

Spotlight Sessions



Nik Rouda

Director of Product Marketing

Cloudera

@nrouda

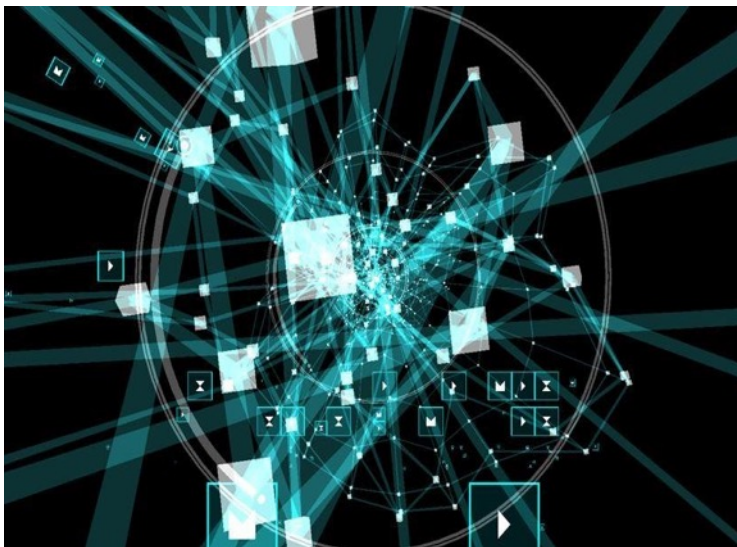
cloudera®

cloudera

Spotlight: Protecting Your Data

Nik Rouda | Product Marketing

Key drivers for secure your data



Stop Advanced Cyber Threats

Has this IP address ever touched my enterprise?

How can I discover unknown threats?



Detect Costly Fraud

How can I decrease my time to fraud response?

How long has this fraud been effecting business?



Meet Dynamic Compliance Regulations

How can I accelerate compliance reporting?

When regulations change will I have to re-architect my infrastructure?

Secure Your Data - Solutions

Cybersecurity



Primary
Data
Source

- Network data
- User data
- Endpoint data

Fraud



Primary
Data
Source

- Credit Card Transactions
- Customer data
- GEO location

Compliance



Primary
Data
Source

- Application logs
- Transactions
- HR data

Secure Your Data – Use Cases



Securing your data, a big data and analytics problem

Risk related data is hard to wrangle...

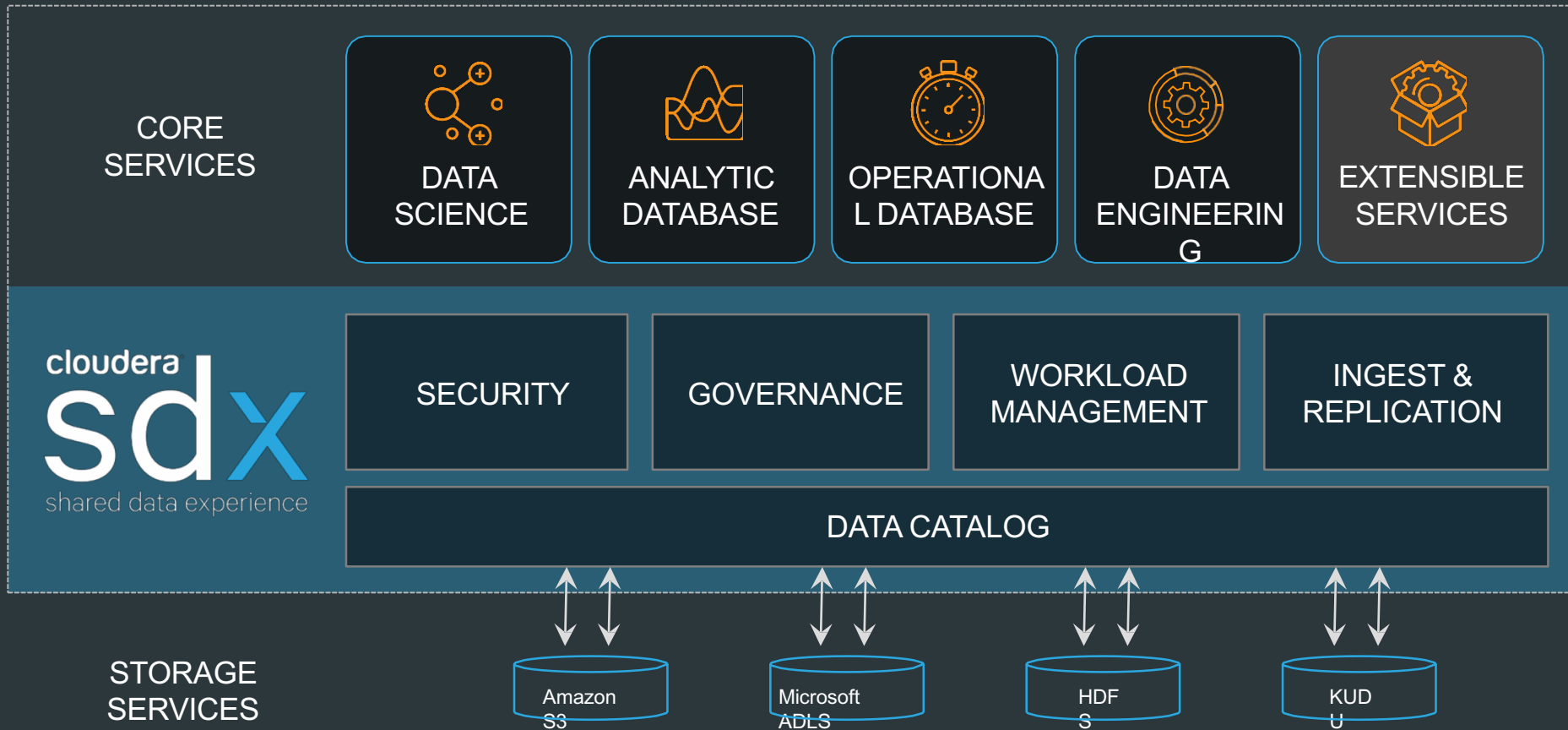
- Massive volumes of data streams are too large
- Managing data security across all of the siloed data
- Data comes in multiple formats (structured and unstructured)
- Historic data needs to be deleted or archived offline

Analytics is not easy either...

