



THE FUTURE OF HADOOP IN YOUR ENTERPRISE: STRATEGY AND NEXT STEPS

David Leichner, CMO

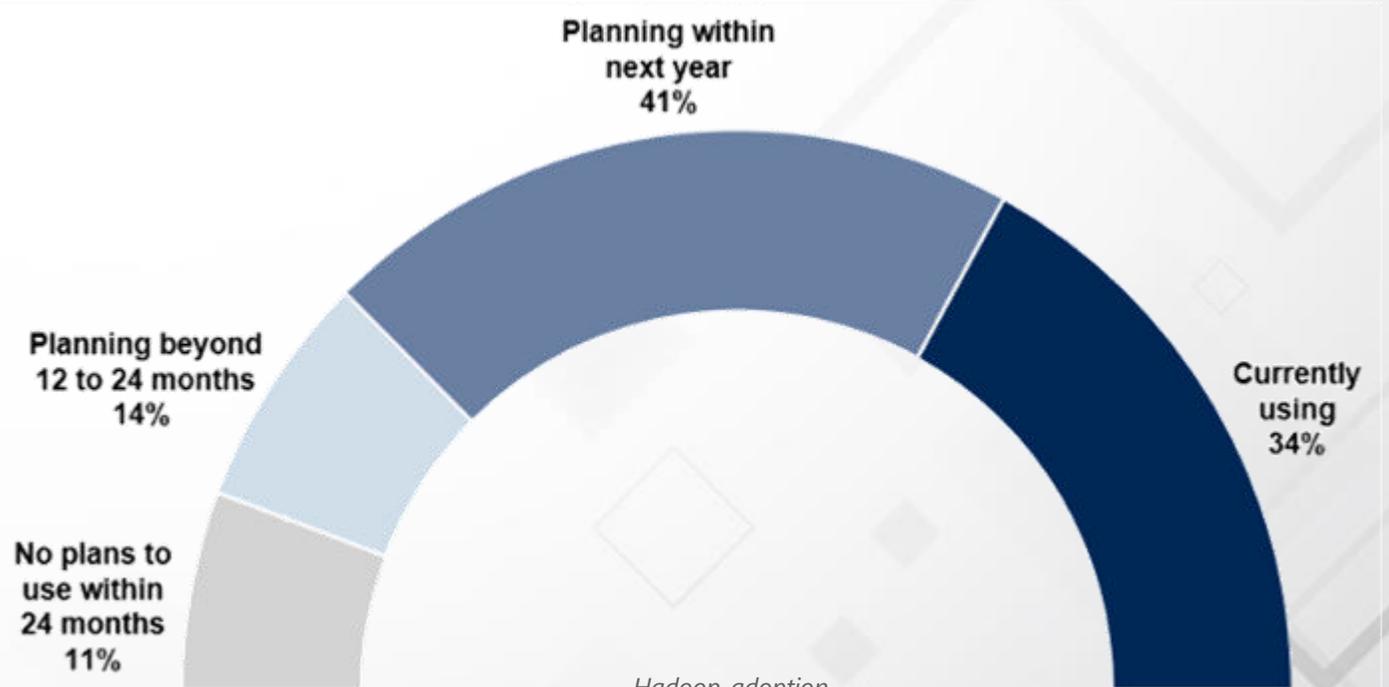
Arnon Shimoni, Product Manager



HADOOP STATUS - 2020

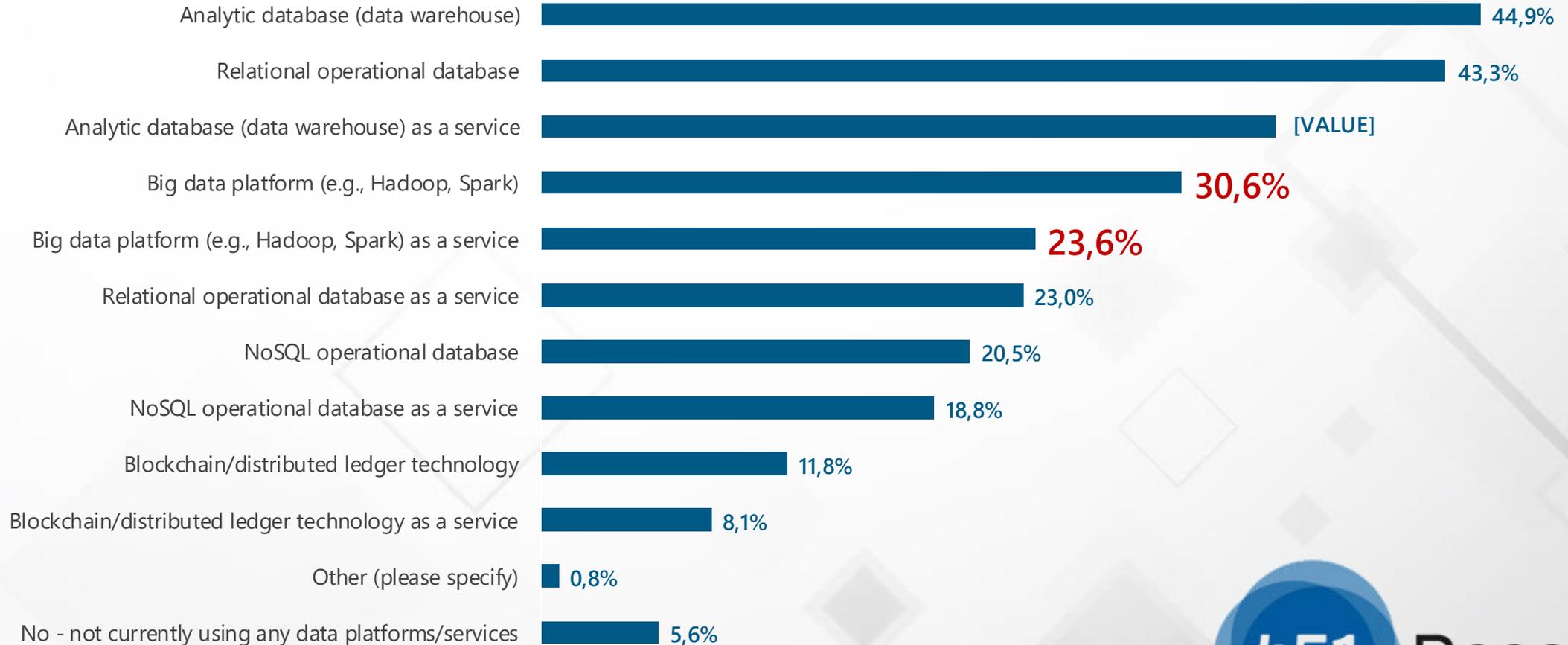
1. "Hadoop is firmly entrenched in the landscape - **it is deployed in thousands of organizations** - but using it effectively has proved to be a key challenge, and **it is not being deployed with widespread success** in many organizations."

2. "Hadoop *can* be used as an effective queryable archive for structured data - albeit **for generally less performant and [less] complex workloads.**"



