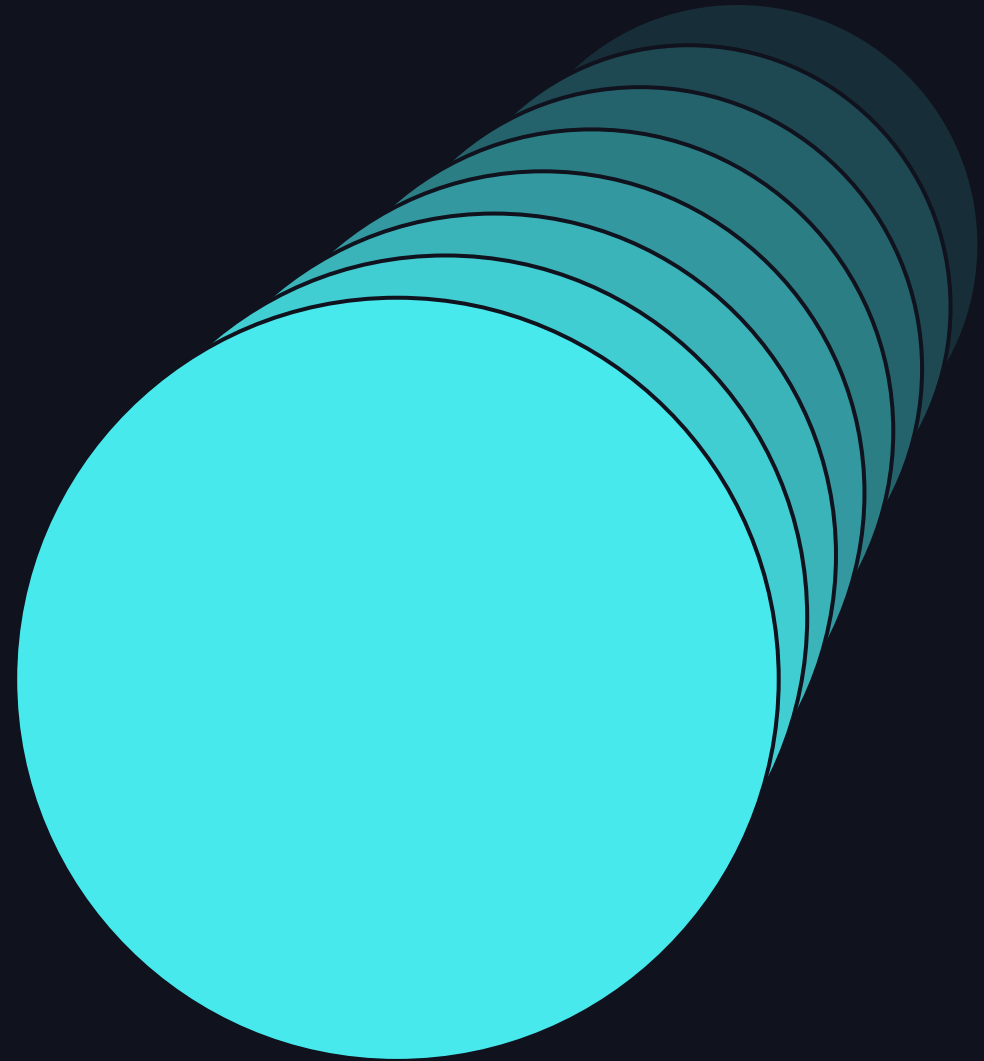


A LEAN DATA FACTORY

Optimizing the Power of a Data
Intelligence Platform

Kythera Labs
June 2024

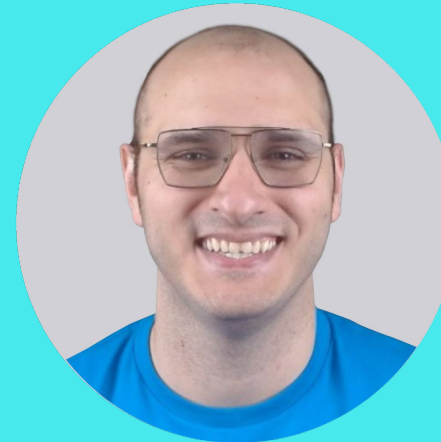


WE LOVE DATA



MATT RYAN

Kythera Labs
Co-Founder, CTO



JOSUE BOGRAN

Kythera Labs
Solutions Architect Manager

AGENDA

- Introductions
- Addressing the 7 Wastes in a Data Intelligence Platform
 - What is a Data Intelligence Platform
 - The 7 wastes defined
- Applying these principles
- The Future
- Q & A