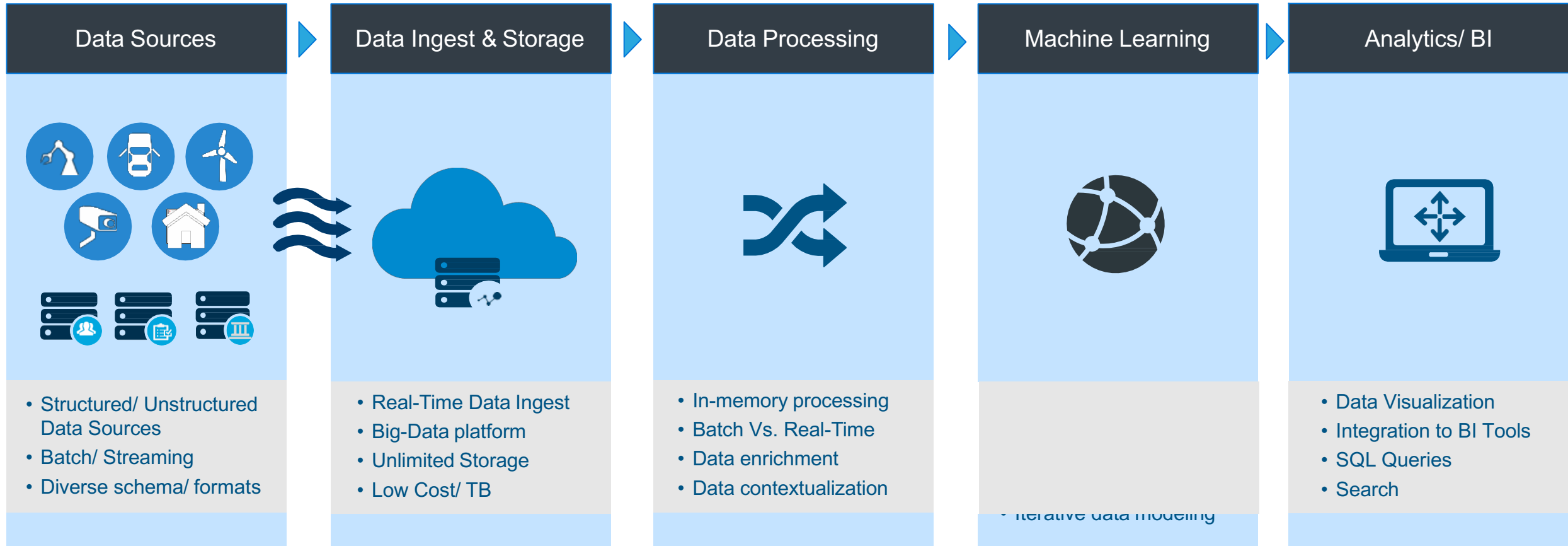
- Large scale search or SQL queries are impossible
- Finding anomalies with machine learning is limited to sample data
- Open source analytic libraries don't work with proprietary applications

Cloudera Enterprise

The modern **platform** for machine learning and analytics optimized for the cloud



The data value chain



Cloudera – Key platform capabilities



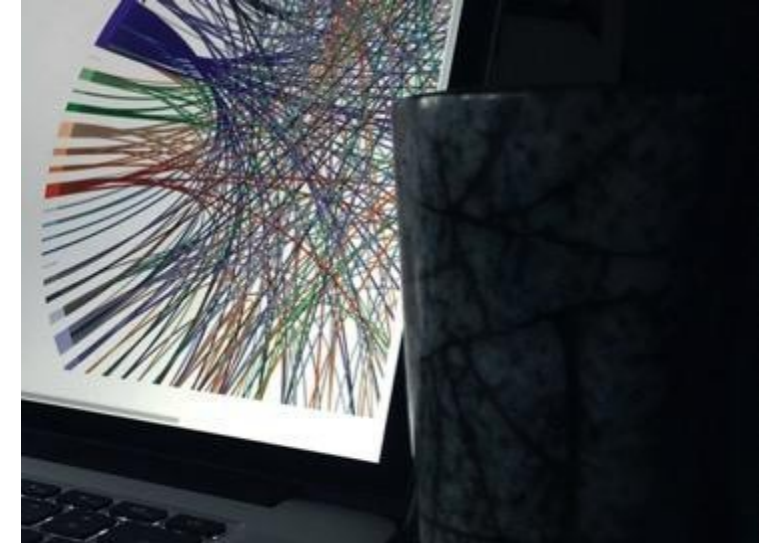
Infinite Scalability

Building on open source innovation in the big data space allows us **economically scale** data storage, processing, and analytics



Multi-Cloud Portability

Preserve business flexibility and data portability and **minimize cloud lock-in** by running in any one of the three major public cloud providers or in private cloud



Data Science

Collaborative hub for **enterprise data science** and an integrated development environment for running **Python, R, & Scala** with support for Spark

