

**DATA+AI
SUMMIT**

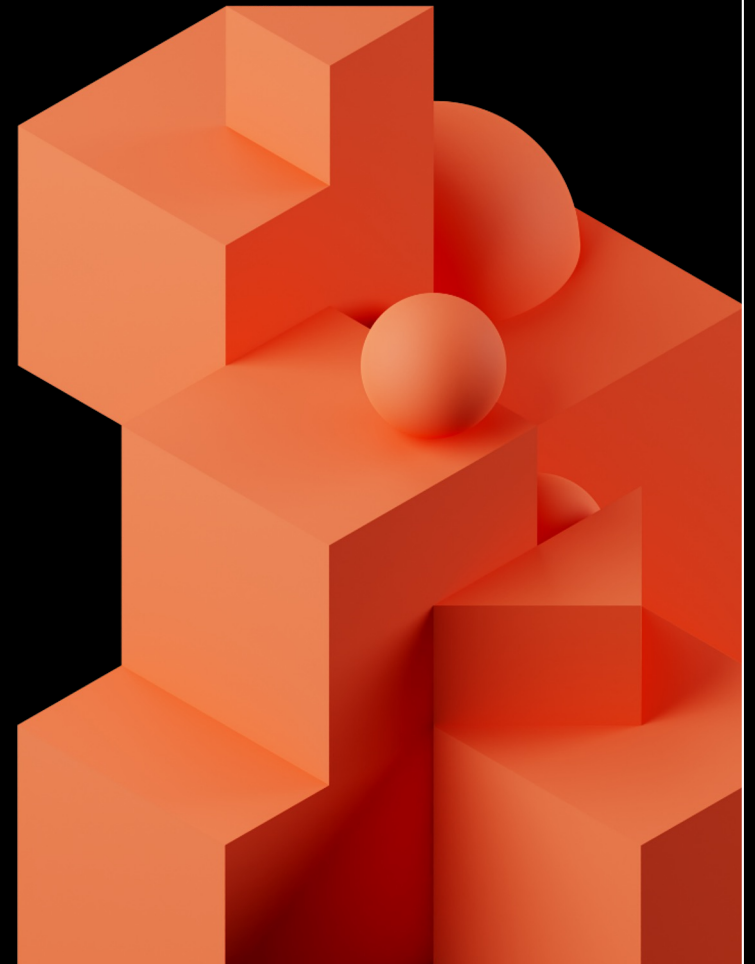
BY  databricks

USING LAKEHOUSE TO FIGHT CANCER

Ontada's Journey to Establish a Real-World
Data (RWD) Platform on Databricks
Lakehouse

Databricks
2023

ontada®



About The Speaker



Donghwa Kim is the Senior Director of Architecture at **Ontada**, a McKesson company.

He is responsible for delivering the next generation Data and Analytics platform using Databricks Lakehouse.

Prior to joining Ontada, Donghwa worked as an enterprise architect, migrating a large-scale on-prem data warehouse onto Databricks at Veterans Affairs (VA) and at Centers for Medicare and Medicaid Services (CMS).

Donghwa has over 20 years of IT experience within the healthcare and finance industries.



ontada®

Agenda

01

Ontada Introduction

02

RWD, RWE and Oncology

03

Ontada Data Platform (DMI)

