



cloudera

MODERN DATA WAREHOUSE FUNDAMENTALS

Part III: Healthy in Production: Optimizing, Managing, and Troubleshooting

December, 2018

SPEAKERS



Raman Rajasekhar

Product Manager
rr@cloudera.com



David Dichmann

Director, Product Marketing
ddichmann@cloudera.com

TYPICAL MODERN DATA WAREHOUSE FOOTPRINTS

LARGE BANK

FRAUD PREVENTION

247

Users

2M

Queries / Month

500

Terabytes

167

Databases

GLOBAL PHARMA

NEW PRODUCT DEVELOPMENT

8700

Users

8 PB

Diverse Data

200+

Use Cases

200x

Fewer Silos

MAJOR TELCO

BUSINESS OPTIMIZATION

7

New Sources

2M

Users

<1S

Query
Responses

>100 M

Min. Data Sets

Managing all this requires comprehensive
Workload Management

Workload Management is about proactively **assisting** and **de-risking** every phase of the big data application lifecycle, supporting **all stakeholders** involved

Cloudera aims to **empower** our customers through a
self-service intelligent workload-centric
software to deliver

Workload XM

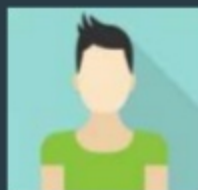
PERSONAS: WHO USES THE MODERN DATA WAREHOUSE?



Hadoop / System
/ Database Admin



Data Architects



Data Engineering
/ BI Developer



System
Integrator



Central Support
Team



Data
Consumers

KEEPING YOUR DATA WAREHOUSE HEALTHY



Visibility into clusters



Deep health checks



Baseline, performance tune & troubleshoot Impala/Hive/Spark



Assist workload migration



Identify rogue applications



Self-service Analysis

Fix **data health** issues like small files

Better **metadata handoff** between ETL & BI

Enhanced visibility within and across clusters

How do I share **petabytes** of **verified** data across **thousands** of users with **varied skill-sets** while maintaining **SLAs** and **cost**?

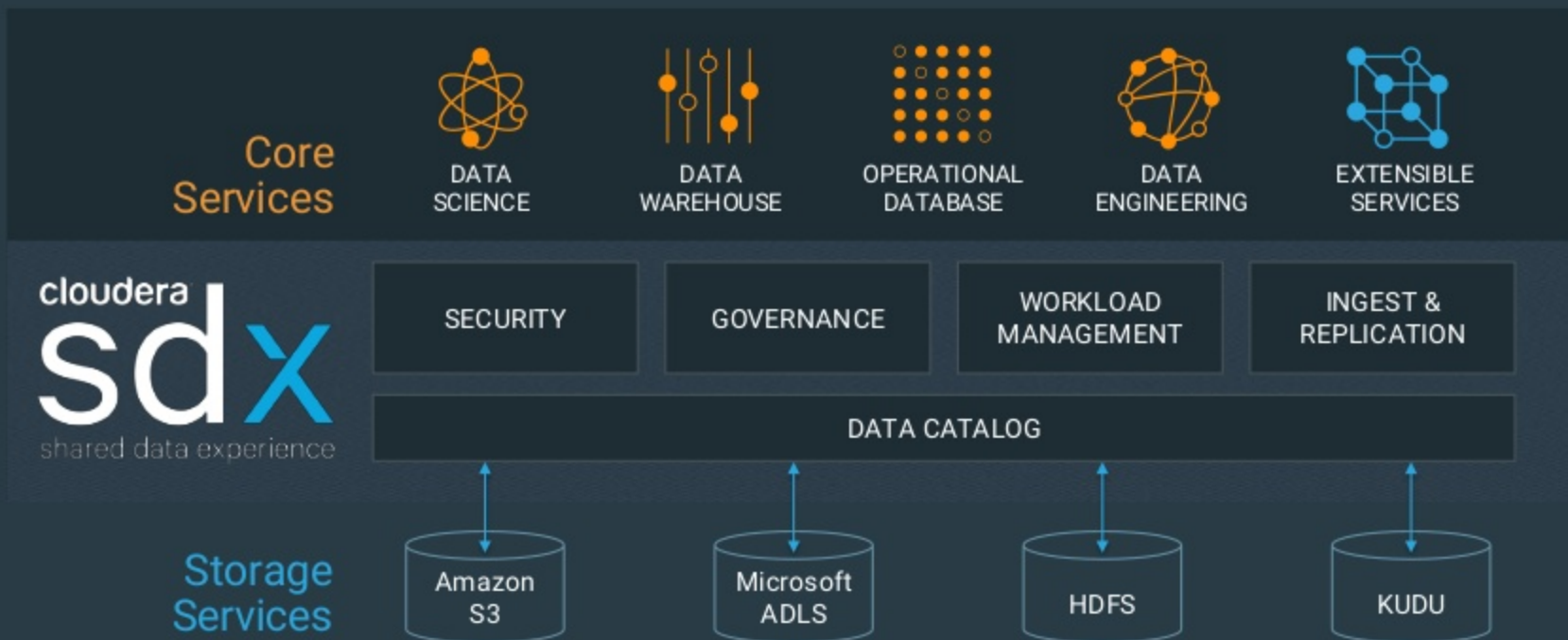
Self-service analytics through prescriptive recommendations

