
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

PRT085

SPONSORED BY DREMIO

Build your open data lakehouse with Dremio and AWS

Brett Roberts

Principal Solutions Architect
Dremio

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

About Dremio



The easy and open data lakehouse platform

- Data warehouse performance directly on the lake
- Query acceleration to eliminate copies and business intelligence (BI) extracts
- Semantic layer to enable governed self service
- Database connectors to enable queries on other sources

Enterprise adoption

- Thousands of companies across all industries
- 5 of the Fortune 10

Open source and large community

- Apache Arrow (60M+ downloads/min), Apache Iceberg, Project Nessie
- Creator and host of Subsurface LIVE conference

BARK&BOX DIAGEO TELENAV

Comerica Bank JPMorgan Chase & Co. NCR

TransUnion FACTSET NUTANIX

New York Life RBC software AG

Moody's Wells Fargo Honeywell

GM EY Fannie Mae

Adobe FACTSET standard chartered

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.


What is a lakehouse?

Data warehouse



Data lake



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

dremio



Data warehouse

- ✓ Supports transactions
- ✓ Provides data quality
- ✗ Potentially costly
- ✗ Proprietary data formats
- ✗ Provides limited scope of data

Really good at BI and reporting based on structured data

Data lake



- ✓ More cost effective
- ✓ Provides full scope of data
- ✓ Built on open formats
- ✗ Can be hard for data consumers to use
- ✗ May lack critical data quality features