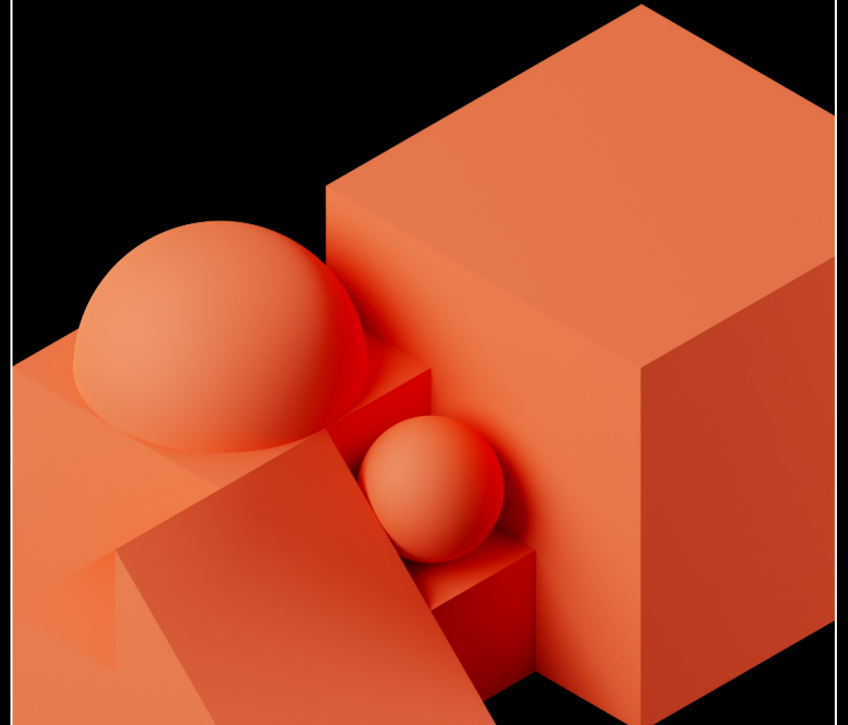
04

Oracle to Databricks Migration

05

Benefits of Lakehouse

Ontada Introduction



Ontada – Introduction

Transforming the fight against cancer

Ontada, a McKesson business, is an oncology real-world data and evidence, clinical education, and provider technology business dedicated to transforming the fight against cancer.



Ontada Products

Provider Solutions



iKnowMed

Premier Oncology EHR

- Comprehensive clinical support for patients
- Updated library of regimens
- Drug availability
- Diagnosis-level biomarker testing
- And more for the best clinical outcomes



Clear Value Plus

Decision Support Application

- Intelligent matching of each patient to best evidence-based treatment options
- Integrated clinical and financial information at the point of need



Ontada Health

Patient Portal

- Enables patients to:
- View their health records and lab results
 - Communicate with their providers



Practice Insights

Performance Analytics Tool

- Proactive and actionable insights into:
- Quality initiatives
 - Value-Based Care (VBC) programs
 - Performance metrics
 - Productivity measures
 - Peer/industry benchmarks

Ontada Products

Life Science



Real-World Data

Fit-for-purpose data products to support oncology decisions and research. Provides the clearest view into the full cancer patient journey, including:

- Diagnosis, staging, and treatment decision
- Therapy initiation and maintenance/disease progression
- Treatment decision, adverse events, and survivorship



Provider Engagement

Customized educational programs and unique channels to help providers achieve their commercial goals:

- Provider education
- Patient education
- Market research



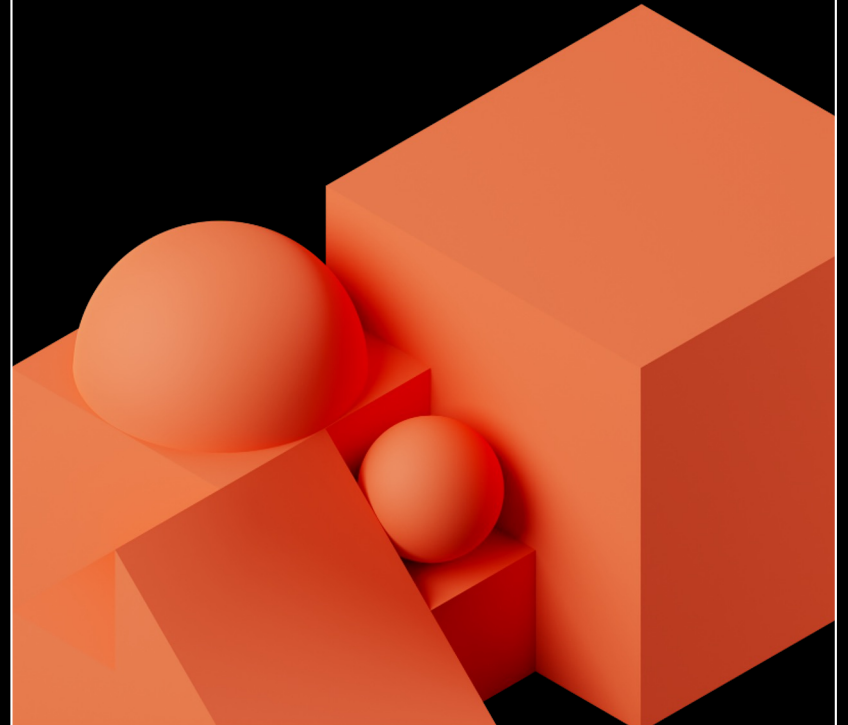
Real-World Research

Ontada's offering to help providers achieve outcomes-based value with:

- Real-World Evidence (RWE)
- Health Economics Outcomes Research (HEOR)
- Ontada's real-world data (RWD) and extensive oncology expertise to help providers to advance clinical research, expand access, and improve outcomes.



