ALTINITY

Scale ClickHouse® Queries Infinitely with 10x Cheaper Storage

Introducing Project Antalya

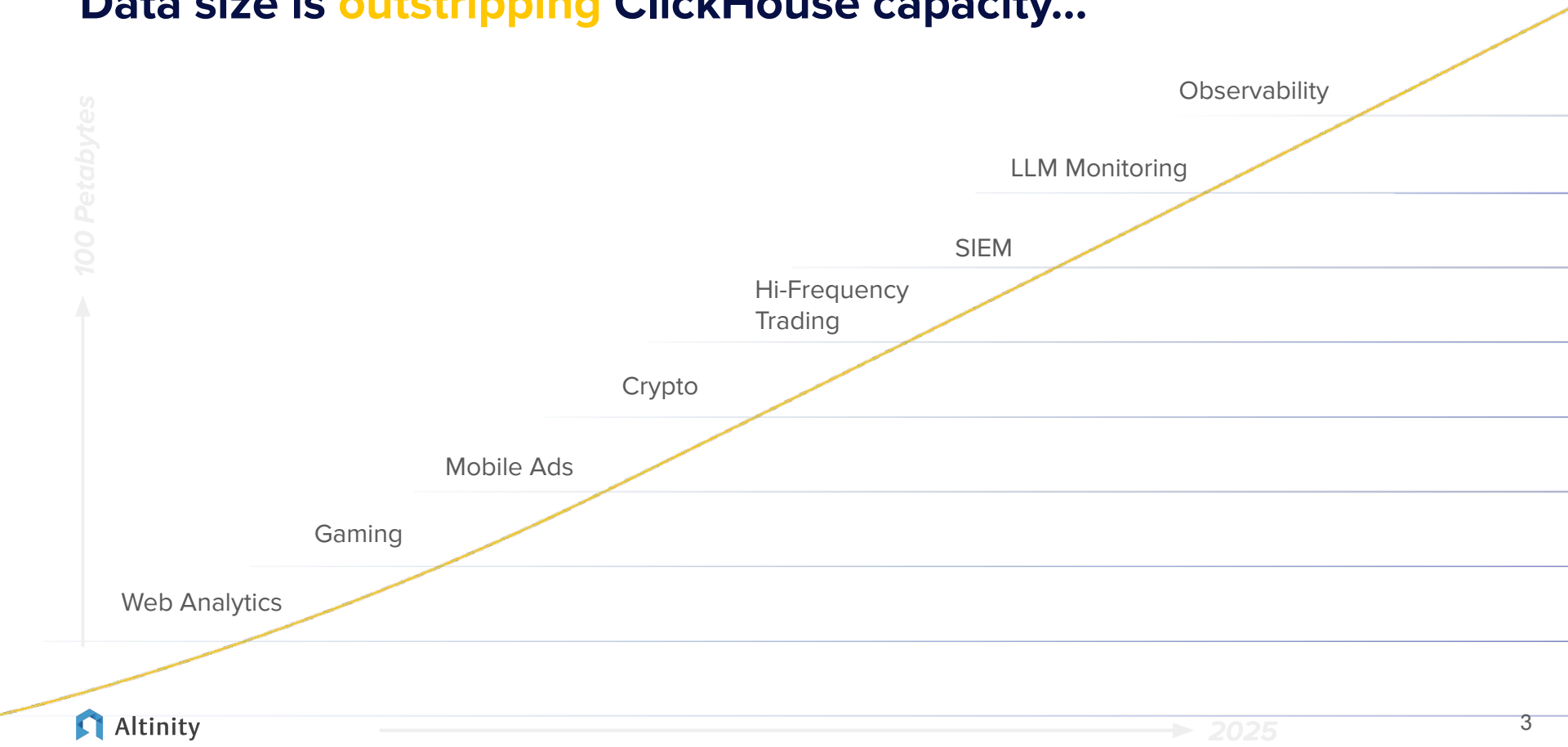
Introductions

Robert Hodges
CEO @ Altinity

Alexander Zaitsev
CTO @ Altinity

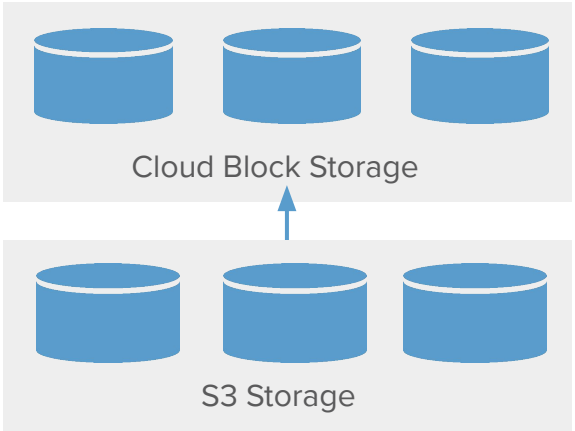
Tim Glaser
Co-CEO @ PostHog

Data size is outstripping ClickHouse capacity...

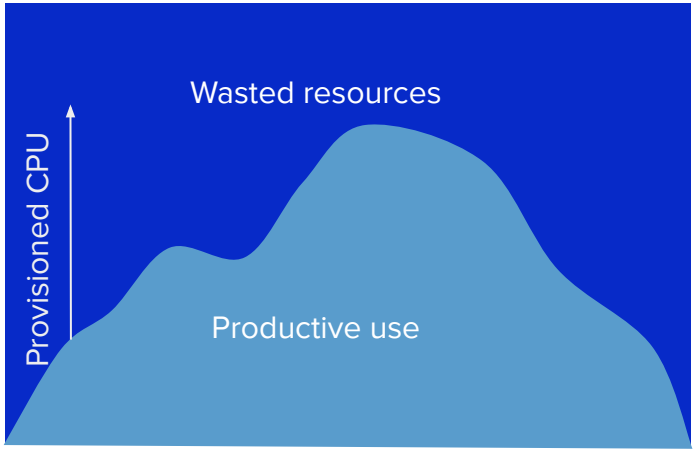


...Leading to pressure on storage and compute cost

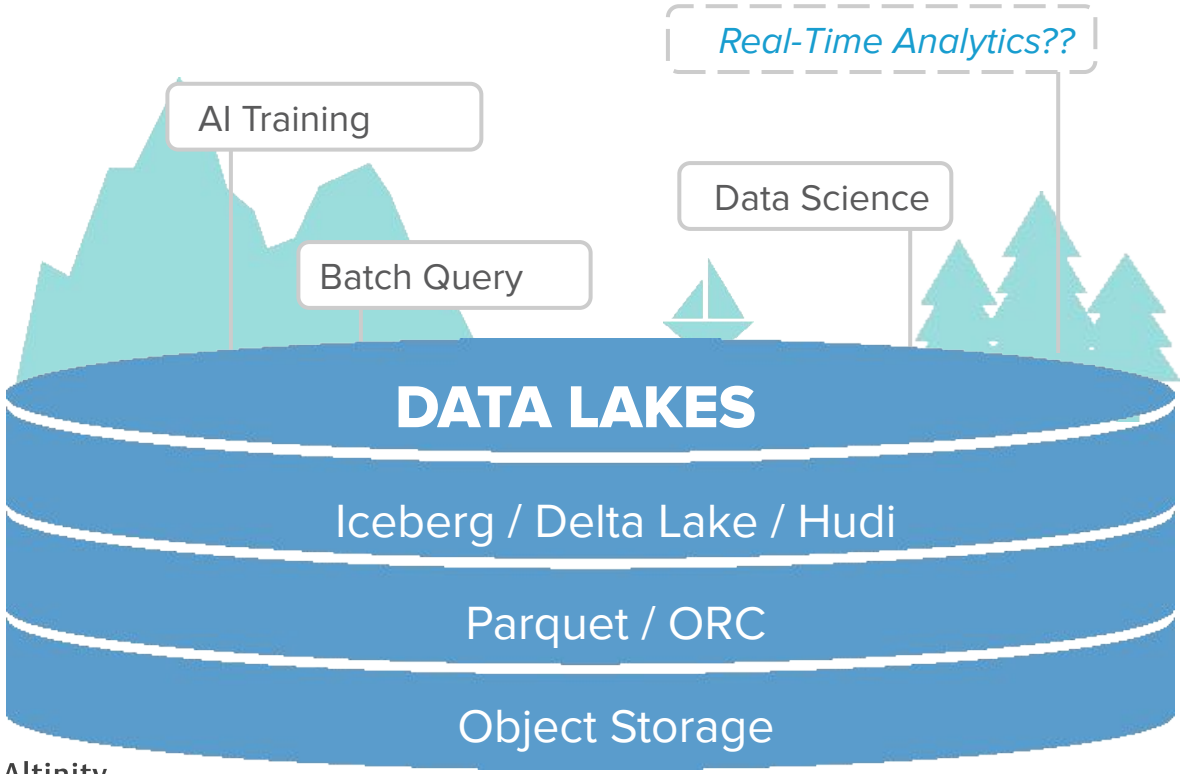
Block storage with replication is 10x more expensive



Overprovisioning wastes compute

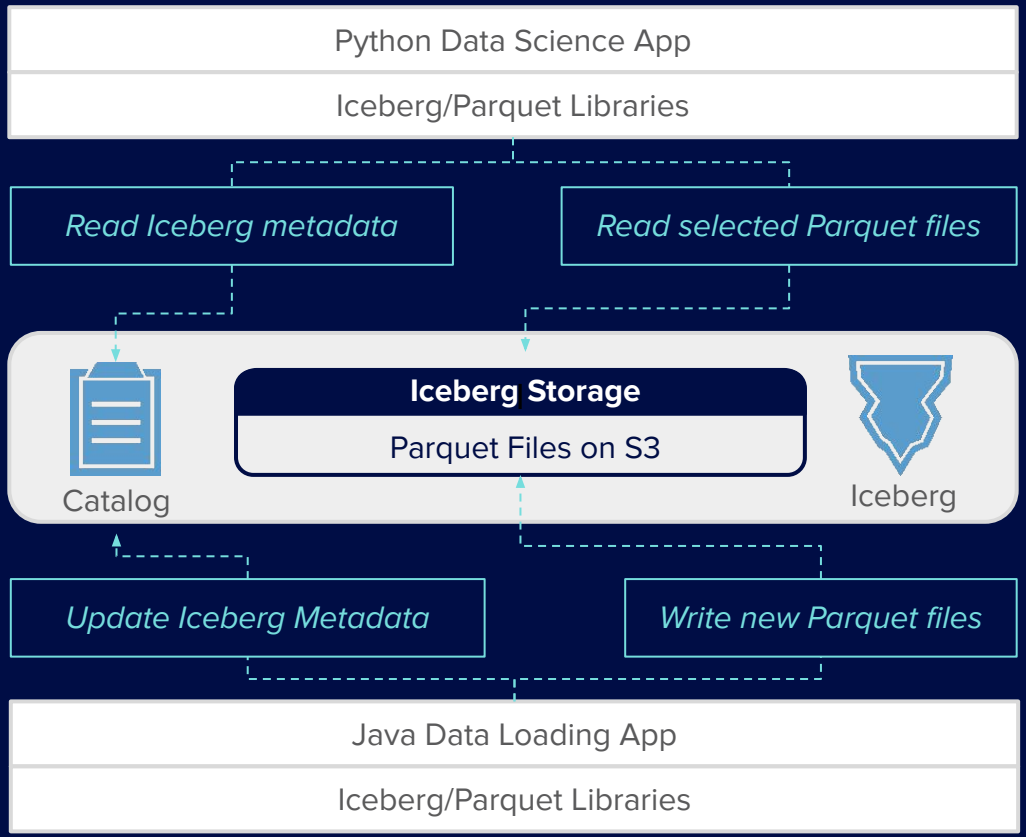


Data lakes share large tables on cheap object storage



- Shareable
- Single copy of data
- Open formats
- Run in user account

How data lakes work - there's no database!