INTRODUCTION - MANUFACTURING ANALOGY

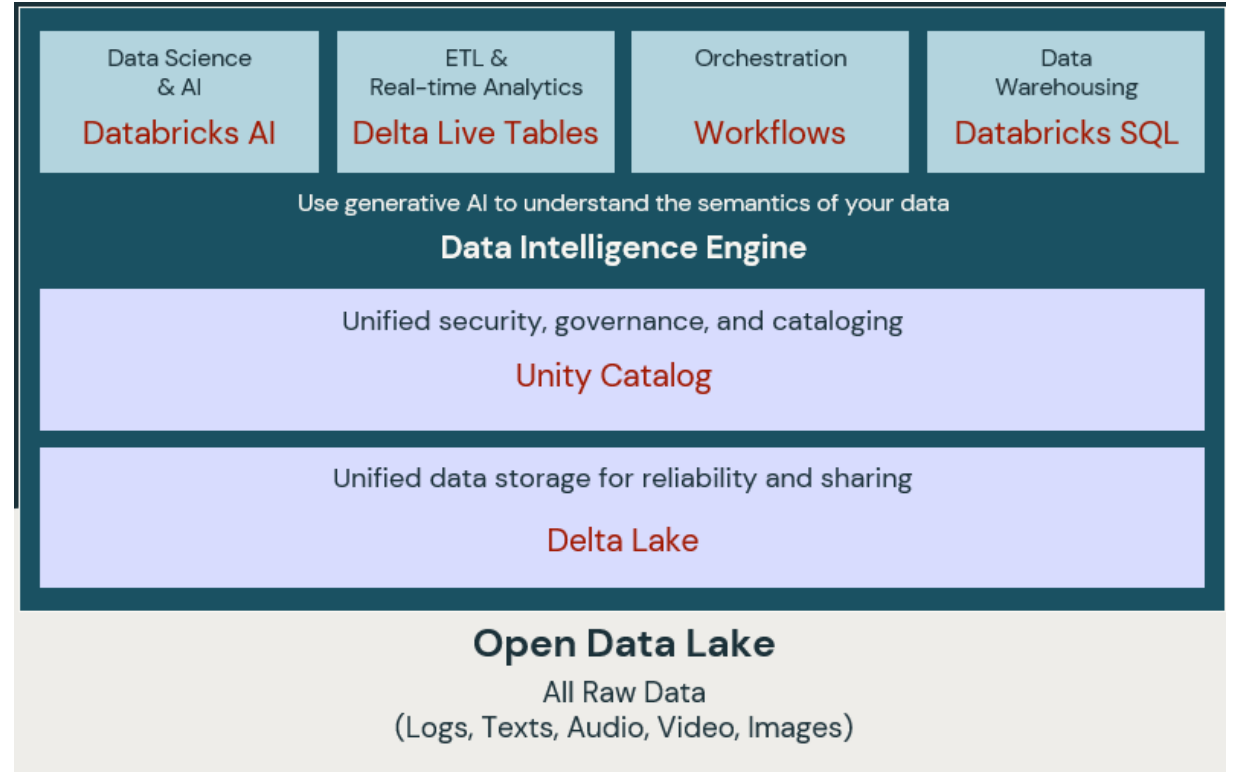


How It's
Made

WHAT IS A DATA INTELLIGENCE PLATFORM

Data Intelligence as Data Factory

- Our use of Unity Catalog
 - Transport, Inventory, Motion, Waiting, Defects
- Built-On
 - Waiting, Production, Overprocessing
- Our Common Data Model / Analytics Assets
 - Waiting, Overproduction, Overprocessing, Defects
- Serverless
 - Waiting, Over Processing



DATA INTELLIGENCE PLATFORM

Is in a Cloud data center



Not in a Cloud

The big decision - Kythera's original data center



LEAN MANUFACTURING

A Fundamental Concept In Reducing Waste in Manufacturing

- Lean manufacturing is a systematic approach to identifying and eliminating waste through continuous improvement.
- The primary objective of Lean Manufacturing is to enhance efficiency, reduce costs, and improve overall production quality.
- Waste in manufacturing refers to any activity or process that consumes resources without adding value to the product or service.
- Excessive waste leads to increased production costs, longer lead times, and lower customer satisfaction.
- Recognizing and addressing waste is crucial for streamlining processes, maximizing productivity, and maintaining a competitive edge