RWD, RWE, and Cancer Care



U.S. Cancer Statistics

- 1,958,310 new cancer cases (2023 projection)
 - 609,820 cancer deaths (2023 projection)
 - The second-leading cause of death after heart disease
 - The leading cause of death among women, 40 to 79 years
 - The leading cause of death among men, 60 to 79 years
-

The Silver Lining

Overall cancer mortality **continues to decline**

33% decrease since 1991

But we need to do more!

RWD and RWE

Real-World Data (RWD)*

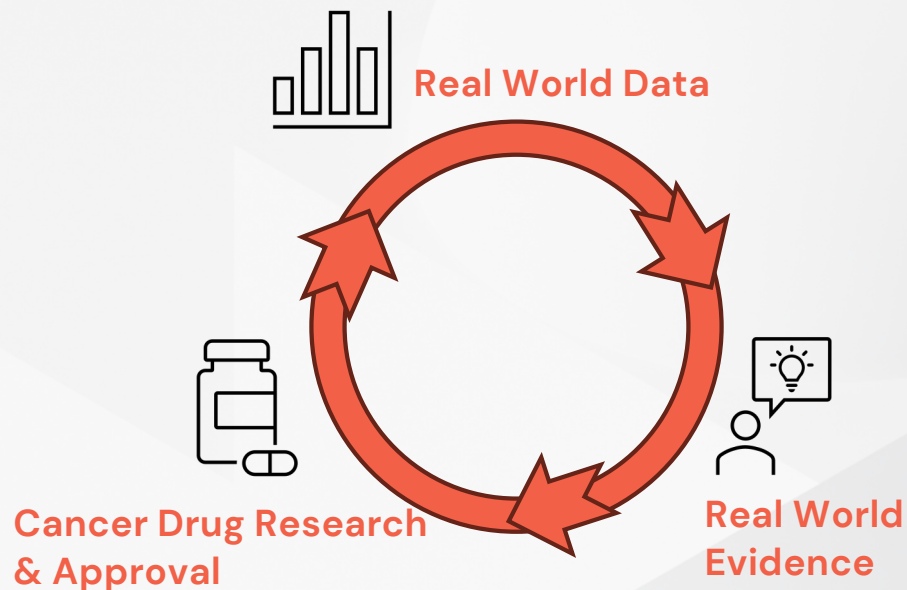
"Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources."

Real-World Evidence (RWE)*

"The clinical evidence about the usage and potential benefits or risks of a medical product derived from the analysis of RWD."



RWD, RWE and Oncology

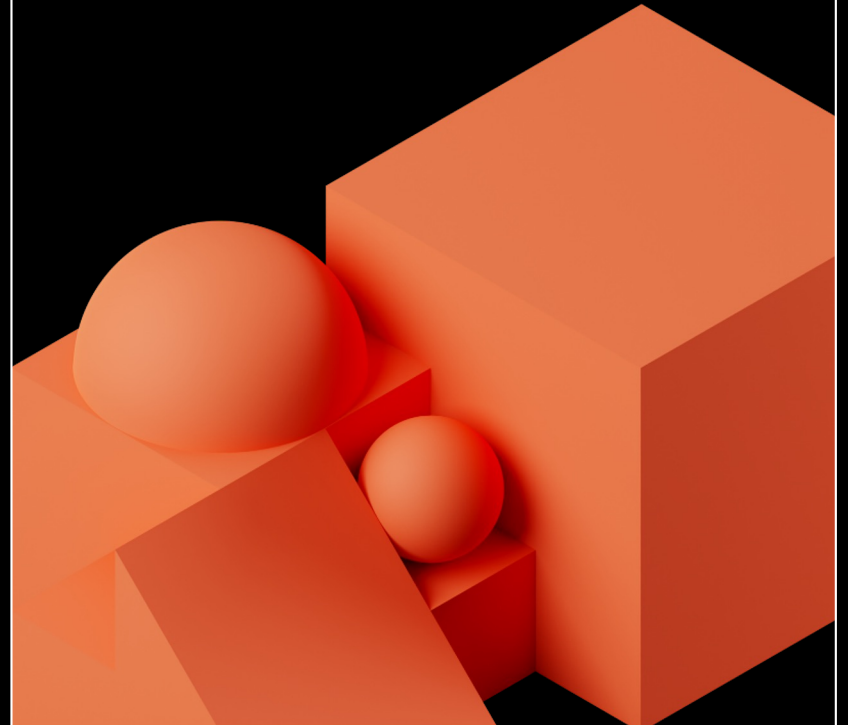


"Each patient's cancer journey, strengthened by real-world data and evidence, paves the way for safe and effective prevention, detection, and treatment, turning personal battles into collective transformative victories."