TELECOMMUNICATIONS

- » PRODUCT IMPROVEMENT
- » FRAUD PROTECTION
- » PREDICTIVE ANALYTICS



Unifying Operational and Network Data for New Service Insight

- Increased customer ratings through improved service, identifying join time delays in real time
- Identified 17x more fraud to help prevent costly fraudulent meetings
- Delivered platform at 1/10 the cost of traditional data warehouse and BI environments



cloudera®

cloudera

Spotlight: IT Modernization

Trends driving analytic modernization



Self-Service Flexibility



Real-Time Analysis



Hybrid Cloud

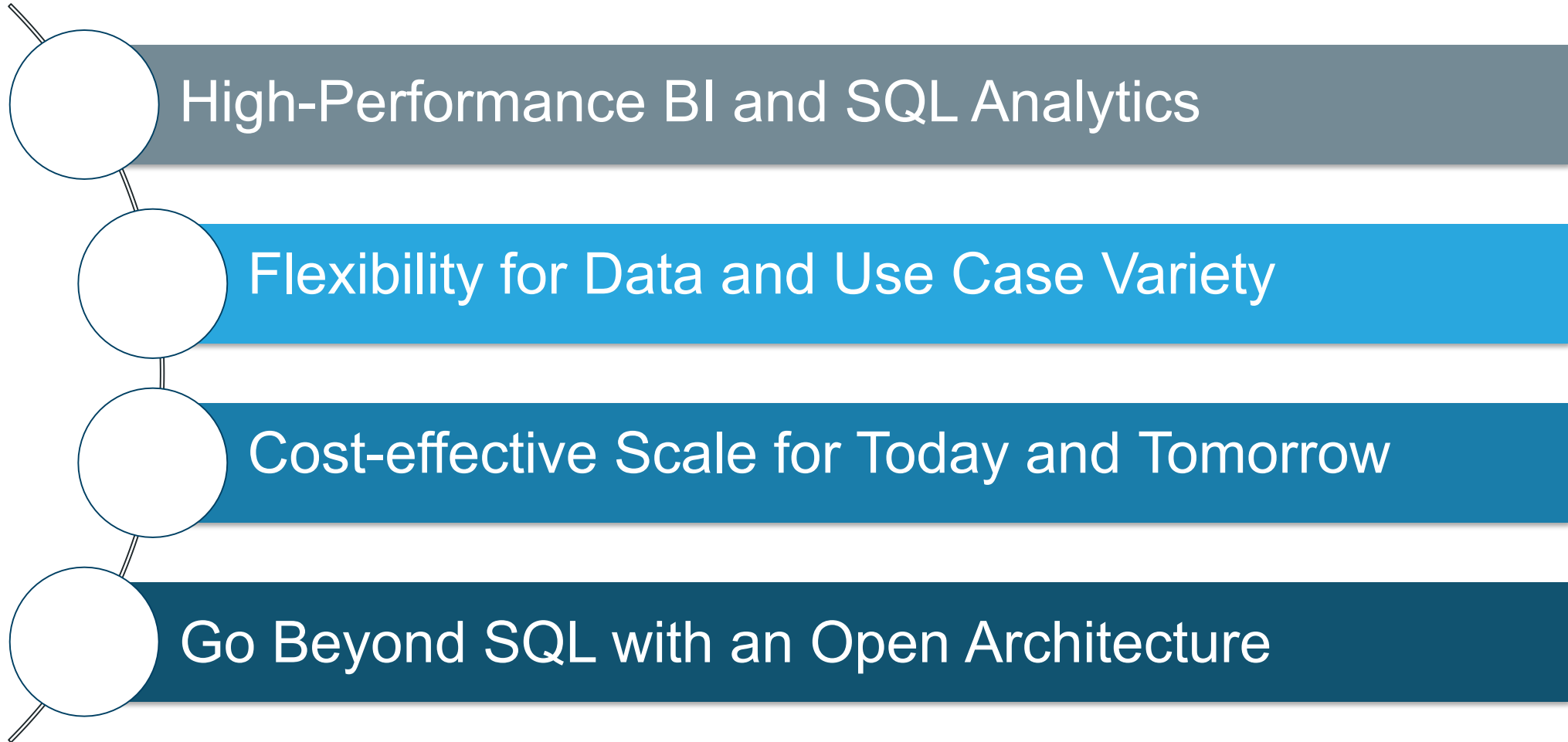


Converged Workloads

Key applications

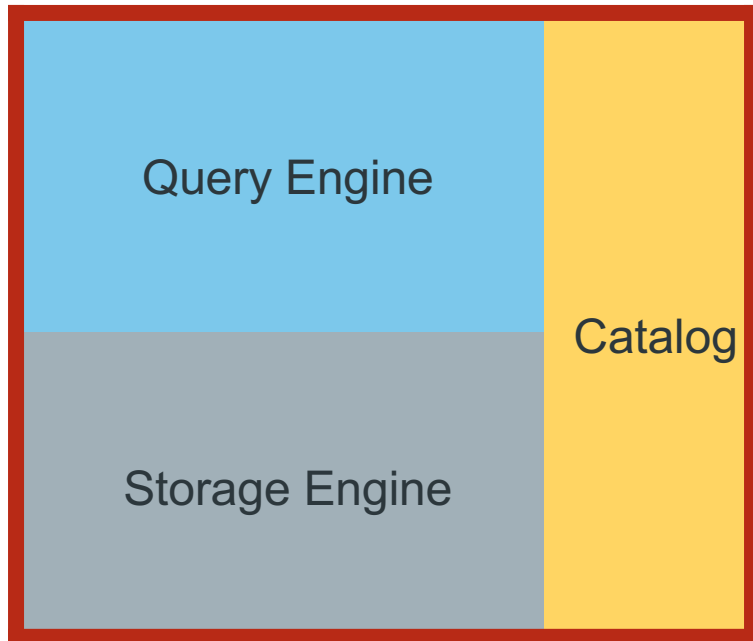


The modern analytic database key benefits

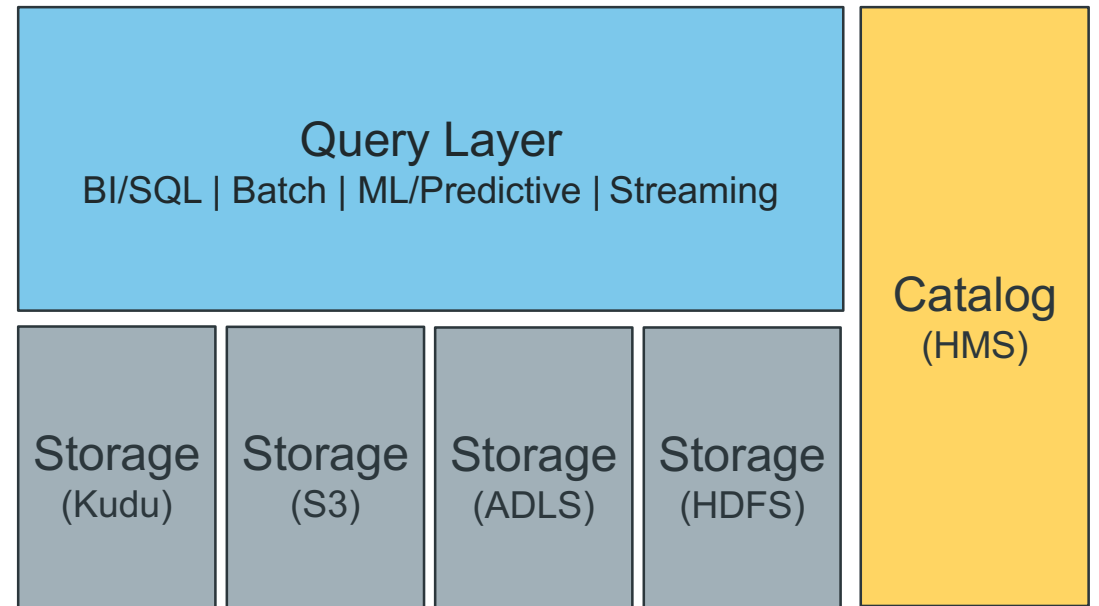


Anatomy of an analytic database decoupled by design

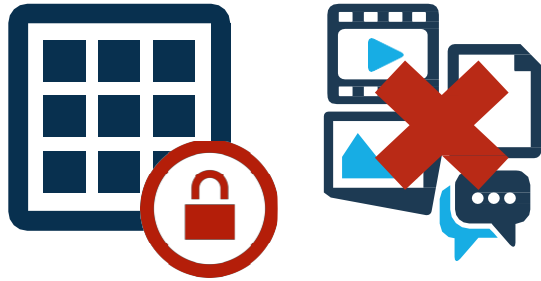
