



# Unlock the Value of Open Data Lakes with presto.

**Girish Baliga**, Chair | Presto Foundation  
Sr. Engineering Manager | Uber

**Dipti Borkar**, Chair | PrestoCon & Community Team, Presto Foundation  
Cofounder & CPO | ahana



December 9, 2021

Virtual Event

#prestocon

## Speakers from



# Technology Cycles Rhyme: Data Lake Analytics

## FDBMS Challenges RDBMS

FDBMS Paper by McCleod / Heimbigner (1985)  
FDBMS Paper by Sheth / Larson (1990)  
Composite Software (founded 2001)  
Garlic Paper by Laura Haas (2002) → DB2 Federated

80's

## Cloud & Big Data

Google File System Paper (2003)  
MapReduce paper (2006)  
Spark Paper (2010)  
Too many Data Sources, No one uber schema

2000's

## New Cloud DW w/ Data Lakes

Based on SQL  
Self Service Platforms which enable Self-Service Analytics

2020's

## OLTP to DW Wins

Data Warehouse becomes the source of truth  
Star schema becomes sacred

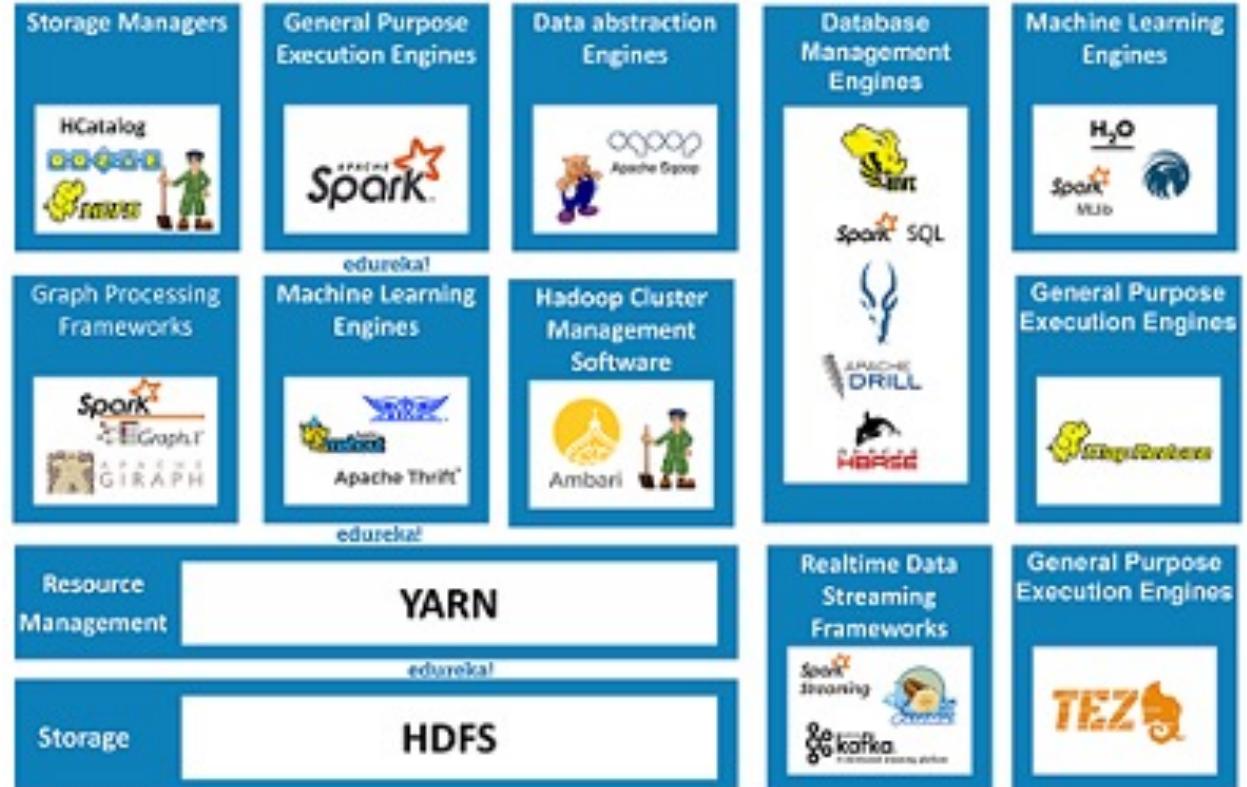
2010's

## SQL Federation Makes Comeback

Dremel Paper (2010) → Drill paper (2012)  
SQL ++ paper (2014) → Couchbase SQL++ engine (2018)  
Presto paper (2019), PartiQL (2019)

# The Traditional Data Lake

- File System Data Store
- Structured / Semi-Structured Data
- Ingestion
- Batch
- Less expensive, but...
- Complex to use
- Hard to maintain