– Sagraan Moodley, Chief Innovation and Technology Officer for Ontada

Ontada Data Platform (DMI)

Migration Journey



What is the Ontada DMI?

The next generation analytics platform built on Azure Databricks



Replaces the legacy on-prem data warehouse and cloud-based genomic data pipeline infrastructure



Provisions real-world data and evidence (RWD/E) generation to accelerate life science research, inform decisions at the point of care, and provide data-driven patient treatment



Improves end-to-end data capture, transformation, curation, standardization, and observability



Acts as a single access point for all Ontada's data and computation needs:

- Reporting
- Data analysis
- Advanced analytics
- Ad-hoc explorations

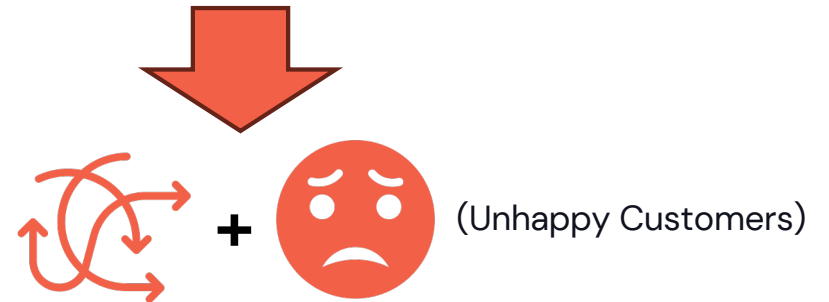
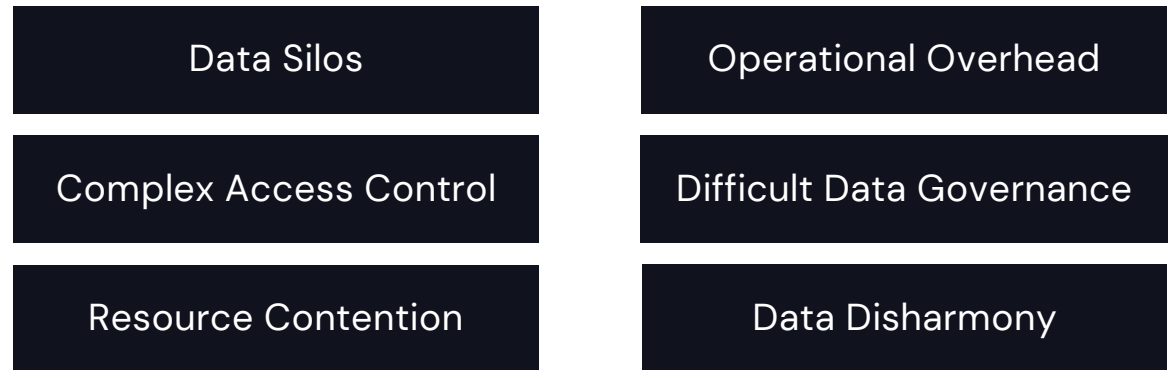
Agility