Altinity is innovating at multiple levels to deliver shared, low-cost data with real-time response

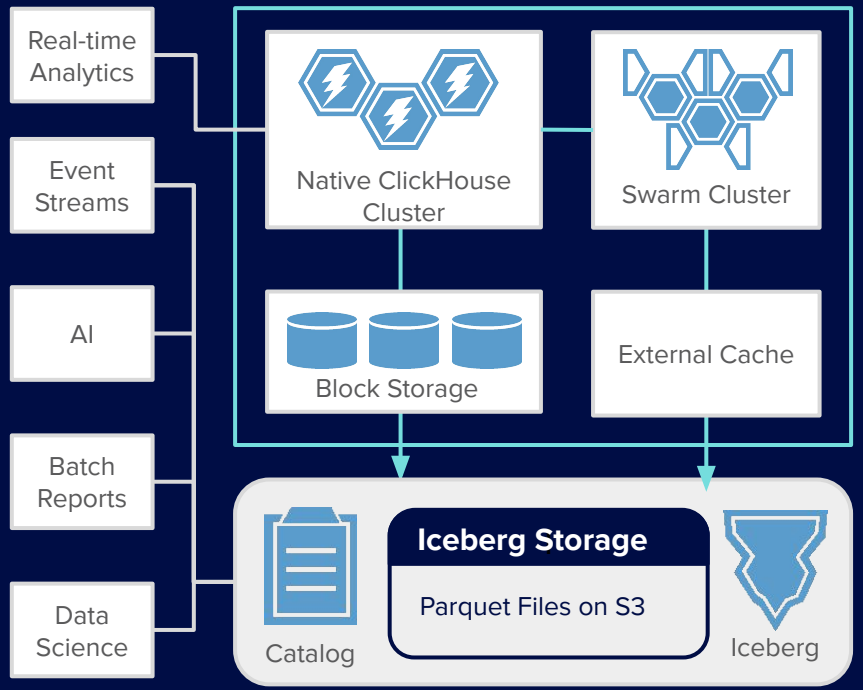
Project Antalya

BUILDS - CONTAINERS - CLOUD NATIVE BLUEPRINTS

Make I/O on Parquet very fast
Bloom filters, PREWHERE on Parquet columns, metadata caching, S3 file system caching

Apply cheap compute easily
Stateless compute swarms, distributed caching, tiered storage to Iceberg

Run in any cloud or on-prem
Kubernetes blueprints for auto-scaling, cache deployment, tooling, monitoring

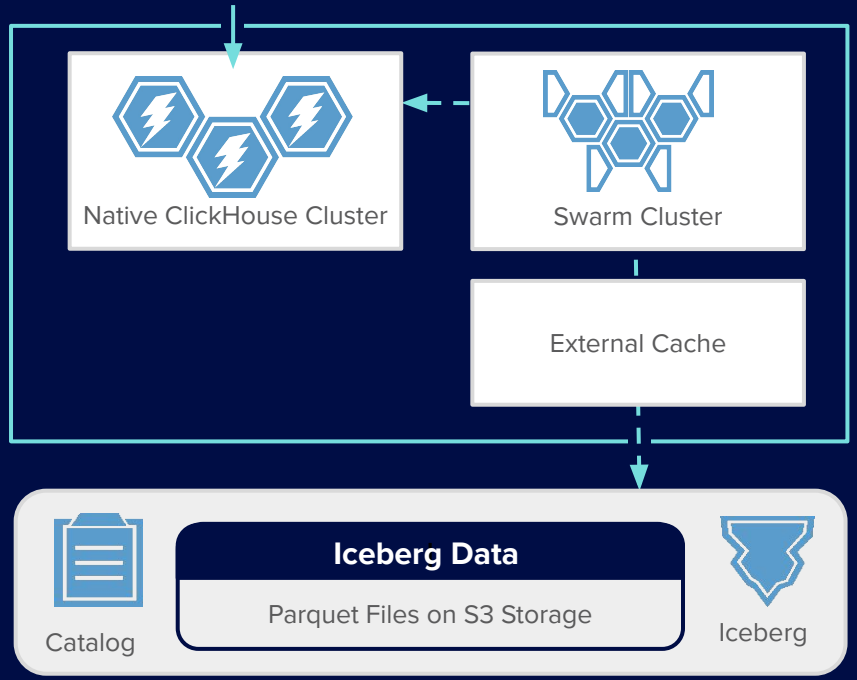


Project Antalya makes ClickHouse run fast and cheap on shared data

- All data visible from a single SQL connection
- Native tables extend seamlessly to single copy on Parquet / Iceberg
- Flexible scaling of compute on cheap spot instances
- Deploy anywhere - vendor account, your account, on-prem

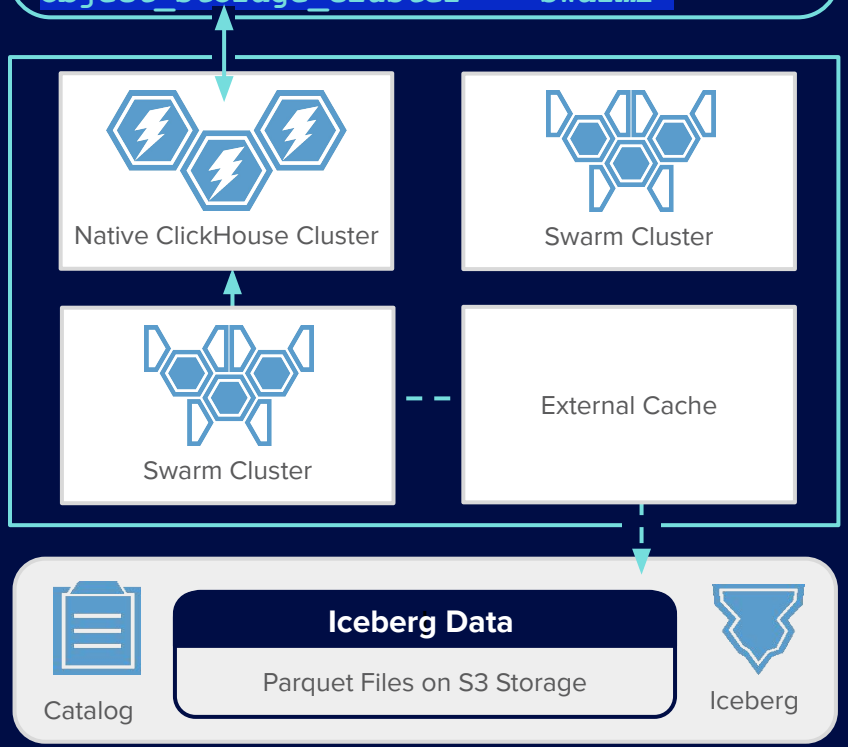
Stateless swarm clusters deliver compute storage separation

```
SELECT data, sum(output_count)
FROM iceberg.`btc.transactions`
WHERE date >= '2024-01-01'
GROUP BY date ORDER BY date
SETTINGS
object_storage_cluster = 'swarm1'
```

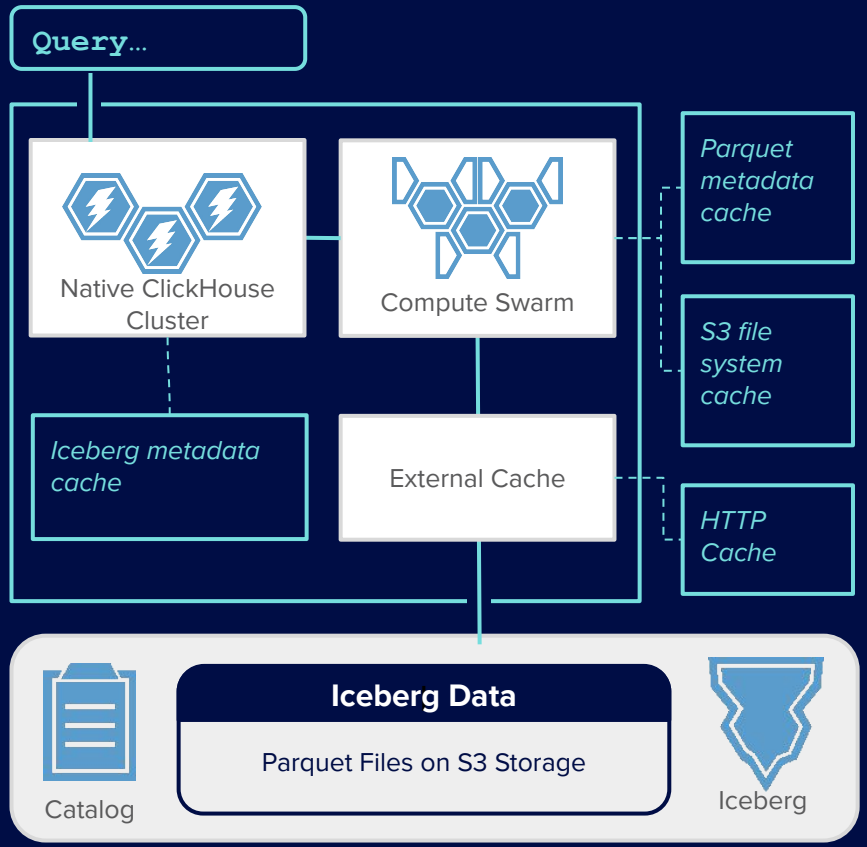


Add as many swarms as you like!