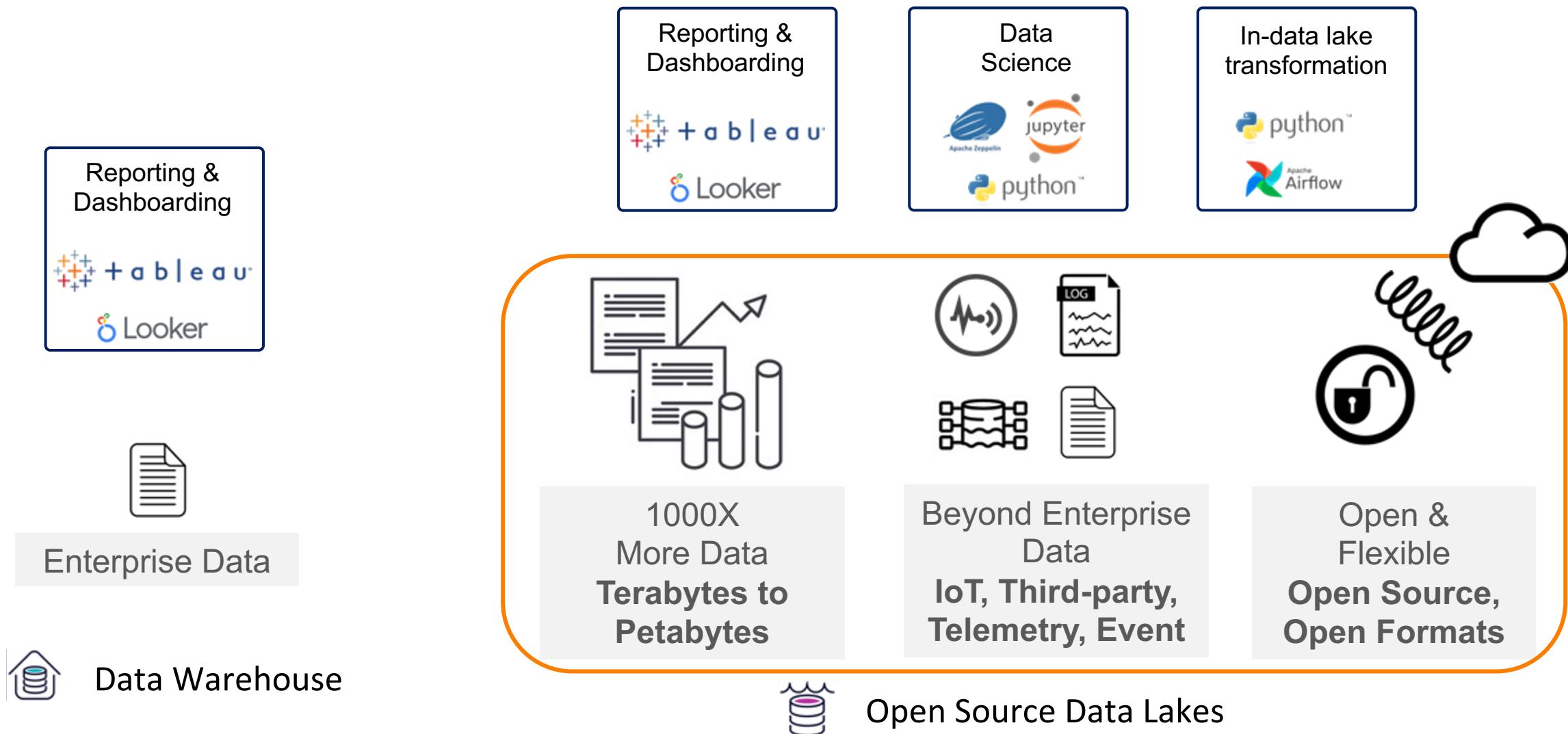


# The Cloud Data Warehouse

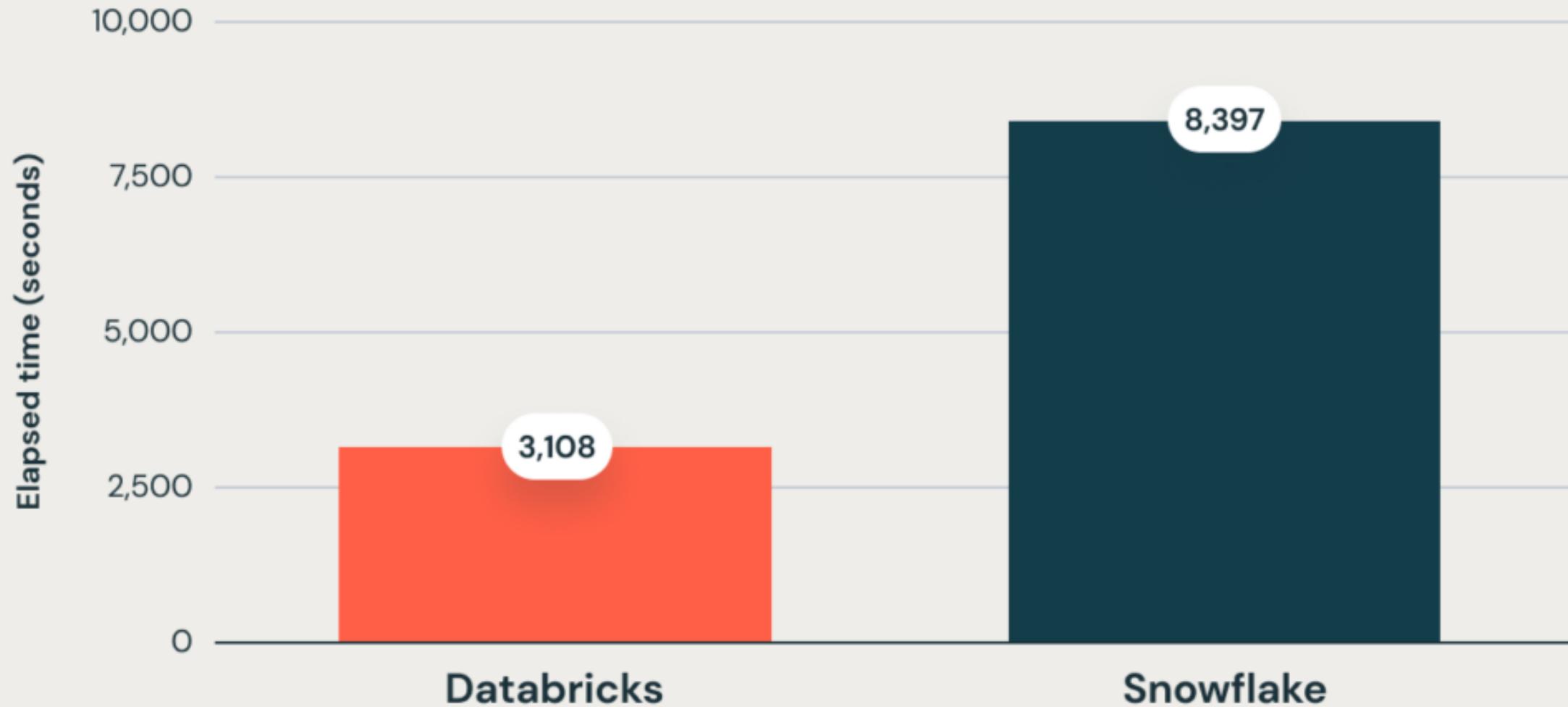
- **Cloud-First**
- **Fully-integrated**
- **In-Memory Capabilities**
- **Complex Data Types**
- **Proprietary formats**
- **TB scale**
- **Expensive**