Monolithic Analytic Database



Modern Analytic Database

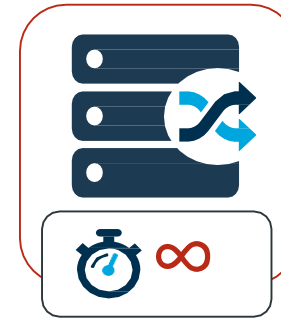


Traditional monolithic analytic databases



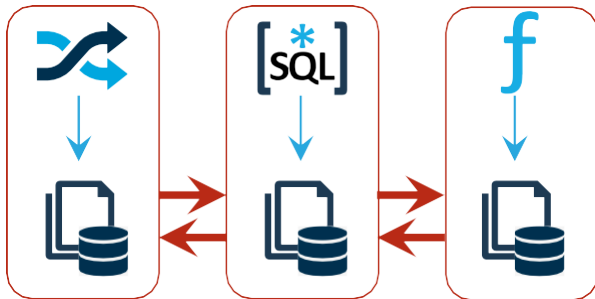
Rigid Data Model

- Tightly coupled storage and compute



Static Sizing

- Major maintenance to add capacity/nodes



Limited to SQL only

- Maintain data copies for non-SQL



Poorly Designed for Cloud

- No elasticity or integration with object storage

Advantages of a modern approach decoupled for cloud and on-premises



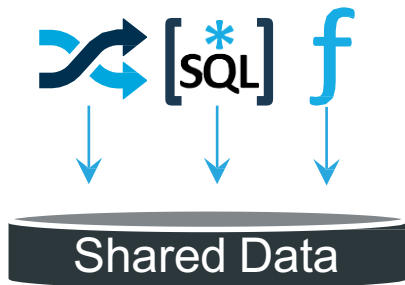
Data Flexibility

- Iterative modeling and self-service accessibility
- Portability: No proprietary formats or storage lock-in



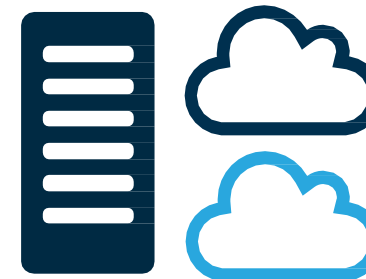
Cost-Effective Scalability

- Elastic scale in any environment
- Cloud-native integration for optimized pay-per-use costs
- Proven at massive scale



