

cloudera

CLOUDERA ALTUS

Mark Samson, Principal Solutions Engineer UK
Johannes Muselaers, Solutions Engineer Nordics

Agenda

1. Welcome & introductions
2. Description of Maersk Use Cases of CDH
3. Cloudera Altus
 - What is Altus
 - Demo of spinning up a cluster with Altus
 - Questions on Altus
 - multiple cluster deployment on ADLS
 - Logging and code deployment
 - Security: kerberos, AD/LDAP federated Identity
4. Cloudera Strategy and Roadmap
5. Altus vs CDP
6. Next steps

PDF of the presentation

<https://github.com/jmuselaers/Maersk>

CLOUDERA ALTUS

Platform-as-a-Service
with Shared Data
Experience

MULTI
FUNCTION



SECURITY

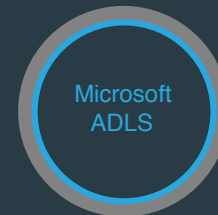
GOVERNANCE

WORKLOAD
MANAGEMENT

INGEST &
REPLICATION

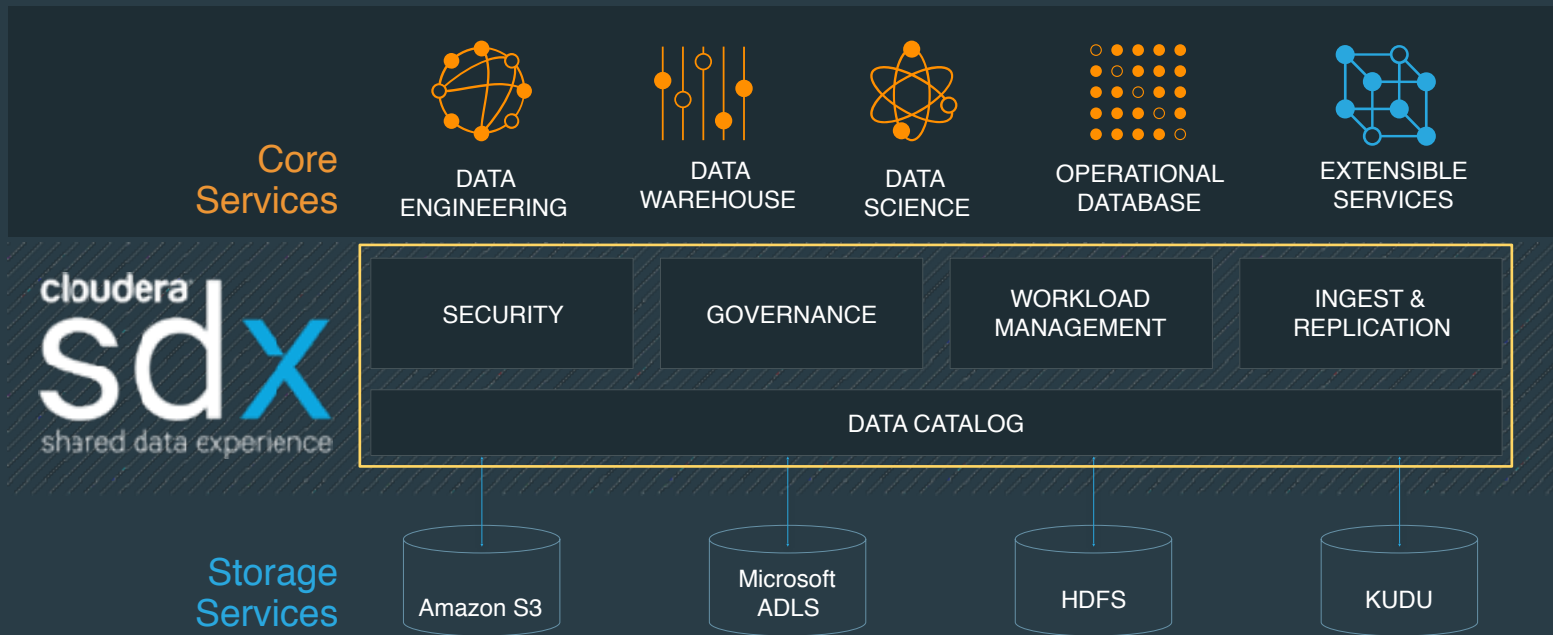
DATA CATALOG

MULTI
CLOUD

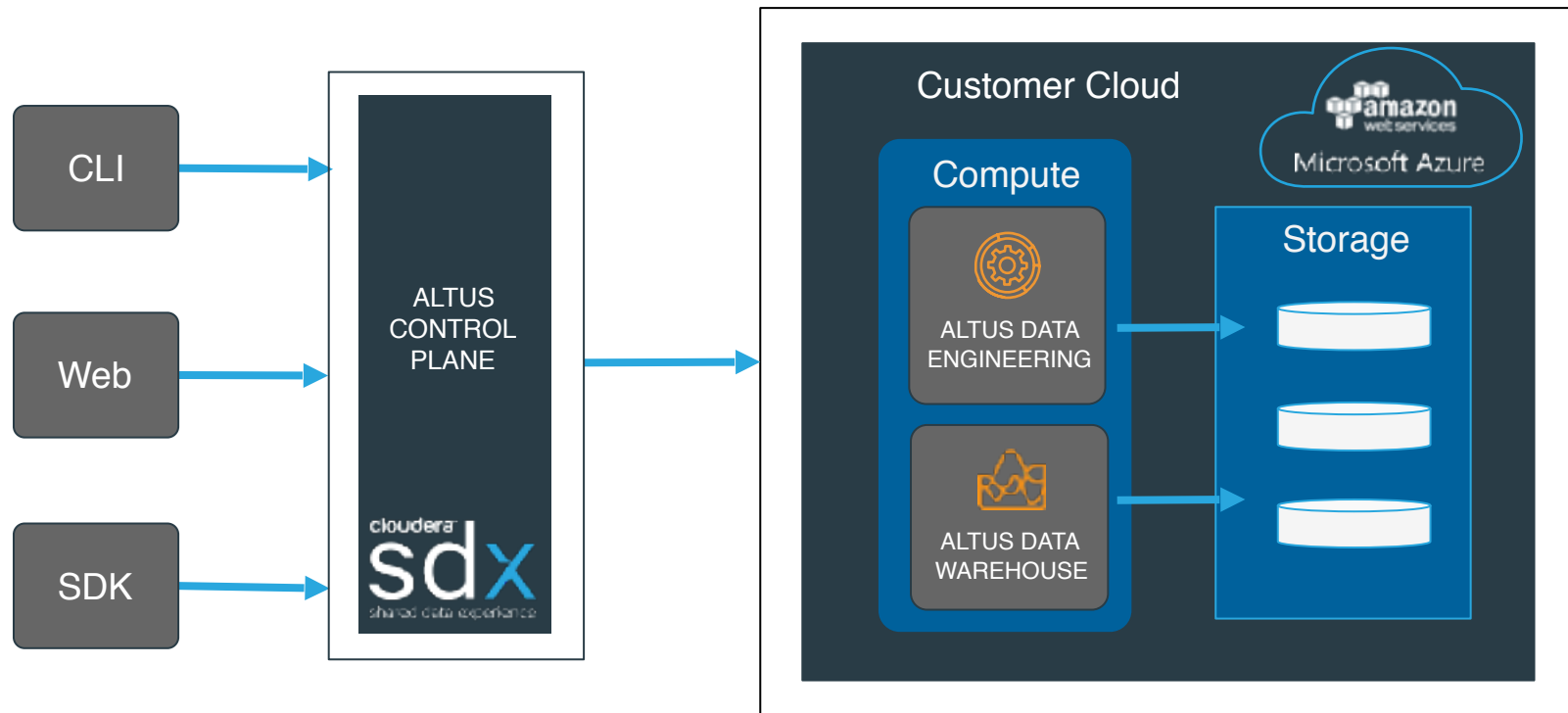


CLOUDERA PLATFORM

Integrated data, workflows, metadata, security, governance, ...



Cloudera Altus Architecture



Altus Data Engineering

What is it?

- Short-lived, single tenant cluster
- Spark, Hive, Hive-on-Spark, and MapReduce jobs
- Processes data in S3 or ADLS

What is it used for?

- ETL jobs
- Batch processing

Why?