Hadoop adoption
From 2018 Gartner Data and Analytics adoption Tends Study

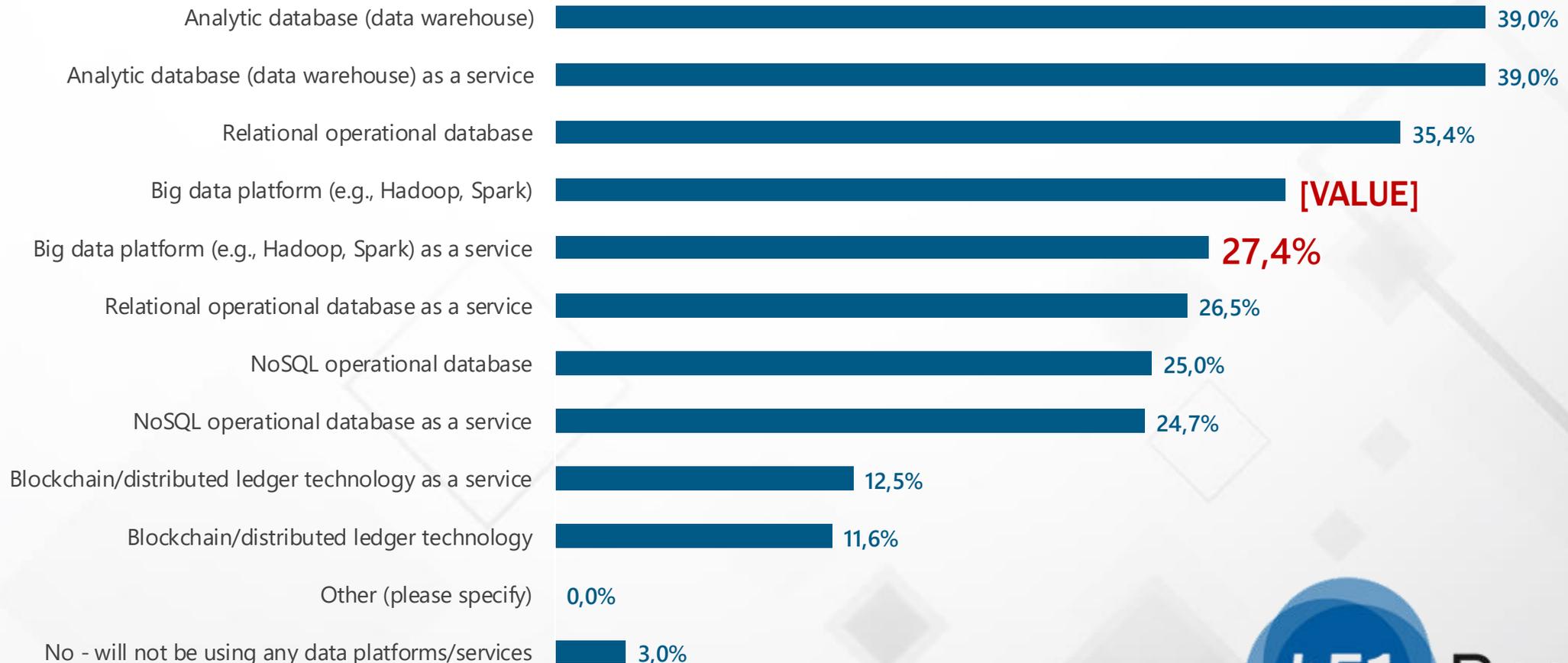
DOES YOUR ORGANIZATION CURRENTLY USE ANY OF THE FOLLOWING TYPES OF DATA PLATFORMS OR SERVICES?



- Sample Size = 356
- Base: Organization uses on-premises relational operational database providers, abbreviated fielding



DOES YOUR ORGANIZATION PLAN TO USE ONE OF THESE DATA PLATFORMS OR SERVICES 2 YEARS FROM NOW?



- Sample Size = 336
- Base: Organization will use on-premises relational operational database providers, abbreviated fielding