Scalability

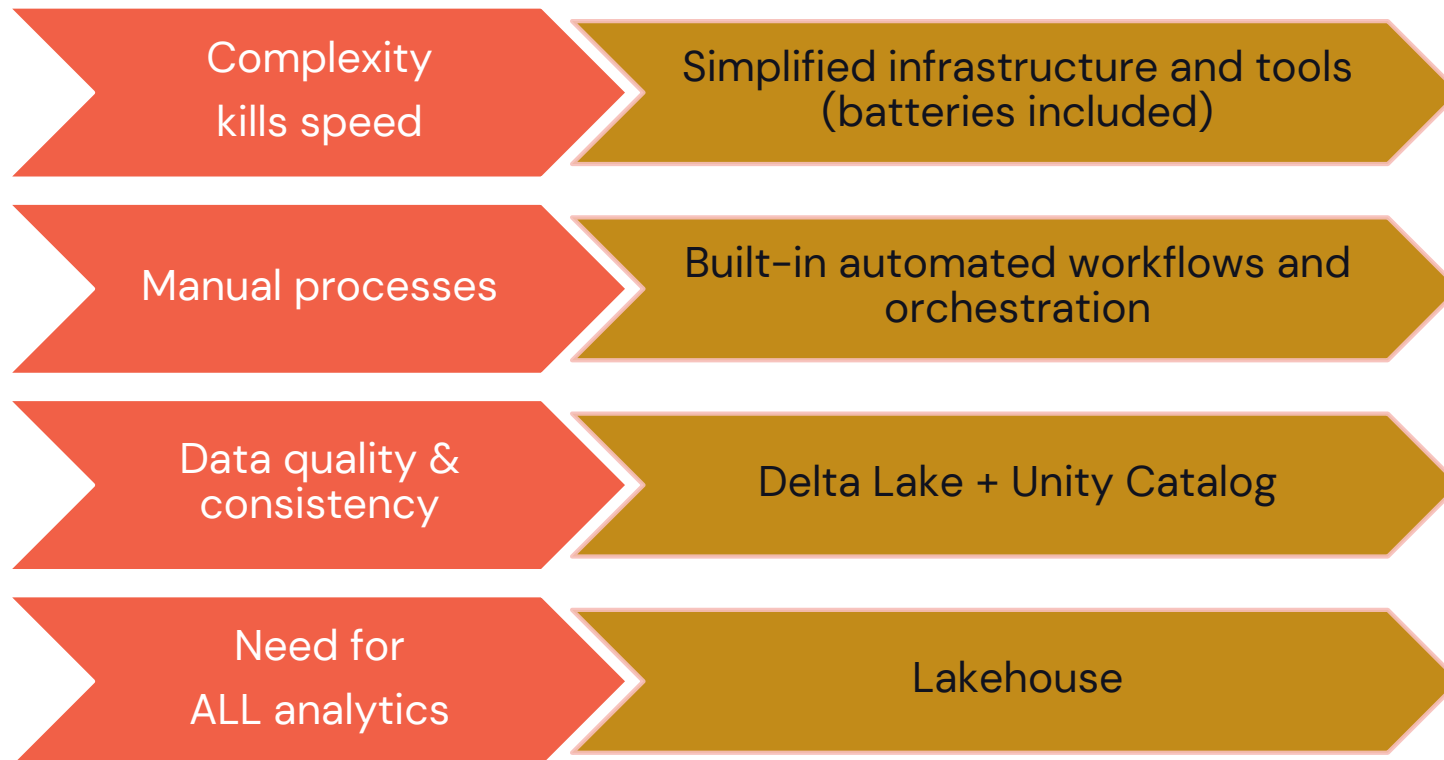
Governance

Common Challenges For Data Platforms

(Without Lakehouse)



Why Databricks?



Ontada Lakehouse

Data Source

On-Prem &
Cloud

Oracle DB

Mongo DB

SFTP

S3 Bucket

+ Other
Data
Sources

Data Ingestion

Batch &
Streaming

Batch

Streaming

CDC

Qlik

debezium

kafka

DMI Lakehouse



Bronze

- "Raw" data from internal and external data sources

Autoloader



Silver

- Common Data Model (Silver)

dbt



Gold

- Common Data Model (Gold)