- Provides fast and easy job submission without cluster management



Altus Data Warehouse

What is it?

- Long-lived, multi-tenant cluster
- Impala
- Processes data in S3 or ADLS

What is it used for?

- Data Warehousing
- Analytics

Why?

- Provides fast and easy analytics without moving data or worrying about cluster management



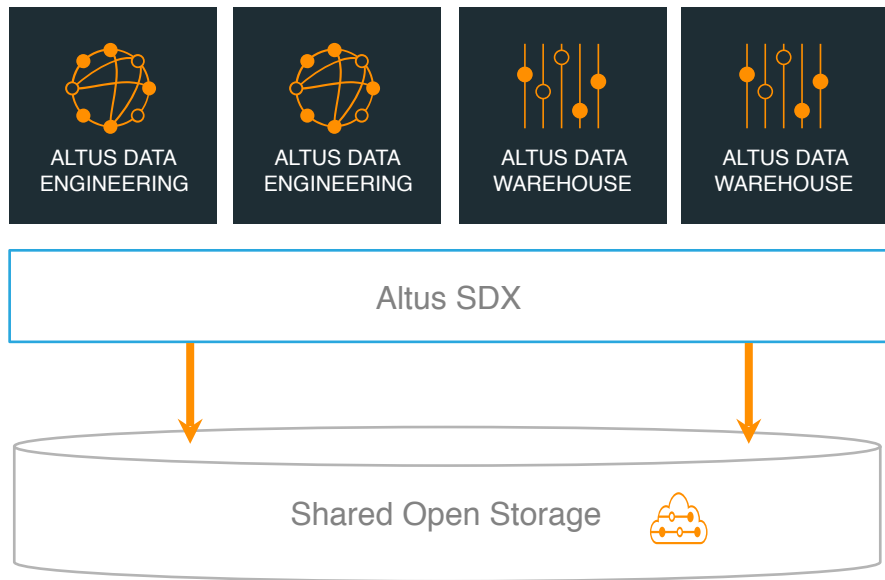
Altus SDX

What is it?

- Cloud native shared metadata store with metadata living in S3 or ADLS

Used for things like

- Shared cataloging to define and preserve structure and business context of data
- Provides unified security across transient and recurring workloads
- Enables consistent governance across all data to increase compliance



EVERYTHING YOU DON'T HAVE TO DO

- Install any software
- Install any hardware
- Worry about cluster configuration
- Upgrade/re-configure clusters
- OS upgrades/patching
- Resource Management

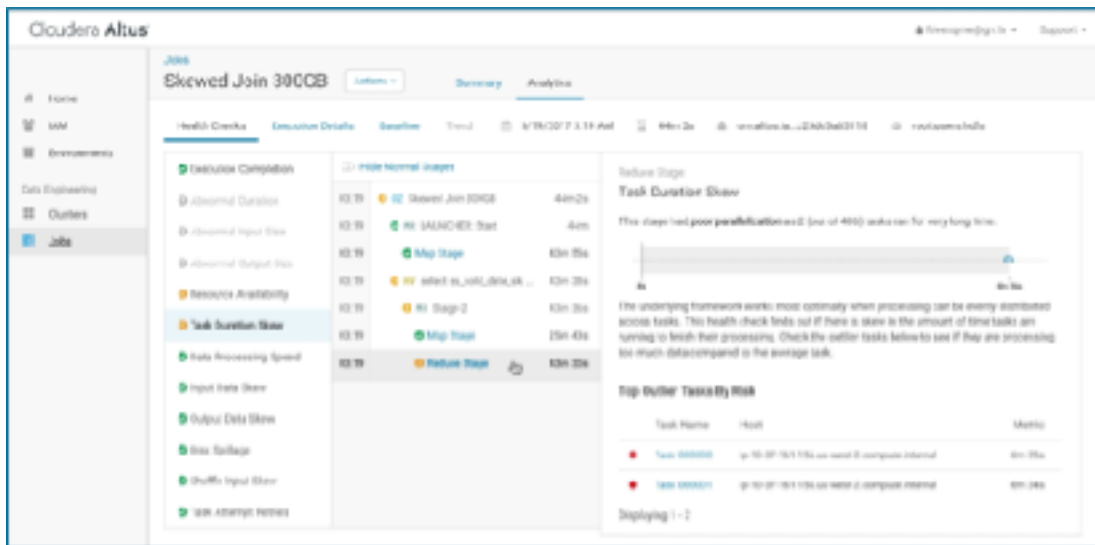


30 day free trial*

Altus Workload Analytics

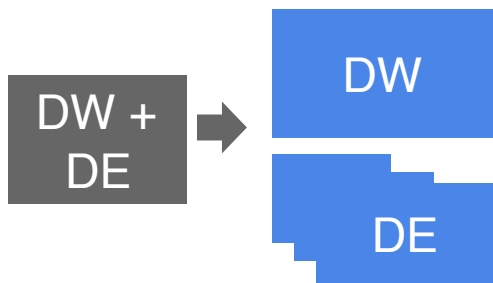
Troubleshoot Issues Faster and Easier

- Get insight into causes of job failure
- Size clusters and optimize job performance
- Identify issues even when they don't show up as errors



