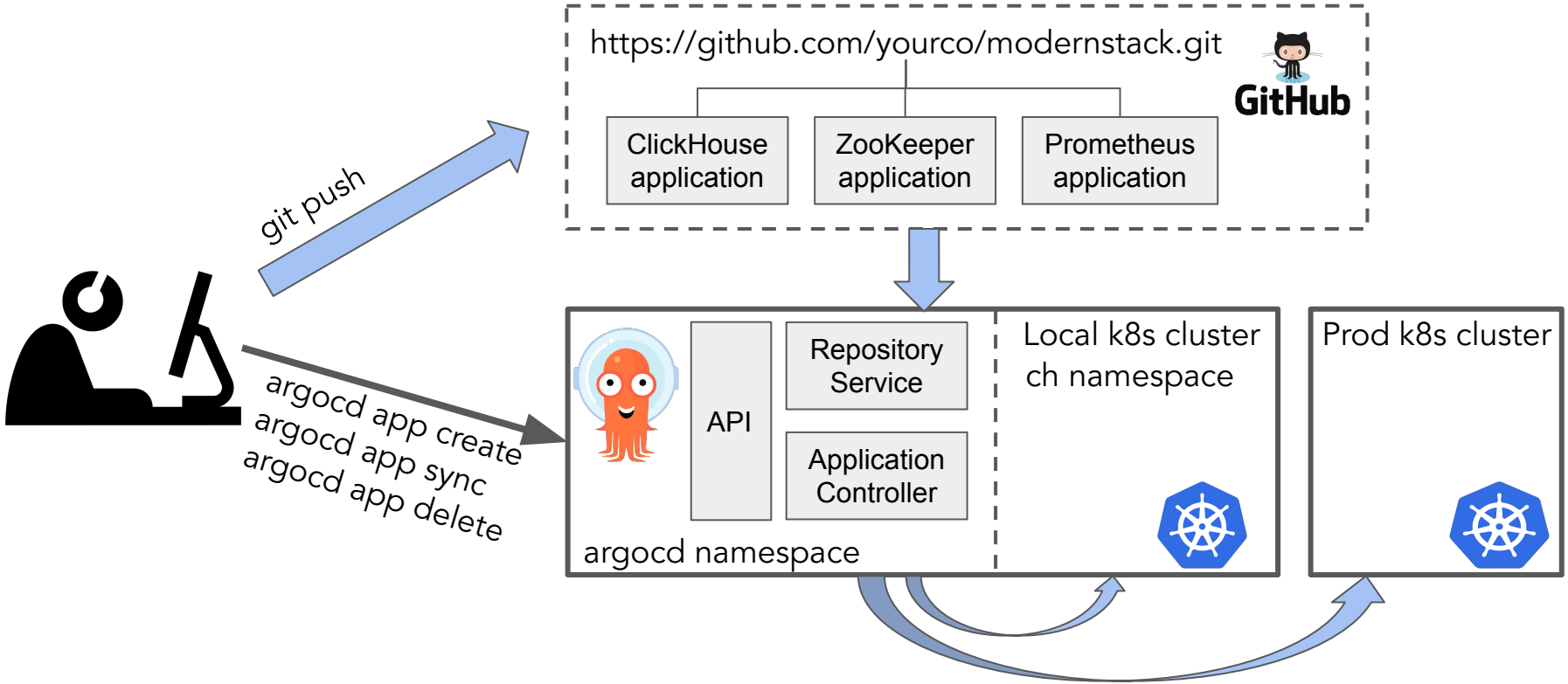
Step 4: Choose observability platform



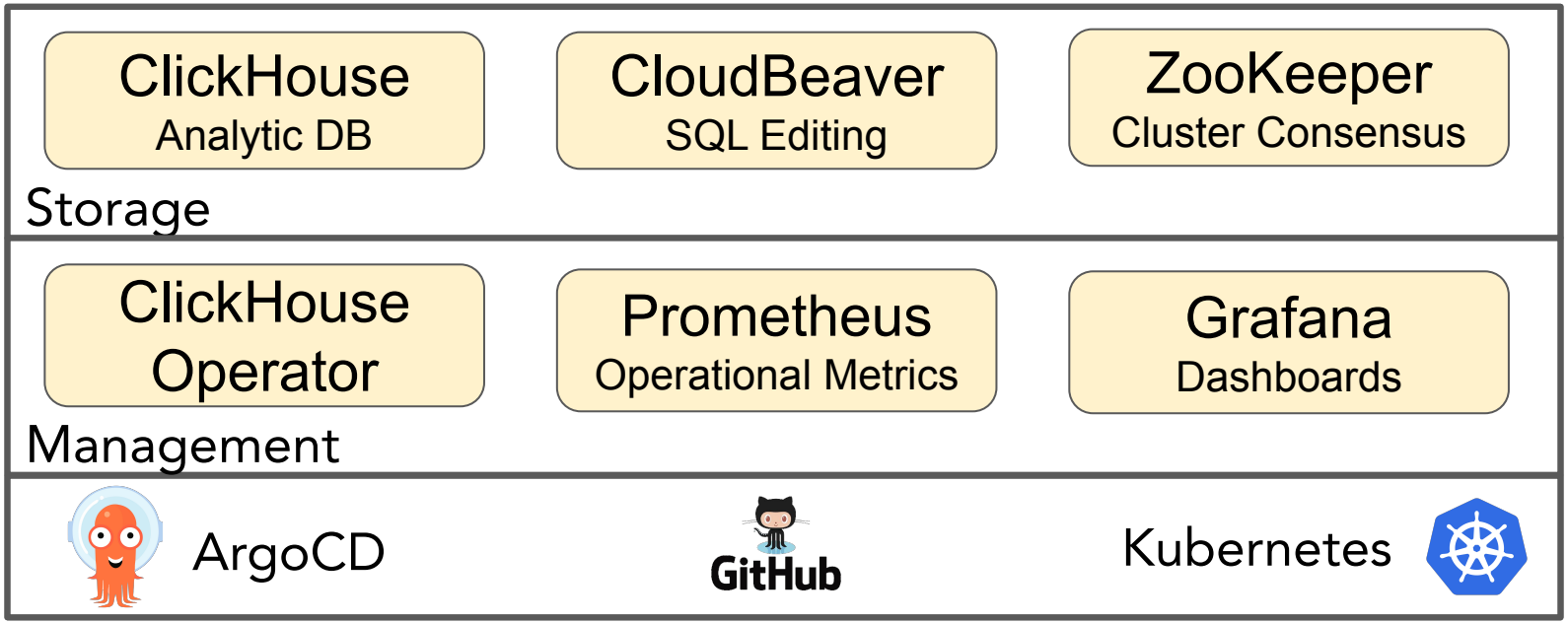
Step 5: pick a Kubernetes GitOps implementation



