



cloudera

MODERN DATA WAREHOUSE FUNDAMENTALS

Part II: Exploring the Move to Cloud and Maintaining a Common Data Context

December, 2018

SPEAKERS



Greg Rahn

Director of Product Management
grahn@cloudera.com

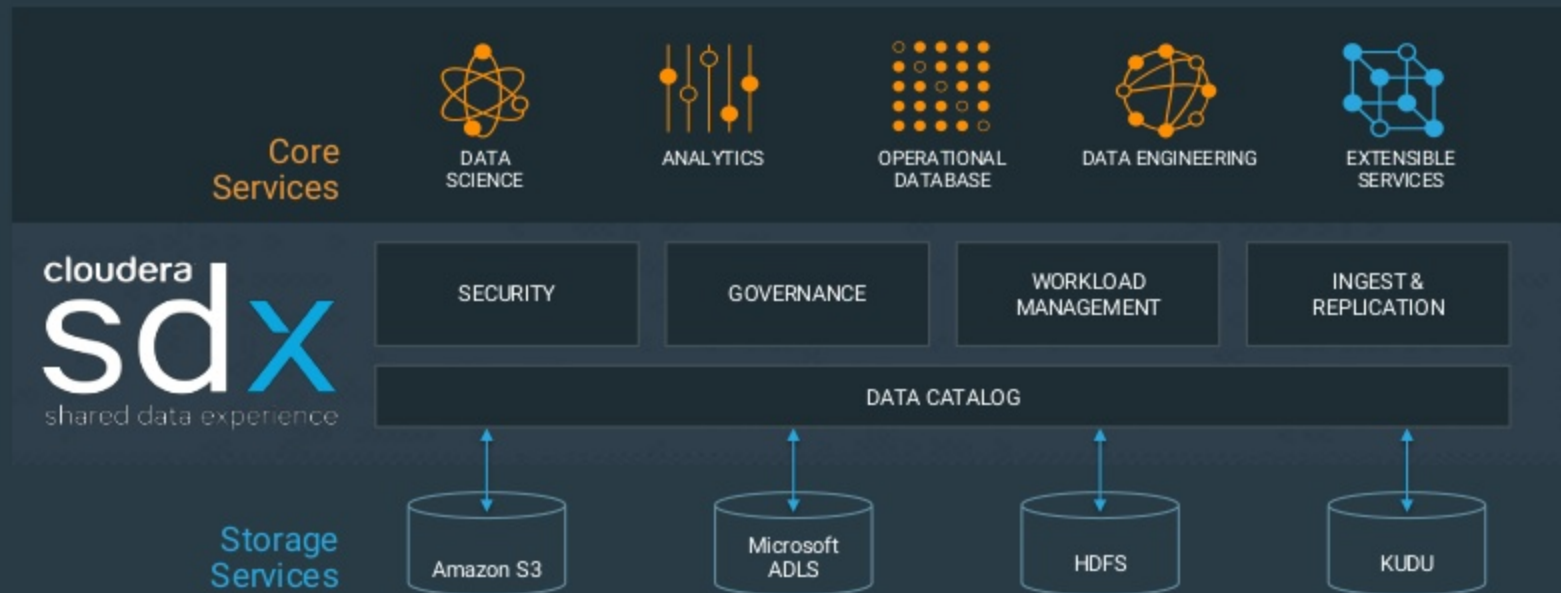


Santosh Kumar

Senior Product Manager
skumar@cloudera.com

BEYOND DATA WAREHOUSING

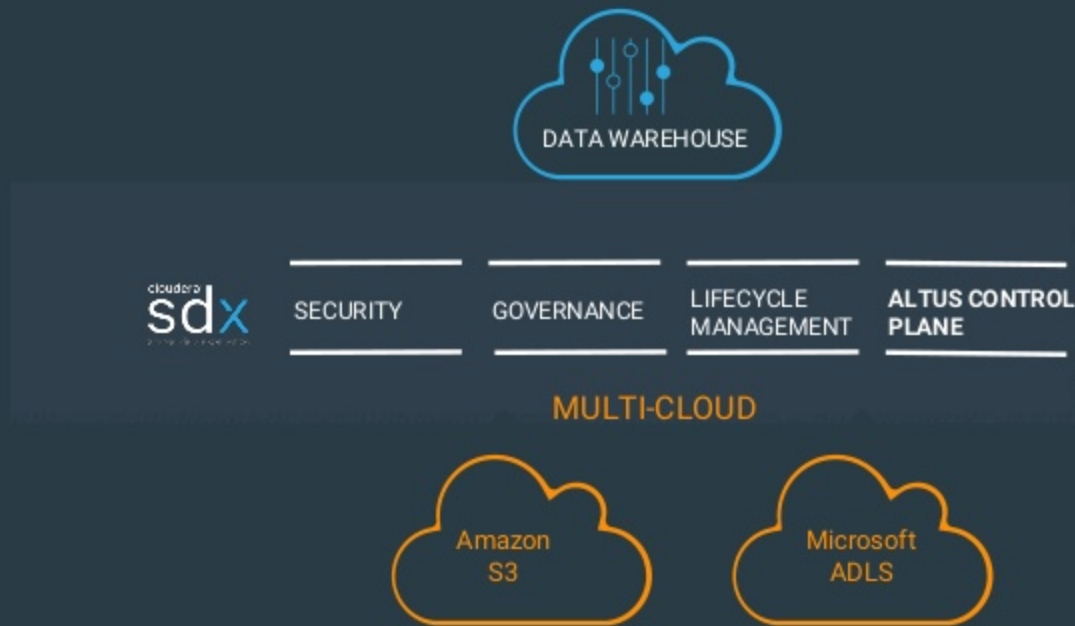
The modern platform for machine learning and analytics optimized for the cloud