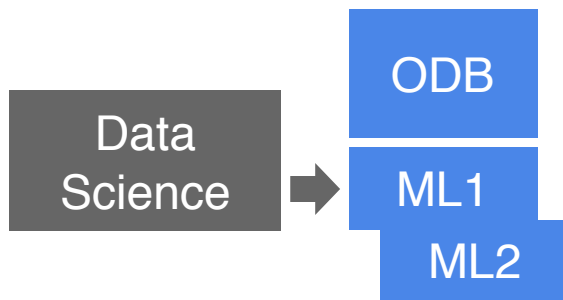
Three Use Cases for Altus

ELASTIC WORKLOADS



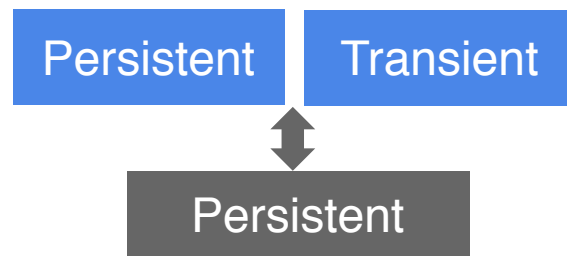
Move from persistent clusters to elastic clusters to be more agile and optimize usage

DATA MART EXPANSION



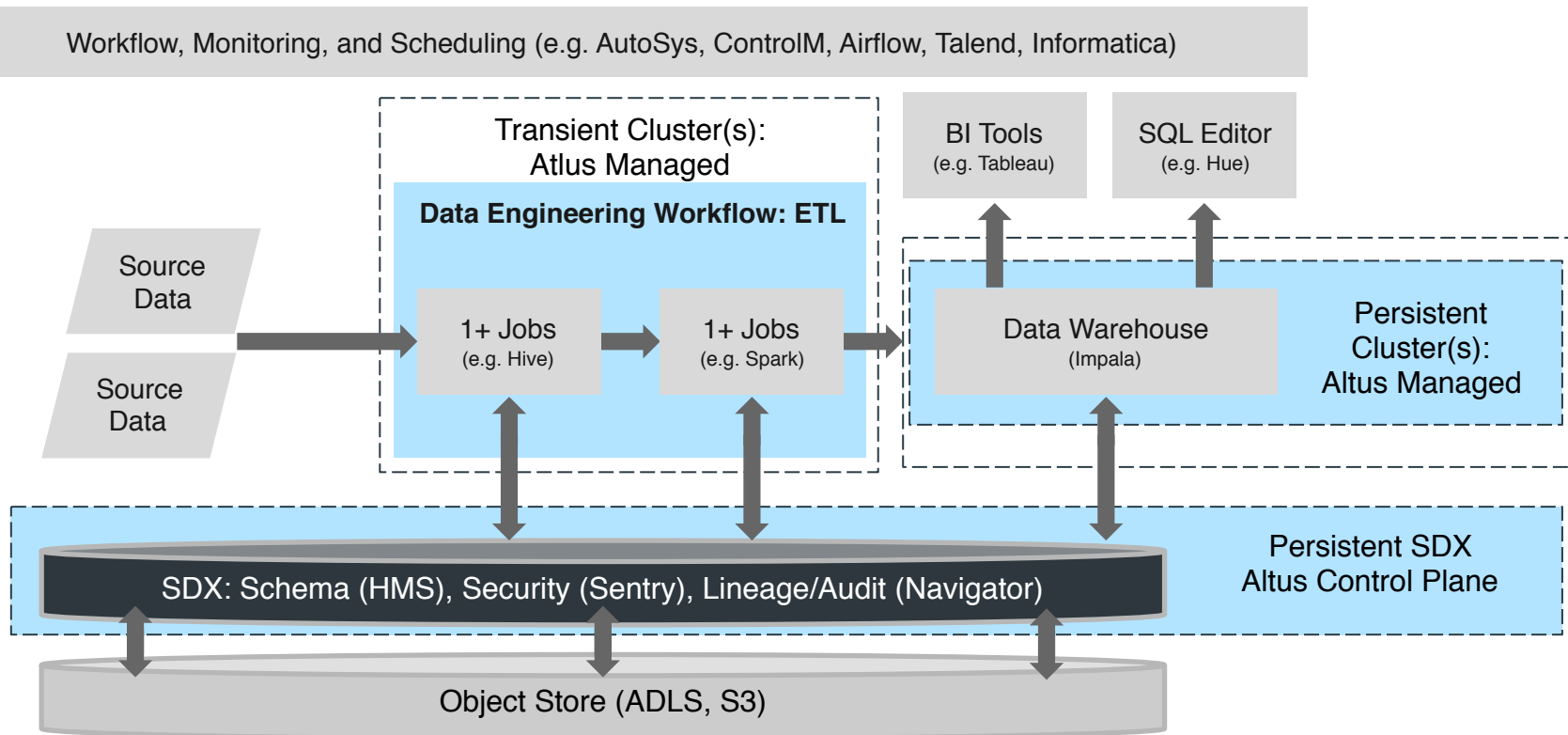
Save labor and/ upfront expense required to accommodate new teams & environments