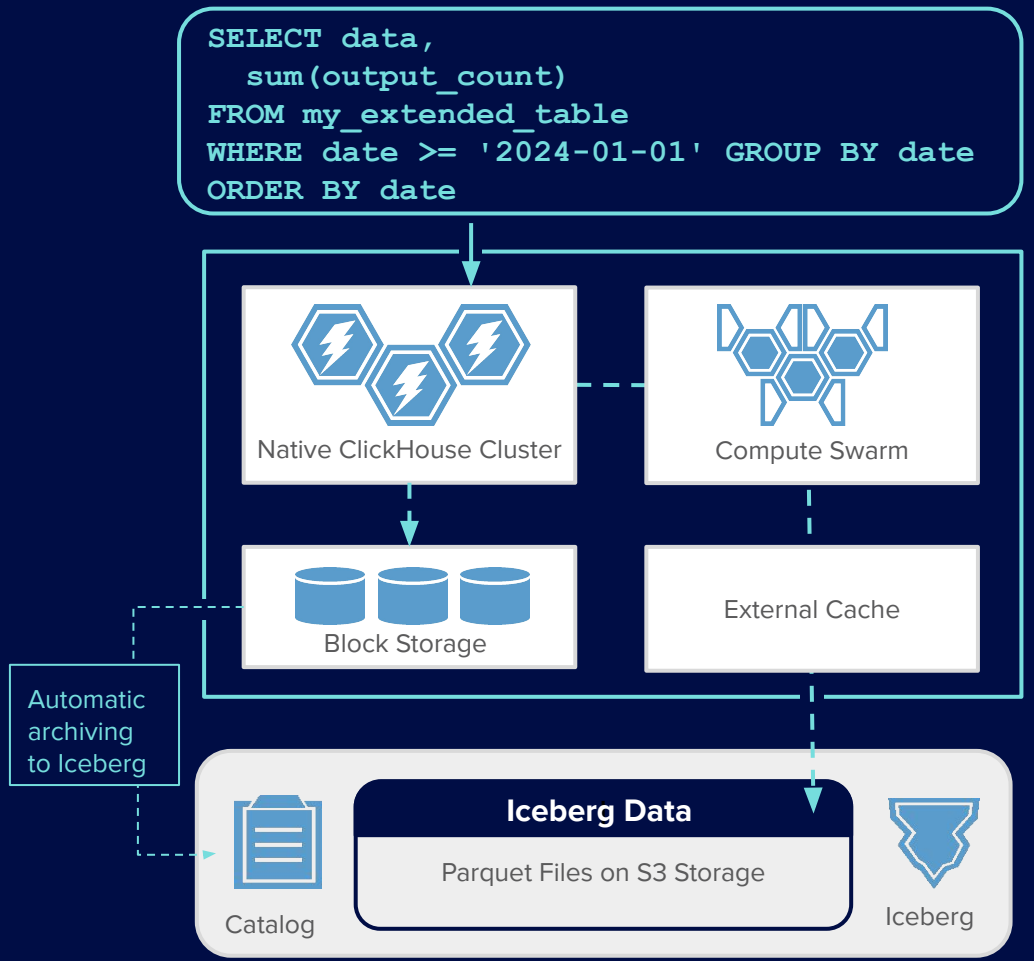
```
SELECT data, sum(output_count)
FROM iceberg.`btc.tranactions`
WHERE date >= '2024-01-01'
GROUP BY date ORDER BY date
SETTINGS
object_storage_cluster = 'swarm2'
```



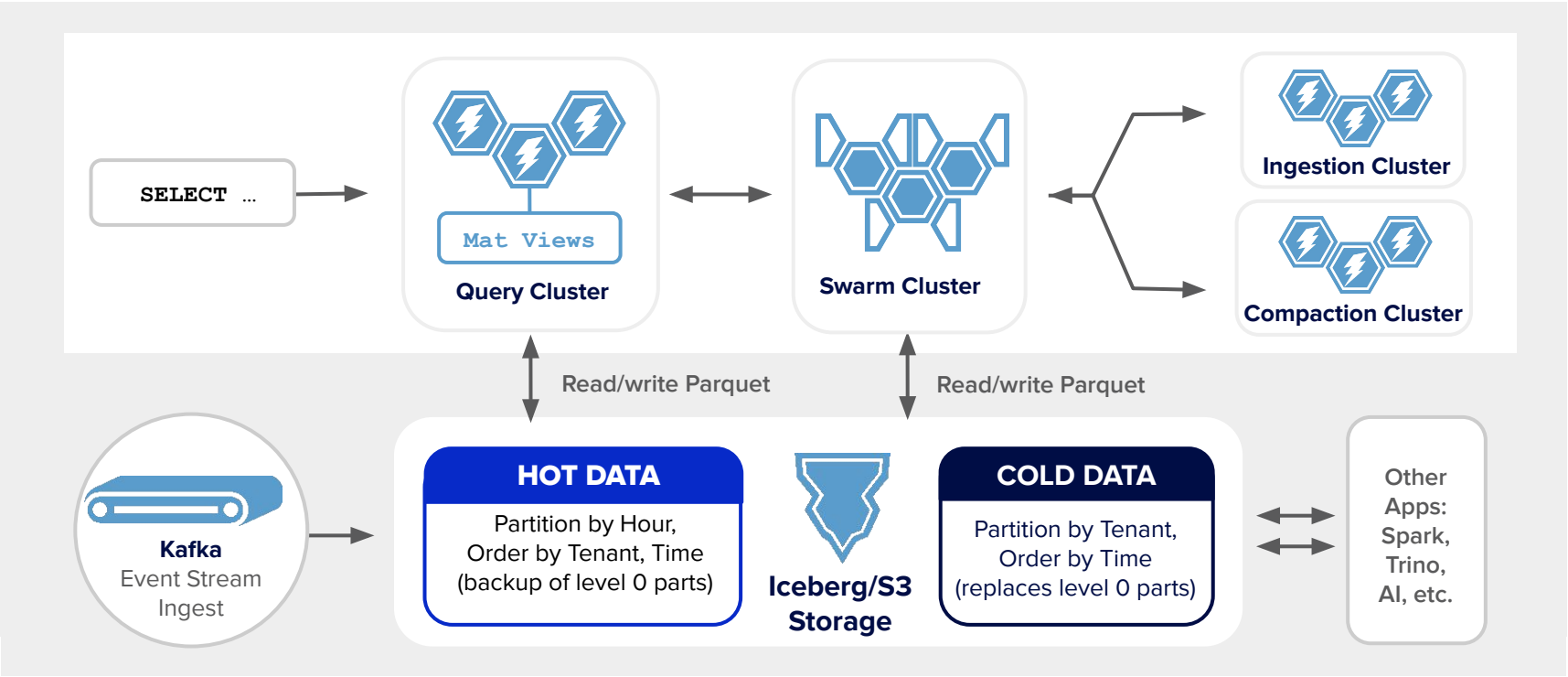
Caches make distance to Iceberg data disappear



Tiered storage to Iceberg will slash ClickHouse storage costs



Our end goal: run anywhere at any price/performance point



Swarm queries work today...

Let's see a demo!

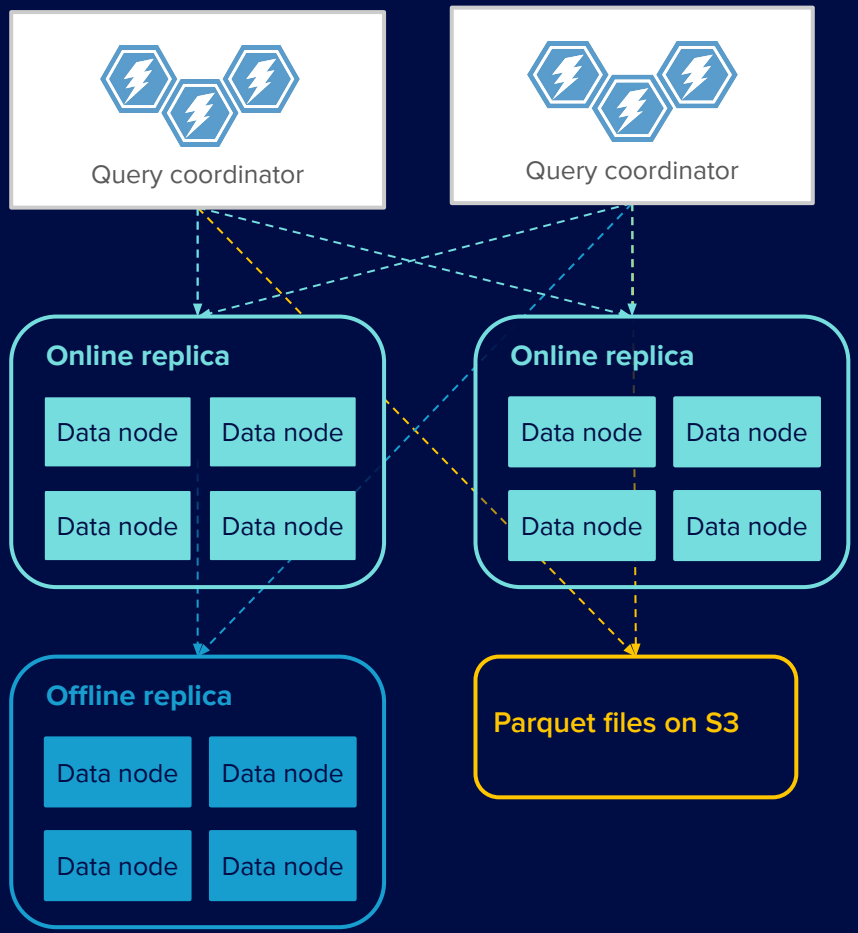
What is  PostHog?

product analytics + web analytics + session replay + data warehouse + error tracking + feature flags + experimentation + open source