WHAT HADOOP USERS ARE SAYING



Hadoop ecosystem is **complex and expensive** to manage



Data management – **2X data engineers** for every data scientist



Steep learning curve, inefficient with small datasets, **slow** analytics



Long response times on real-world data volumes



Needs hard to find and costly **skilled resources**



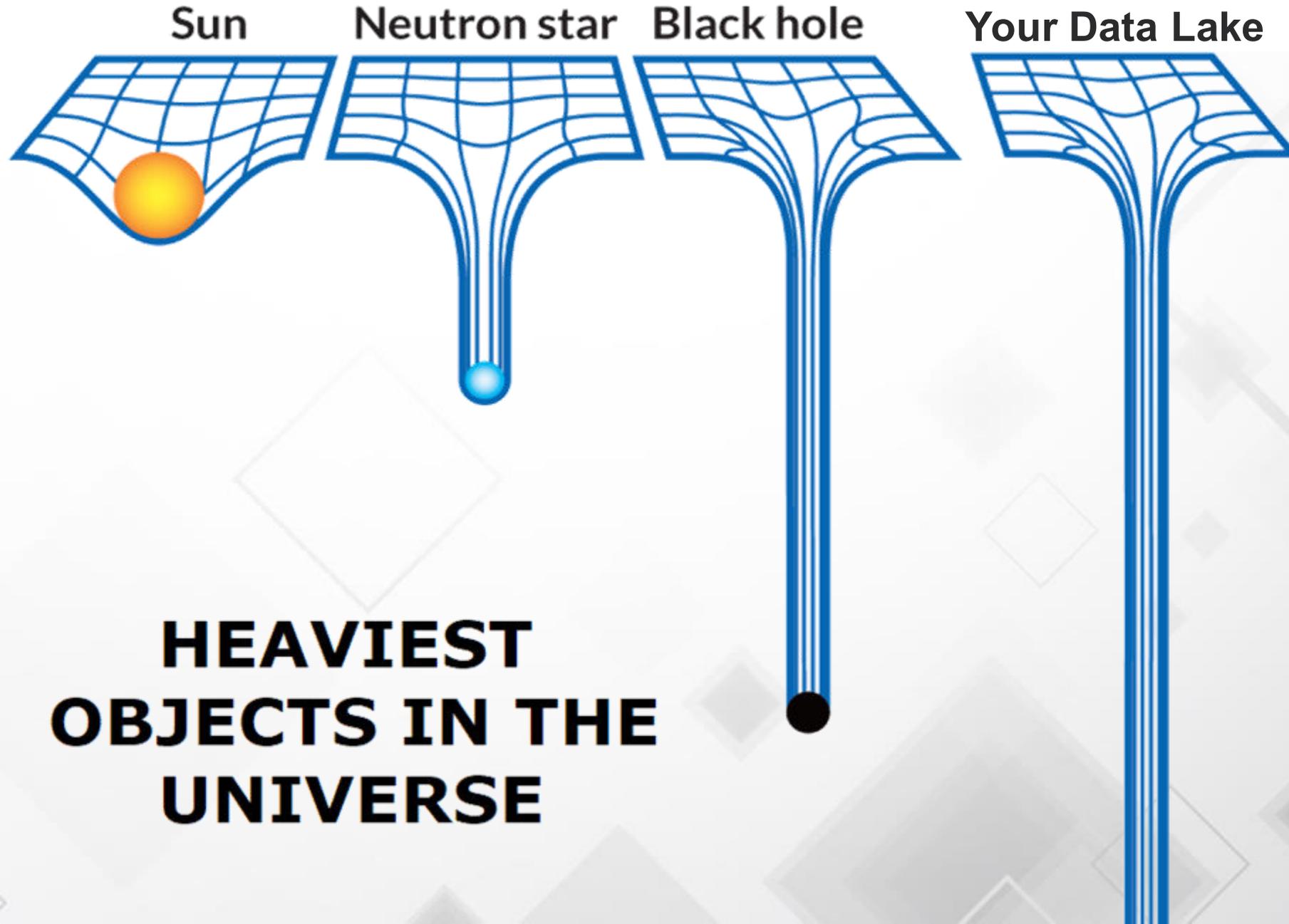
Configuration and scaling extremely difficult and **expensive**



Hard to **secure and govern**



Messy mix of roles, authentication, permissions



**HEAVIEST
OBJECTS IN THE
UNIVERSE**

DATA IS LIKE A GRAVITY FIELD

- Lots of data is hard to move
- Lots of data attracts even more data

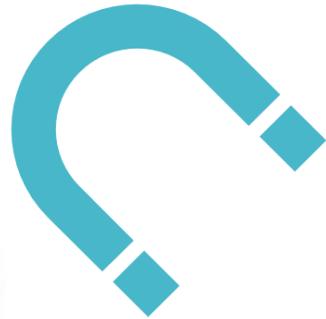
So, what if you have a big Hadoop setup?

