KYTHERA

DECIPHERING HEALTHCARE

Estimating Compute Costs: A Guide for Kythera Wayfinder WorkSpace Usage

Exploring how Kythera's advanced data technology delivers greater precision, efficiency, and predictability in cloud compute performance.

Executive Summary

Health systems today rely on advanced data analytics platforms to drive clinical, operational, and financial insights. However, the adoption of cloud-based solutions like Kythera Wayfinder WorkSpace, our advanced data technology platform built on Databricks, introduces scalable compute capabilities that can vary based on usage patterns. These costs arise from the elastic, on-demand nature of cloud computing, where resources are provisioned in real-time to handle data processing demands.

This white paper explores the primary drivers of compute costs, including user count, data volume, query complexity, and code efficiency, and illustrates how Kythera's training and support help control expenses. Using real-world modeling, it projects annual costs for health systems ranging from five to one hundred users, demonstrating predictable, linear scalability. By understanding and optimizing these factors, organizations can maximize ROI while taking full advantage of Wayfinder WorkSpace's powerful analytics capabilities.

KYTHERA

Introduction

In an era of value-based care and precision medicine, health systems generate vast amounts of data from electronic health records (EHRs) and claims to genomic sequences and population health metrics. Platforms like Kythera Wayfinder WorkSpace enable seamless analysis of this data, empowering providers to identify trends, reduce readmissions, and optimize resource allocation.

At the core of WorkSpace is a scalable compute engine which processes queries across distributed clusters. While this delivers unparalleled speed and flexibility, it incurs compute costs tied directly to resource consumption. These costs are not arbitrary; they reflect the value of on-demand infrastructure that scales with need and eliminates the sunk costs of on-premises hardware.

This paper demystifies compute costs, outlines key influencers, and provides a cost model to forecast expenses based on user scale.

Why Compute Costs Exist in Platforms Like Kythera Wayfinder WorkSpace

Cloud compute platforms such as Databricks and Snowflake operate on a pay-as-you-go model, charging for the virtual machines, storage, and network resources consumed during data operations.

In WorkSpace, compute costs are derived from three sources: cluster provisioning, data movement and storage, and elastic scaling.

- **Cluster Provisioning:** When a user runs a query, WorkSpace spins up ephemeral clusters of compute nodes. These nodes perform parallel processing on large datasets, and charges accrue per second of runtime.

- **Data Movement and Storage:** Ingesting, transforming, and querying terabytes of health data requires bandwidth and I/O operations, billed based on volume.

- **Elastic Scaling:** Unlike fixed hardware, cloud resources auto-scale to handle peak loads (e.g., end-of-month reporting), ensuring performance but tying costs to actual usage.

These advanced capabilities enable health systems to analyze petabytes of data without upfront capital expense. This democratizes advanced analytics for non-technical users, including strategists and analysts, by making scalable computation accessible as needed.

KYTHERA

Key Factors Influencing Compute Charges

Compute costs in WorkSpace are not fixed. They fluctuate based on several interconnected variables. By understanding how these factors interact, health systems can proactively manage spend and maintain consistent performance.

1. Number of Users Accessing Compute Functions

The most direct driver is concurrency and volume of user activity. More users running simultaneous queries increase cluster utilization. For instance, a team of 10 analysts reviewing large data parameters will consume more resources than a solo data scientist.

2. Size of Data Being Reviewed

Health datasets often span billions of records. Querying 1-2 states of data versus a national dataset demands less compute. Platforms like WorkSpace optimize by partitioning data, but large, unfiltered scans amplify bills.

3. Query Design and Code Efficiency

The structure of SQL, Python, or Spark queries profoundly impacts resource use:

- **Efficient Code:** Uses indexing, filtering early in the pipeline, and broadcast joins to minimize data shuffling.

- **Inefficient Code:** Full table scans, nested loops, or unoptimized aggregations can balloon runtime from minutes to hours, multiplying costs exponentially.

For example, a poorly written query aggregating claims data across 5 years might scan redundant partitions repeatedly, incurring 10x the compute of a refactored version with window functions.

Real-World Implications

In health systems, inefficient code can occur during exploratory analysis (e.g., an analyst iteratively querying variants in a national set of data). A single misguided query on a large dataset could cost $50–$200 in compute, scaling to thousands across tens of users.

KYTHERA

Cost Modeling: Projected Expenses for Kythera Wayfinder WorkSpace Users

To provide transparency, Kythera has modeled compute costs based on real-world usage patterns from several health systems across 189 active days. The tables below project monthly and annual costs, accounting for dataset size, user count, and activity level. Projections are derived from observed daily rates, extended to 30-day and 360-day periods, providing a reliable baseline for budgeting.