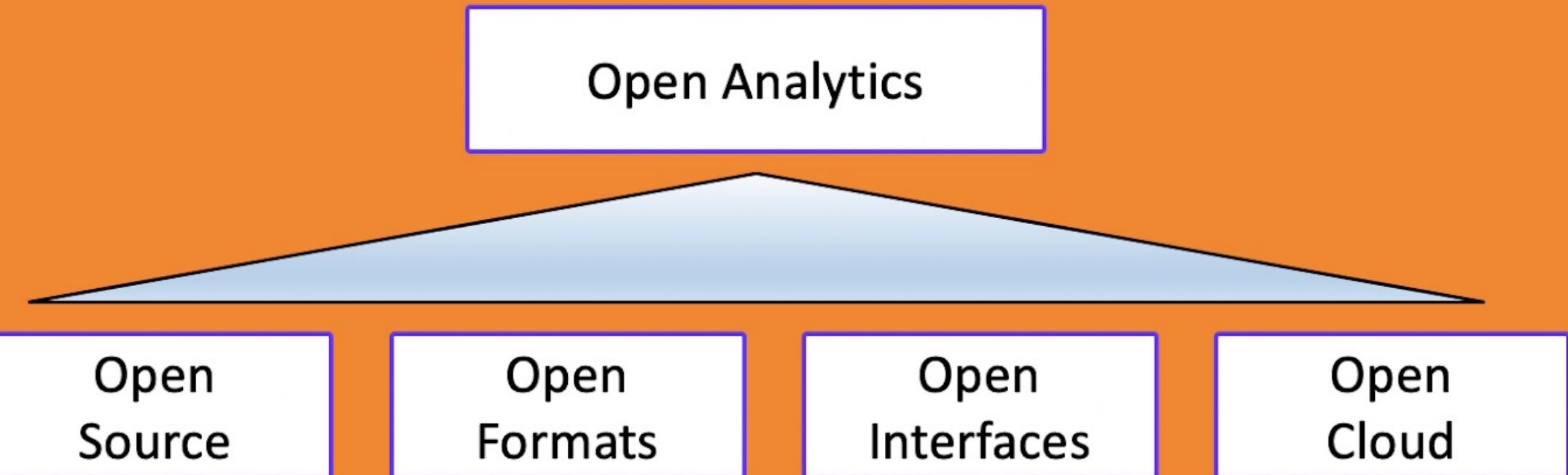
# Why Open Source Data Lake Analytics?



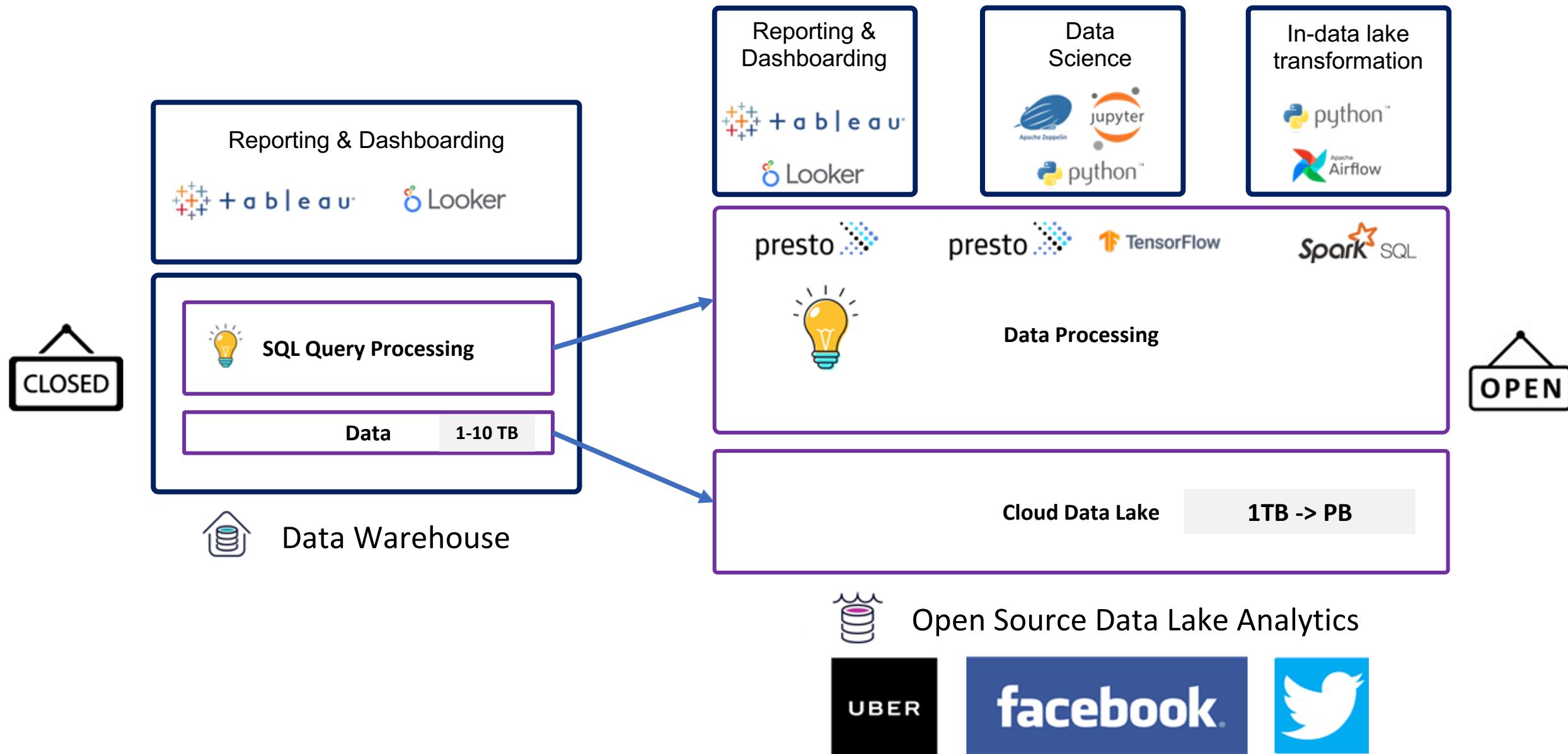
## Barcelona Supercomputing Center test derived from TPC-DS 100TB Power run (lower is better)



