



Lake House with Dremio

Wolfgang Disselhoff
Dremio, Solution Architect

#OSW2021



RETE ITALIANA
OPEN SOURCE



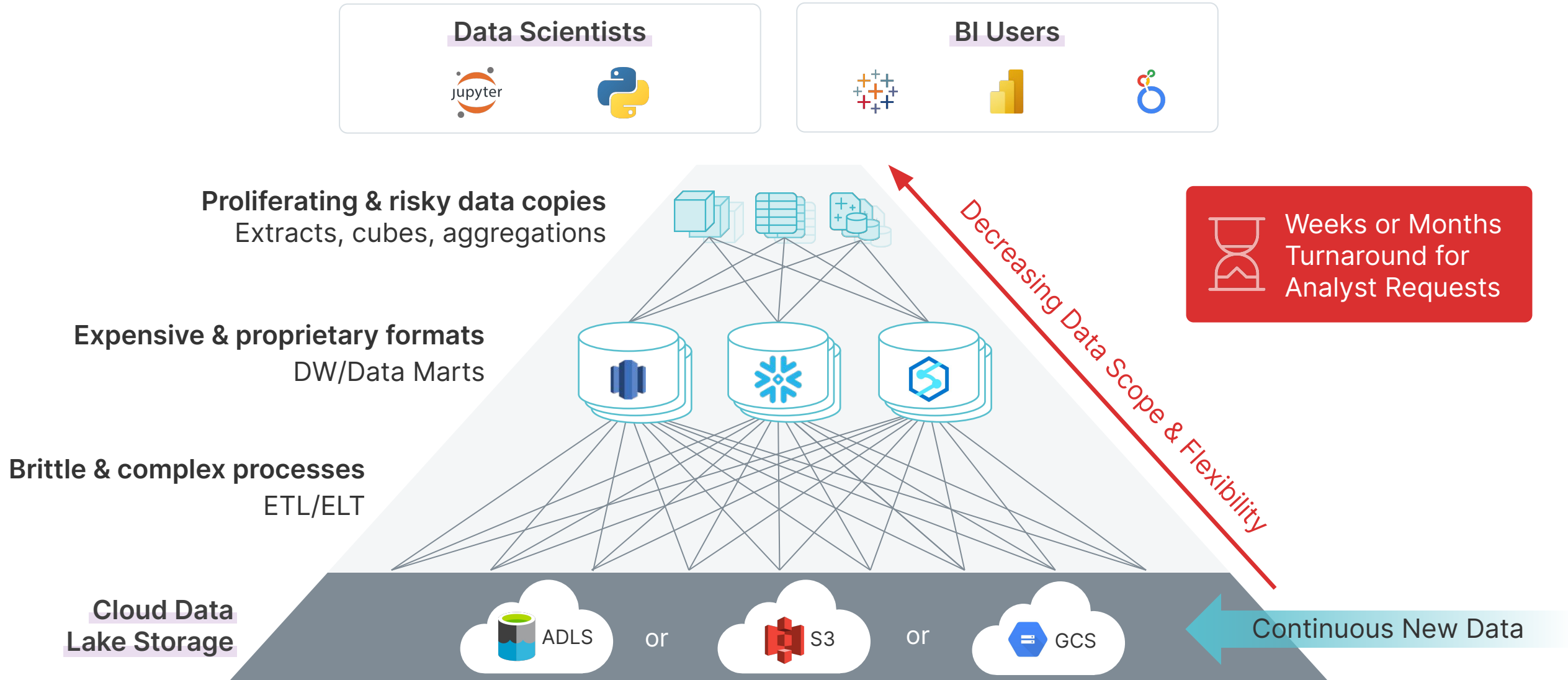
Agenda

What problem do we solve ?
Where do we fit in the OSS world ?
Demo

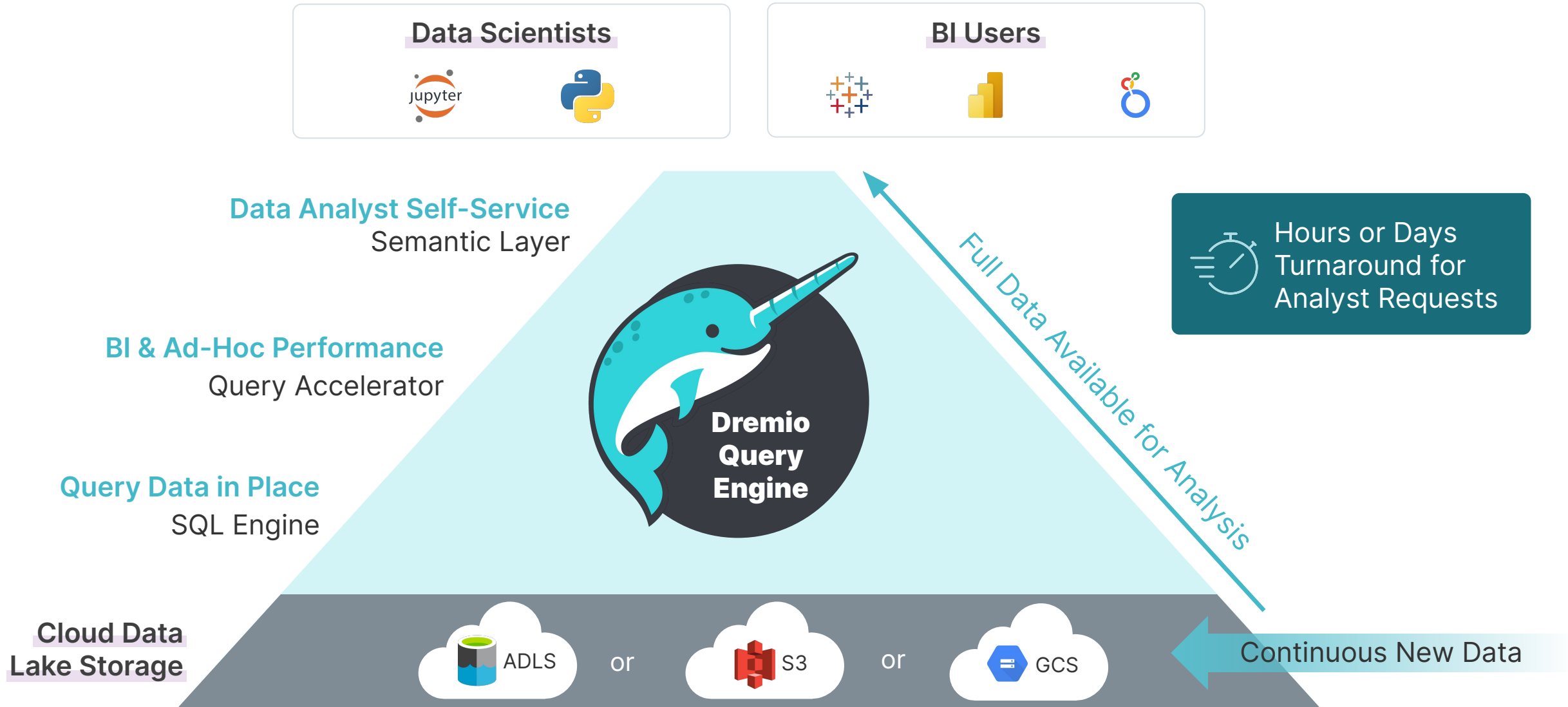


RETE ITALIANA
OPEN SOURCE

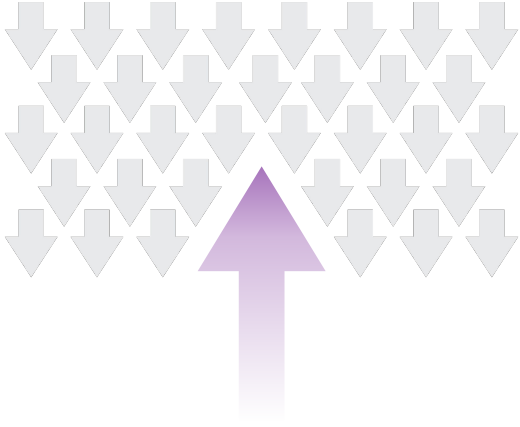
What is the Problem ?



How to solve ? A No Copy Architecture



Who is Dremio ?



Unicorn Status

Over \$200M Raised in 2020

\$1B Valuation

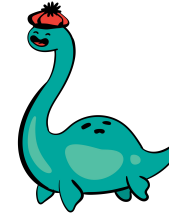
Strong Strategic Partnerships
Like Amazon, Microsoft, Tableau