=

How developers build **successful products**

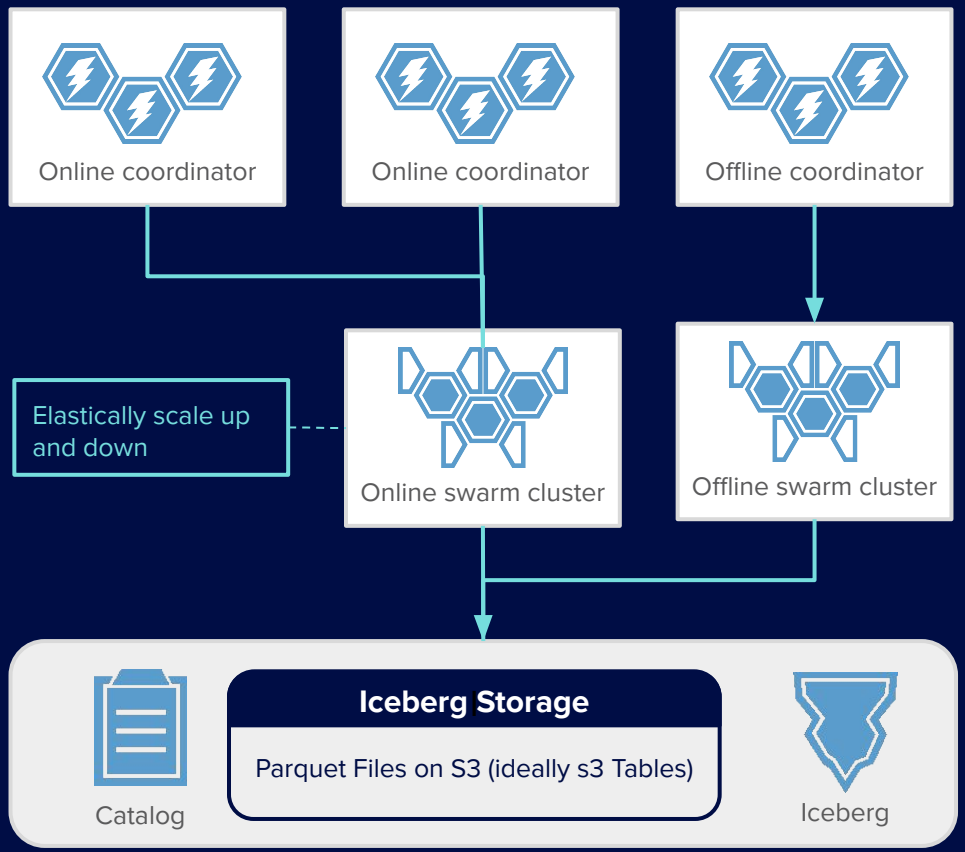
PostHog use case



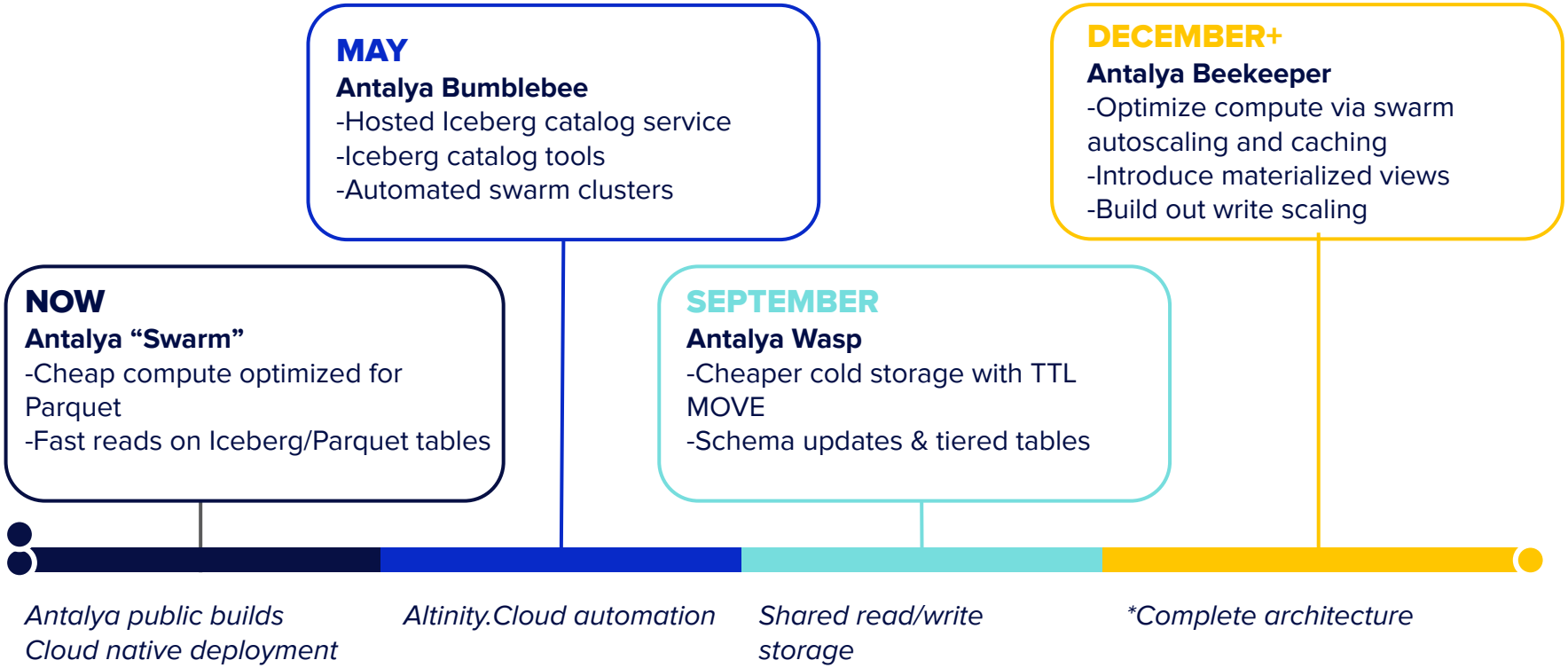
Problems/limitations of Clickhouse

- Tight coupling of compute and storage means we can’t easily scale up/down depending on workloads
    - Either have to scale every node (which means 2x increase in cost), reshard or add another replica
- All data is replicated 3x on expensive NVME and EBS disks
- No reliable caching of block storage results

PostHog future state with Antalya swarm clusters



Project Antalya 2025 implementation plan



Altinity

How is Project Antalya different from Official ClickHouse® Builds?

Open Source Features	Project Antalya	Official ClickHouse®
Fast Parquet and S3 reads	Yes	Yes
Iceberg catalog support	Yes	Yes
Swarm clusters	Yes	**No**
S3 HTTP caching	Yes	**No**
Tiered storage to Iceberg	Coming soon	**No**
Cloud native deployment blueprints	Yes	**No**

Getting started

- Clone the antalya-examples project

```
git clone git@github.com:Altinity/antalya-examples.git
```

- Set up the docker compose example and start playing around

```
cd antalya-examples/docker
(Peruse the README.md and set up Docker Compose.)
./x-up.sh
```

- Build and contribution instructions are also covered there

Contributing to the project

- **Try out Antalya** (and tell your friends!)
- Join our Slack workspace to ask questions (https://altinity.com/slack)
- Raise issues on GitHub
- Better still, submit Pull Requests (We love them.)

Summary

- Exponential data growth == **excessive costs** for ClickHouse apps
- Iceberg solves storage and delivers a single copy of data for all workloads
- Project Antalya enables fast, scalable access to Iceberg & Parquet
- The code is 100% open source and runs anywhere
- Future releases will:
  - Extend ClickHouse tables automatically to Iceberg, dropping existing storage costs
  - Integrate ClickHouse applications seamlessly with data lake storage
- Try it out and contribute. It's a community effort!

Credits

The engineering team at work

Altinity

PostHog

jumptrading

...and hundreds of customers and community friends!

Sign up for our May
Project Antalya Webinar

Thank you!
Questions?

Contact us to learn more and join our community:

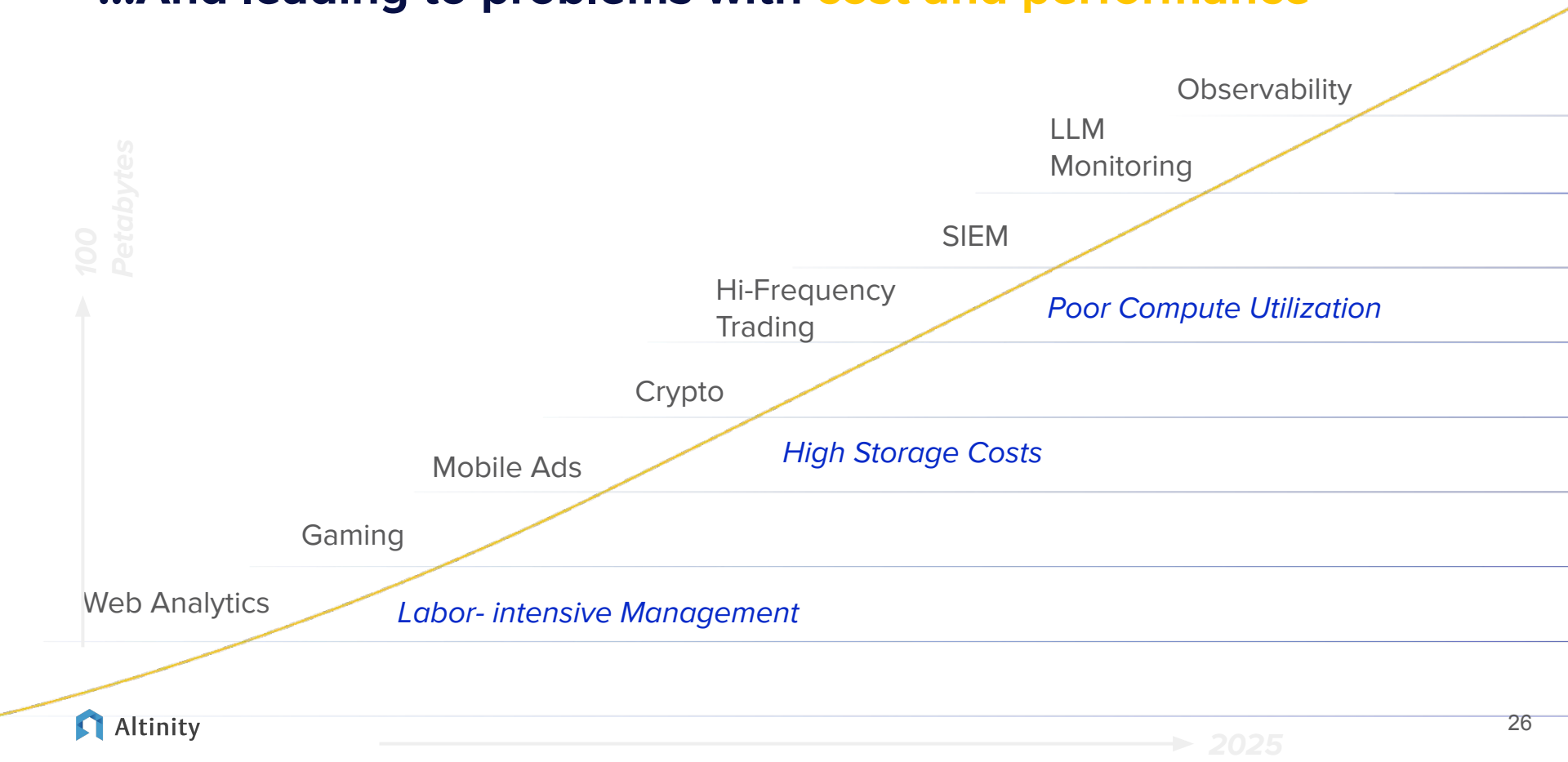
https://altinity.com

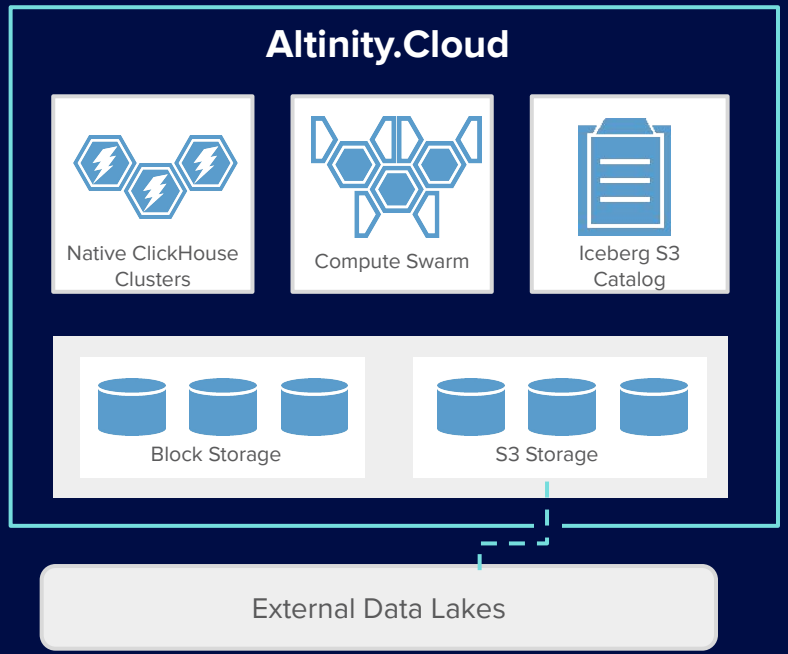
https://altinity.com/slack

https://github.com/Altinity/antalya-examples

ALTINITY

...And leading to problems with cost and performance

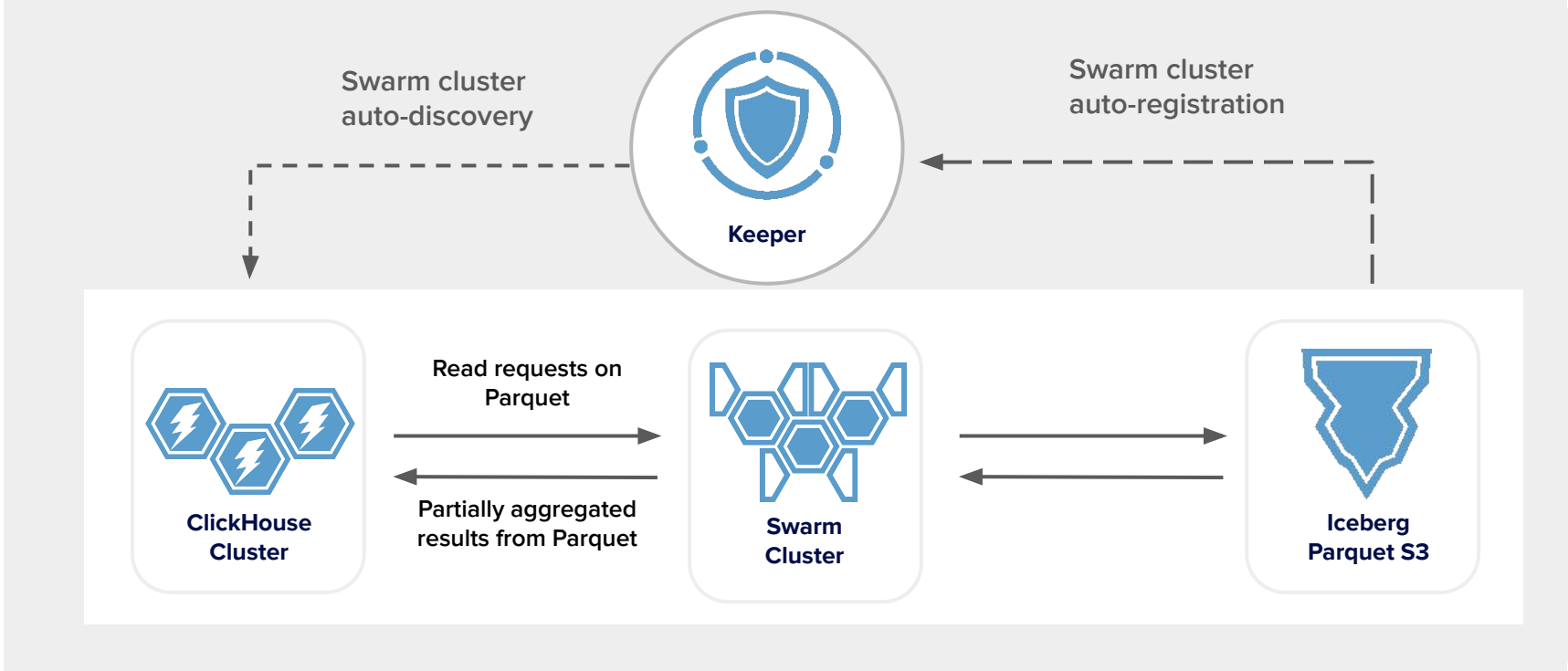




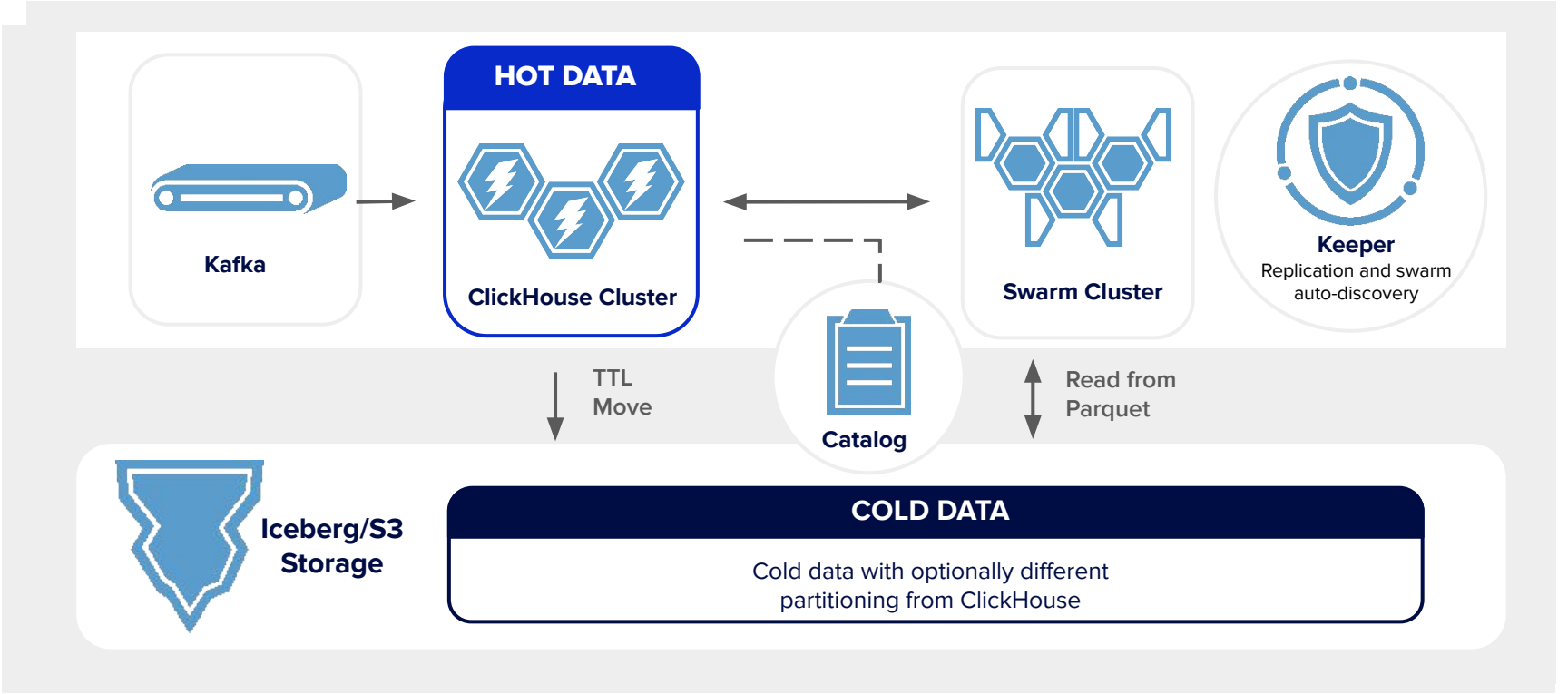
Project Antalya unites ClickHouse with flexible compute and Iceberg storage

- All data visible from a single SQL connection
- Native tables extend seamlessly to single copy on Parquet / Iceberg
- Flexible scaling of compute on cheap spot instances
- Data lake shared with external applications

Swarm clusters enable deployment of cheap, scalable compute



Tiered storage to Iceberg reduces ClickHouse costs

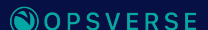






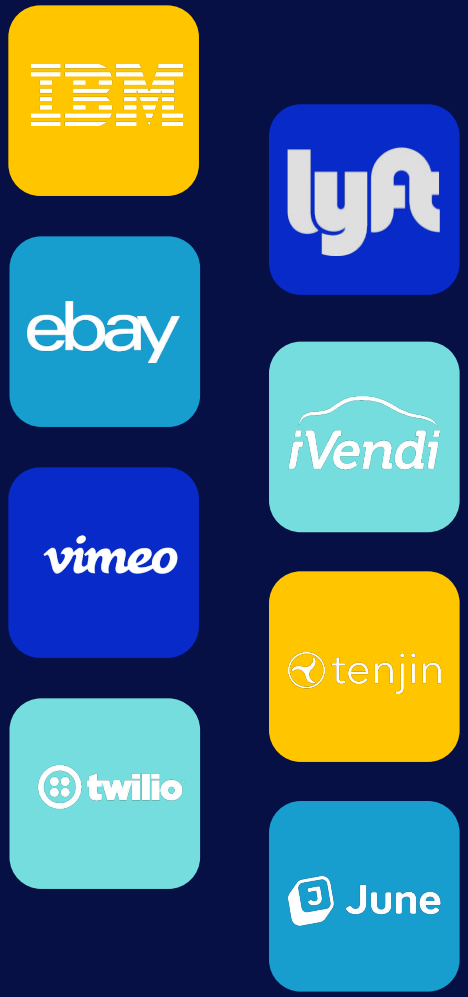




Altinity is ready to bring real-time analytics to data lakes

- Altinity.Cloud is a pioneering platform for managing ClickHouse
  - SaaS/BYOC deployment, monitoring, upgrade
  - Runs on AWS, GCP, Azure, Hetzner
- Extensive experience in Kubernetes and public cloud
- Deep expertise in ClickHouse
  - ~600 merged PRs covering Parquet, storage, security, ...
  - Maintainers of K8s operator, etc.
- 230+ customers with many use cases
  - Financial services, observability, security, gaming, LLM monitoring, ...
- We're profitable and here to stay

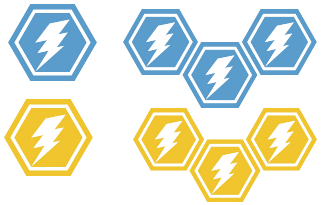


Altinity is ready to bring real-time analytics to data lakes

- Altinity.Cloud is a pioneering platform for managing ClickHouse
  - SaaS/BYOC deployment, monitoring, upgrade
  - Runs on AWS, GCP, Azure, Hetzner
- Extensive experience in Kubernetes and public cloud
- Deep expertise in ClickHouse
  - ~600 merged PRs covering Parquet, storage, security, ...
  - Maintainers of K8s operator, etc.
- 230+ customers with many use cases
  - Financial services, observability, security, gaming, LLM monitoring, ...
- We're profitable and here to stay

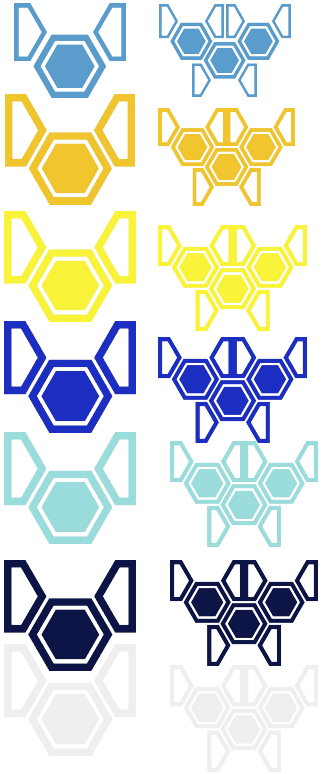
Icon Revisions

Clickhouse (Native) Cluster



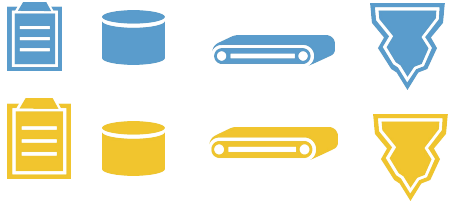
Director Cluster

Swarm Cluster



Keeper

Other



Altinity is innovating at multiple levels to deliver shared, low-cost data with real-time response

Project Antalya

- Builds
- Containers
- Cloud Native Blueprints

Iceberg, Parquet, and S3

Bloom filters, PREWHERE on Parquet, Parquet metadata caching, Iceberg catalog utilities