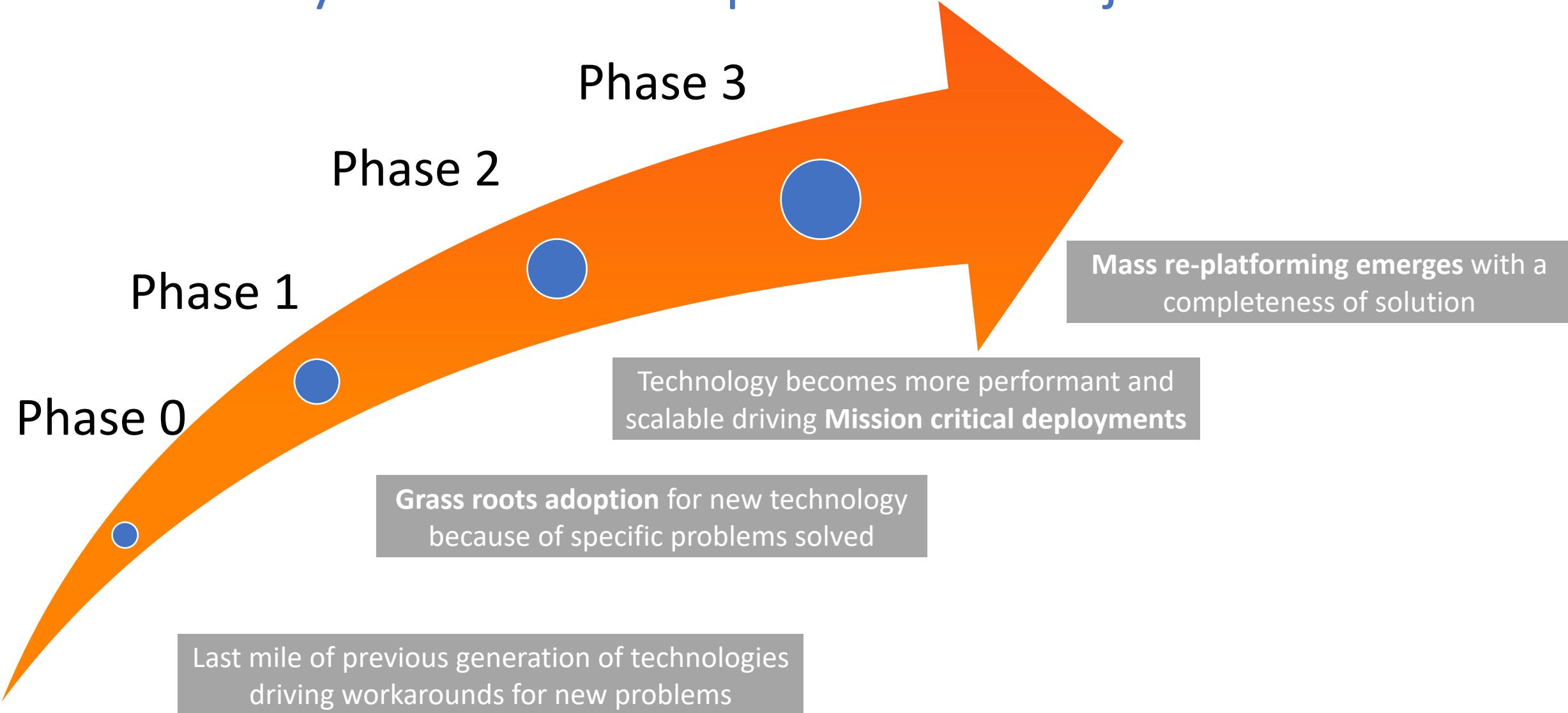
# Why Open Matters?