PEAK SHAVING



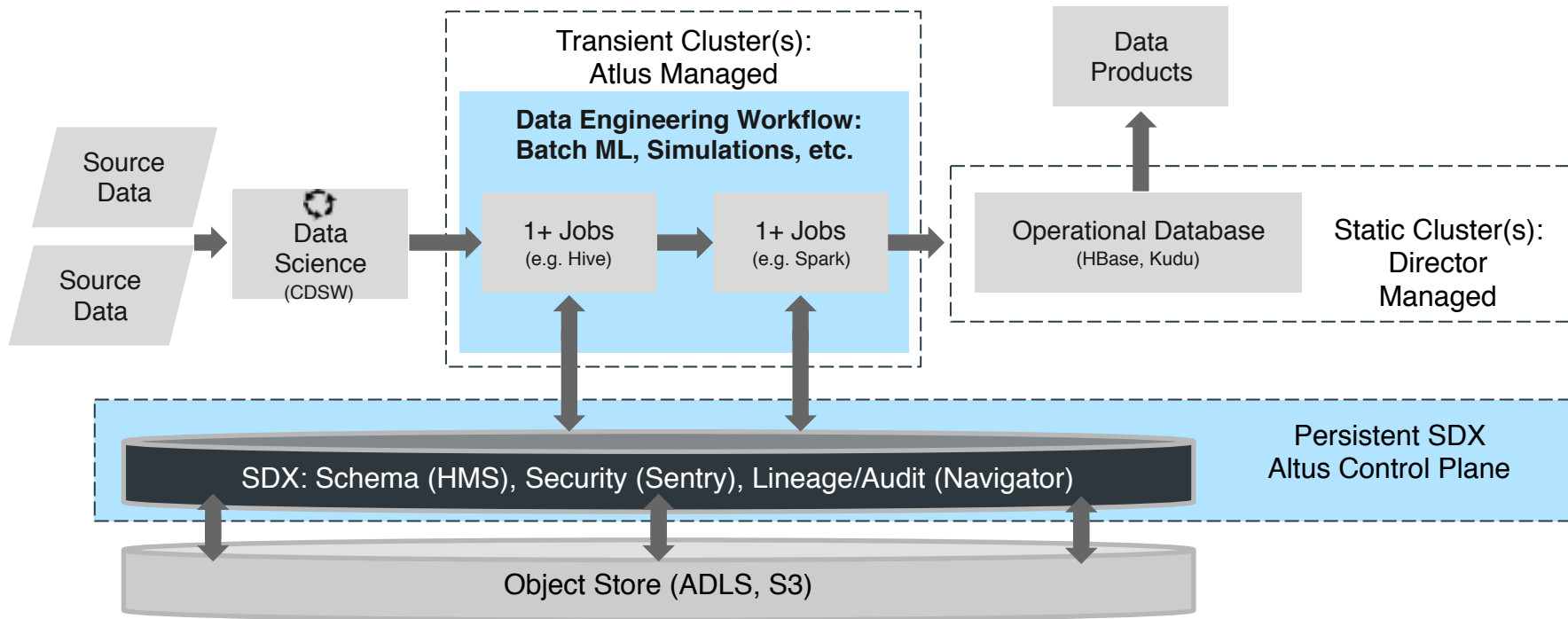
Save hardware/software costs by using transient nodes for peak workloads as opposed to always-on nodes.

Elastic Workloads

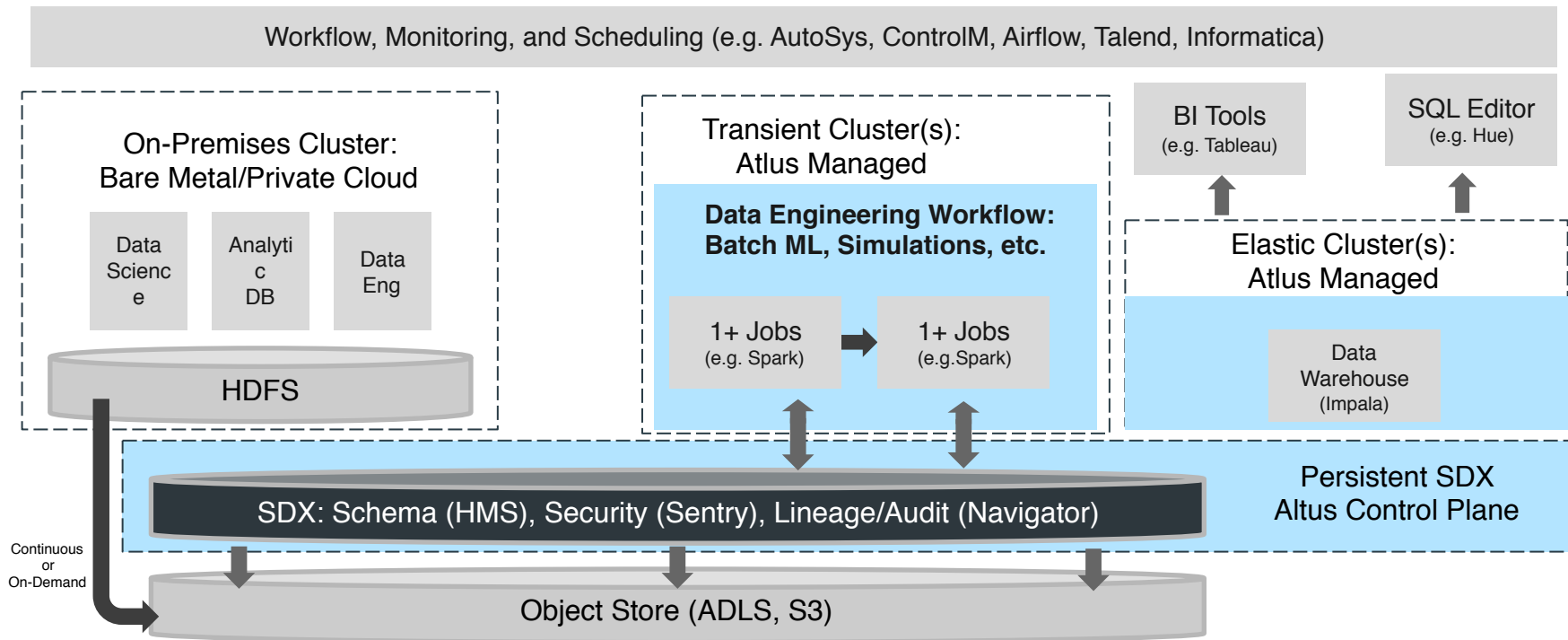


Batch Machine Learning

Workflow, Monitoring, and Scheduling (e.g. AutoSys, ControlM, Airflow, Talend, Informatica)



On-Premises Cloud Bursting



DEMO

RETAIL DEMO

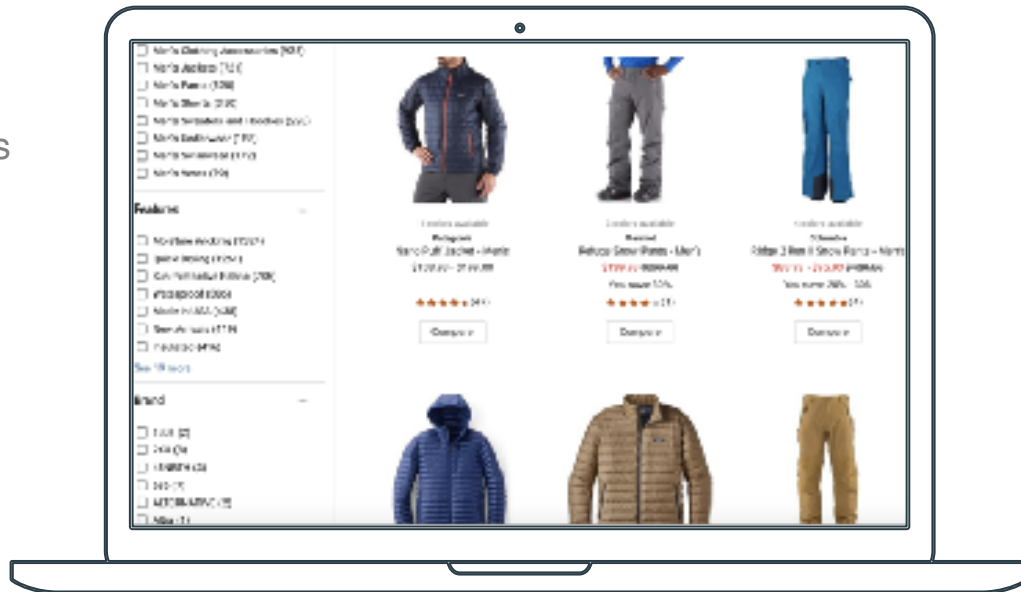
The Problem

- Website sales are struggling
- How to figure whether sales orders correlate with website visits?