WITH A CLOUD NATIVE OPTION - ALTUS DW

MULTI-CLOUD PAAS SOLUTION

- Quick time to value - no software or clusters to manage
- Bring warehouse to the data with zero copy simplicity
- Use your security policies with your data - no proprietary stacks
- Apply enterprise governance to transient workloads
- Shared data experience with SDX
- Optimized for Azure & AWS



KOMATSU MINING: Optimize Machine Performance

CHALLENGES

Create an Industrial IoT (IIoT) solution for optimizing mining equipment utility and build better next-generation products

Current system couldn't handle:

- Scale of IoT data
- Demand for new users and use cases
- 30TB/month data growth

SOLUTION

Cloud-based IIoT analytics for a full view of mining operations

- Quickly and easily analyze huge volume and variety (time-series, sensor, event, and more) of data
- More use cases and users: "democratizing analytics for different user groups"
- Scale quickly and easily in the cloud

RESULTS

- 2X Increase in production hours on key equipment
- Design next-generation equipment: environmentally smarter, more productive, at lower cost
- Meet or exceed all KPIs: "Deliver all of the data with less complexity and significant cost savings"

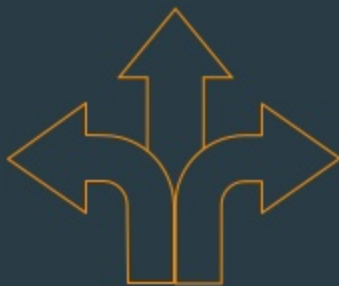
<https://www.cloudera.com/more/news-and-blogs/press-releases/2017-11-15-komatsu-helps-improve-mining-performance.html>

MOTIVATIONS FOR THE CLOUD

BUSINESS DRIVERS FOR DATA WAREHOUSING IN THE CLOUD



Scalability (51%)



Flexibility (41%)



Business agility /
reduce IT involvement
(39%)



Cost (37%)

*BI, Analytics, and the Cloud: Strategies for Business Agility, TDWI, 2016

TECHNOLOGY DRIVERS IN THE CLOUD

3. Open and modular architectures



2. Access to limitless utility-based compute



Amazon
EC2



Azure
Virtual Machine

1. Cost-effective, scalable storage in a single, shared repository



Amazon S3



Azure Data
Lake Storage

KEY STAKEHOLDERS



KNOWLEDGE WORKERS

Instant, self-service
access to data and
resources

Application performance

Job-oriented tools

Choice



DATA TEAM

Advance strategic
initiatives

Link analytics to business

Reduce admin burden

Integrated solutions



INFRASTRUCTURE TEAM

Secure, controlled
provisioning

Predictable costs

Systems-oriented tools

Standards and portability

KEY BENEFITS

Modern Data Warehouse



ADVANTAGES OF A MODERN DATA WAREHOUSE

High-Performance SQL and ...