Proactive **De-risking** through early detection & elimination of bottlenecks

Reduced time-to-production through quick troubleshooting & tuning

CLOUDERA ENTERPRISE

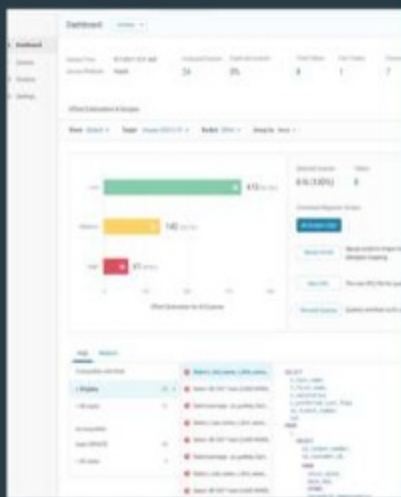
The modern **platform** for machine learning and analytics optimized for the cloud



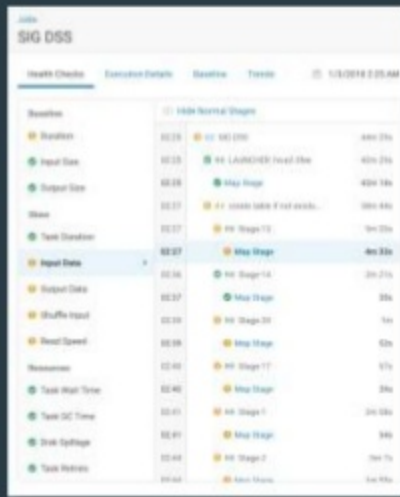
DEMO Video

CLOUDERA WORKLOAD XM

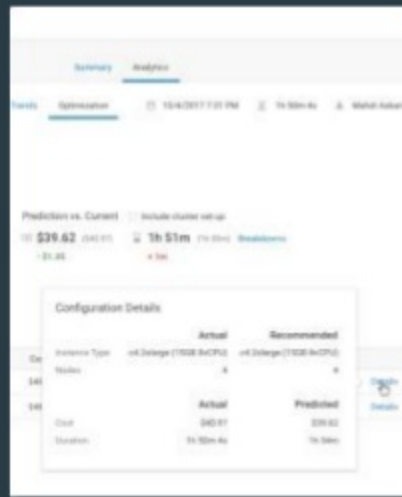
Proactively optimize Workloads, Application Performance, and Infrastructure Capacity for
Data Warehousing, Data Engineering, and Machine Learning Environments