THE 7 WASTES OF LEAN MANUFACTURING

Date Intelligence as a Data Factory

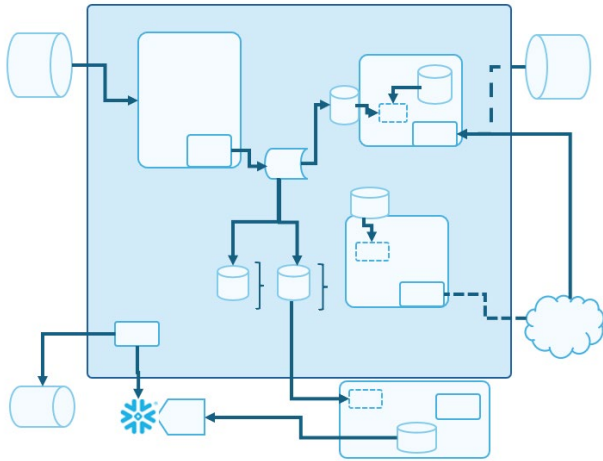


LEAN PROCESSING IN DATA FLOWS

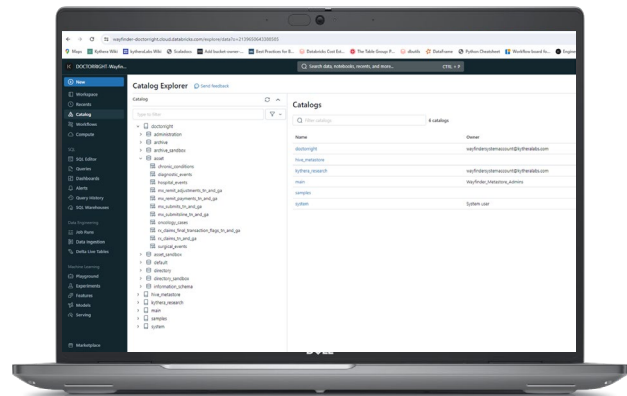
Addressing the 7 wastes



Batch Pipeline



Built On

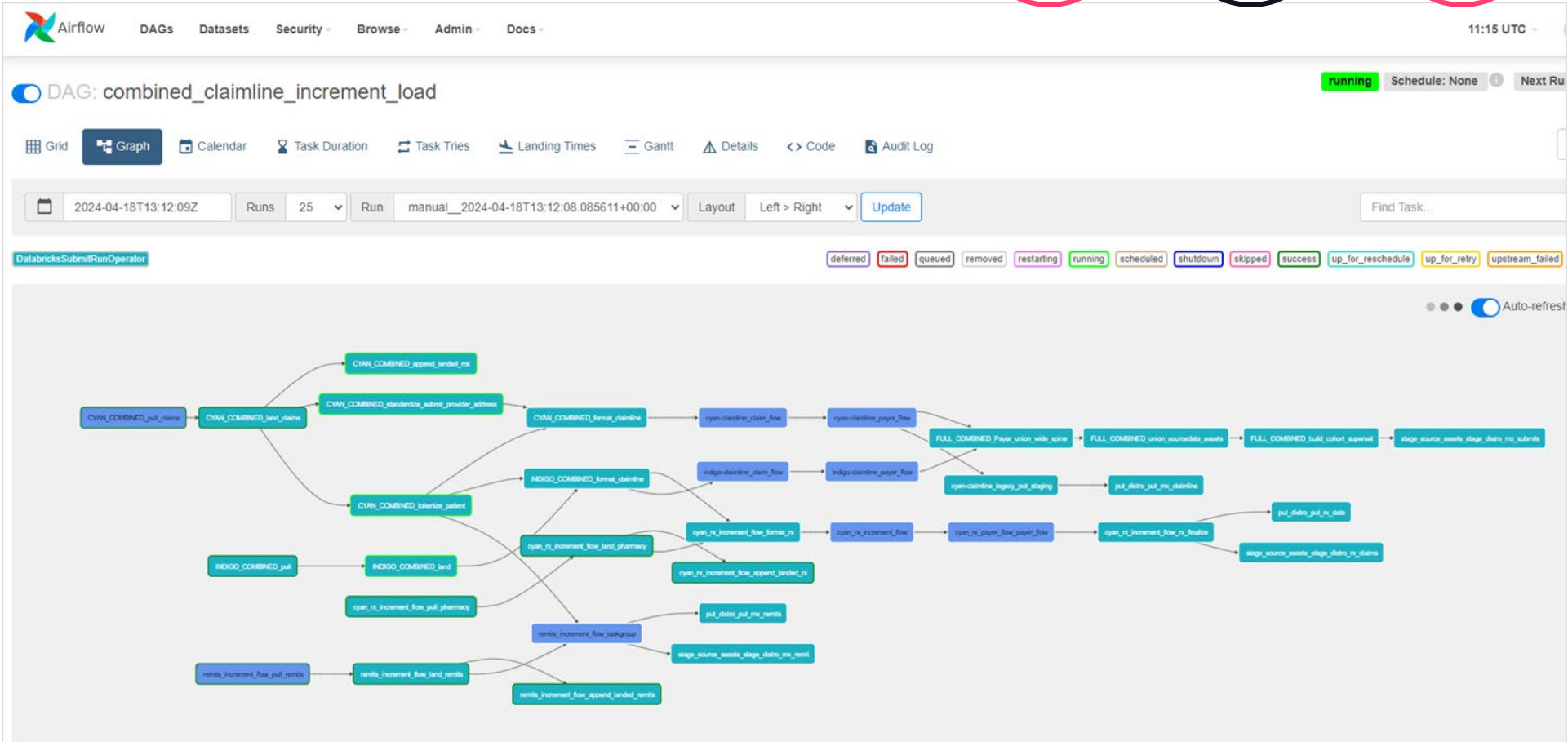
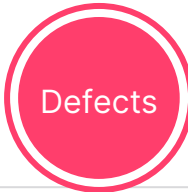