Business Analyst/Data Scientist

- Analysis
- Develop insights
- Build models
- Solve business problem

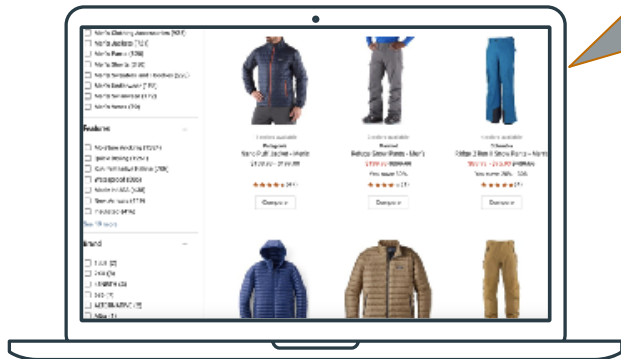


RETAIL DEMO

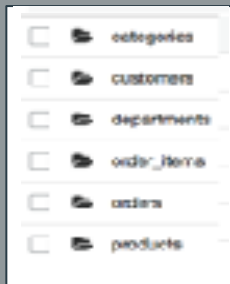


Data Engineer

Develops, tests, maintains large scale data processing systems



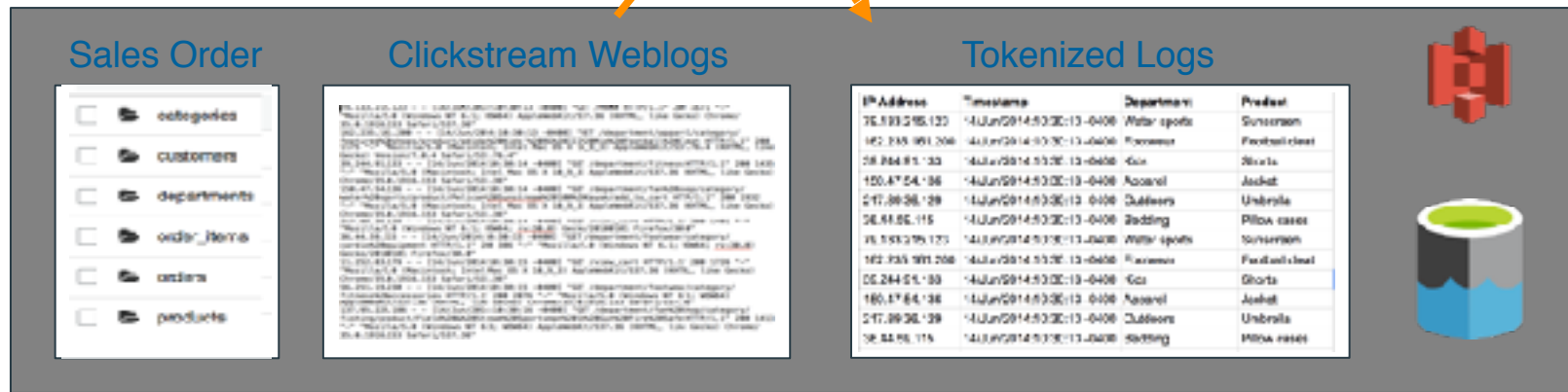
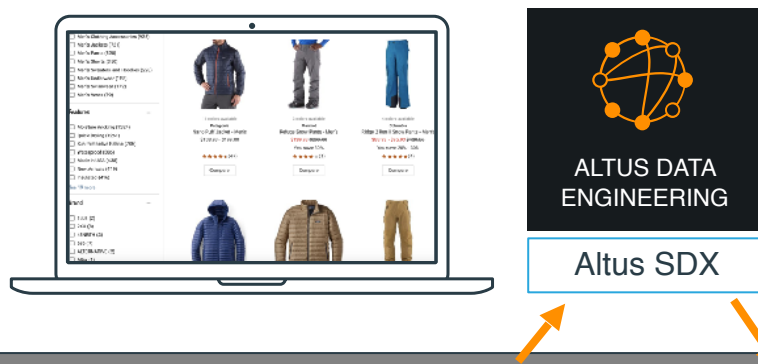
Sales Order



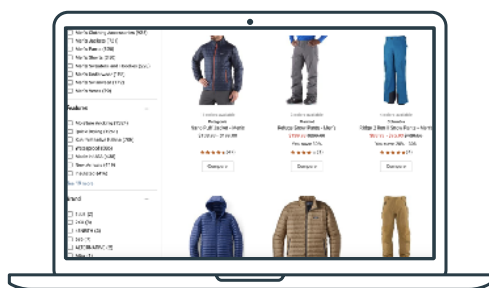
Clickstream Weblogs



RETAIL DEMO



RETAIL DEMO



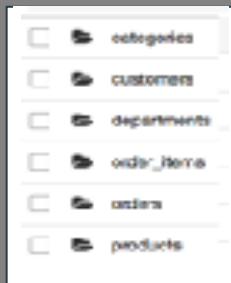
Altus SDX



Sales Order

Clickstream Weblogs

Tokenized Logs



IP Address	Timestamp	Department	Product
76.593.295.123	4Jun2014:5:30:13-0400	Water sports	Sunscreen
192.135.191.206	4Jun2014:5:30:13-0400	Accessories	Facemask/head
58.864.81.33	4Jun2014:5:30:13-0400	Gas	Shirts
100.67.54.36	4Jun2014:5:30:13-0400	Accessories	Jacket
217.89.36.29	4Jun2014:5:30:13-0400	Outdoors	Underwear
56.81.66.115	4Jun2014:5:30:13-0400	Bedding	Pillow cases
76.593.295.123	4Jun2014:5:30:13-0400	Water sports	Sunscreen
192.135.191.206	4Jun2014:5:30:13-0400	Accessories	Facemask/head
58.864.81.33	4Jun2014:5:30:13-0400	Gas	Shirts
100.67.54.36	4Jun2014:5:30:13-0400	Accessories	Jacket
217.89.36.29	4Jun2014:5:30:13-0400	Outdoors	Underwear
56.81.66.115	4Jun2014:5:30:13-0400	Bedding	Pillow cases



WHAT'S NEXT?