Data Access Improvements

Stateless compute swarms, distributed caching, tiered storage to Iceberg

Cloud Native Runtime

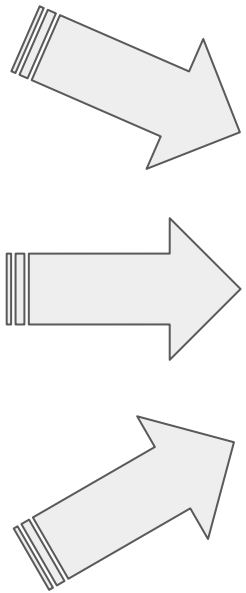
Swarm auto-scaling, cache deployment, Kubernetes cluster setup, monitoring

Altinity is innovating at multiple levels to deliver shared, low-cost data with real-time response

Iceberg, Parquet, and S3
Bloom filters, PREWHERE on Parquet, Parquet metadata caching, Iceberg catalog utilities

Data Access Improvements
Stateless compute swarms, distributed caching, tiered storage to Iceberg

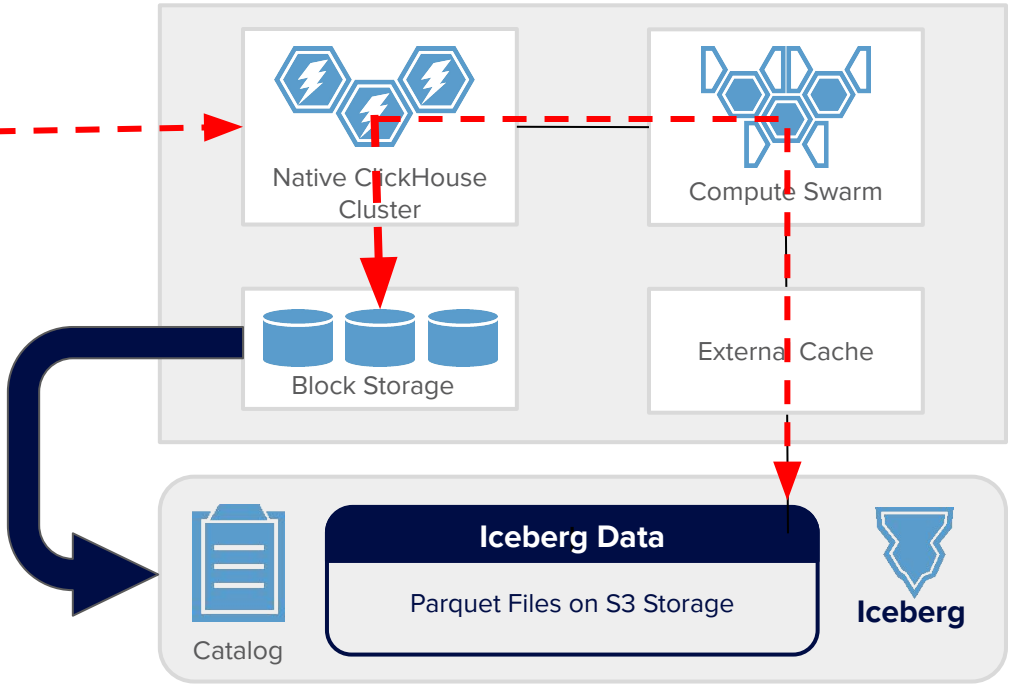
Cloud Native Runtime
Swarm auto-scaling, cache deployment, Kubernetes cluster setup, monitoring



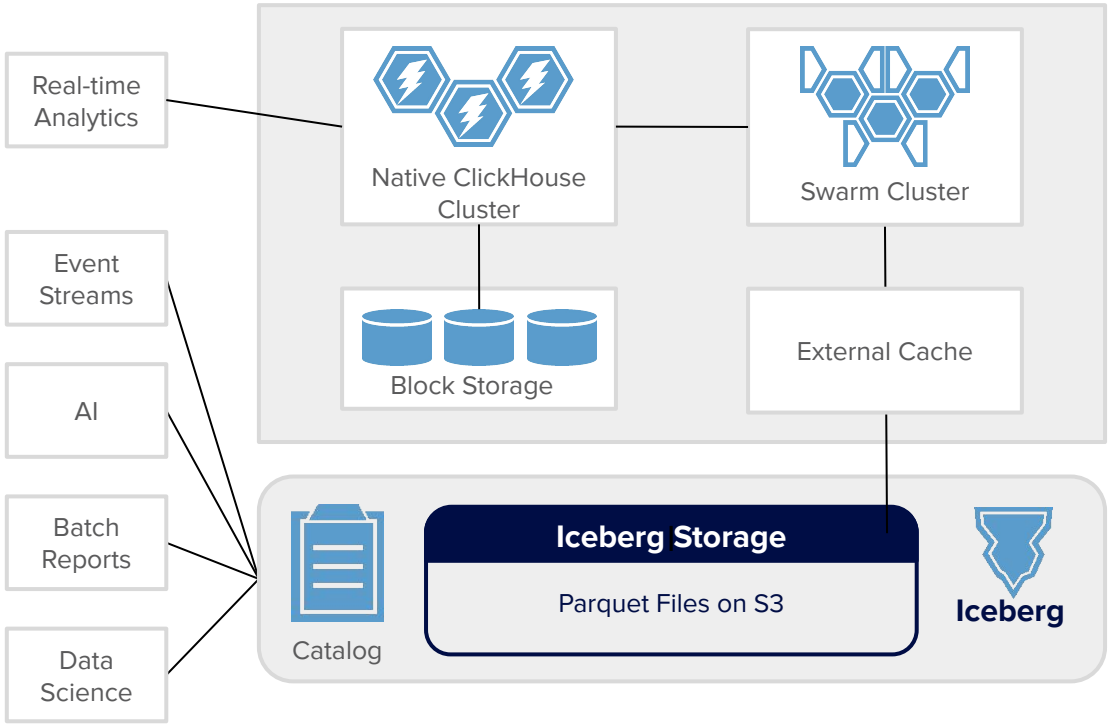
Project Antalya
Builds
Containers
Cloud Native Blueprints

Tiered storage to Iceberg will slack ClickHouse storage costs

```
SELECT data,
  sum(output_count)
FROM my_extended_table
WHERE date >= '2024-01-01'
GROUP BY date
ORDER BY date
```



Project Antalya opens up ClickHouse to cheap, shared data



All data visible from a single SQL connection

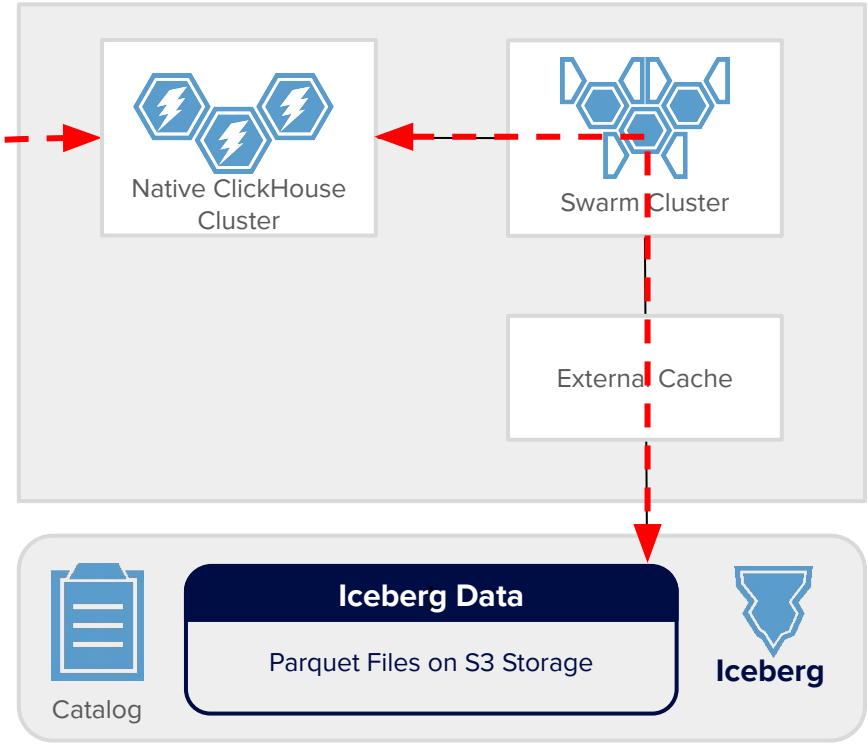
Native tables extend seamlessly to single copy on Parquet / Iceberg

Flexible scaling of compute on cheap spot instances

Deploy anywhere - vendor account, your account, on-prem

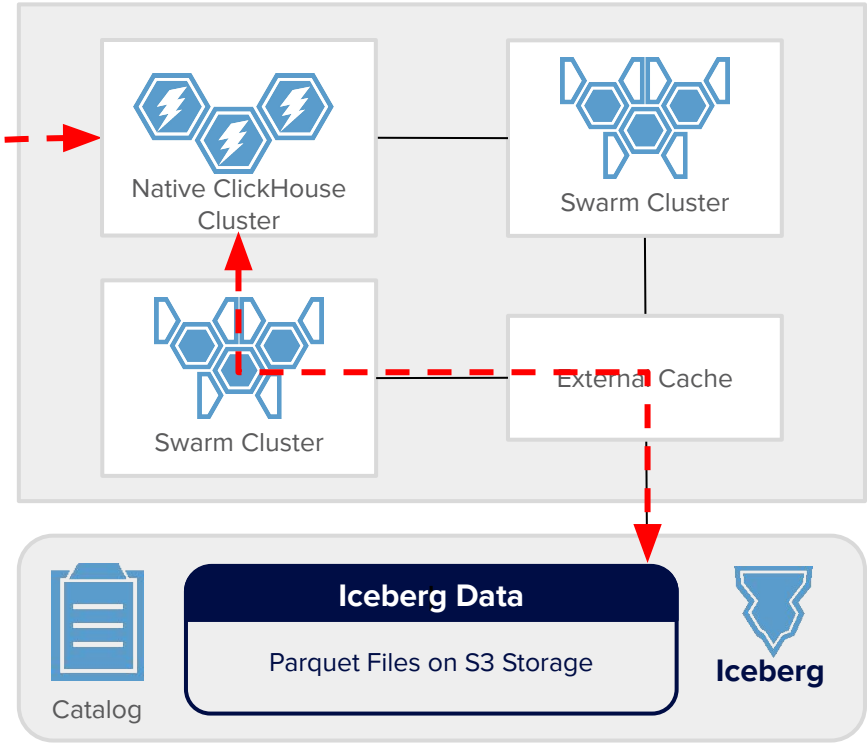
Stateless swarm clusters deliver compute storage separation

```
SELECT data, sum(output_count)
FROM iceberg.`btc.tranactions`
WHERE date >= '2024-01-01'
GROUP BY date ORDER BY date
SETTINGS
object_storage_cluster = 'swarm1'
```