dbt

Jobs Clusters
GP Clusters
SQL Warehouse

Unity Catalog

Delta Sharing



Azure Data Lake Storage Gen 2



Azure Databricks



Data Delivery

BI/SQL

tableau

sas

Advanced
Analytics



John Snow LABS

mlflow

Consumers

BI
Developers

Data
Scientists

Data
Analysts

Self-Serve
Reporting

Data
Stewards

Data
Extraction

ontada®



User Provisioning



Access Control



Master Data Management



Compliance Monitoring



FinOps

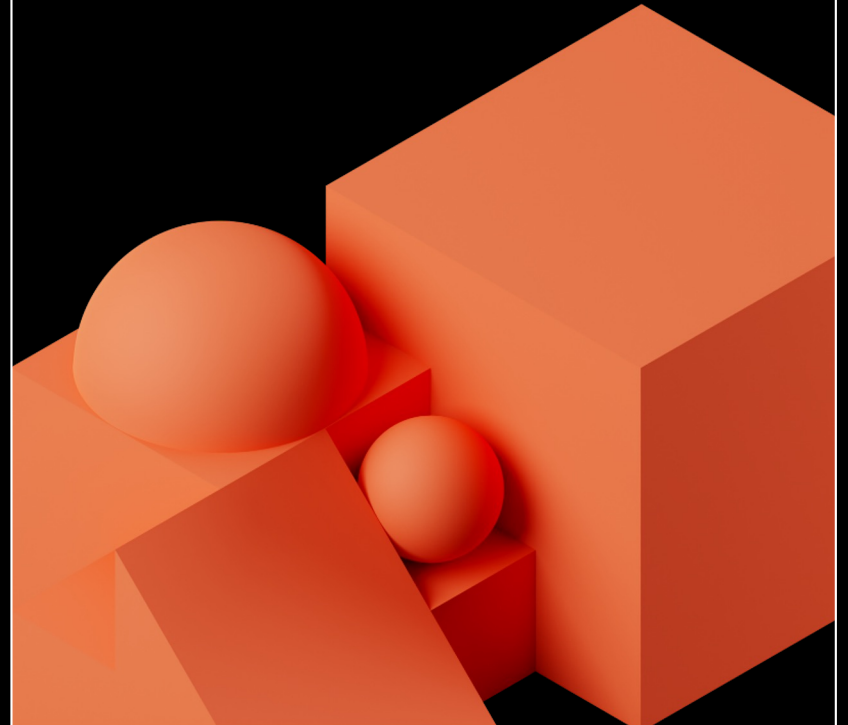


Automation

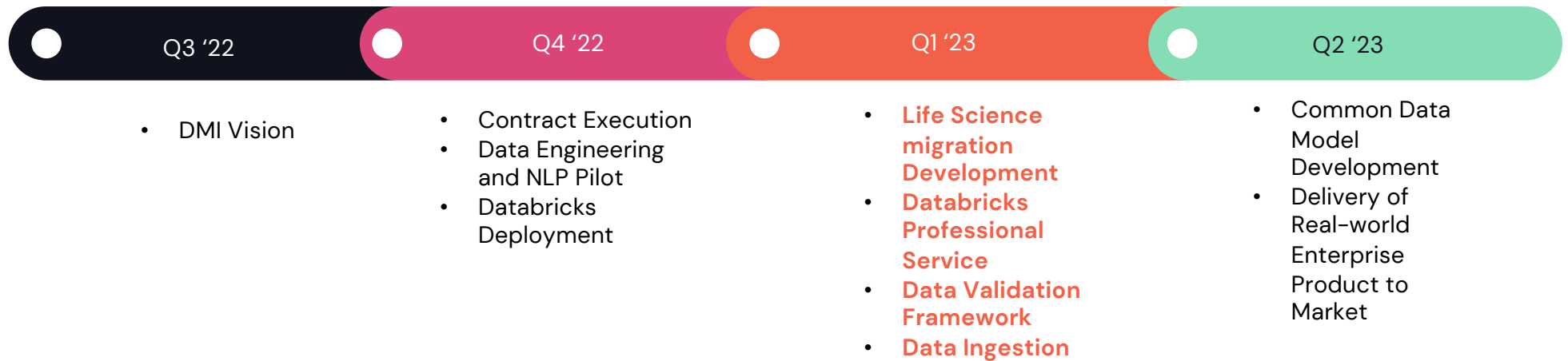
DMI Use Cases – Highlights

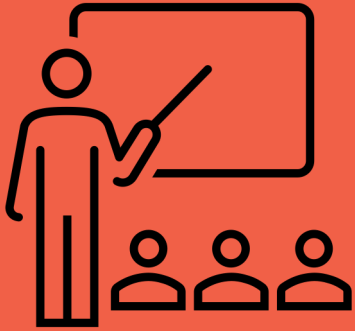
Need	Harmonize data from disparate sources	Abstract information from unstructured data	Expand Clinico-Genomics data product offerings	Generate and share reports + dashboards
	Common Data Model	NLP	Genomics Data Processing	Self-Serve Reporting
Solution	• Driver for Real World Data Enterprise • Oncology-focused, scalable, and interoperable data model • FHIR mCode and Genomics Reporting IG	• Minimize the time and cost associated with chart abstraction • Biomarker extraction from clinical notes • Leverage both commercial and open-source libraries	• Ingest, process and analyze raw genomics files	• Easy functionality to share aggregated data with external clients and self-service access to dashboards and reports. • Allows to published data for supporting self-serve reporting

On-Prem DW to Databricks Migration



Migration Timeline





ontada[®]