“The website team wants to use this data too but I don't want them in my cluster slowing me down. What should I tell them?”

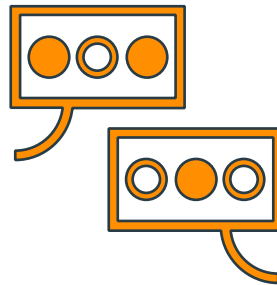
— Analyst

“I had to wait 5 hours for that new weblog data to merge into my data mart. I don't want to wait that long. Can we do this faster? Will it cost me more?”

— Data Engineer

“We've using a new cloud provider. But I don't want to learn a whole new set of tools and concepts.”

— Admin



OBJECTIONS



“This demo looked simple but what did it take to set things up in the first place? Can you handle the complexity of our environment?”

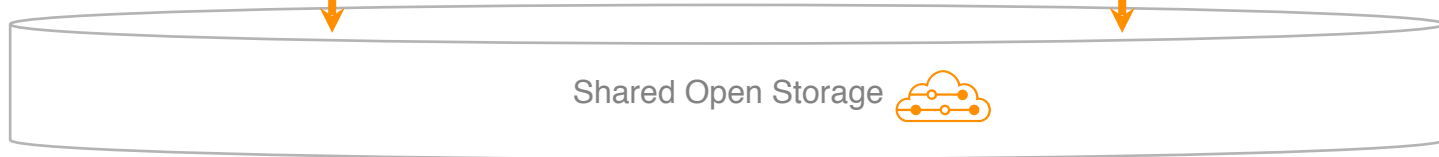
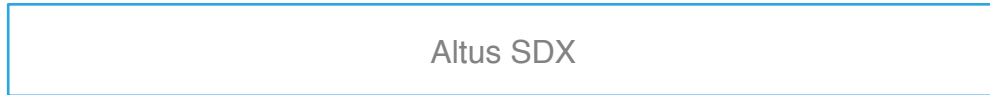
— Admin
“We have strict security requirements. Can you support them? How hard are they to set up?”

— Infosec

PRIMARY WORKLOADS IN THE CLOUD

Only pay for what you need,
when you need it

- Fast, easy job submission
- No cluster management
- Built-in Workload
- Lower costs with transient resources
- Full benefits of resource isolation

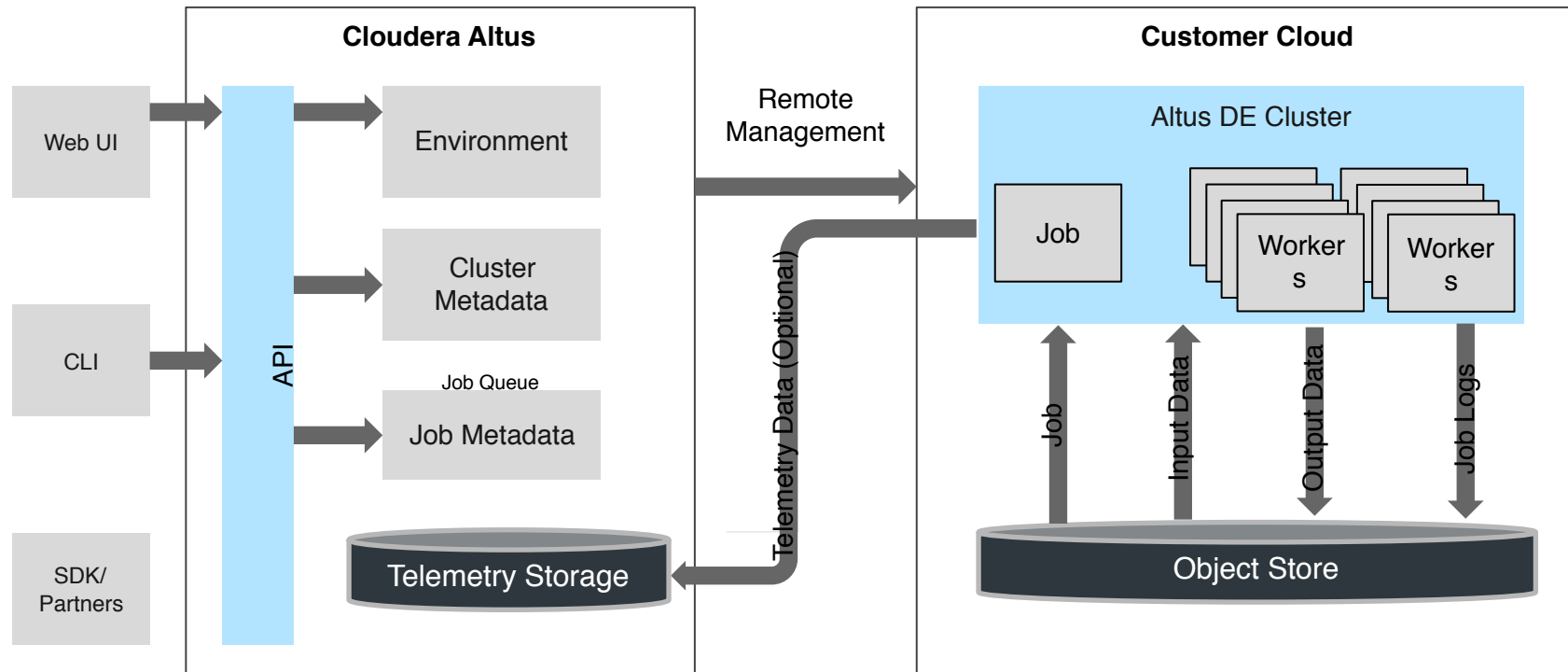


Self-service flexibility
at any scale

- Fast, easy, automated cluster management
- Long running, multi-tenant Impala cluster
- Lower costs with pay-per-use pricing
- Full benefits of resource isolation