Enterprise Customer Adoption

Industry-wide Customer Success

> 2x Revenue and Customer
Count YoY



Nessie

Open Source Innovation

Creator of Arrow & Arrow Flight

20M Downloads/Month

Creator of Nessie



10K+
attendees

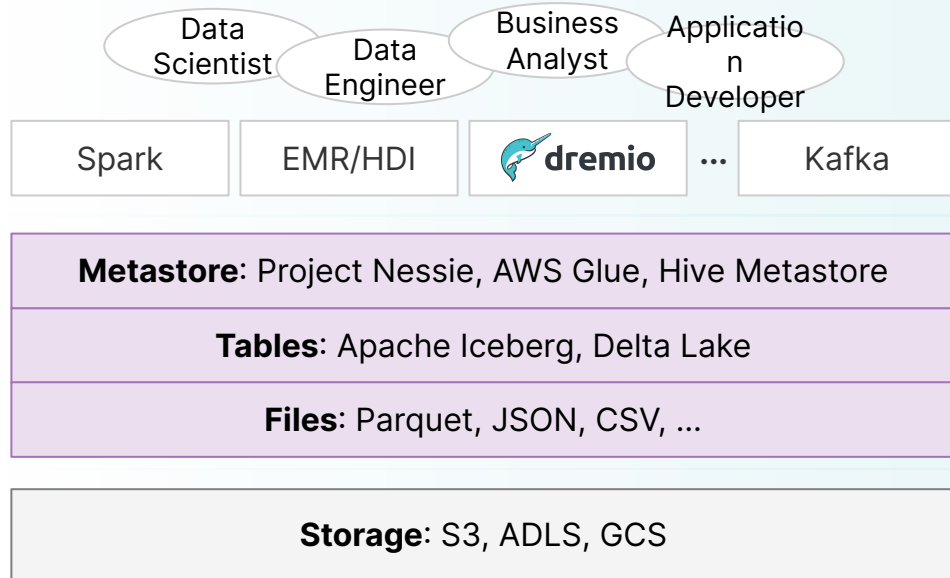
Thought Leadership

Creator & host of Subsurface
Cloud Data Lake Conference

Diamond sponsors:



The Open Data Architecture



- 1 Flexibility to use best-of-breed engines**
Use Dremio (SQL), Databricks (Spark), EMR, Athena, Dask, Flink and many other engines, and enjoy future innovation too.
- 2 No vendor lock-in**
Your data is available to any data lake engine, so it's easy to complement or replace your existing vendor.
- 3 Multi-cloud**
Use the same engine in multiple clouds to reduce cloud lock-in and administration overhead.



“With a data lake, data is stored in an **open format**, which makes it easier to work with different analytic services. Open format also makes it more likely for the data to be compatible with tools **that don't even exist yet**.

Various roles in your organization, like data scientists, data engineers, application developers, and business analysts, can access data with **their choice** of analytic tools and frameworks.”



Werner Vogels
CTO, Amazon

Demo



RETE ITALIANA
OPEN SOURCE

Demo Scenarios

1



Data Consumer Scenario

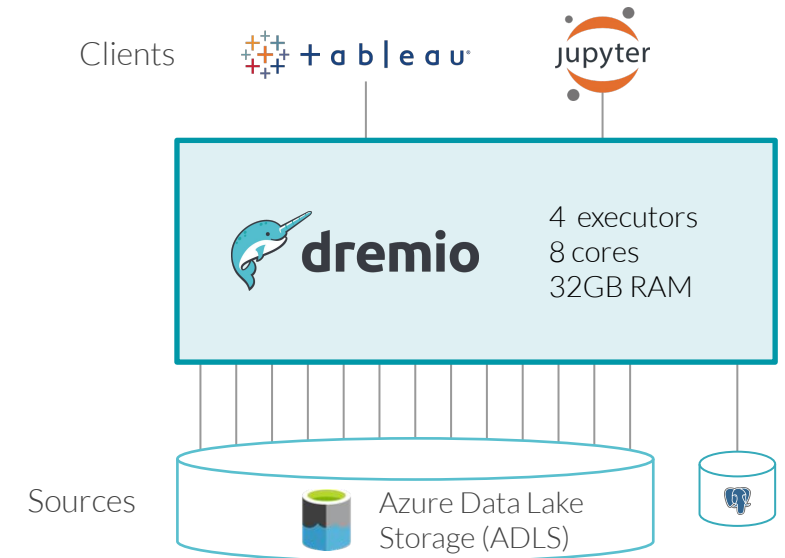
- Goal: trend analysis on taxi rides in NYC
- Quickly find and validate dataset
- Perform exploratory analytics in Tableau

2



Data Engineering Scenario

- Goal: prep a dataset for machine learning
- Augment dataset for data science team
- Share with data scientist team



**Lightning-fast queries and a self-service semantic layer,
directly on your data lake**

The Value of Dremio

- Faster time to value
- Make data engineers more productive
- Make data consumers more self sufficient
- Reduce data infrastructure cost & complexity





Grazie mille !

Feel free to contact me
wolfgang.disselhoff@dremio.com



RETE ITALIANA
OPEN SOURCE