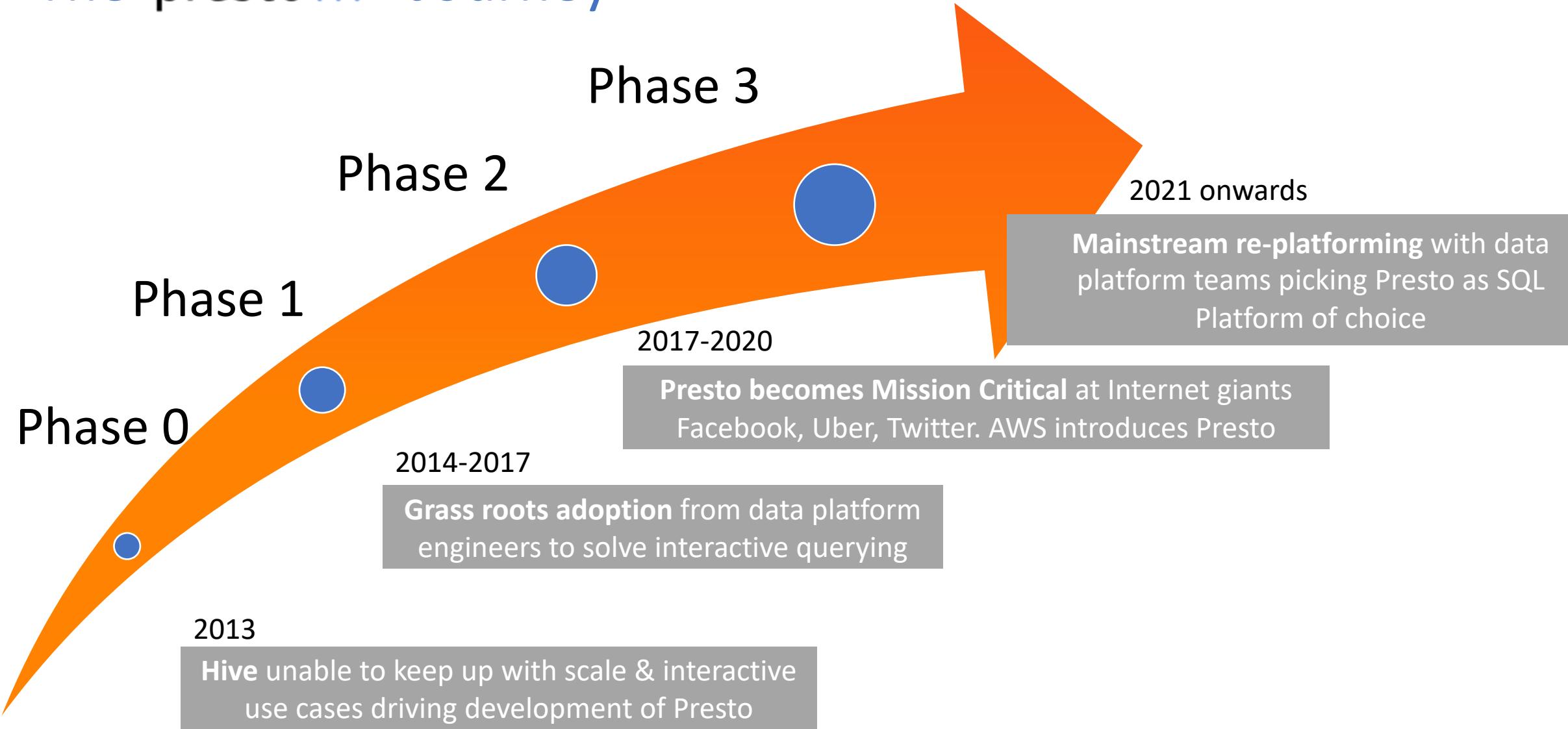
# Open Source Data Lake Analytics



# The Journey of Successful Open Source Projects



# The presto Journey



# presto is a SQL platform

Interactive  
querying  
use cases



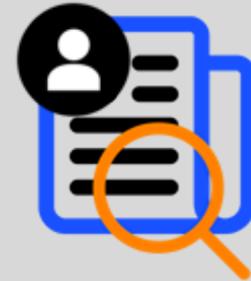
Reporting &  
dashboarding



SQL  
Data Science



Federated access  
across data sources



Customer-facing  
app analytics

Emerging  
use cases



Data Lakehouse  
Analytics



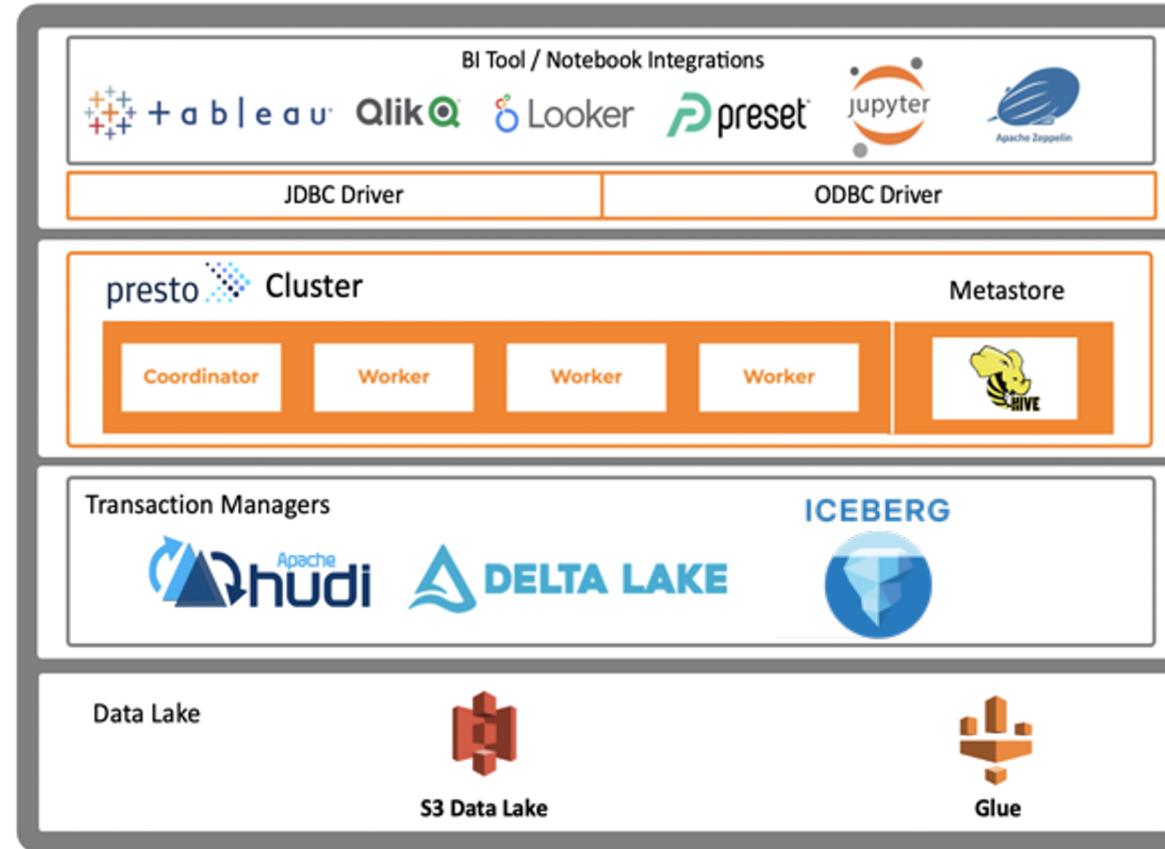
Transformation  
using SQL (ETL)