Basic GitOps using GitHub, ArgoCD, and Kubernetes



The full stack including GitOps



Managing Kubernetes applications with ArgoCD

DEMO TIME!

Best practices for do-it-yourself modern analytic stacks

- Build on managed Kubernetes
- Pick the right open source database for the job (ClickHouse!)
- Use operators for databases (Altinity Operator!)
- Don’t forget observability and other management services
- Pick a GitOps implementation: ArgoCD or Terraform
- Mix and match cloud services with Kubernetes services!!

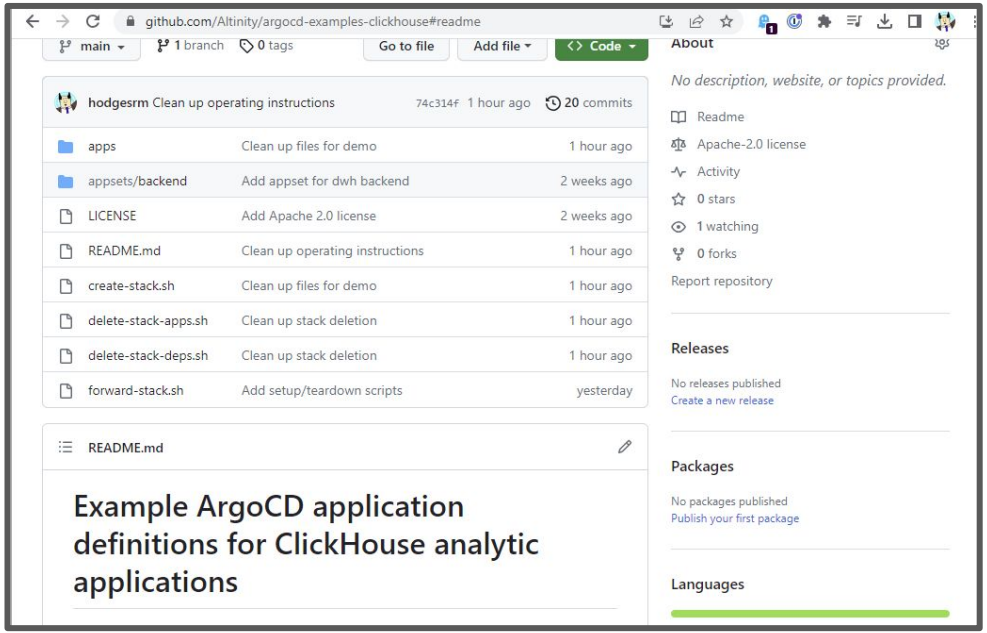
Altinity and modern analytic stacks

Three ways Altinity helps you build modern analytic stacks

- Altinity.Cloud Platform for ClickHouse
  - Altinity.Cloud Anywhere can even manage ClickHouse inside your Kubernetes clusters!
- Software
  - Altinity Kubernetes Operator for ClickHouse
  - Altinity Stable Builds (including FIPS releases)
  - Clickhouse-backup
  - Community Vertamedia Grafana Plugin
  - Altinity Tableau Connector for ClickHouse
- Services
  - Expert DBA and Kubernetes support with enterprise SLAs
  - Proof-of-concept / co-design
  - Feature development for ClickHouse and ecosystem projects

How to get started with the example analytic stack

```
git clone https://github.com/Altinity/argocd-examples-clickhouse
```



Projects that went into the stack

- ArgoCD: https://argo-cd.readthedocs.io/en/stable/
- Altinity Projects
  - ArgoCD Examples
  - Altinity Kubernetes Operator for ClickHouse
  - Altinity Stable Builds for ClickHouse
- The rest of the stack
  - ClickHouse: https://github.com/ClickHouse/ClickHouse
  - Prometheus: https://github.com/prometheus-community/helm-charts
  - Grafana: https://github.com/grafana/grafana
  - CloudBeaver: https://github.com/dbeaver/cloudbeaver

Thank you and good luck!

Any questions?

Robert Hodges - Altinity
https://altinity.com

Altinity.Cloud
Altinity Stable Builds for ClickHouse
Altinity Kubernetes Operator for ClickHouse