Data Flexibility

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



Go Beyond SQL

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



Cost-Effective Scalability

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



Hybrid Decoupled Architecture

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

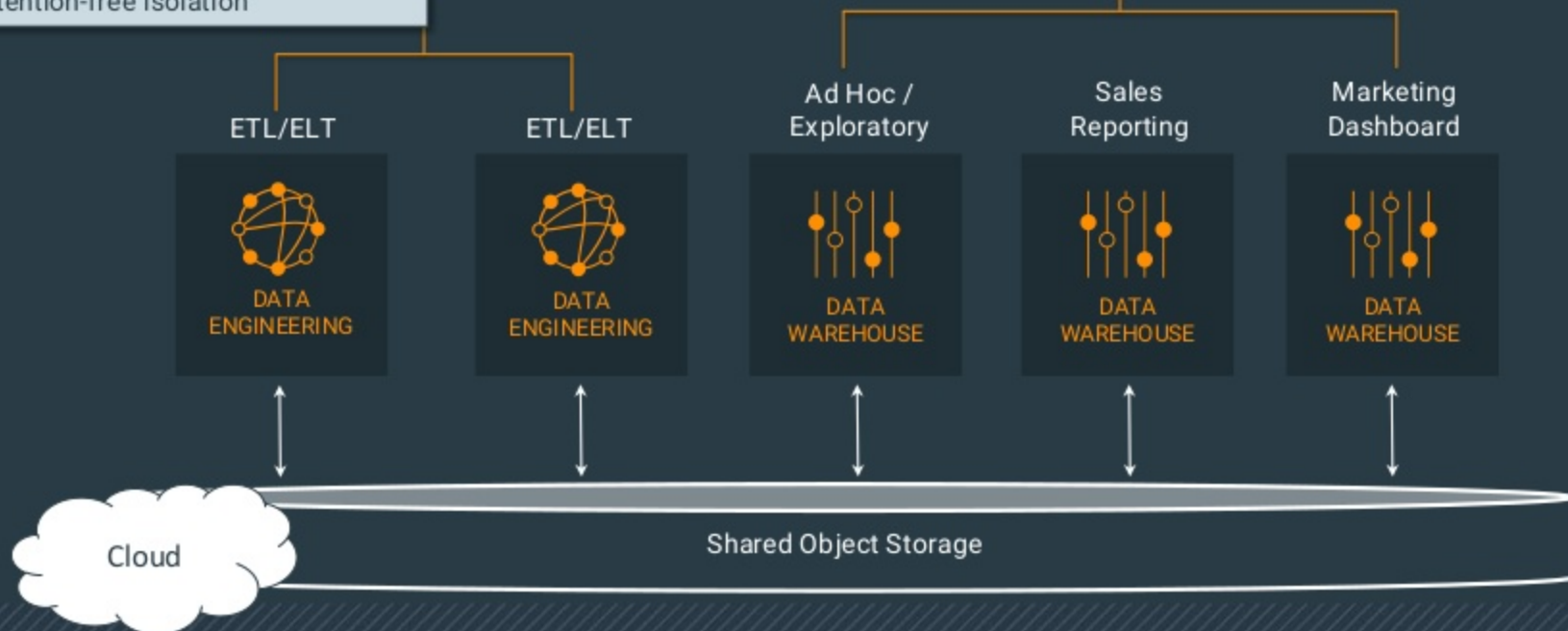
COMMON CLOUD ANALYTIC PATTERNS

Only pay for what you need,
when you need it

- Transient workloads
- Contention-free isolation

Self-service flexibility at any
scale

- Elastic scale on-demand
- Multi-tenant isolation



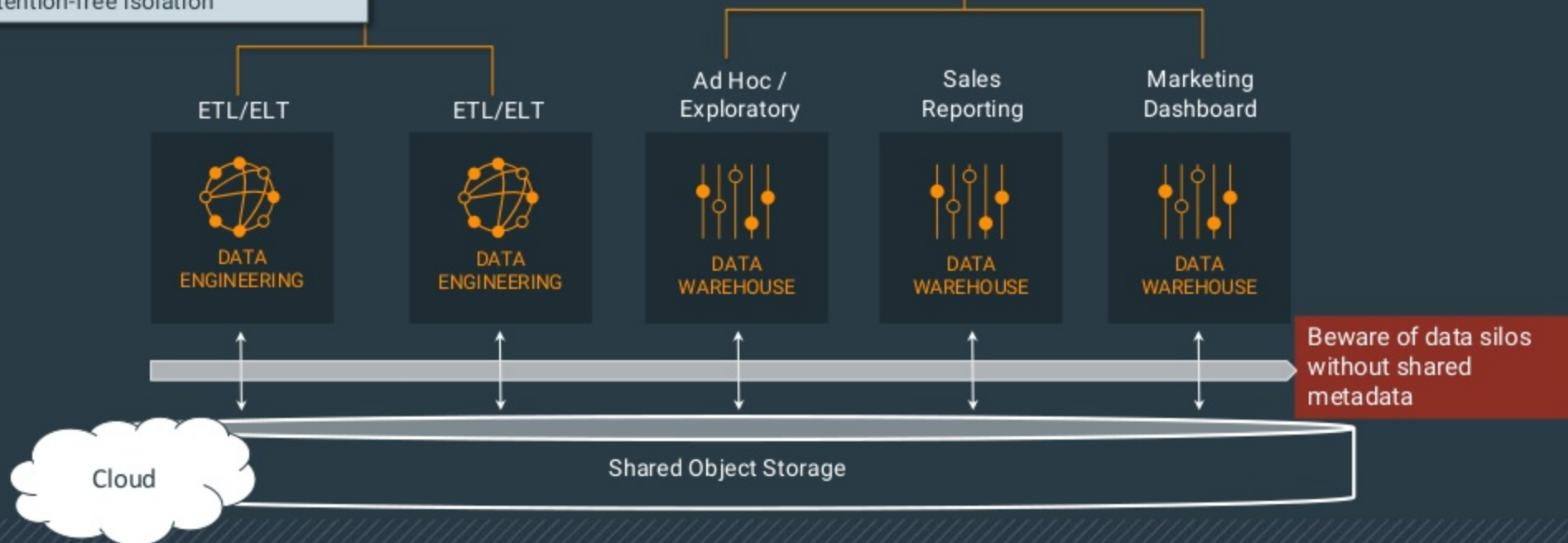
COMMON CLOUD ANALYTIC PATTERNS

Only pay for what you need,
when you need it

- Transient workloads
- Contention-free isolation

Self-service flexibility at any
scale

- Elastic scale on-demand
- Multi-tenant isolation



INTELLIGENT DATA CONTEXT - SHARED DATA EXPERIENCE

INTELLIGENT DATA CONTEXT



Stateful Context, Shared Experience



With Cloudera Altus Data Warehouse and SDX running on Microsoft ADLS, we were able to establish our Telekom Data Intelligence Hub: a trusted, fully governed platform and ecosystem where our users are empowered to exchange and analyse data and develop multi-function, data-driven applications easier and securely. - Sven Löffler, BizDev Executive

CUSTOMER STORIES



NAVISTAR®



Couldn't solve predictive maintenance goals

EDH delivers:

- Ingest telematics in real-time
- Machine learning to predict failures
- Analytics to minimize service downtime
- Protect sensitive and regulated data
- Consistent security and governance
- "SDX is the key to making that happen" - CIO



Drug R&D too slow and expensive

EDH delivers:

- Self-service analytics
- Meet HIPAA regulations
- >5 petabytes from 2100 silos
- Using Spark, Impala, & Search side-by-side
- With Anaconda, AtScale, Cloudwick, Kinetica, StreamSets, Tamr, Trifacta, & Zoomdata

CHALLENGES WITH MULTIPLE DEPLOYMENT MODELS

How are you managing your Data Warehouse today?

How do you share datasets?