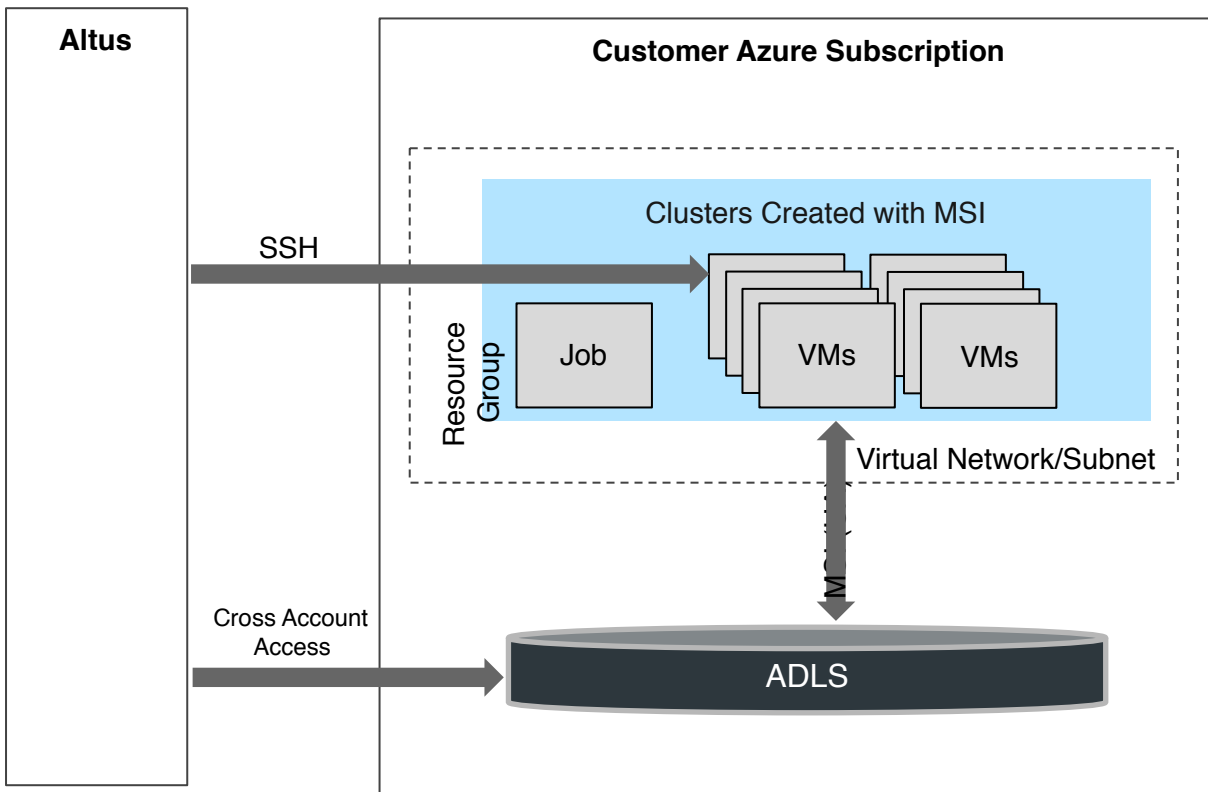
ALTUS ARCHITECTURE AND SECURITY

Altus Architecture



AZURE

Azure Security



Needed Permissions

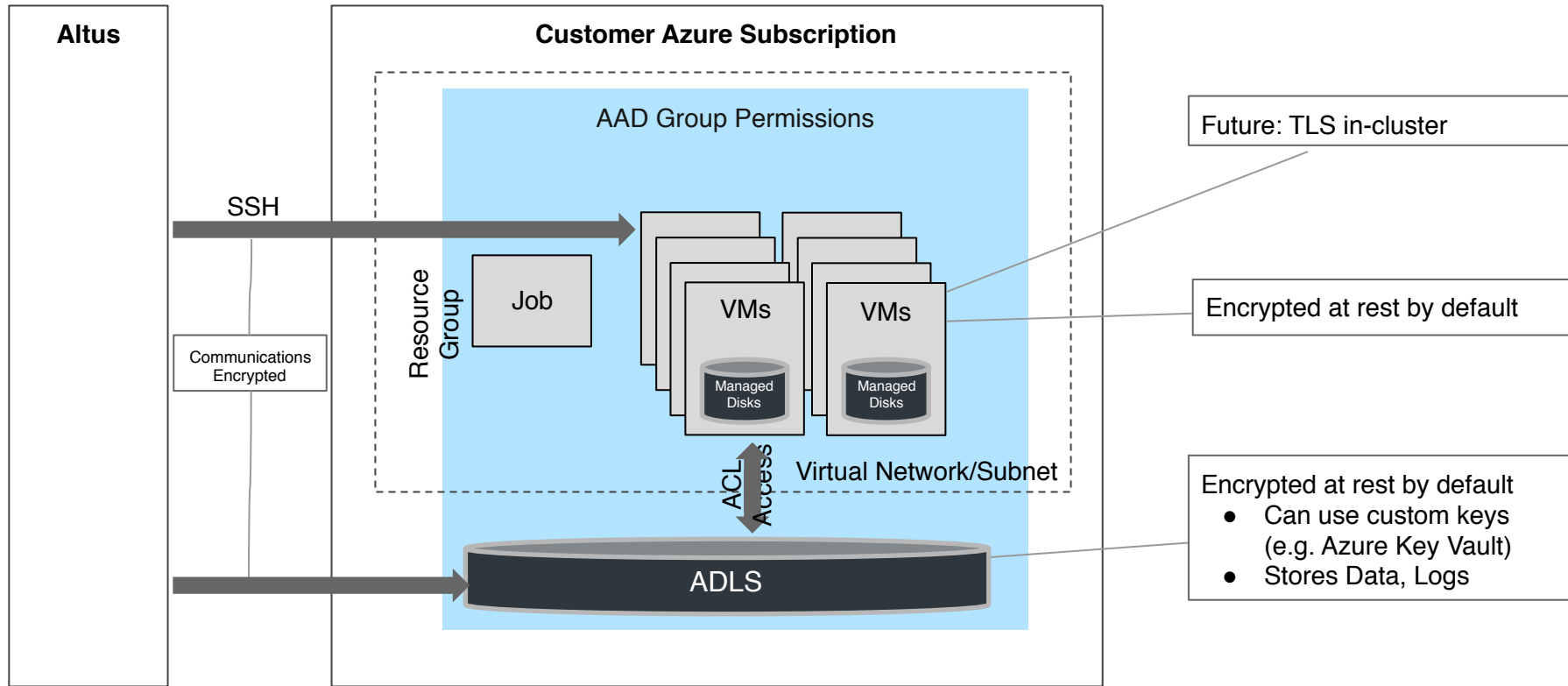
Create a User Assigned MSI with permissions to

- Create resources
 - Can be restricted to Resource Group
 - For POCs, recommended to have contributor access to subscription
- Read/write to ADLS folders via User Assigned MSI