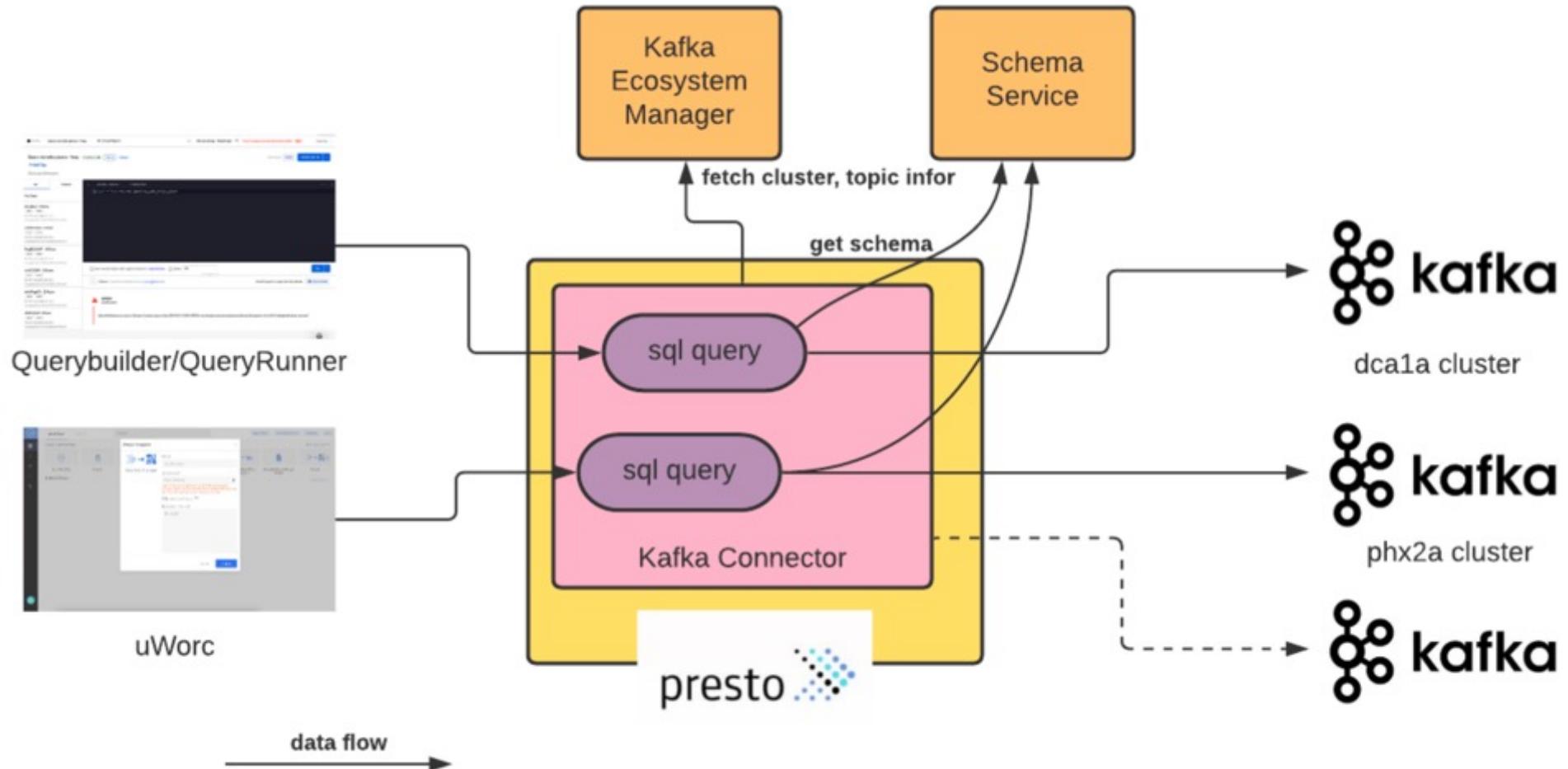
Streaming  
Analytics

# Innovations that change the game

## Data LakeHouse Analytics with Table Formats

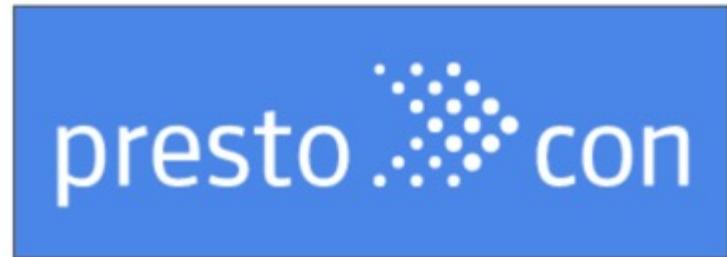


# Innovations that change the game



# Innovations that change the game

## Presto on Spark



Presto on Spark - Scaling not Failing with Spark

Shradha Ambekar & Ariel Weisberg

**Intuit**<sup>®</sup>

 Meta

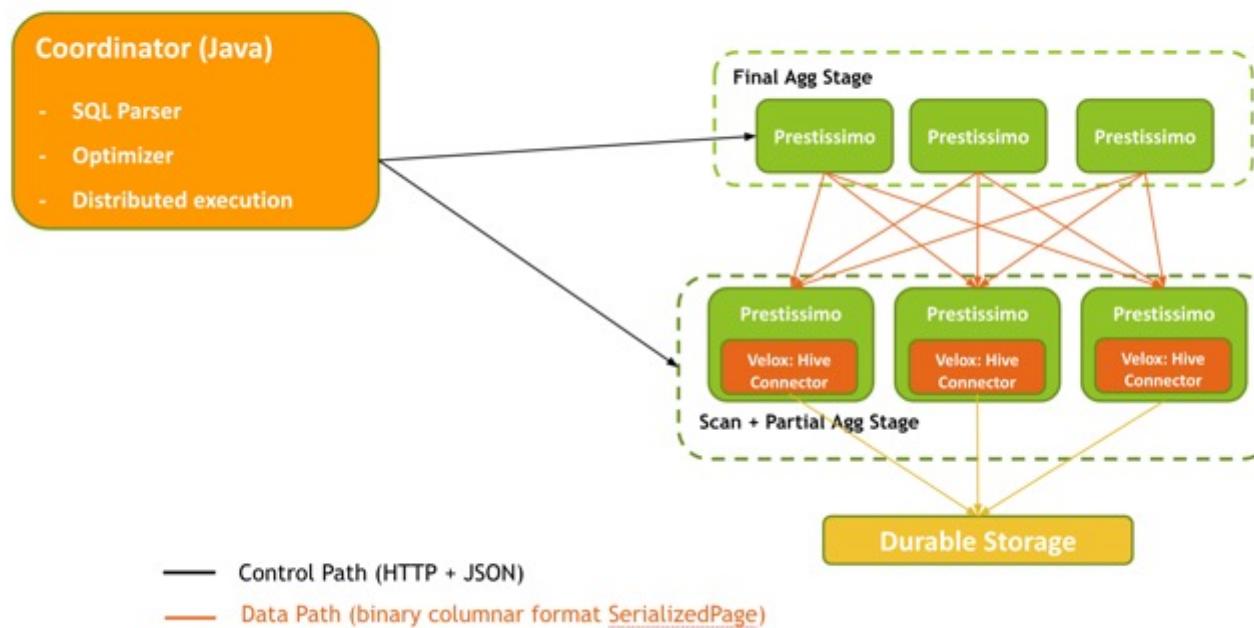


# Innovations that change the game

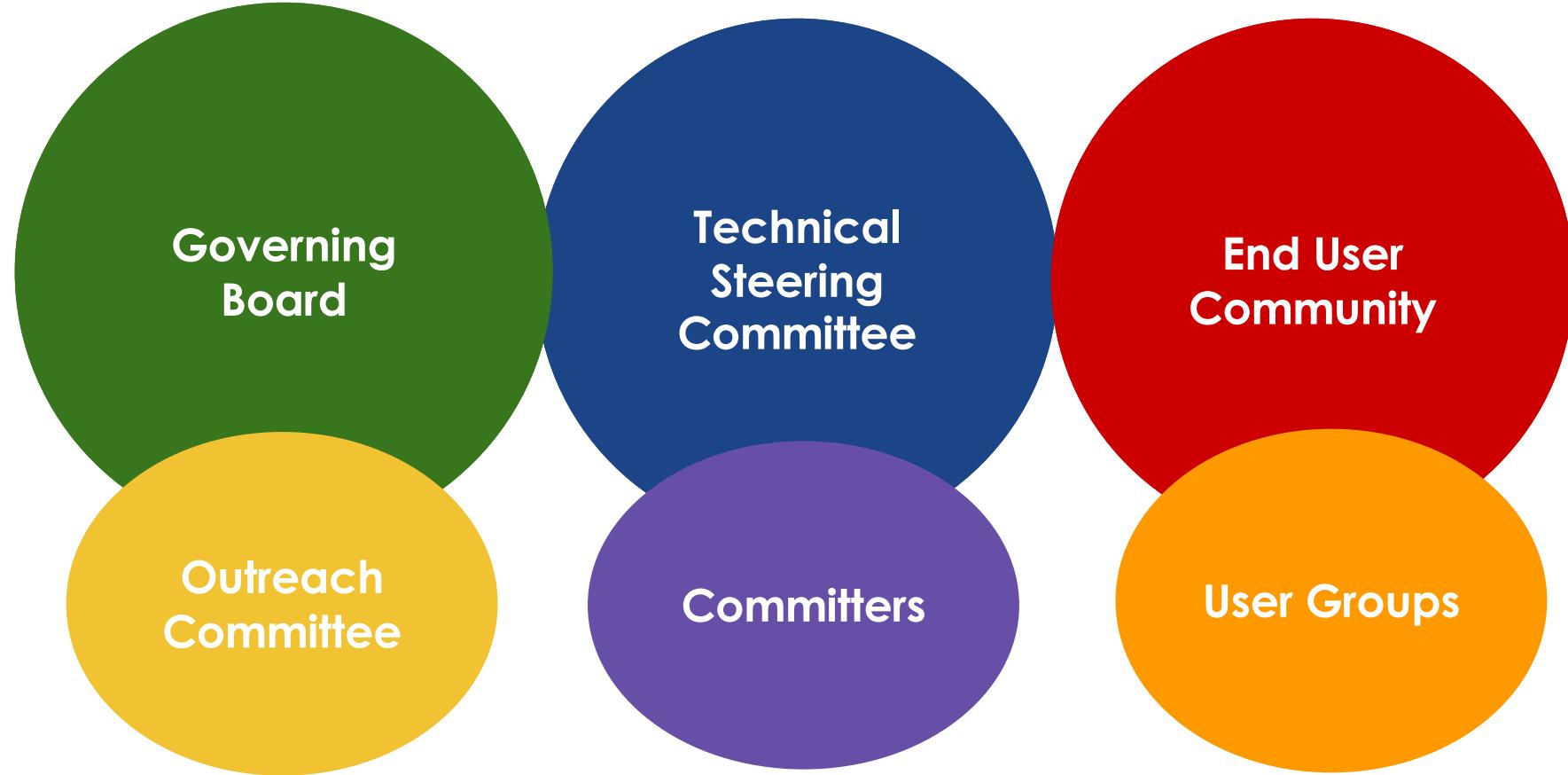
Presto – Native C++ Worker

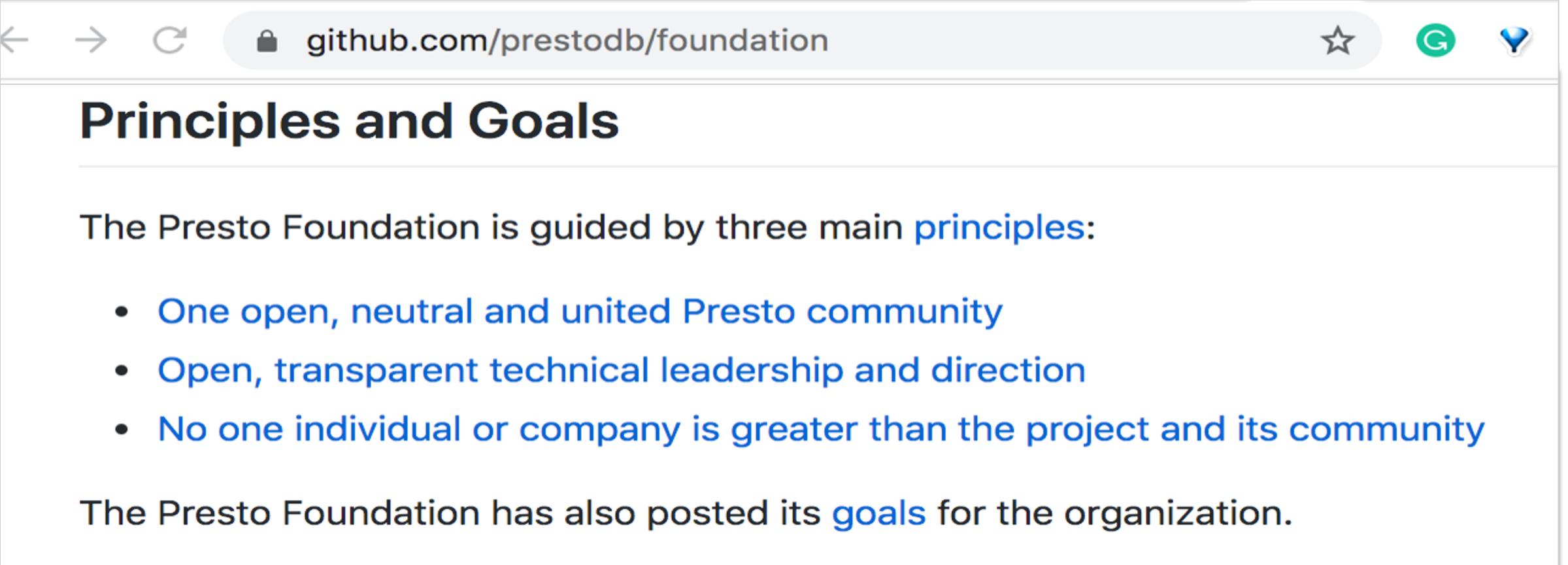
Prestissimo - Presto-on-Velox for Faster More Efficient Queries