Do you copy things around?

How do you audit accesses across copies?

Have you lost the track of the Source of the Truth?

How do you propagate access permissions on copies?

Have you ended up with multiple silos in the process?

BUSINESS IMPACT OF SILOED SYSTEMS



Lost Revenue

Inaccurate and duplicated data directly impacts bottom line of 88% of all companies.



Costly Compliance

By 2023, regulated organizations will spend over 5% of revenue on compliance.



Limits of Legacy

Legacy limits organizations from taking advantage of data-driven opportunities.

Cloudera Enterprise with SDX

provides **maximum cloud flexibility**, enabling enterprise IT to **control workloads anywhere, managed any way**, and deliver a **shared data experience** business and data professionals demand



**Charles,
SVP, Emerging Businesses**

Charles: With increased focus on
... business insights.. dashboard
... FAST...



**Alan,
Internal EDH Data Platform
Manager**

Adding more workloads to Internal
EDH clusters is risky and adds
uncertainty to existing SLA-
sensitive workloads.

May be separate cluster with
"required" data?

Of course! We have our
internal EDH cluster. That
would be easy!

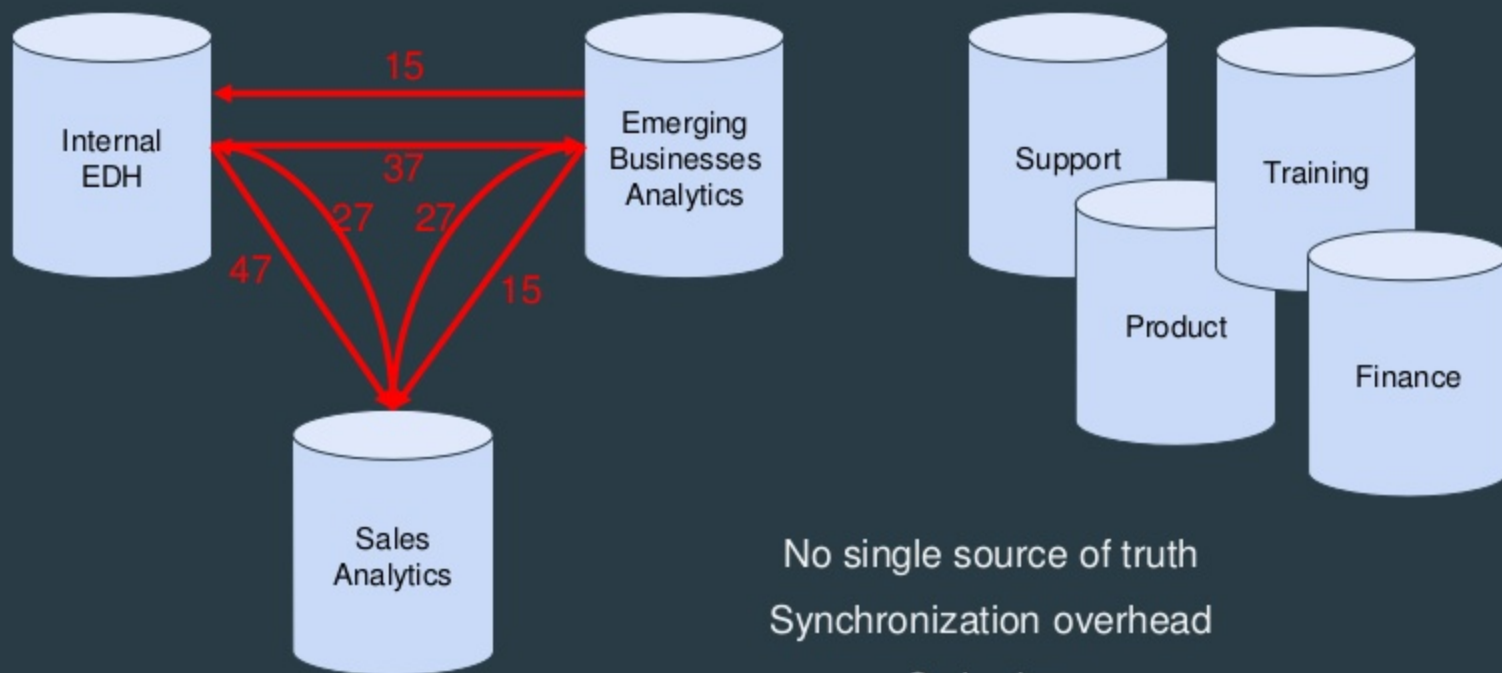
Pipelines! Workloads!
Queries! More pipelines.
More workloads! More
queries! Even more....