Go Beyond SQL

- Consolidate data silos with an open architecture
- Shared data across SQL and non-SQL workloads

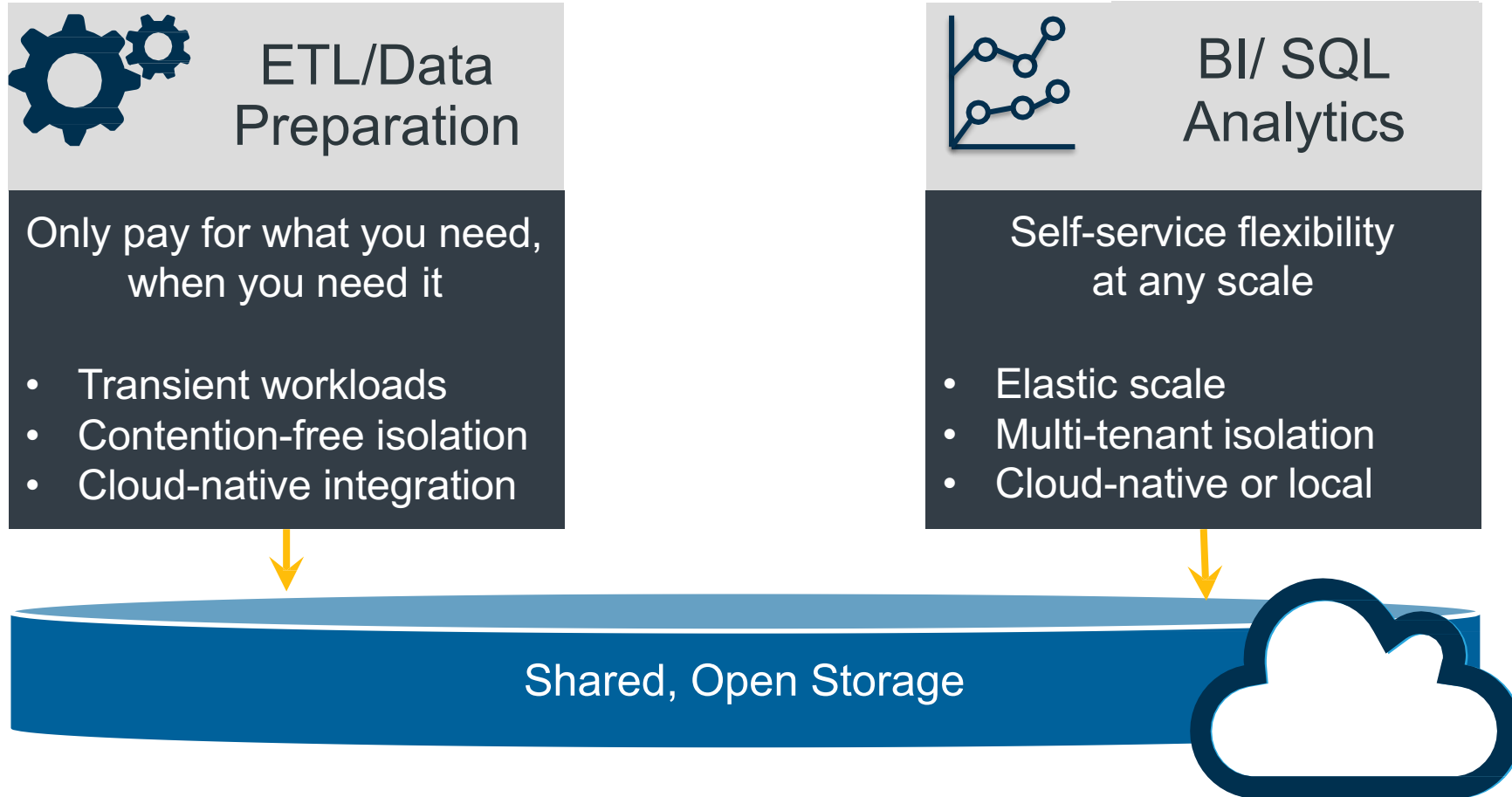


Hybrid

- Runs across multi-cloud & on-prem for zero lock-in
- Multi-storage over S3, ADLS, HDFS, Kudu, Isilon, etc