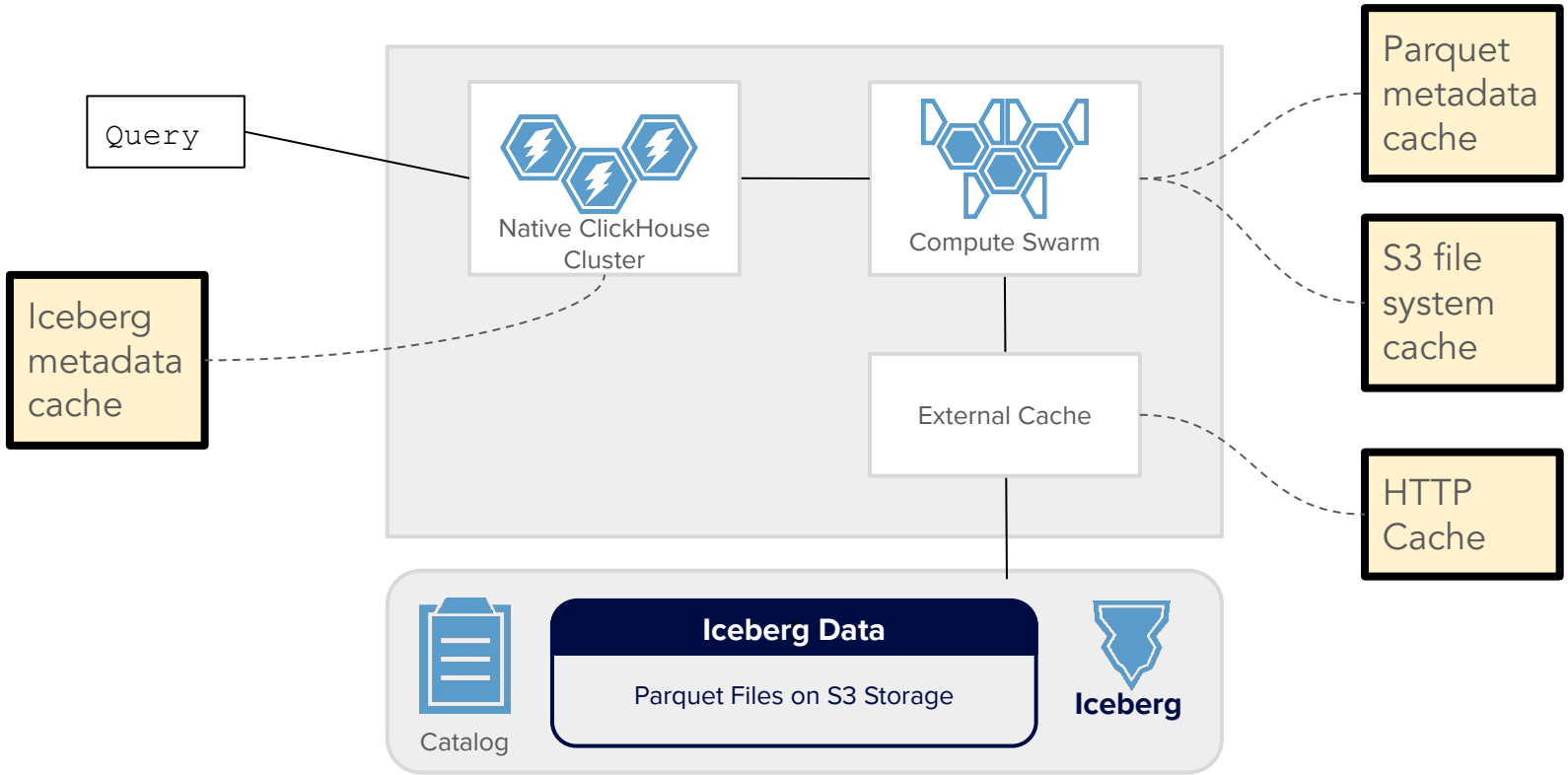


Add as many swarms as you like!

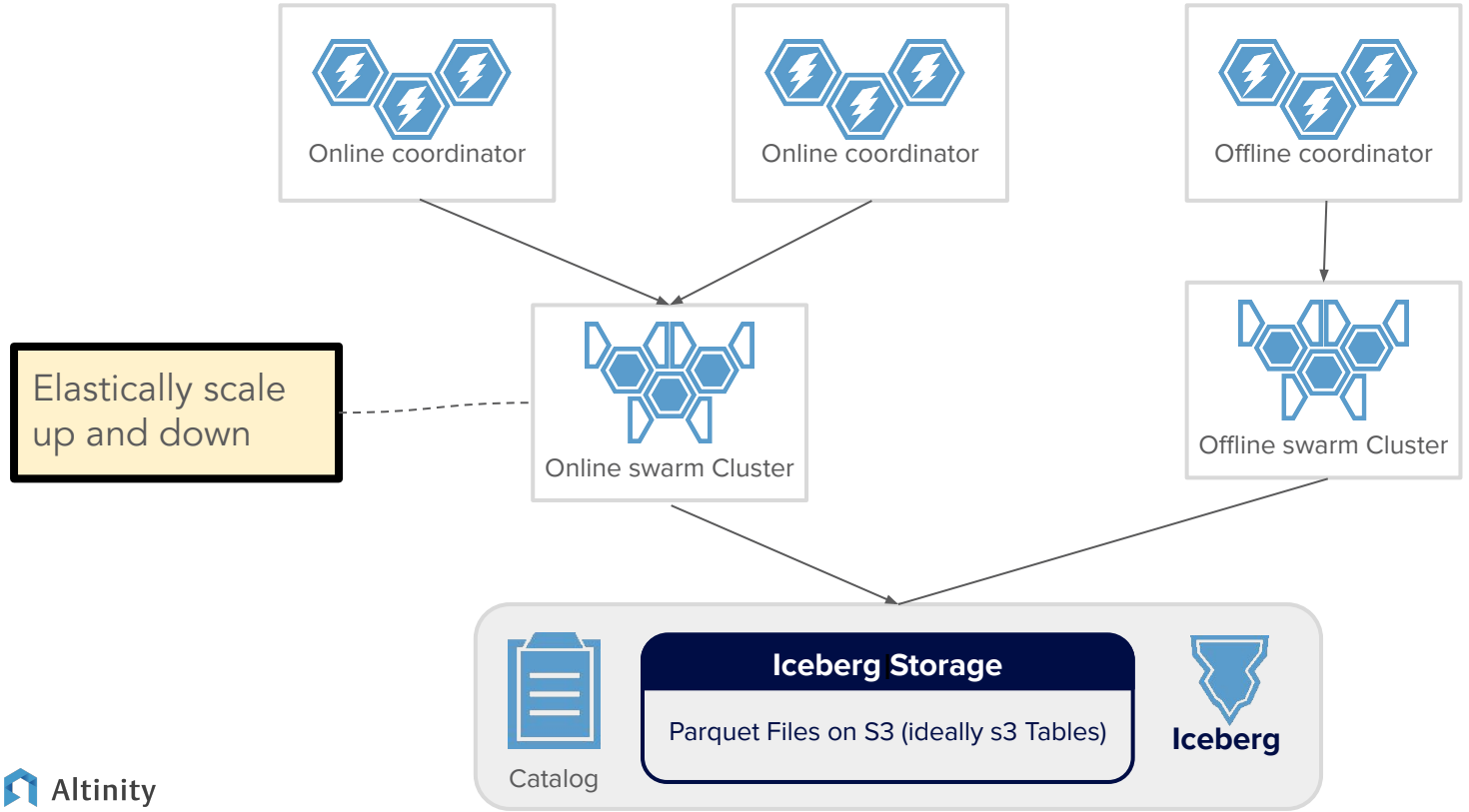
```
SELECT data, sum(output_count)
FROM iceberg.`btc.tranactions`
WHERE date >= '2024-01-01'
GROUP BY date ORDER BY date
SETTINGS
object_storage_cluster = 'swarm2'
```



How do caches work?