Migration Lessons Learned

01

Get executive sponsorship

02

Parallel execution

03

Start conversations with DBAs early

04

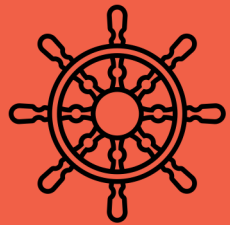
Beware of code conversion limitations

05

Set validation strategy

06

Monitor, early feedback and repeat



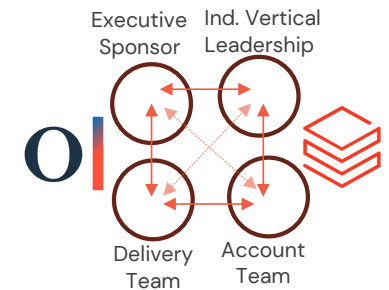
ontada®

Lessons Learned – 01

01

Get executive sponsorship

- Financial Commitment
 - Initial and continued
 - Provide clarity on the total budget.
 - Long term journey and investment
- Frequent Vision and Execution Alignment
 - Ontada Executive Sponsor & DMI Delivery Team
 - Ontada Executive Sponsor & Databricks Industry Vertical Leadership
 - DMI Delivery Team & Databricks Account Team
- Databricks Business Value Consulting





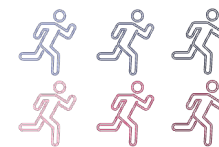
ontada®

Lessons Learned - 02

02

Parallel execution

- Have a strategy for parallel execution
 - Current workload
 - Migration
- Be ready to juggle
 - Competing priorities
 - You cannot predict unknown unknowns
- RAID
 - Risks, Assumptions, Issues, and Dependencies
- (Seriously) Consider Databricks Professional Service or their SI partner ecosystem



Don't let those risks encroach on you!



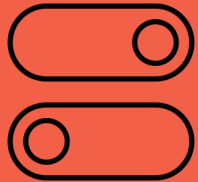
Lessons Learned – 03

03

Start conversations with DBAs early

- Workload schedule
 - Understand the schedule of existing jobs (competing for the resource)
 - Find slots for new ingestion job(s)
- Resource contention – CPU, Memory, Disk
 - Ensure current critical jobs are not impacted
 - CDC requires supplemental logging turned on
 - On-prem disk space and cost
- Testing data ingestion in QA environment
 - Running production-like workloads (via data refresh)
 - Need collaboration with DBAs and QAs

Lessons Learned - 04



04

Beware of code conversion limitations

- DDL Generation
 - Manual generation of table schema
- Time Zone Conversion
 - Align on proper time zone strategy
 - Have early discussion with the product team
 - Session vs. Global Configuration
- Oracle PL/SQL Procedures and Functions
 - Spark SQL limitations



Lessons Learned – 05

05

Set validation strategy

- Testing Data
 - Static Data – T (Transformation) Pipeline Validation
 - Catalog-to-catalog match in Databricks
 - Iterations and cycles – multiple catalogs
 - Live Data – (EL) Ingestion Validation
 - Oracle-to-Databricks match
- Validation Automation Framework
 - Level 1: Count and checksum validation
 - Level 2: Logic validation – need to work with data users
- Implement Quality check at every layer



Lessons Learned – 06

06

Monitor, send feedback and repeat

- Financial
 - Monitor and prevent run-away cost
 - Provide early feedback on consumption
 - Frequently validate the usage
 - Leverage FinOps Principles
- Compliance
 - Who, What, When, and Why
 - Establish a matrix of user groups and data assets
 - Develop an access approval process
 - Create regular compliance reports