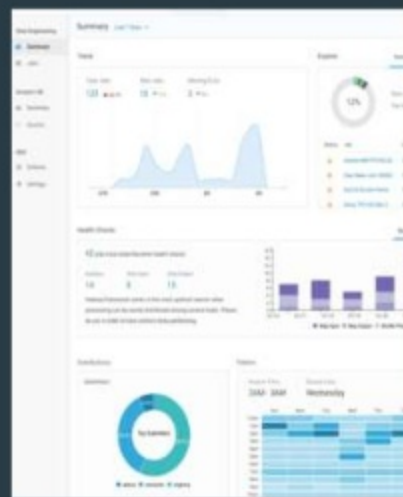
Migrate



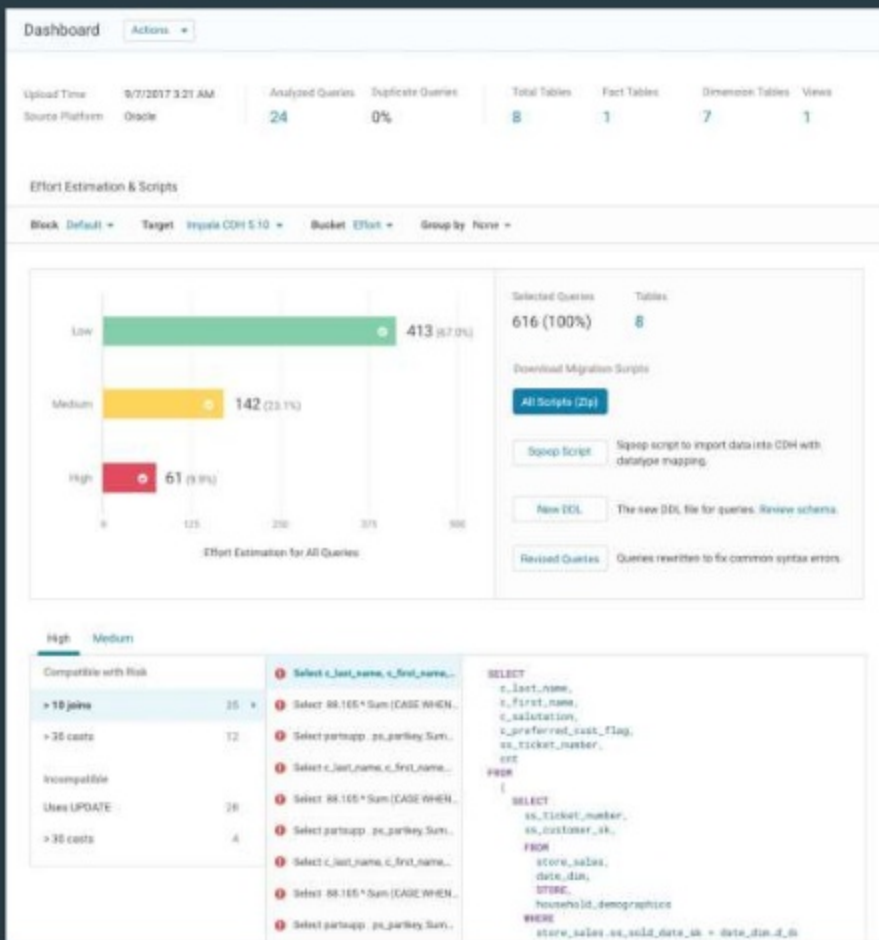
Analyze



Optimize



Manage

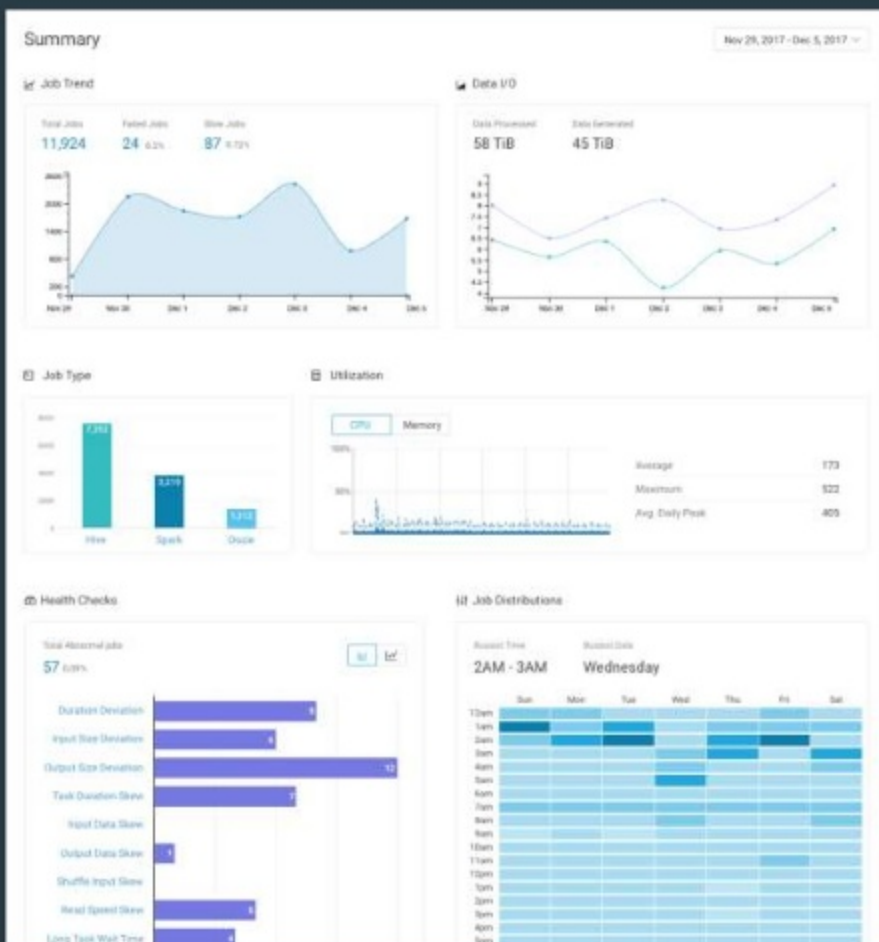


MIGRATE

Comprehensive Risk Assessment

Easy Onboarding of New Applications

Reduced Time to Production

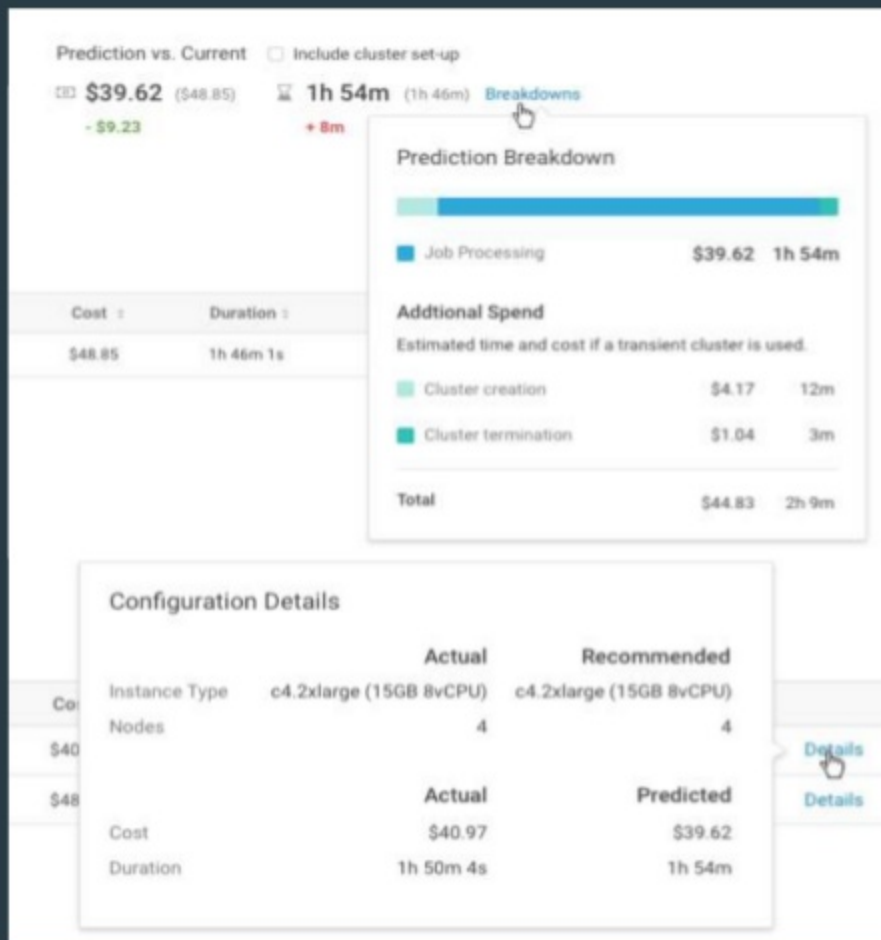


ANALYZE

Monitor Application Health

Self-service Analytics

Identify Anti-patterns & Bad Hardware

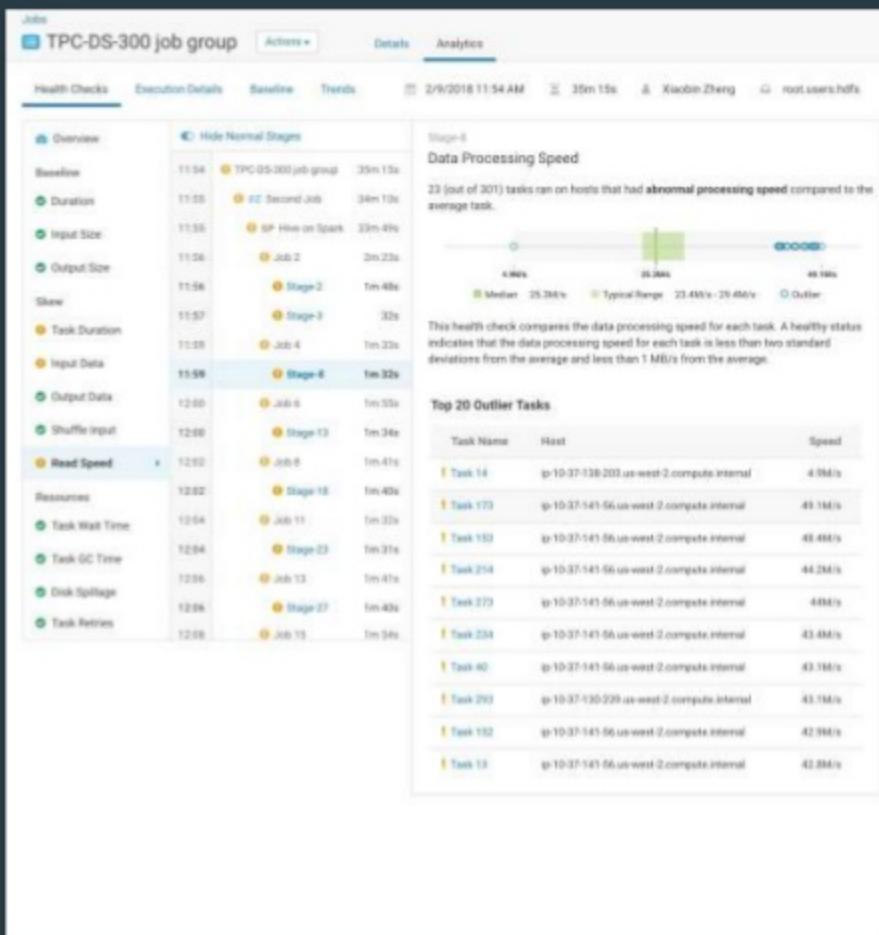


OPTIMIZE

Achieve Performance Predictability

Prescriptive Tuning & Workload Balancing

Resource-to-Cost Optimizer



MANAGE