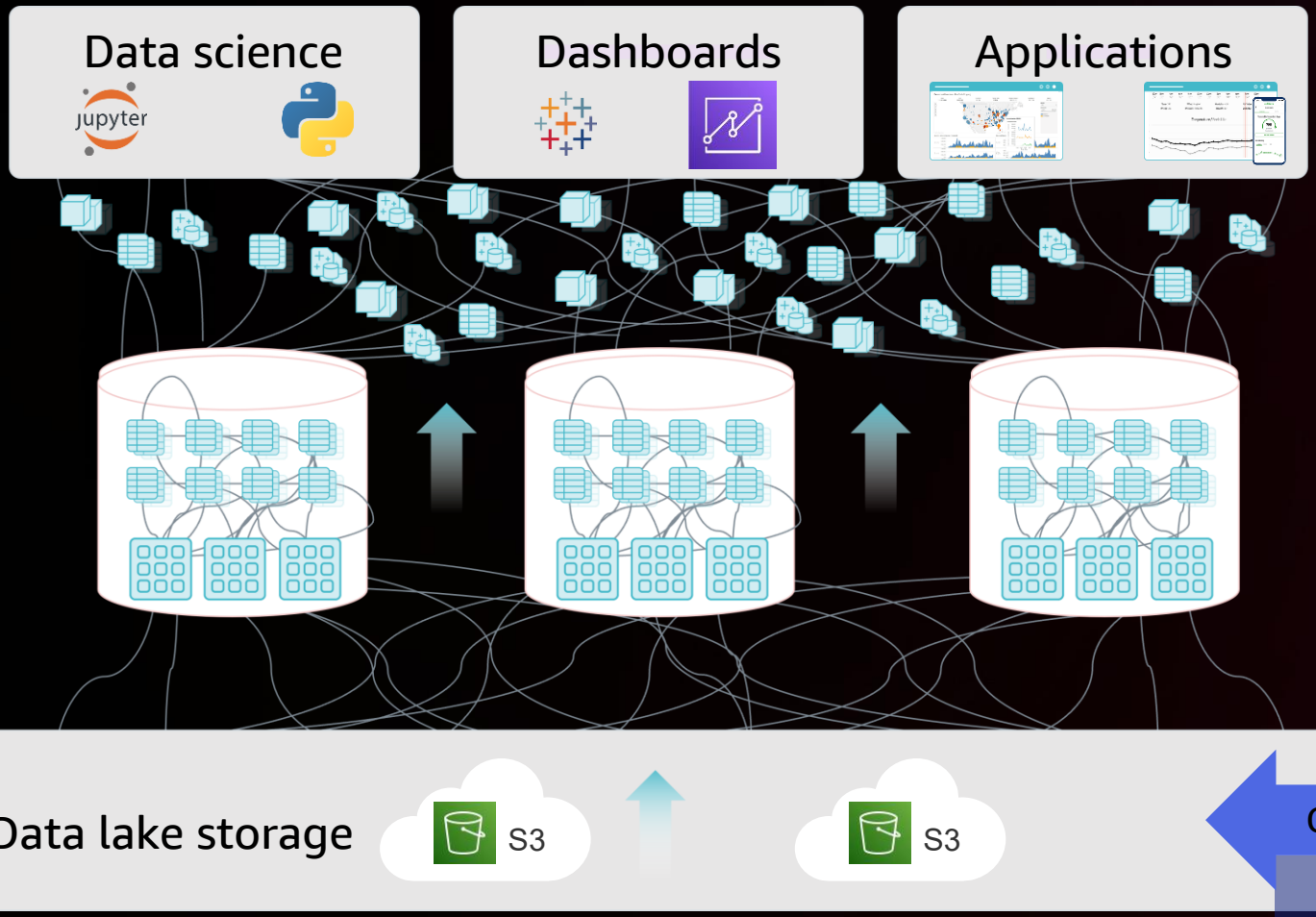
Really good at exploratory data science

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

dremio

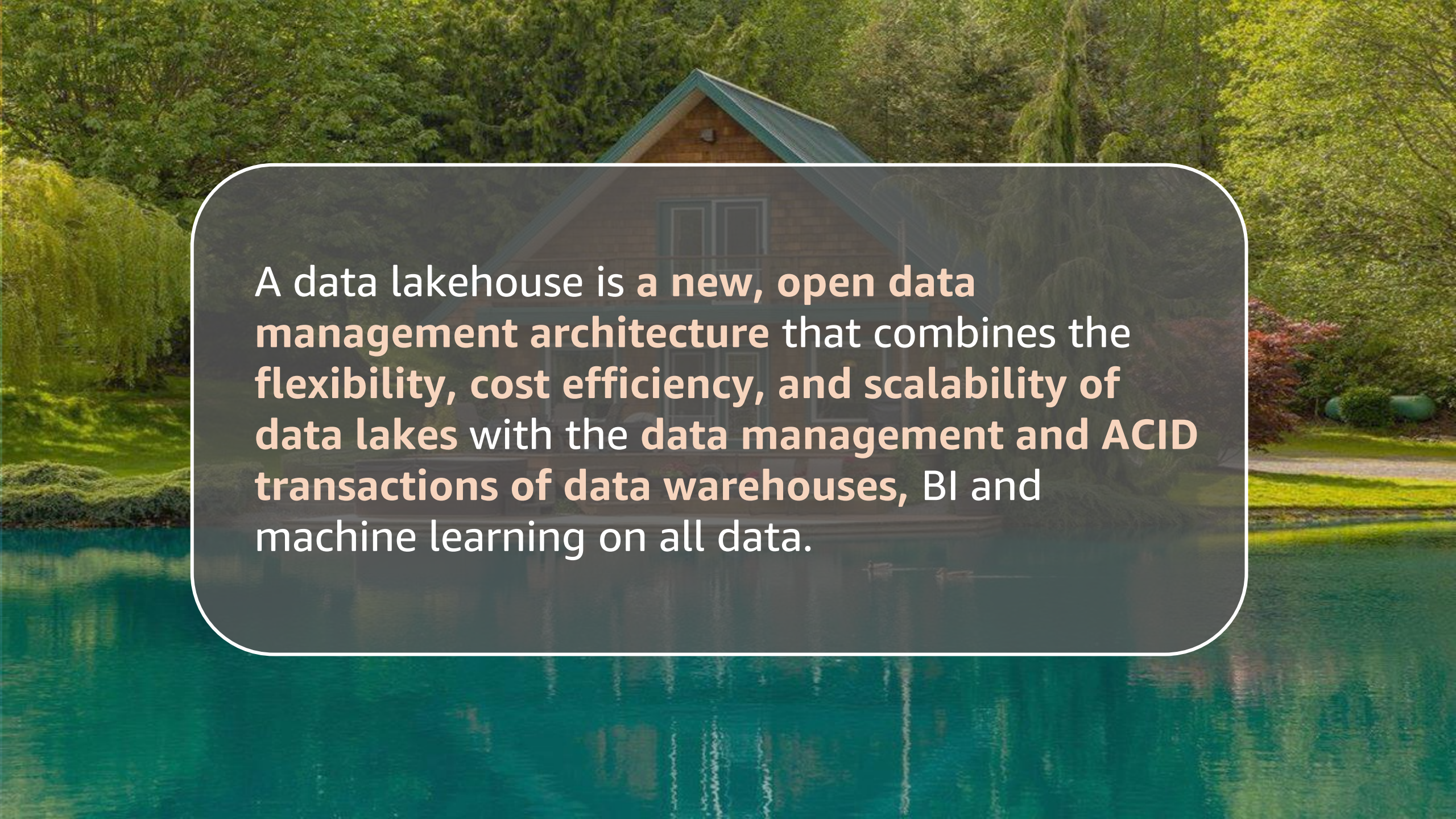
Traditional data architectures can be expensive, proprietary, or complex