CHOOSING THE RIGHT MODERNIZATION APPROACH



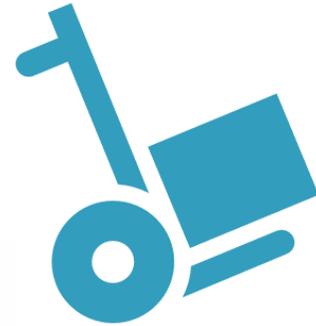
Recreate

Re-engineer all applications to match the new infrastructure – for example, cloud-native.



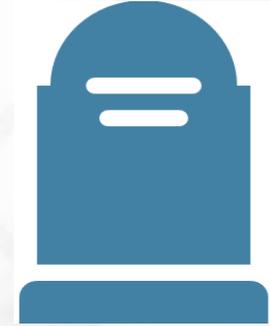
Hybrid

Some modernization, like accelerated hardware, improved applications, some SaaS and some microservices are brought in.



Lift & Shift

Applications are rehosted as-is on a new platform.



Retire

Old applications are retired.

Business use-cases are moved to other applications entirely.

Let's start here

MIGRATE HADOOP TO THE CLOUD? SOUNDS GOOD...



Cost-efficiency

Availability of latest infrastructure technology at a fraction of the cost of on-premises infrastructure



Flexibility

Data can stream seamlessly from sources to the cloud processing services



Scalability

Up or out during heavy traffic periods, making it easier to process large amounts of data

SOME COMMON MISCONCEPTIONS

- ❑ **Most companies** are moving all data to the cloud.
- ❑ Deployment of cloud resources is **quick and always works**
- ❑ Cloud **scales indefinitely** and without resource limits
- ❑ Easy to **lift and shift** infrastructure from on-premise to cloud
- ❑ Porting of Hadoop/Spark jobs to the cloud is **effortless**
- ❑ Moving to the cloud **reduces operational expense and staff**

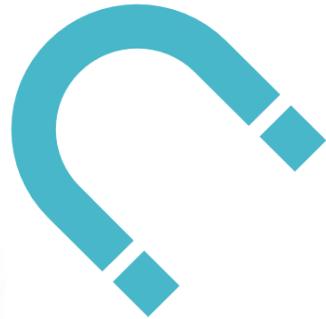


CHOOSING THE RIGHT MODERNIZATION APPROACH



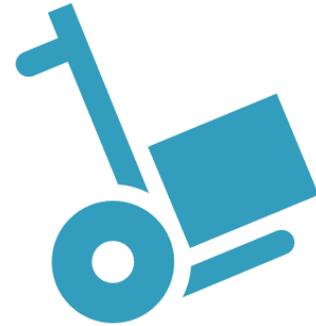
Recreate

Re-engineer all applications to match the new infrastructure – for example, cloud-native.



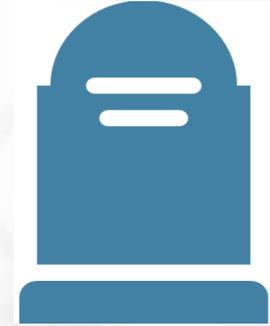
Hybrid

Some modernization, like accelerated hardware, improved applications, some SaaS and some microservices are brought in.



Lift & Shift

Applications are rehosted as-is on a new platform.



Retire

Old applications are retired.

Business use-cases are moved to other applications entirely.

a sweet spot?

UNDERSTANDING YOUR WORKLOADS

Bulk data loads

Overnight data prep

Ad-hoc queries

Daily reports



DATA IS DESIGNED TO GIVE INSIGHTS

Reconsider the “one-stop-shop” if all you do is prepare data

- **Bulk data loads** – usually loaded at fixed times, but:
 - Loads aren’t “ready” – JSONs are notorious for requiring more work
 - Data needs re-organization afterwards
- Workloads **constantly change**
- **Resource contention** – leaves data consumers unhappy
- Hadoop in the cloud – **expensive scaling**
(e.g. EC2 instances that need expensive tightly-coupled storage)

CAN YOU DO THIS WITH HADOOP?