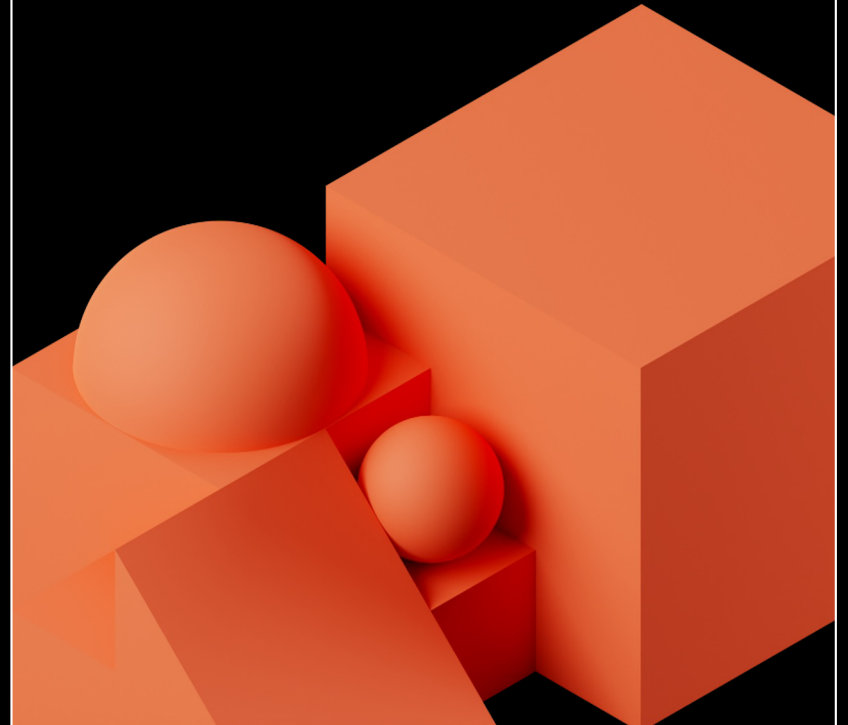
User Access Control and Compliance

Getting Most Out of Unity Catalog

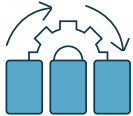
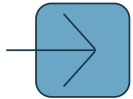


- Automated User Provisioning using SCIM
 - Windows AD -> Azure AD -> Databricks User Groups
 - Integrate with ServiceNow workflows
- Fine Granular Access Control
 - Appropriate privileges on data objects based on group association
- Ongoing compliance report generation using Python SDK
 - `databricks_cli.groups.api`
 - `databricks_cli.unity_catalog.api`

Benefits of Lakehouse



The Outcomes

Key Impacts	Description
 Data Availability	• Faster data ingestion from disparate data sources • Quality data via automated data validation framework
 Delivery Speed	• Improved speed to market • Expedited product development and delivery • Ability to run parallel workstreams via dedicated compute availability
 Enhanced User Experience	• Exceptional performance improvement (10x in certain use cases) • User collaboration via notebook, GitHub integration • One-stop shop – Data Engineering, Data Science, ETL, SQL analytics
 Future Ready	• Ready to develop and deliver new LS products • All data types (structured, semi-structured, un-structured), all analytics (descriptive, predictive, prescriptive)

Find Us on Databricks Marketplace

Databricks Marketplace Preview

powered by  DELTA SHARING

Q ontada

Provider

Category

1 item

Sort: Newest

O| Ontada

Ontada RWD - Clinical Data Views

Databricks Marketplace >

Ontada RWD - Clinical Data Views Preview

O| Ontada

Ontada RWD - Clinical Data Views

- Homepage: <https://www.ontada.com/>
- Life Science Solution: <https://www.ontada.com/Life-Sciences-Solutions/>
- Real-World Data: <https://www.ontada.com/Life-Sciences-Solutions/Real-World-Data/>
- Real-World Research: <https://www.ontada.com/Life-Sciences-Solutions/Real-World-Research/>
- Provider Engagement: <https://www.ontada.com/Life-Sciences-Solutions/Provider-Engagement/>

Details

Access

By request

Visibility

Public

About this provider

Request access

ontada®


30

Databricks Data for Good Award Finalist

Impacting the World with Data & AI: Announcing the Finalists for the 2023 Databricks Data for Good Award



by Mike Sanky

June 16, 2023 in Customers

Share this post



McKesson/Ontada

Ontada is an oncology data, research and technology business dedicated to improving the lives of patients and transforming the fight against cancer. The Databricks Lakehouse Platform enhances Ontada's deep learning natural language process (NLP) models and algorithms, improving the efficiency of the computation as well as improving the efficacy and accuracy of the algorithm. With the Databricks Lakehouse Platform, Ontada is able to quickly ingest and process millions of unstructured medical documents. One such example is the ability to accelerate the extraction of biomarker data from unstructured notes. This is important because access to and accuracy of biomarker information is critical for physicians to provide targeted therapies for patients, harness the promise of precision medicine, and ultimately, improve cancer care outcomes for patients. Combining our in-house data science and research expertise with Databricks' data platform architecture allows us to scale and accelerate the speed of clinical insights while ensuring high-quality results, thereby acting as a cornerstone for Ontada's differentiation and data value offerings.

<https://www.databricks.com/blog/impacting-world-data-ai-announcing-finalists-2023-databricks-data-team-good-award>

Q & A