Network

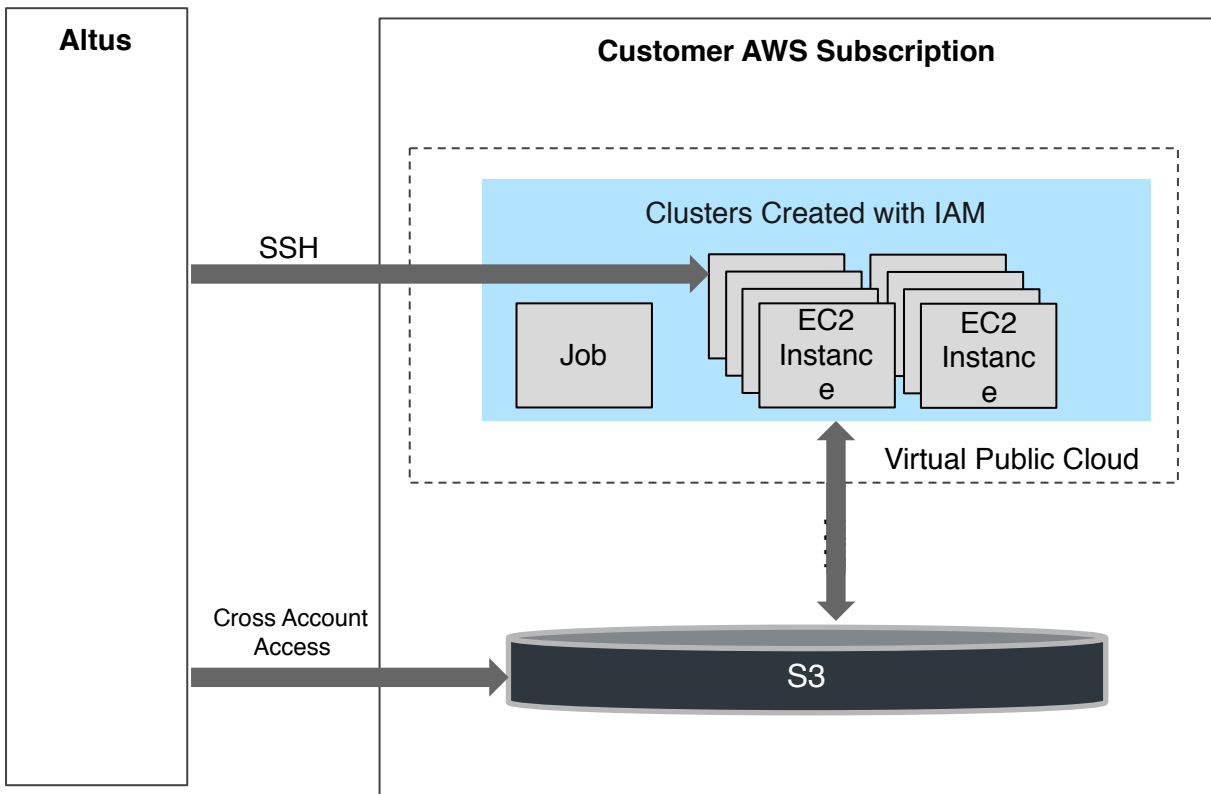
- Ability to SSH from Altus (Cloudera subscription) to VMs for Management
 - Only available from Altus IPs

Azure Encryption



AWS

AWS Security



Needed Permissions

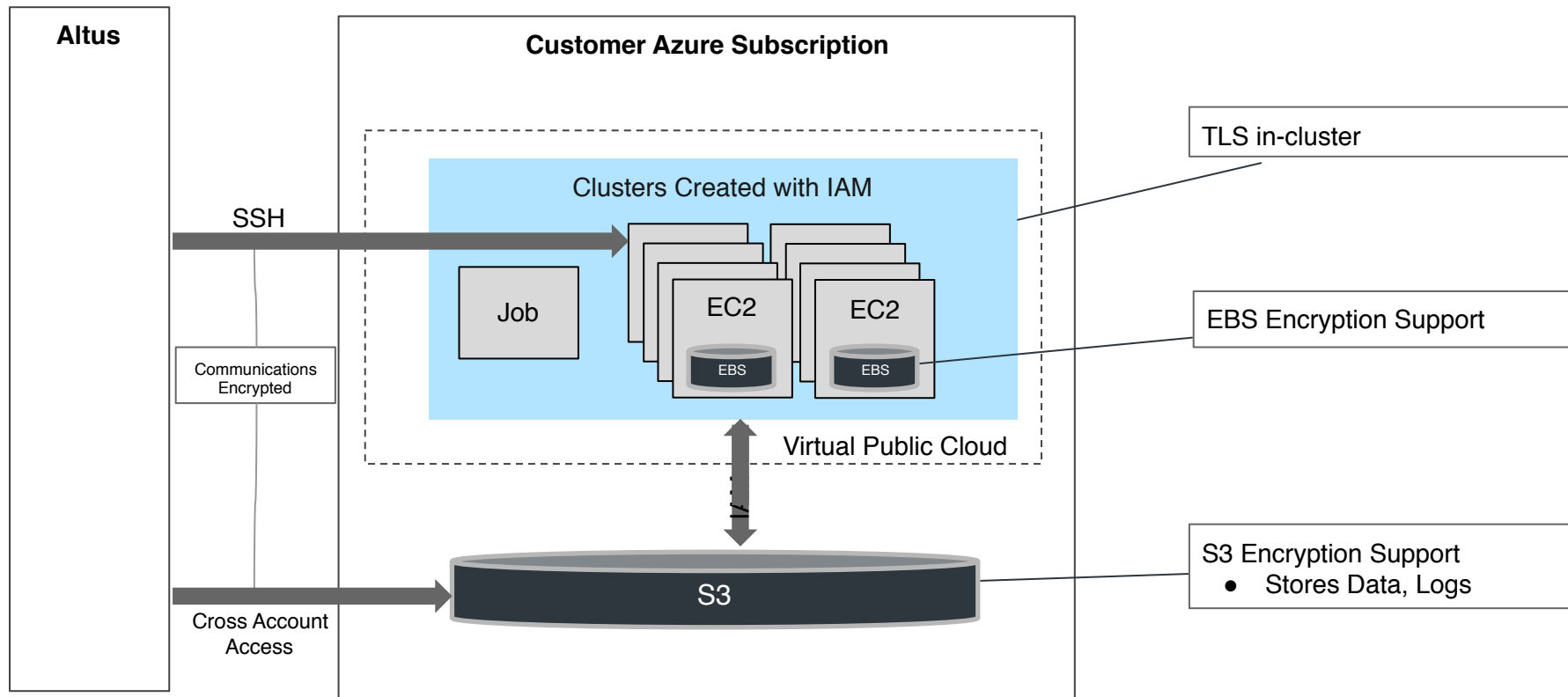
Cross Account Delegation Role

- Create resources
- Read/write to S3 buckets via IAM roles

Network

- Ability to SSH from Altus (Cloudera subscription) to VMs for Management
 - Only available from Altus IPs

AWS Encryption



THANK YOU

CLOUDERA ALTUS - SELF-SERVICE FREE TRIAL

Our customers want choices

Everything you don't have to do:

- Install any software
- Install any hardware
- Worry about cluster configuration
- Upgrade/re-configure clusters
- OS upgrades/patching
- Resource Management



30 day free trial*