Ad-Hoc SQL analytics

Exploring data, correlating and joining different sources is fragile with Hadoop



Empower data consumers

Letting BI analysts, data scientists, data engineers and other domain experts query without coding or deploying complex jobs



Data consolidation

Hadoop struggles to serve data from all LoBs, with a single, widely accepted interface like SQL

SPLIT THE WORKLOADS

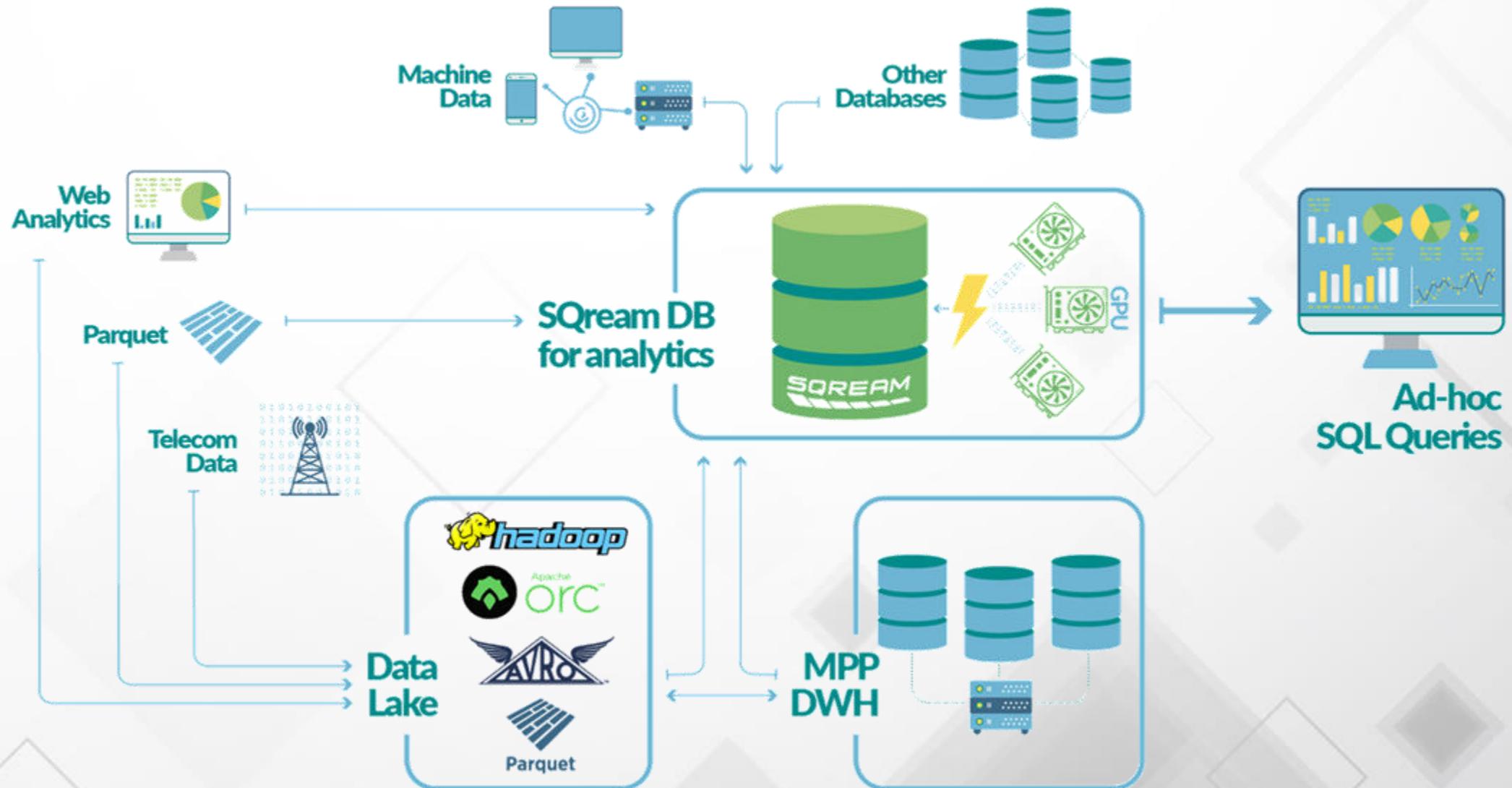


- 🐘 Distributed, cost-effective storage
- 🐘 Well established as a data lake
- 🐘 Already deployed
- 🐘 Well-integrated



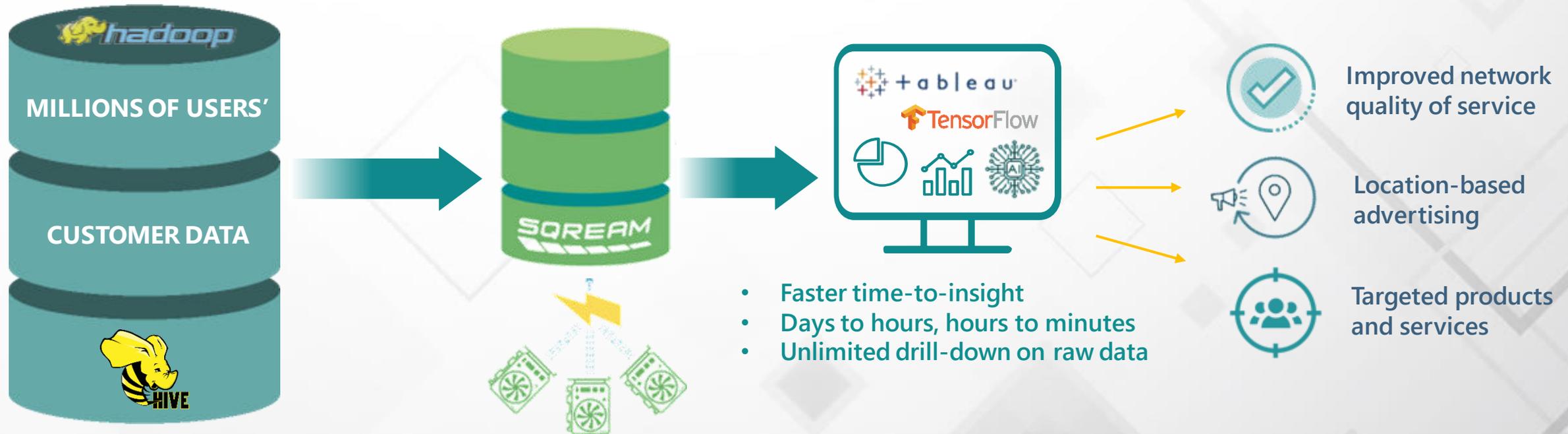
- ✓ Scalable in both compute & storage
- ✓ High-performance Ad-hoc queries
- ✓ Leverages existing SQL skills
- ✓ Easier to use