Primary analytic patterns in the cloud

scale, agility, and cost-efficiencies



Key benefits translated to the cloud

High-Performance BI and SQL Analytics

Same SQL engine native across any cloud and on-prem

Flexibility for Data and Use Case Variety

Self-service access directly on object stores, without the silos

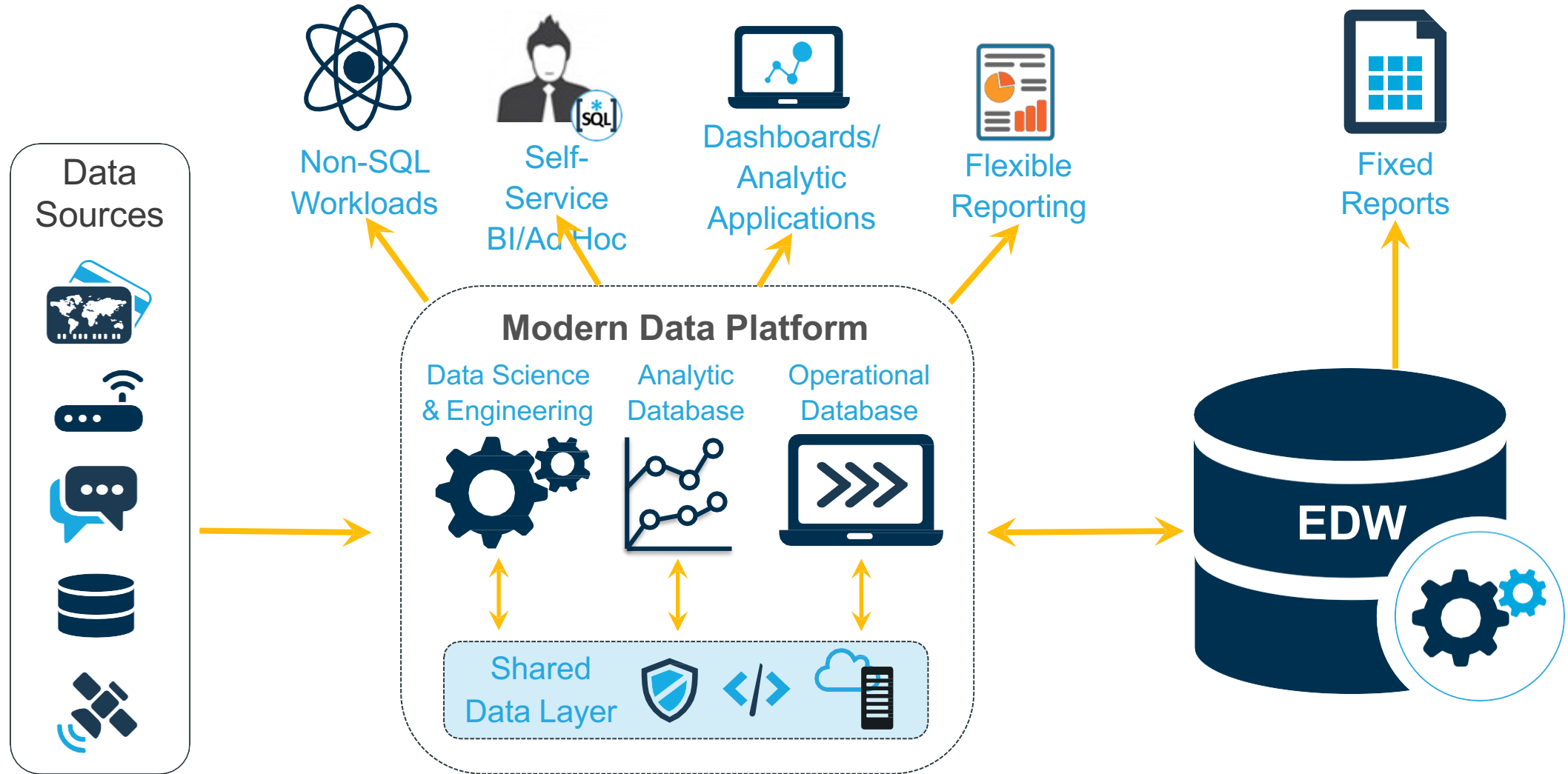
Cost-Effective Scale for Today
and Tomorrow

Elasticity on-demand through decoupled compute and object storage

Go Beyond SQL with an Open
Architecture

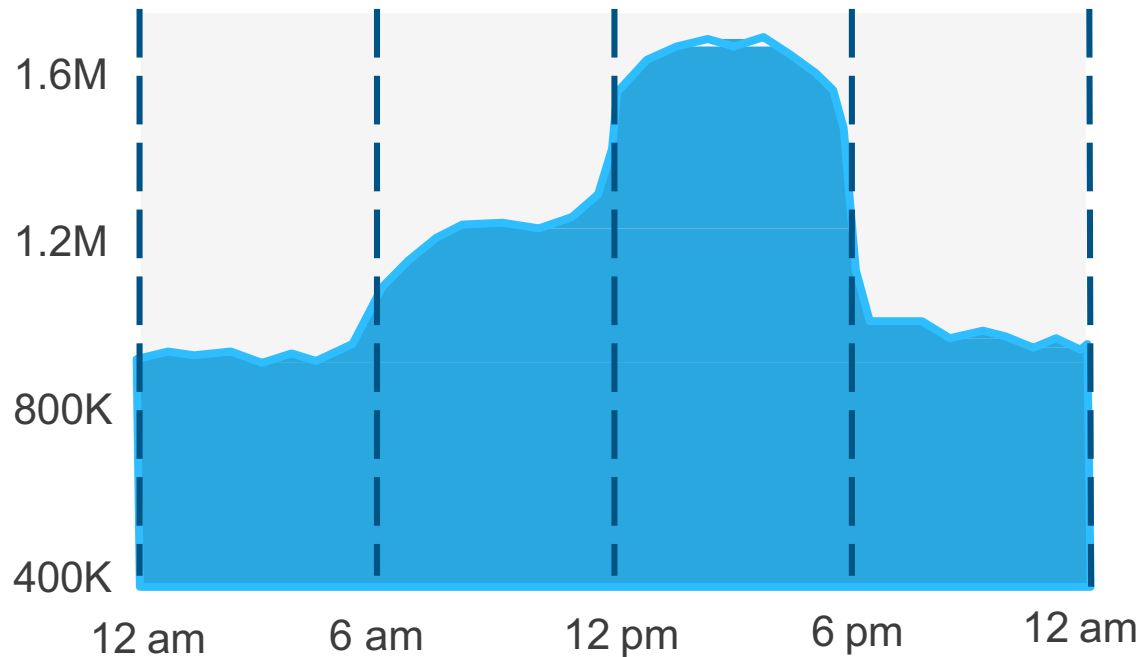
Converge workloads over shared data, with zero lock-in

Modern Data Warehouse Landscape



But where do you start?