- Skyrocketing costs
- Vendor lock-in
- Growing requests backlog
- Difficulty exploring data
- Lack of self-service

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.




A data lakehouse is a new, open data management architecture that combines the flexibility, cost efficiency, and scalability of data lakes with the data management and ACID transactions of data warehouses, BI and machine learning on all data.

Key components of a lakehouse



Query engine

Query data in place

Warehouse performance on the data lake

Self-service semantic layer







Transactional tables

ACID transactions on the data lake

Time travel/versioning

Easy schema/partition evolution







Data lake storage

Highly scalable

Stores data in open file and table formats

Always available



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Transactional tables on the data lake

Record-level data mutations with SQL DML

```
INSERT INTO t1 ...
UPDATE t1 SET col1 = ...
DELETE FROM t1 WHERE state = 'CA'
```

Automatic partitioning

```
CREATE TABLE t1 PARTITIONED BY (month(date),
bucket[1000](user))
```

Instant schema and partition evolution

```
ALTER TABLE t1 ADD/DROP/MODIFY/RENAME COLUMN c1 ...
ALTER TABLE t1 ADD/DROP PARTITION FIELD ...
```

Time travel

```
SELECT * FROM t1 AT/BEFORE <timestamp>
```

ICEBERG

- Created by Netflix, Apple, and other big tech
- INSERT/UPDATE/DELETE with any engine
- Strong momentum in OSS community

DELTA LAKE

- Created by Databricks
- INSERT/UPDATE with Spark, SELECT with any engine
- Primarily used in conjunction with Databricks

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

dremio

The Open Data Lakehouse with Dremio and AWS

Data science

Dashboards

Applications



Consumer self-service
Semantic layer

BI and one-time performance
Query accelerator

Query data in place
SQL engine

Automatic data management
Intelligent metastore

Data lake storage
Scalable and available storage

Dremio Sonar

An open lakehouse engine powered by Apache Arrow

Dremio Arctic

An intelligent metastore for Apache Iceberg

RDBMS

Amazon S3

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The Open Data Lakehouse with Dremio and AWS