National Data Set Usage of 30 Billion Claims with 50 Billion of Claimline records (doesn't include Rx data). User Average Daily Cost assumes a very active analyst working within the system for 8-10 hours per working day.

User Count	Avg Daily Cost ($)	Avg Monthly Cost ($)	Annual Cost ($)	Cost per User per Day($)
1 User	$70.49	$2114.70	$25,376.40	$70.49
5 Users	$352.45	$10,573.50	$126,882	$70.49
10 Users	$704.90	$21,147	$253,764	$70.49
25 Users	$1,762.25	$52,867.50	$634,410	$70.49
50 Users	$3,524.50	$105,735	$1,268,802	$70.49
100 Users	$7,049	$211,470	$2,537,640	$70.49

KYTHERA

The first table represents national-level usage scenarios, while the second illustrates a smaller, single state of claims data.

Single State Usage of 1 Billion Claims with 2 Billion of Claimline records (doesn't include Rx data). User Average Daily Cost assumes 0-4 hours a day of an analyst working within the system per working day.

User Count	Avg Daily Cost ($)	Avg Monthly Cost ($)	Annual Cost ($)	Cost per User per Day($)
1 User	$0.70	$21.10	$253.80	$0.70
5 Users	$3.50	$105.70	$1,268.80	$0.70
10 Users	$7	$211.50	$2,537.60	$0.70
25 Users	$17.60	$528.70	$6,344.10	$0.70
50 Users	$35.20	$1,057.40	$12,688.20	$0.70
100 Users	$70.50	$2,114.70	$25,376.40	$0.70

Optimization Strategies: Kythera's Commitment to Efficiency

Kythera recognizes that sustainable analytics, not raw power, drive value. To promote efficiency from day one, we embed optimization practices into every implementation.:

- **Training Programs:** Onboarding workshops teach query best practices, including query optimization, data partitioning, and caching. Users learn to profile queries with built-in tools, identifying bottlenecks before they use compute.

- **Dedicated Support:** Our team provides code reviews, performance audits, and custom scripts.

- **Compute Governance:** Kythera's Wayfinder WorkBoard, a companion analytics environment to WorkSpace, provides pre-developed BI visualization tools that enable users to monitor pre-set dashboards without creating one-off analysis. Because WorkBoard operates without compute charges, using these visualization tools instead of running direct queries in WorkSpace can help reduce overall compute costs.

- **Proactive Guidance:** Quarterly reviews assess usage trends, recommending data pruning or ML-based query tuning.

KYTHERA

Conclusion

Compute costs in Kythera Wayfinder WorkSpace are a natural byproduct of elastic, high-performance analytics which are essential for health systems to thrive. By factoring in user scale, data volume, and query efficiency, organizations can anticipate expenses with precision, as shown in our user-based model.

Kythera's training and support help organizations control costs and realize the full value of WorkSpace, avoiding unexpected budget surprises. When paired with efficient solutions such as WorkBoard, organizations can extend this value even further by gaining additional visibility without added compute expense. We invite health leaders to engage with our team for a tailored cost assessment, turning data insights into measurable outcomes.

About Kythera

Kythera Labs is a data technology company built to reduce the uncertainty in the use of healthcare data and to accelerate time to value for our customers. Unlike traditional data vendors, Kythera delivers analysis-ready data from day one, providing immediate access through interoperable, cloud-based environments such as Databricks, Azure, and Snowflake. Our technology architecture is designed for efficiency, minimizing waste and enabling customers to begin meaningful analysis within days of implementation.

Our differentiation lies in the fidelity and usability of our data. Kythera's mastering and quality assurance processes ensure that each dataset reflects real healthcare encounters and patient activity, rather than inflated or fragmented volumes. Supported by deep expertise in healthcare data sourcing and enrichment, Kythera helps organizations gain faster, more confident answers from their data while addressing long-standing challenges such as token mismatches and the usability of raw claims data.

Connect with Kythera. Kythera is a data technology company that brings unprecedented clarity and structure to complex real-world healthcare data. Kythera's Wayfinder Technology Platform, supported by pre-configured pipelines, processing libraries, analysis tools and remastered datasets, helps Healthcare and Life Sciences organizations work with greater speed, scale and confidence. Learn more at www.kytheralabs.com.

KYTHERA
DECIPHERING HEALTHCARE

346 Main Street
Franklin, TN 37064
www.kytheralabs.com