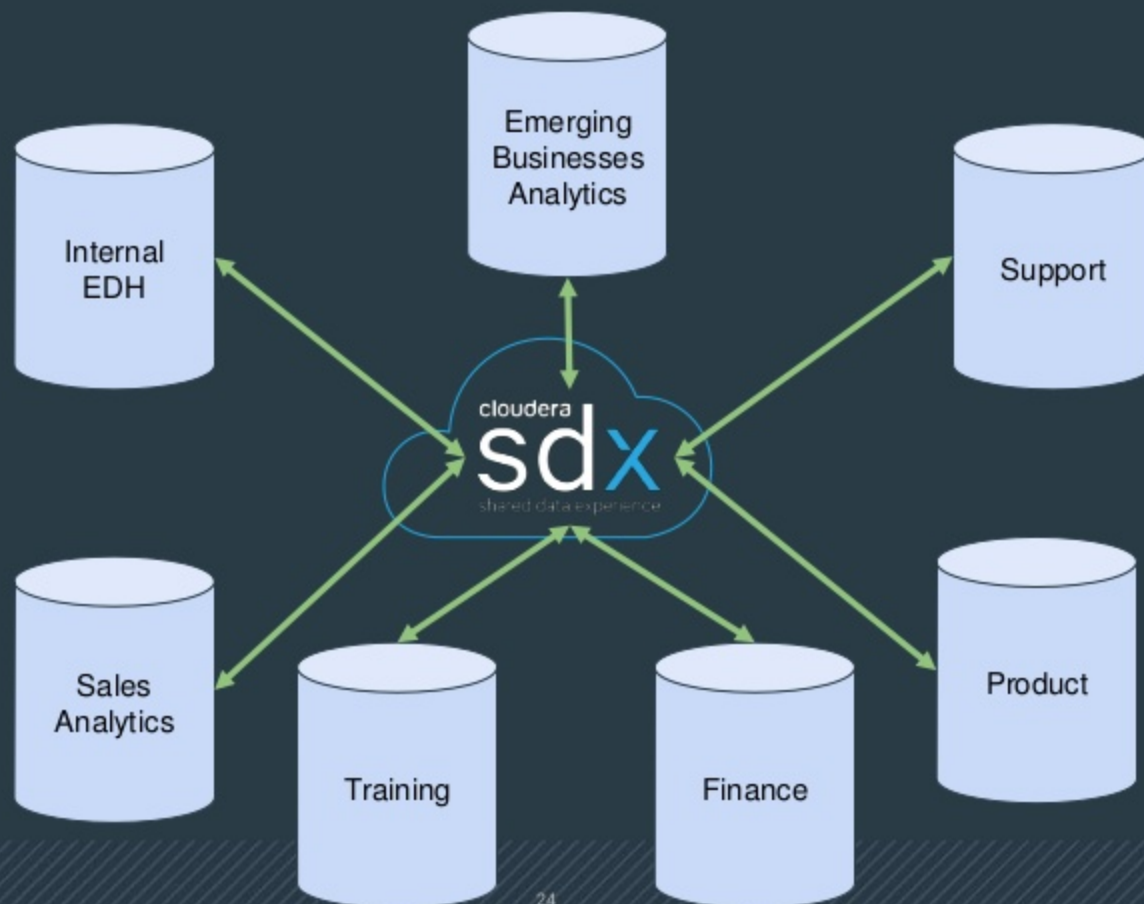
**Mulyadi,
Data Scientist**

Why not!!