Unity Catalog / Distribution



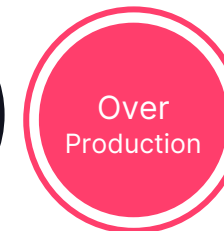
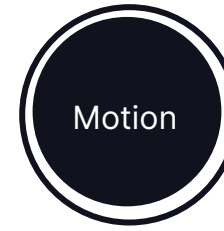
BATCH PIPELINE

Airflow

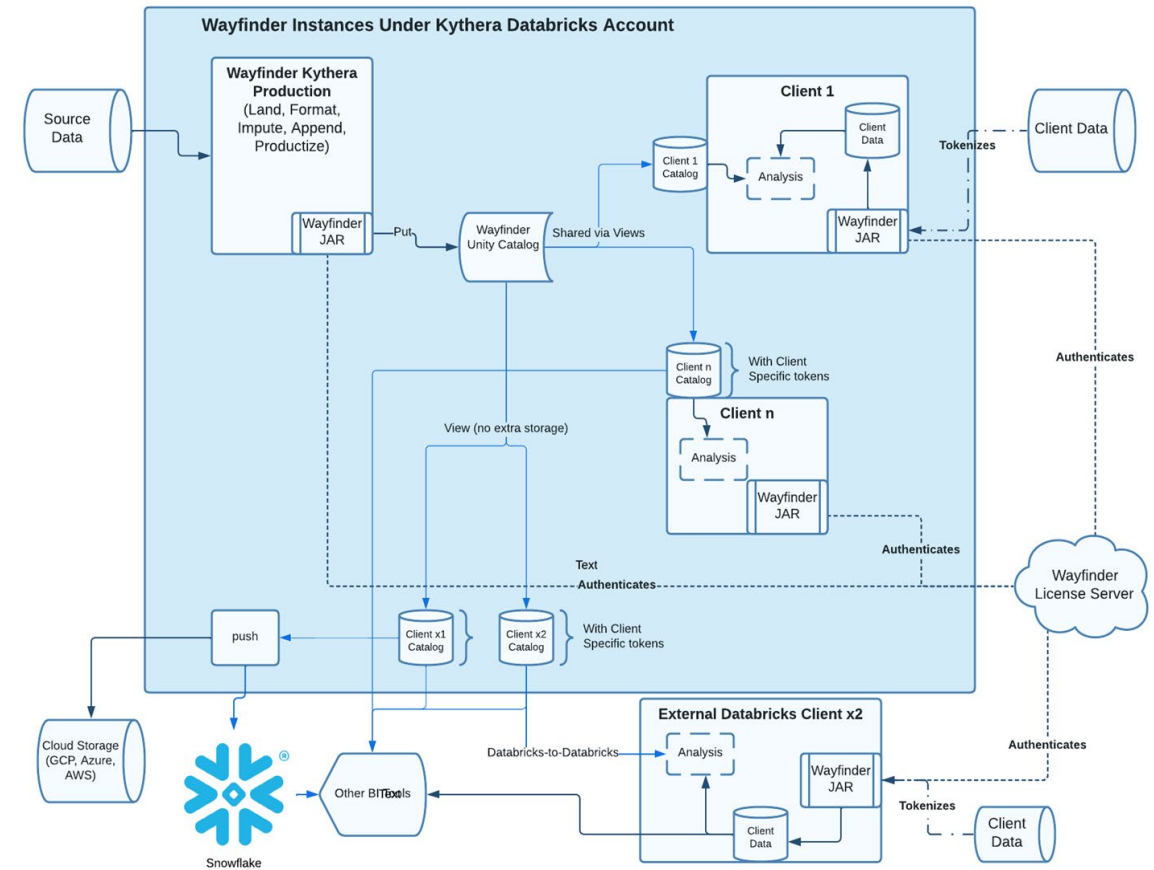


BUILT ON

Kythera Wayfinder

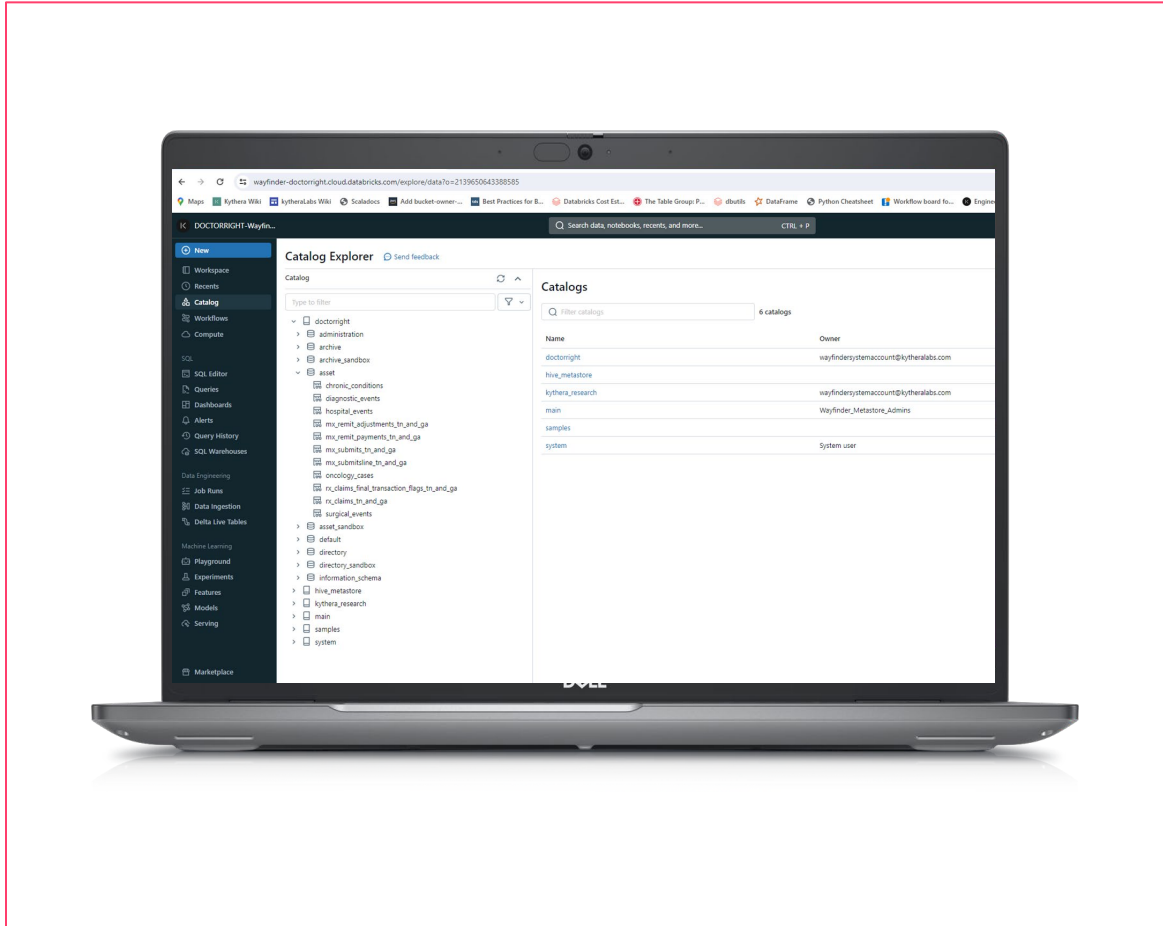
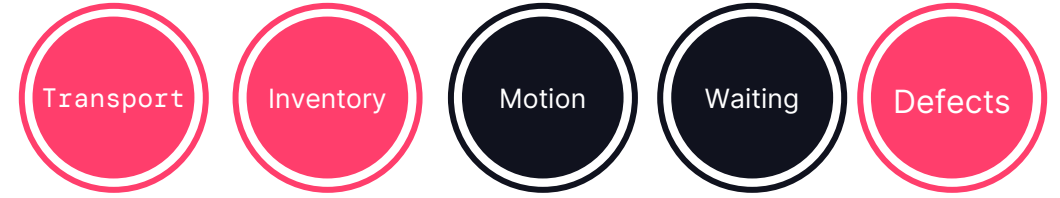