Data governance and security

Consumer self-service
Semantic layer

BI and one-time performance
Query accelerator

Query data in place
SQL engine

Automatic data management
Intelligent metastore

Data lake storage
Scalable and available storage

Data science



Dashboards



Applications



Dremio Sonar
An open lakehouse engine powered by Apache Arrow

Dremio Arctic
An intelligent metastore for Apache Iceberg

RDBMS

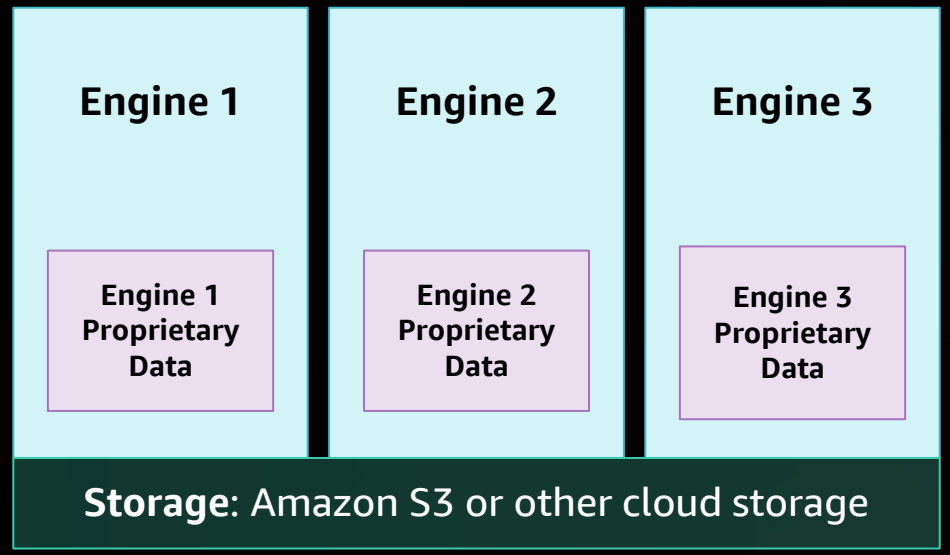
Amazon S3

AWS Lake Formation

aws

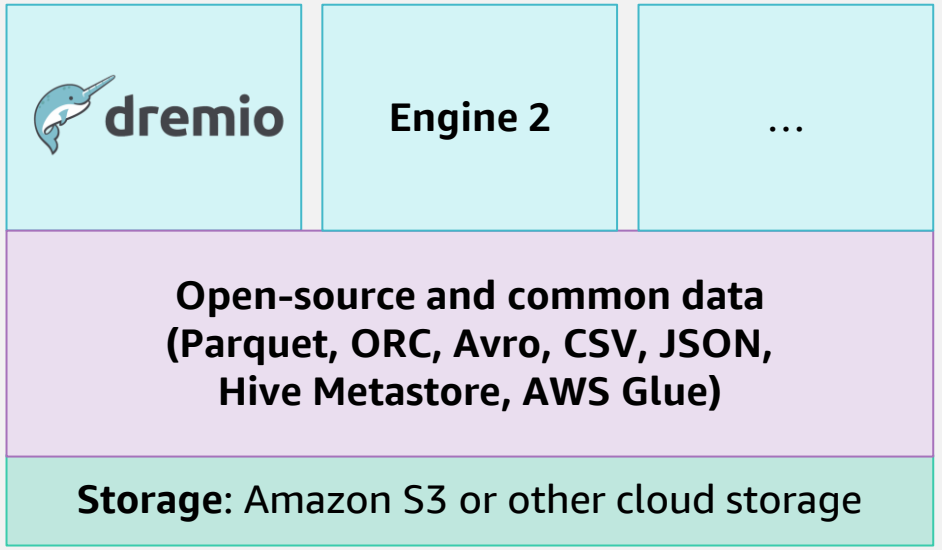
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Closed data architecture



- ✗ Can't use multiple engines on the same data or use cases
- ✗ Limited to functionality of one engine
- ✗ Can't adopt new engines in the future

Open data architecture



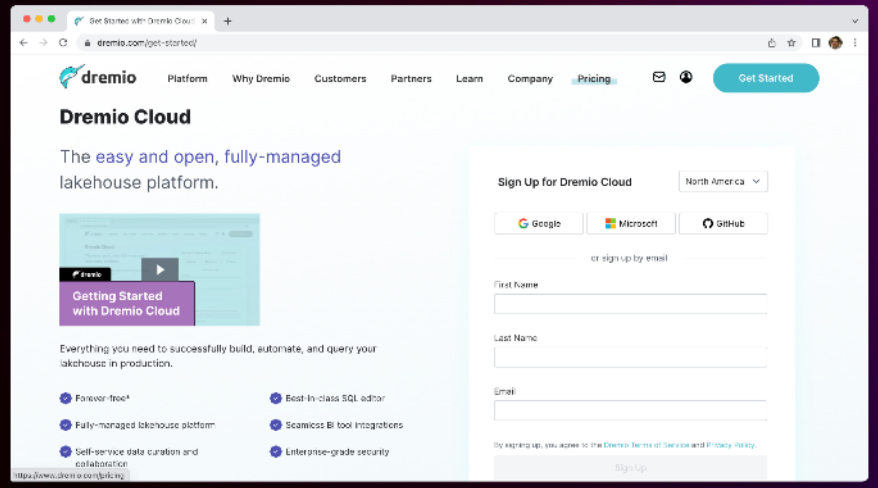
- ✓ Flexibility to use best-in-class engines on the same data and use cases
- ✓ Easy to adopt additional engines
- ✓ Easy to adopt new engines in the future; simply point them to the data

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

See value from Dremio in minutes



Stop by Dremio booth (3827) for a demo
or to access the Dremio sandbox



Sign up for Dremio Cloud and
connect to your cloud account

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.