REINVIGORATING HADOOP



IN-DEPTH CUSTOMER USAGE ANALYSIS

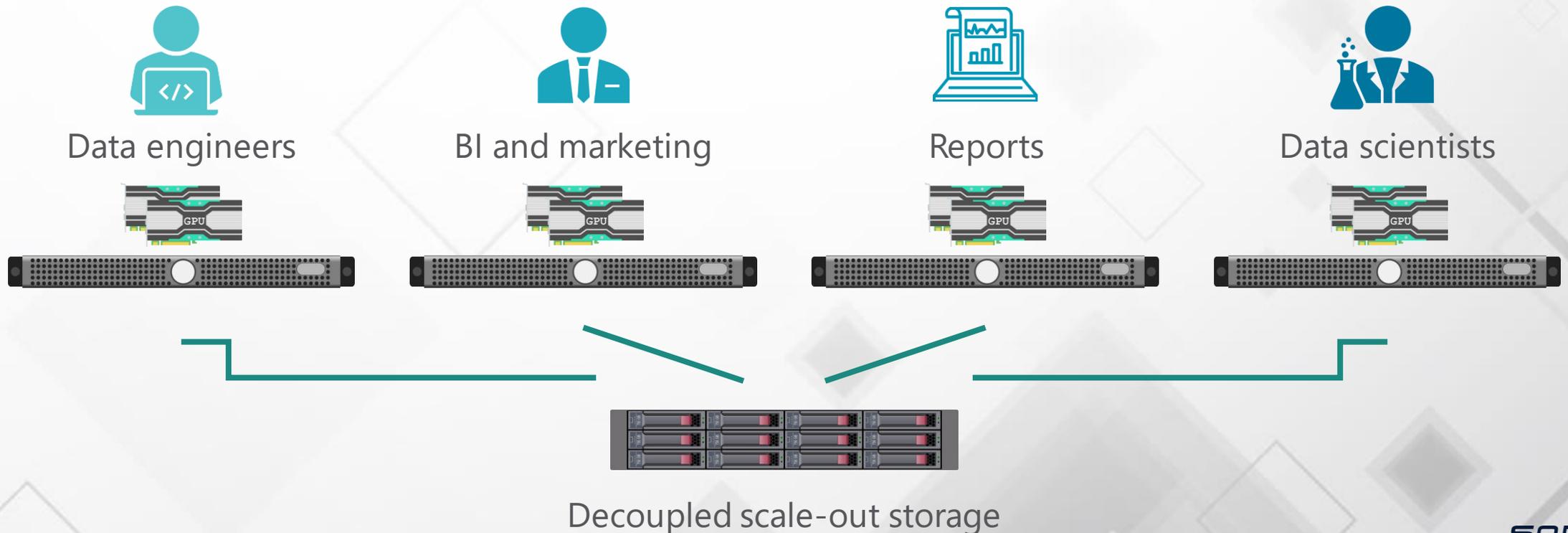
Customer mobility and CDR analysis at a large multinational telecom



Faster time to insight → Better efficiency → Increased revenue, Reduced churn

FLEXIBILITY IN RESOURCES

- Dynamic workload manager and worker groups help assign resources at runtime, but can also be preconfigured
- Assign resources to different users, use-cases, or lines-of-business



SQREAM DB AND HADOOP INTEGRATION



DISCOVER REQUIREMENTS

- Gather information about the current Hadoop environment
- Develop the joint methodology



MATCH SYSTEM COMPONENTS

- Size the solution
- Ensure best practices for Hadoop and SQream DB data management



MIGRATION AND VALIDATION

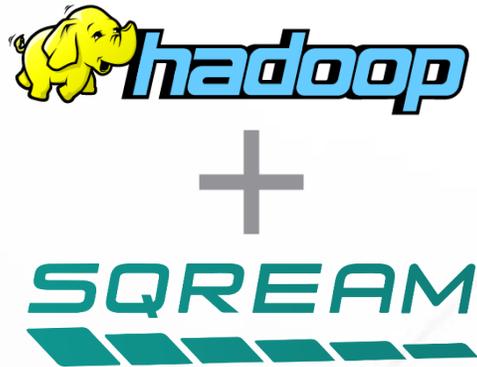
- Set up SQream DB alongside Hadoop
- Migrate workloads
- Data validation



HANDOVER

- Happy and empowered data consumers

CAN YOU DO THIS WITH HADOOP?



Capitalize on Hadoop Investment

Combining Hadoop with SQream DB maximizes economy of Hadoop with performance of SQream DB

Full SQL Support

Fully-featured SQL support to empower data owners and business users

Fast, Performant, Scalable

Analyze more data for critical insights faster - with SQream DB

FEEL FREE TO CONTACT

David Leichner, CMO
david@sqream.com | www.sqream.com

ADDRESS

Headquarters, 7 WTC
250 Greenwich Street
New York, New York

WE ARE SOCIAL