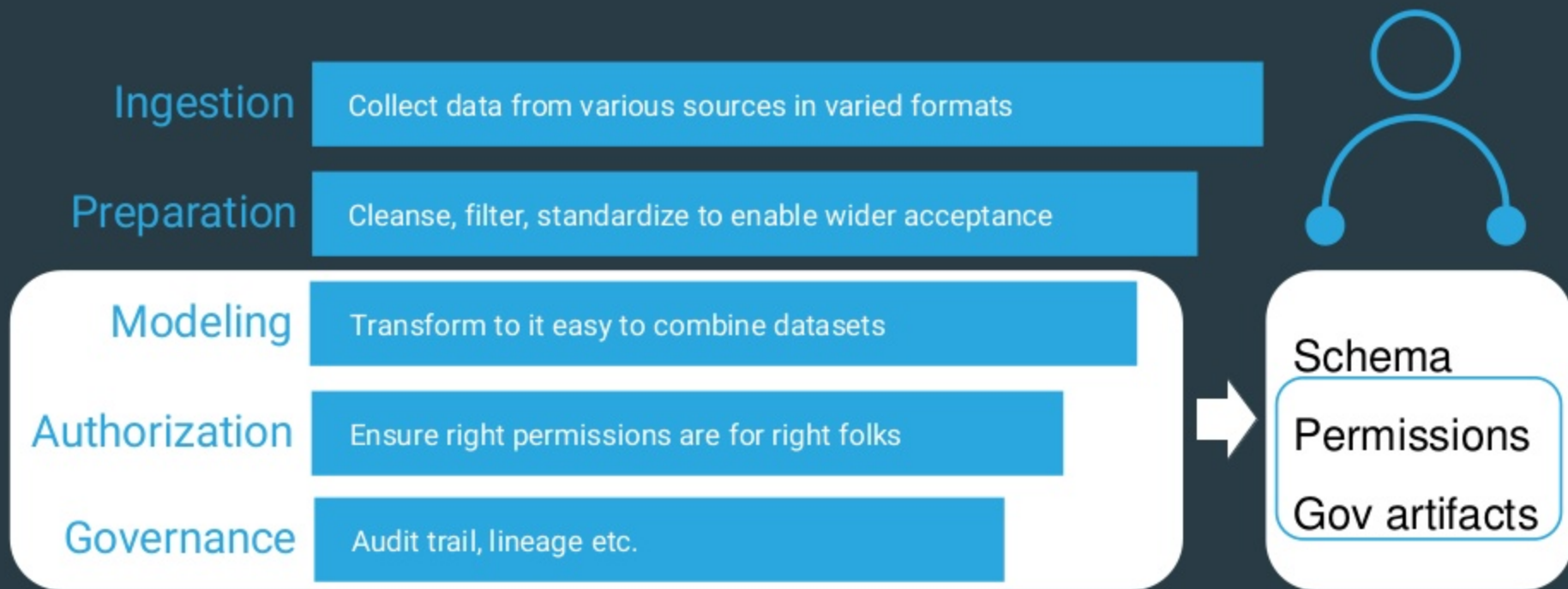
Data Migration Cost Grows Exponentially



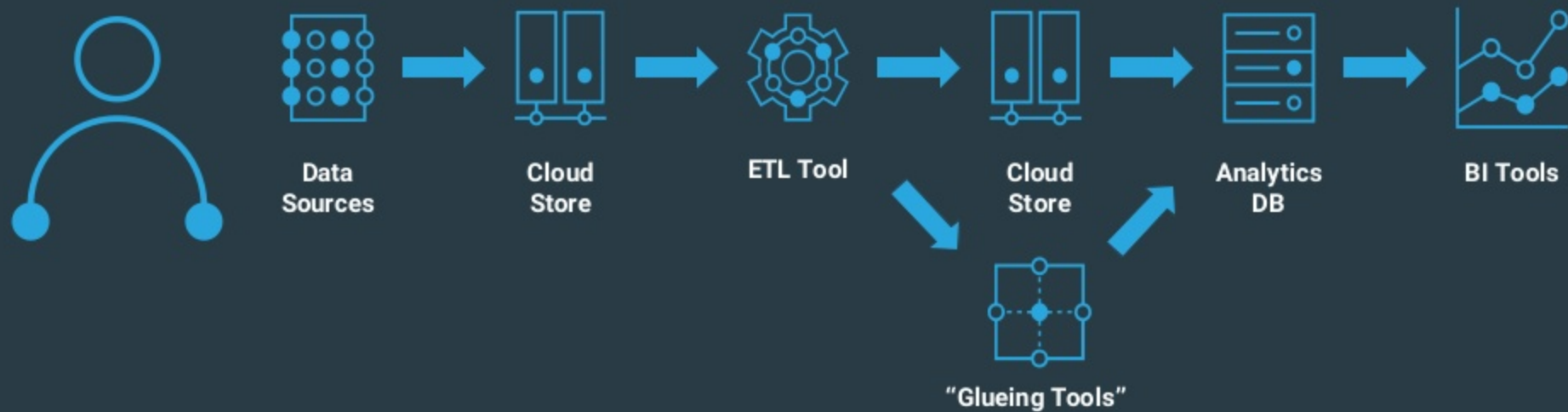
Embrace unification of data and data context via SDX



MODERN DATA WAREHOUSE REQUIREMENTS



DATA WAREHOUSE IN CLOUD DEPLOYMENTS



THREE THINGS TO REMEMBER ABOUT SDX

- **SDX is a differentiated capability offered by Cloudera only**
- **SDX enables a shared data experience across multiple deployment model**
- **SDX provides shared data context essential for global enterprise including schema, access permissions and governance**

CLOUDERA ALTUS FOR DATA WAREHOUSING

CLOUDERA ALTUS DATA WAREHOUSE

BRING THE WAREHOUSE TO YOUR DATA

- ✓ No software to install or clusters to manage
- ✓ Get multiple workloads up and running within minutes
- ✓ Enable self service across your organization
- ✓ Fully secure, automated, with identity preserved across functions
- ✓ Optimized for both AWS and Azure
- ✓ Pay only for what you use

MULTI FUNCTION



DATA ENGINEERING



DATA WAREHOUSE



DATA SCIENCE*



SECURITY

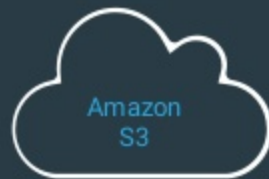
GOVERNANCE

LIFECYCLE
MANAGEMENT

CONTROL
PLANE

DATA CATALOG

MULTI CLOUD



Amazon
S3



Microsoft
ADLS

* roadmap

ALTUS DATA WAREHOUSE

The first data warehouse cloud service to bring the warehouse to the data—delivering instant analytics to anyone

For business analysts:

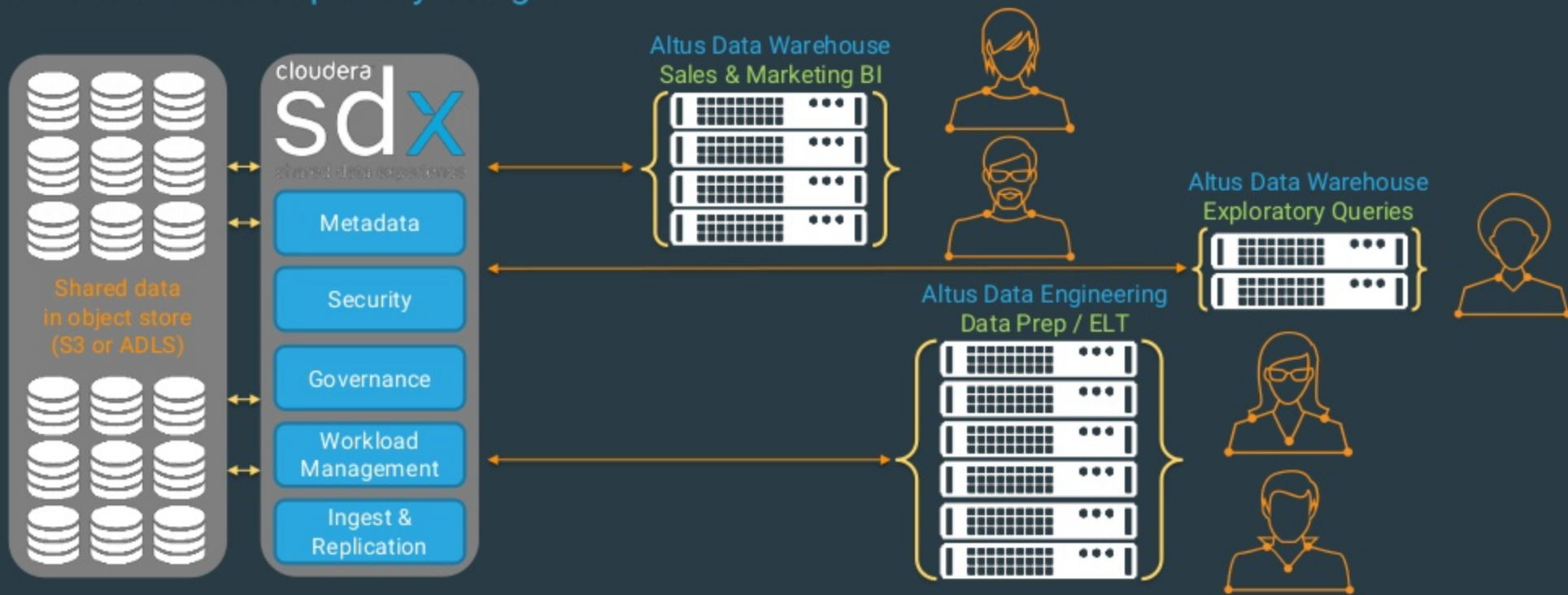
- Run reports and queries at any time, with fast, predictable performance
- Get self service analytic access on demand, using the same preferred tools and SQL skills
- Power reports, BI, exploratory analytics, and ad hoc queries, all over the same shared data and schemas
- Extend insights to data science teams, data engineers, production applications, and more

For IT:

- Eliminate data movement across workloads with lock-in-free open architecture
- Provision isolated resources as they're needed, with just a few clicks
- Easily manage unlimited tenants, and maintain consistent security and governance with the Shared Data Experience
- Support transient and long-running workloads with elastic scale, all with a single view into cloud costs and usage

MODERN DATA WAREHOUSING WITH ALTUS

Elastic and decoupled by design



WHAT'S MISSING FROM YOUR CLOUD DATA WAREHOUSE?



Does data need to be copied/loaded into the database?
Is upfront modeling or a proprietary data format required?

Flexibility



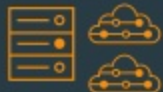
Can you scale compute and storage independently?
What's required to grow/shrink your cluster?

Scale



Is data shared across workloads or do non-SQL workloads require different data silos?

Beyond SQL



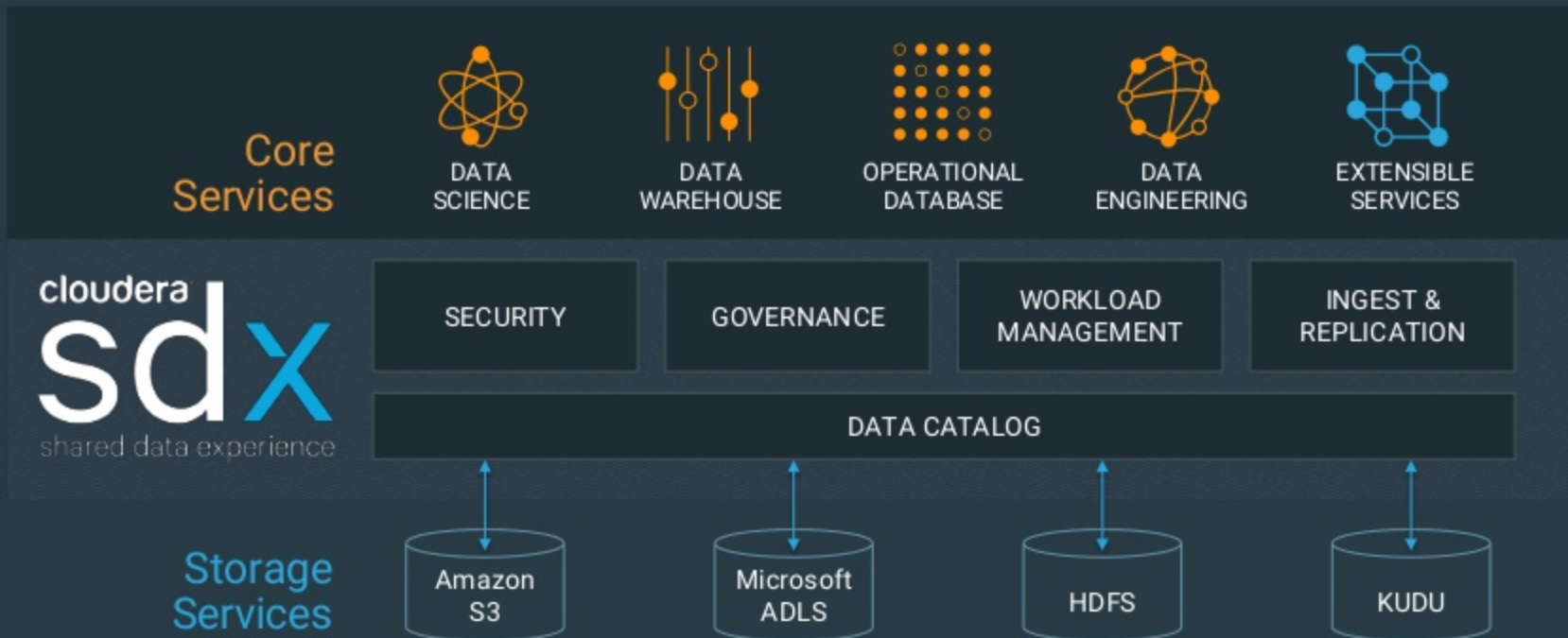
Are object stores a native storage layer?
Can the database span on-prem and multiple cloud environments?