- Kythera spins up Workspaces on behalf of clients under Kythera AWS and Databricks
- Leverage Unity Catalog for sharing of Kythera's large data sets
- Ability for clients to access/process data with a wide variety of industry tools (including their own Databricks)
- Unified cost reporting



UNITY CATALOG

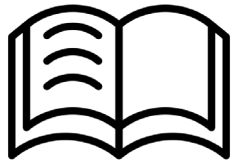
Multiple workspaces, central location



- Each client has their own workspace and catalog
- Only have access to views to centrally stored tables - not to underlying tables
- Shared volume for access to Wayfinder features API
- Leverage Hive Partitioning / Liquid Clustering for efficient joining of client specific patient dimension

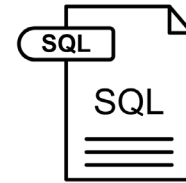
CODE EDITORS

Minimizing defects and motion across diverse coding environments...



Notebooks

- Language Flexibility
- Organized Complex Code
- Familiar Modern Coding Experience



SQL Editor

- Traditional SQL Experience
- Enhanced Capabilities
- Data Exploration with Ease
- Leverages the SQL Serverless Warehouses

COMPUTE

Reducing waiting, overproduction, and excess processing in compute.



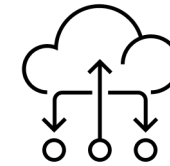
SQL Serverless Warehouses

- SQL Without Management Overhead
- Fast Performance, 4 Second Startup



Job Compute

- Low Cost For Recurring Workloads

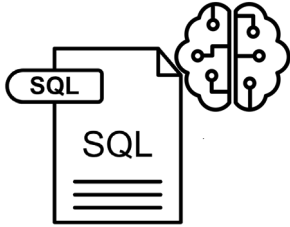


All-Purpose Compute

- Versatility & Interoperability
- Scaling Out & Scaling In As Needed

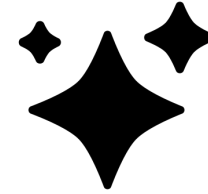
AI

Cutting transportation, motion, and over processing



SQL AI

- Easy to Use
- High-Impact

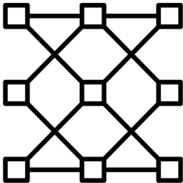


Databricks AI Assistant

- Chat-Like Interface
- Easy Debugging
 - Enriched By Unity Catalog Metadata

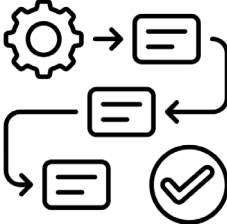
GOVERNANCE & ORCHESTRATION

Optimizing inventory while decreasing over-processing, motion, and defects



Unity Catalog

- Embedded Dictionaries With Comments
- Enriched Data Experience in Databricks



Workflows

- Schedule & File-Arrival Triggers
- Cost Safeguards
- Task Dependencies
- Pipeline Traceability



THE FUTURE (COMING SOON)

- Kythera features in workspaces outside of Built On (Python in Serverless)
- More and more data!
- Sigma dashboards
- Databricks default storage
- Data rooms and AI libraries

RESOURCES



Improving Data Quality Through
the Medallion Architecture



Tame the Costs of Cloud
Computing With the Right Tools



Serverless SQL Compute: Time
Is Money In a Very Real Way

DATA+AI SUMMIT

Thank You