## Prestissimo: Presto on Velox



# Presto Foundation – Community-driven Open Source





The screenshot shows a web browser window with the URL [github.com/prestodb/foundation](https://github.com/prestodb/foundation) in the address bar. The page title is "Principles and Goals". The main content discusses the three main principles of the Presto Foundation:

The Presto Foundation is guided by three main [principles](#):

- One open, neutral and united Presto community
- Open, transparent technical leadership and direction
- No one individual or company is greater than the project and its community

The Presto Foundation has also posted its [goals](#) for the organization.

# The mission of Presto Foundation

Enable and Support  
Broad Re-platforming for data platform teams  
picking presto as engine of choice for analytics

Presto Users



SIEMENS



Tuned Global



moengage



KOREAN AIR



SmugMug



21CF



xylem  
Let's Solve Water



facebook.



MORNINGSTAR



TOYOTA RACING DEVELOPMENT  
TRD

Jemena  
bringing energy to life

TETRASCIENCE



TNG  
TNG Fintech Group



Burt.

Eko.in

metropolis

MS&AD  
Aioi Nissay Dowa USA

elula

Hewlett Packard  
Enterprise

upsolver

ENGIE

Dialog

nearmap

Canva



December 9, 2021

Virtual Event

#prestocon

## Speakers from