Query volume can be huge

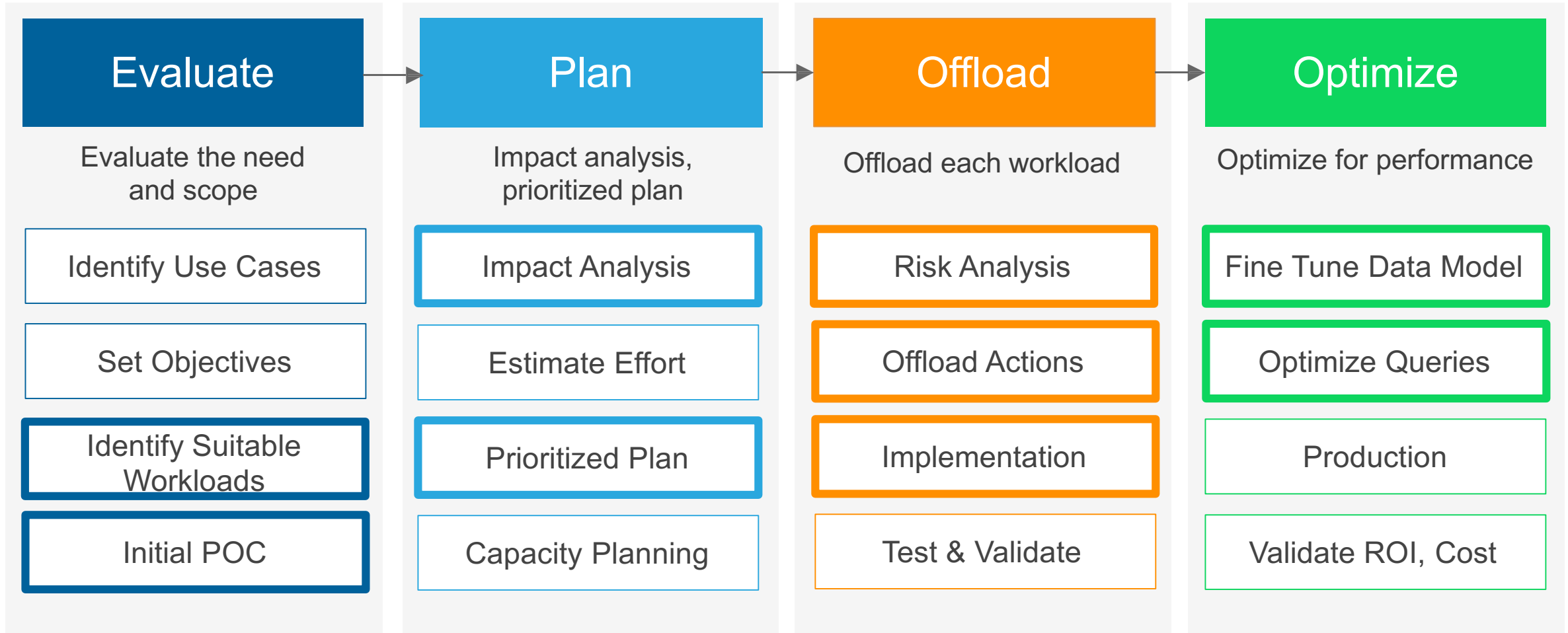


Numerous databases, thousands of tables, many users and applications

Queries can be very complex

- How do you determine what workloads to run on Cloudera's platform?
- Will the queries run efficiently?
- What does it take to migrate?
- How do you prioritize?

Four-step migration process with tools built to get you to success



Global Payments Processor

Challenges

- Queries limited to past year of data
- Analysts needed faster, unconstrained access to better understand fraud
- Already spending >\$1B annually on traditional data warehouse

Solution

- Offloaded ETL, exploration, and other analytic reporting to Cloudera
- Greater flexibility and scale to explore longer histories of data and converge data sets from disparate sources
- Create more robust merchant reports faster for new revenue stream

Results

- \$30M in annual savings from offload
- 1/15th cost per TB vs traditional DW



cloudera®

cloudera

Spotlight: Connected Government
and the Internet of Things

Key Drivers for Instrumenting Your Operations



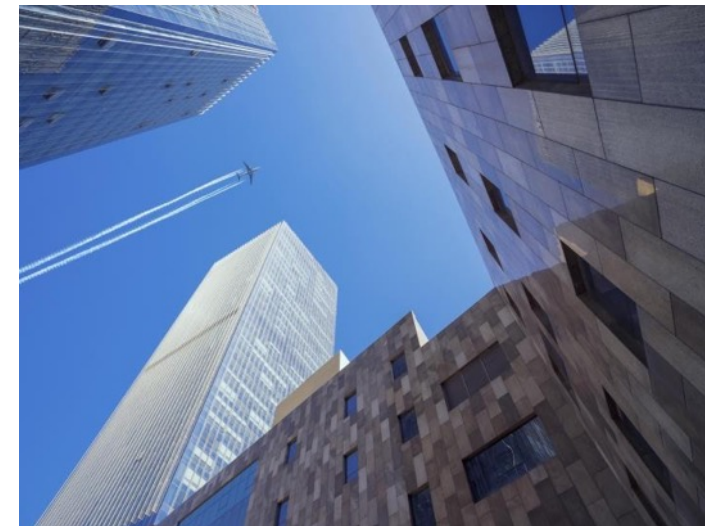
Drive Operational Efficiencies

How can we lower asset downtime?
How can I drive efficiencies?



Improve Citizen Experience

How are people using my products?
How can we instrument them better?