Define, Monitor & Manage SLAs

Proactively Help Eliminate Bottlenecks

Assist in Capacity Planning

WHY WORKLOAD XM?

- No Installation, Restarts, Downtimes
- No Agents or Sensors deployed on cluster
- Instant availability of new features and enhancements
- Extensive Knowledge base on platform usage from customers across the globe
- A non-intrusive, secure and stateless metrics collection
- SDX Advantage, tighter integration into the Cloudera ecosystem & enhanced UX
- No one understands our platform better!

WORKLOAD MANAGEMENT - WHAT'S COMING?



Intelligent Workloads



Prescriptive Recommendations



Actionable Capabilities

THE WORKLOAD XM ADVANTAGE

~70% ↓

Time to Detect Issues &
Bottlenecks - Saving
Time & Costs

50x

Faster RCA &
Troubleshooting -
Reducing expensive
Overheads & Costs

43% ↑

Improved SLA
Adherence - Equates to
XXX Revenue

30%+ ↑

Cluster Utilization -
Increased ROI on
infrastructure

2x ↑

Intrinsic Value of
Node

<0.2%

Minimal Additional
Resource Consumption

THE MODERN DATA WAREHOUSE

Deeper Business Insights at Extreme Speed and Scale While Managing Cost



THANK YOU

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

cloudera