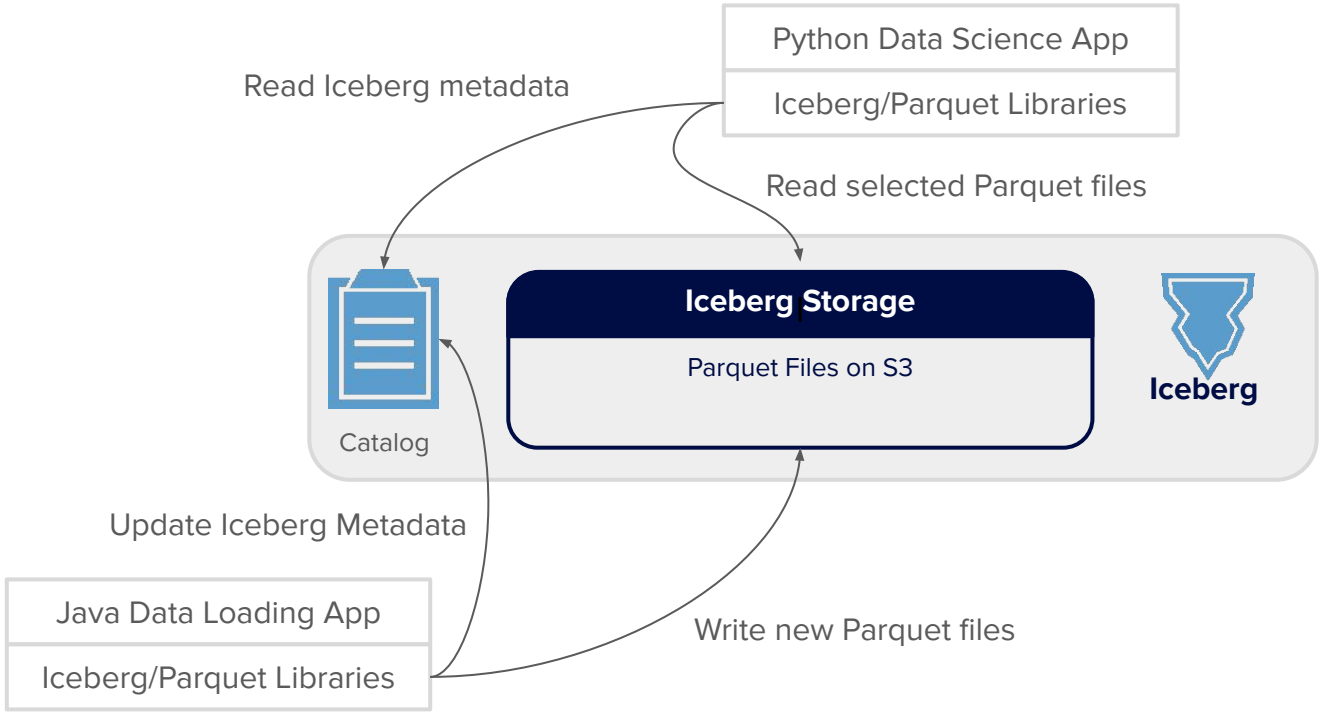
Future state with Antalya



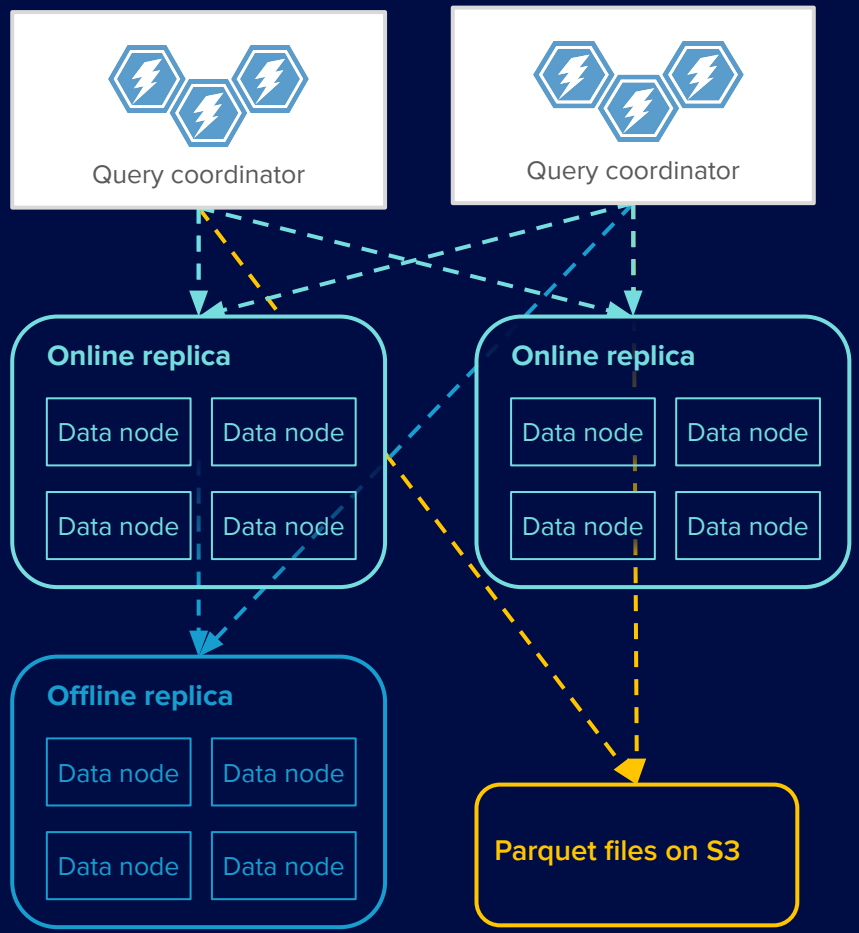
Swarm queries work today...

Let's see a demo!

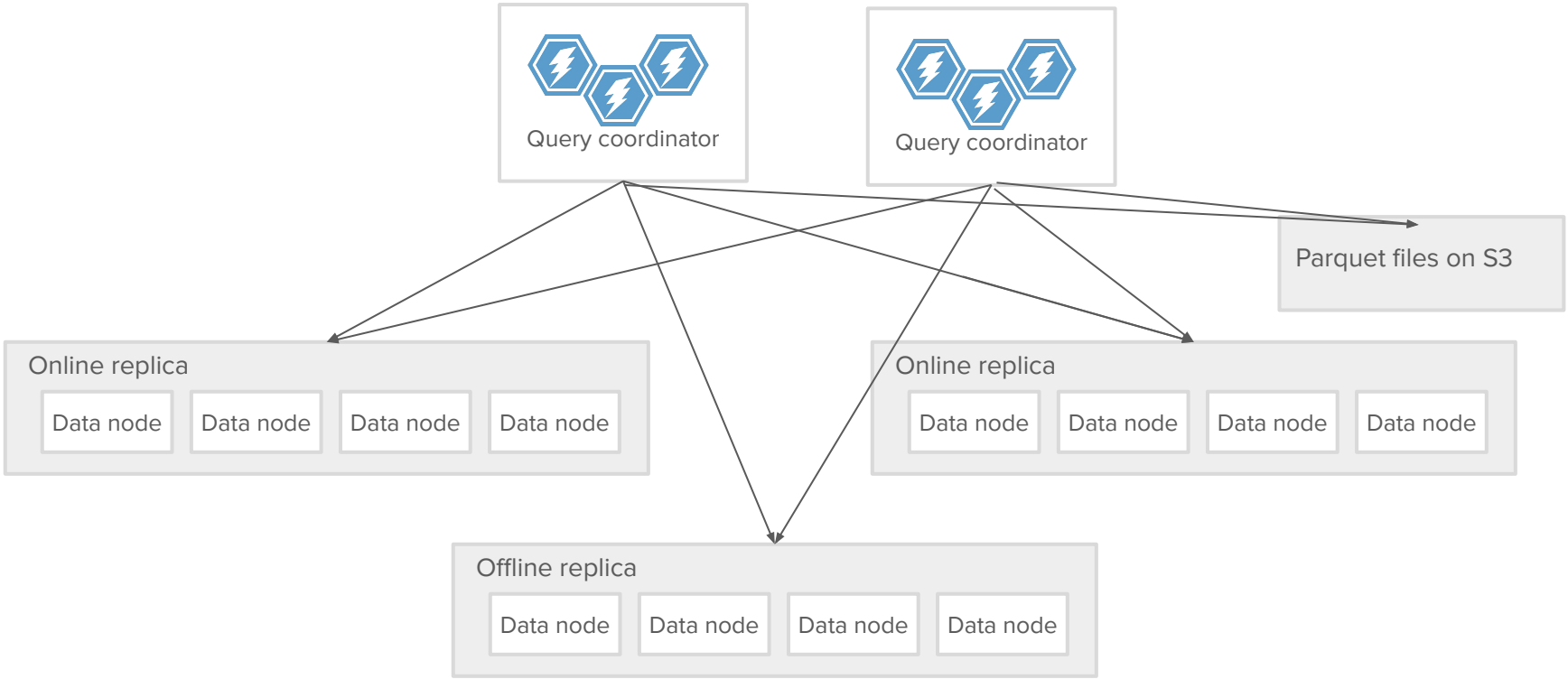
How data lakes work - there's no database!



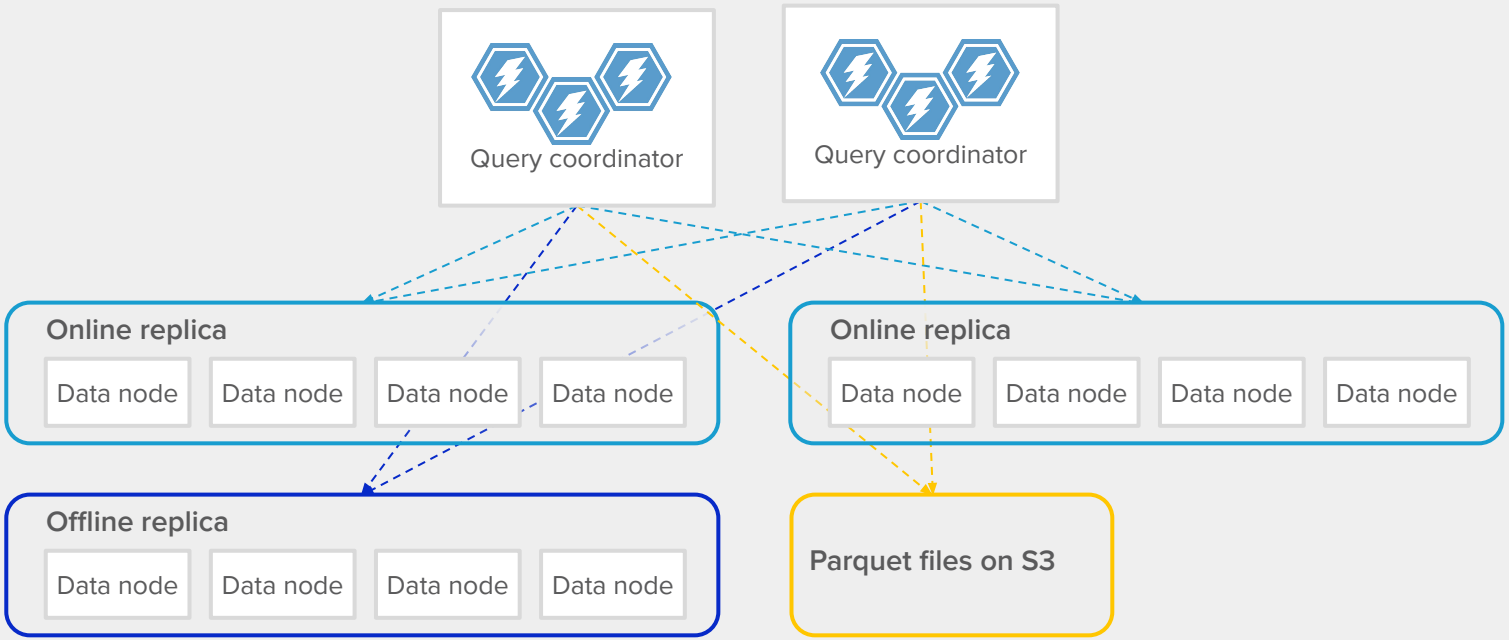
PostHog use case



PostHog use case



PostHog use case



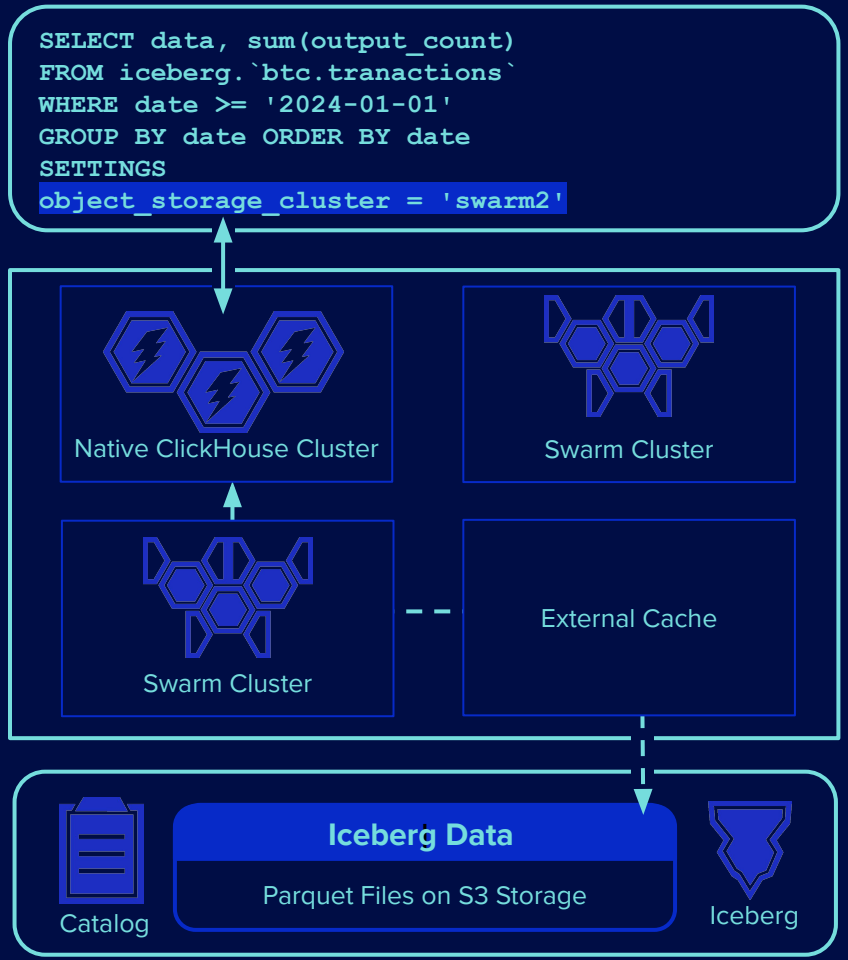
What is PostHog?

PostHog

Product Analytics + web analytics + session replay + data warehouse + error tracking + feature flags + experimentation + open source

= how developers build successful products

Stateless swarm clusters deliver compute storage separation



Our end goal: run anywhere at any price/performance point