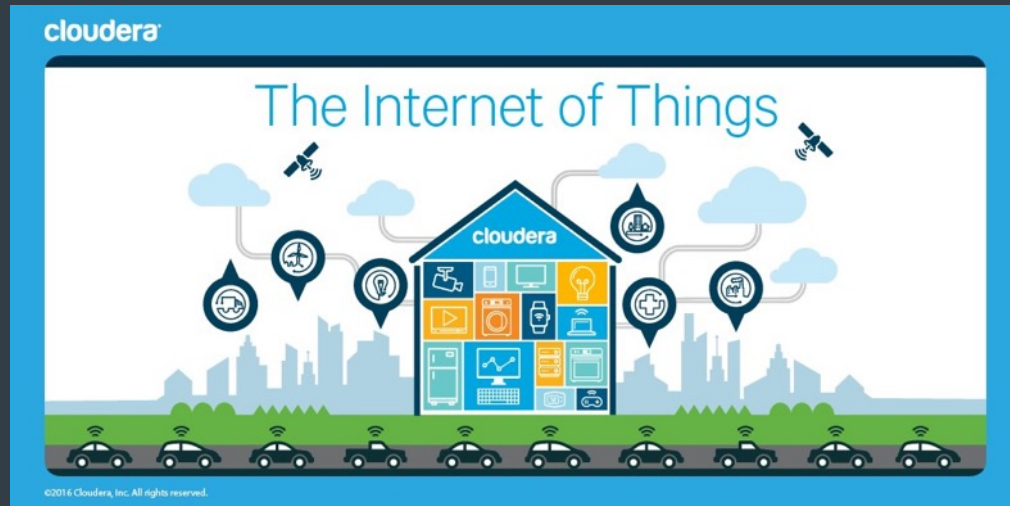


Drive New Operating Models

How do we implement new business models?
How can we launch new services?

Instrument Your Operations – Solutions

IoT Based



Primary
Data
Source

- Streaming Sensor Data Sources
- Connected Devices



Non-IoT Based



Primary
Data
Source

- Non-Sensor Data sources



Instrument Your Operations – Key Use Cases



Instrumenting Your Operations – A Big Data Problem

IoT data comes from a variety of different sources

- Massive volumes of **intermittent data streams**
- Generated from a **variety of data sources**
- Predominantly **time-series**
- Can come in **streams (real-time)** or **batches**
- Diverse data **structures** and **schemas**
- Some of it may be perishable



Combining sensor data with contextual data is the key to value creation from IoT

Cloudera – Key Enabling Capabilities



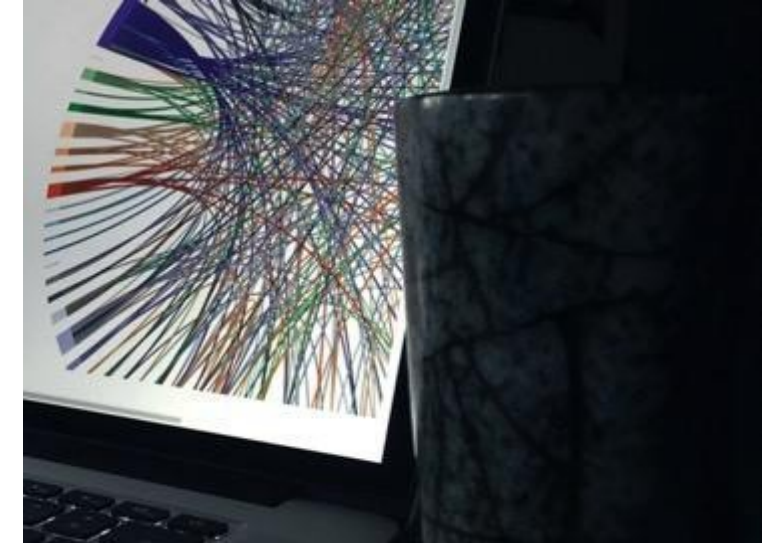
Kudu: Real-Time Analytics

Ideal for real-time analytics on **IoT** and **time series** data. Simplifies Lambda architectures for running real-time analytics on streaming data



Multi-Cloud Portability

Preserve business flexibility and data portability and **minimize cloud lock-in** by running in any one of the three major public cloud providers or in private cloud



Data Science

Collaborative **Workbench** for **Enterprise data science** and an integrated development environment for running **Python, R, & Scala** with support for Spark



IOT &
Connected
Products

- TRAVEL & TRANSPORTATION
- » PREDICTIVE MAINTENANCE
 - » IMPROVED SERVICE
 - » DATA DRIVEN PRODUCTS

NAVISTAR®

Using Predictive Maintenance to Improve Performance and Reduce Fleet Downtime

- Real-time visibility of 300,000+ trucks in order to improve uptime and vehicle performance
- OnCommand Connection is collecting telematics and geolocation data across the fleet
- Reduced maintenance costs to \$.03 per mile from \$.12-\$.15 per mile
- Centralizing data from 13 systems with varying frequency and semantic definitions





Data-Driven
Products

CASE STUDY

TRAVEL & TRANSPORTATION

- » INTERNET OF THINGS
- » PREDICTIVE MAINTENANCE
- » ADVANCED ANALYTICS

Smart Ports

Leading Cargo Handling
Providers in Europe

Using sensors & IoT to improve efficiencies
in cargo handling

Challenge:

- Bring together data streams from **millions** of cargo equipment to enable predictive maintenance

Solution:

- **Sensor Data Analytics platform** based on Cloudera and TCS to collect, store and analyze data collected from port equipment & machinery
- Improve utilization, reduce unplanned equipment downtime

cloudera



Powering a Variety of Use Cases...



Thank you

Nik Rouda
@nrouda