Hybrid

SUMMARY

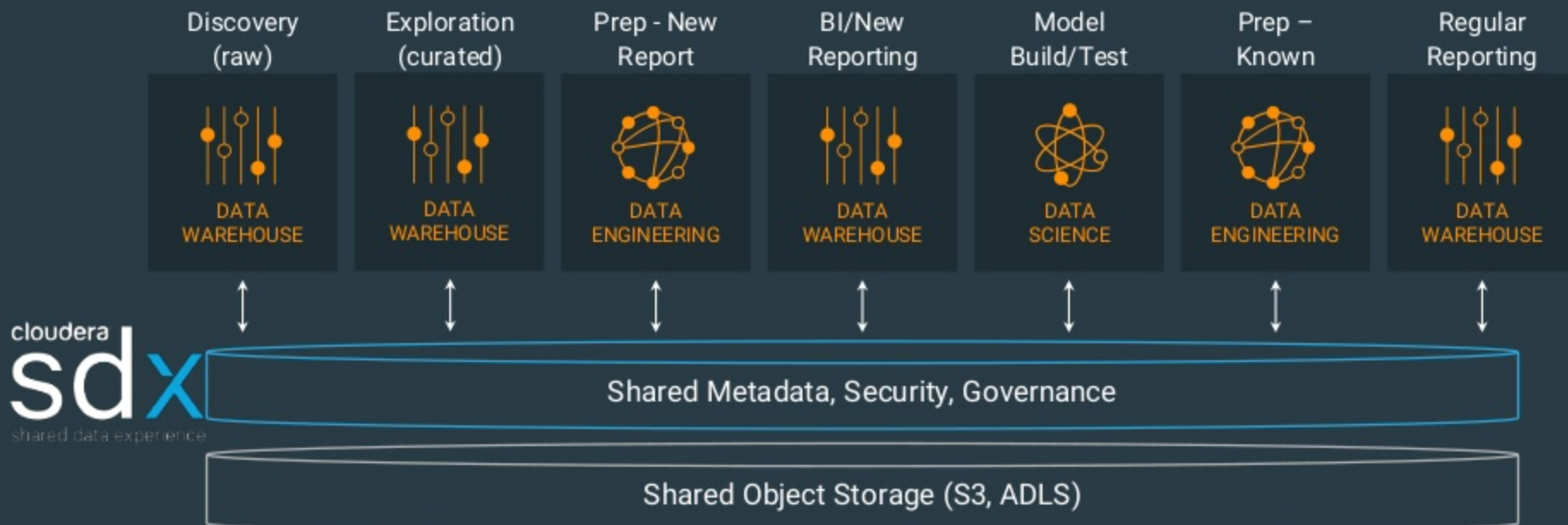
CLOUDERA ENTERPRISE

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



CLOUDERA ALTUS

Data warehousing in the cloud – multiple clusters over single shared data



Q&A

ALTUS FREE TRIAL

<https://cloudera.com/altus>

The screenshot shows the Cloudera Altus website. At the top is the Cloudera logo and a navigation bar with links for PRODUCTS, SOLUTIONS, DOWNLOADS, and MORE. A blue banner below the navigation bar contains the text "Start running your data pipelines in the cloud quickly and easily with Cloudera Altus Data Engineering." and a "FREE 30-DAY TRIAL" button. The main content area features the heading "Cloudera Altus" followed by the tagline "Focus on insights. We handle the big data software." and a sub-headline "The industry's first machine learning and analytics platform-as-a-service built with a shared data experience." Below this are two buttons: "FREE TRIAL" and "LOG IN". A central image shows a laptop displaying the Cloudera Altus user interface, which includes a sidebar menu and a main dashboard with various data visualizations and controls. At the bottom of the page, a section titled "Get control. Unleash the power of the cloud." provides a detailed description of the platform's capabilities, emphasizing its role in automating massive-scale data engineering and analytic database compute workloads in the public cloud, and its core Shared Data Experience (SDX) that eliminates data silos. The footer of the page includes the Cloudera logo, the page number 36, and the copyright notice "© Cloudera, Inc. All rights reserved."

cloudera

PRODUCTS SOLUTIONS DOWNLOADS MORE

Start running your data pipelines in the cloud quickly and easily with Cloudera Altus Data Engineering. **FREE 30-DAY TRIAL**

Cloudera Altus

Focus on insights. We handle the big data software.

The industry's first machine learning and analytics platform-as-a-service built with a shared data experience.

FREE TRIAL **LOG IN**

Get control. Unleash the power of the cloud.

Cloudera Altus lets you automate massive-scale data engineering and analytic database compute workloads in your public cloud, without the headache of managing the infrastructure yourself. At the core of Altus is Cloudera's Shared Data Experience (SDX) that eliminates data silos with persistent metadata, security, and governance.

Altus platform-as-a-service (PaaS) empowers data engineers and BI engineers with self-service capabilities to create and maintain data pipelines on exabytes of data in a shared storage. It eliminates the need for a data lake and the expensive overhead of managing multiple systems for data engineering and analytics.

Cloudera Altus

THANK YOU

<https://www.cloudera.com/products/data-warehouse.html>

cloudera

