



# Charting the Lakehouse Trail

## A Data Migration Adventure

*Migrating from a  
Data Warehouse to a Data Lakehouse*



## About Me:

- B.S. in CS @ Stanford
- Previously SWE on DoorDash's Data Platform Team
- Big Data & Distributed Systems Nerd
- Likes: Surfing & Snowboarding
- Dislikes: Making Slide Decks



**Erick Enriquez**

CEO & Cofounder  
InQuery

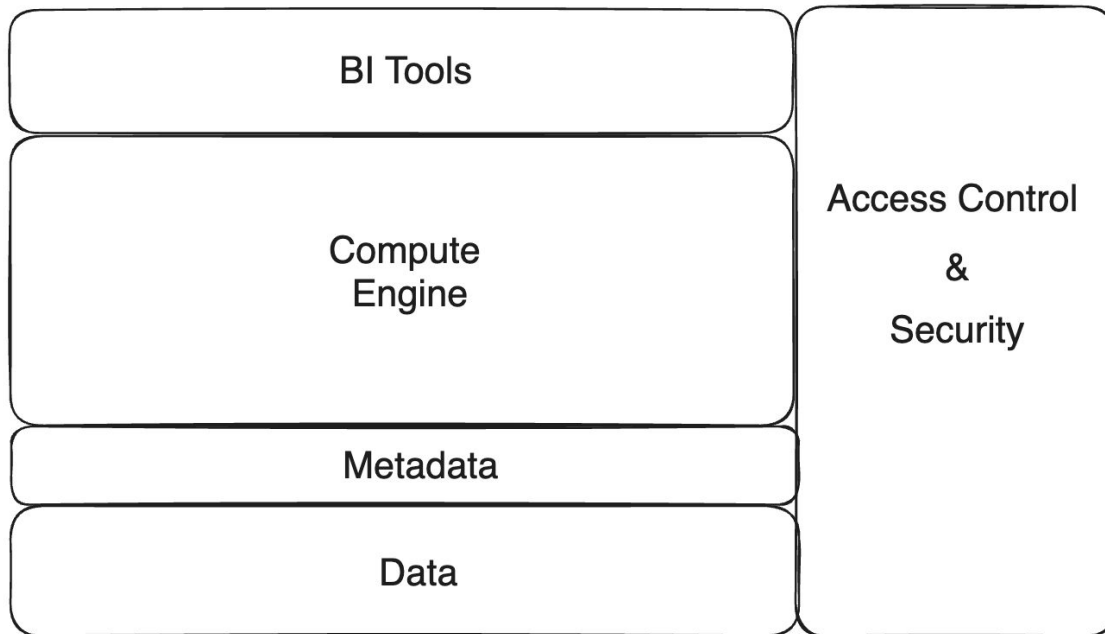
# What is a lake house ?

## Needs

- Storage engine
- File format
- Metastore
- Access control
- Query (compute) Engine
- Data

## Nice to haves

- Table format
- Semantic catalog
- Monitoring & observability
- Thorough documentation



# Why a lake house?

- Cost Efficiency
- Data Freshness
- Improved performance
- Vendor leverage
- Reduced data copies
- Improved data governance
- Data diversity
- Flexibility and future-proofing



# Step 1: Dataset selection

## Workload characteristics

- High value
- Low dependencies
- Strong freshness SLAs
- High volume or growth rate

## Examples

- Interactive analytics
- Event data
- ML pre-processing jobs



# Step 2: Metadata Registration

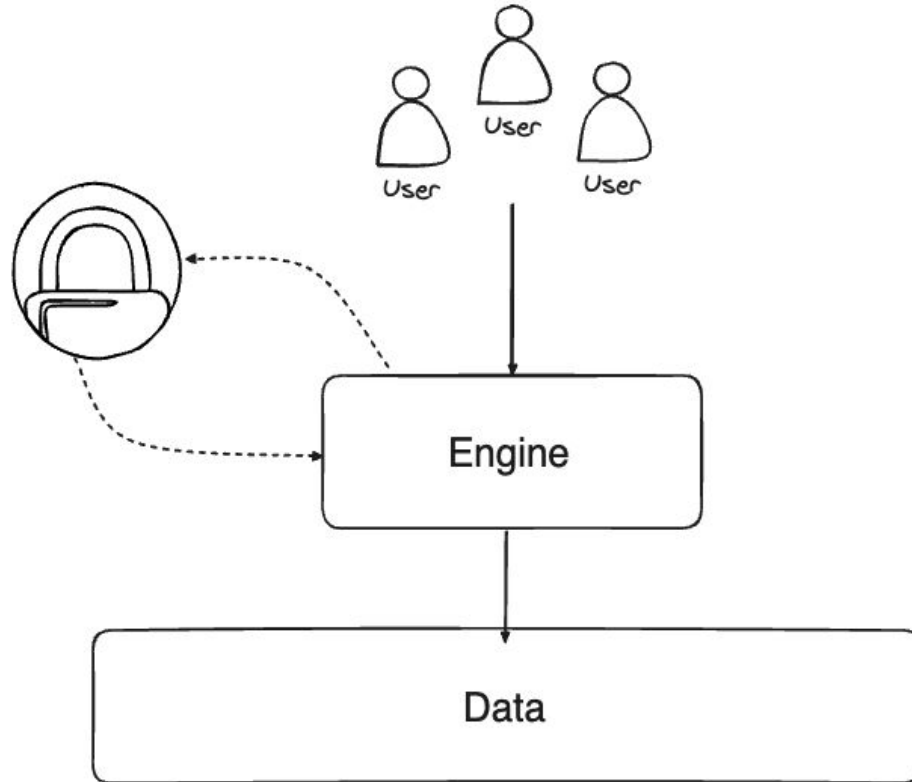
## Table Formats



## Metastores



## Step 3: Securing your data



## Step 4: Establishing workload parity

Snowflake SQL ↔ ANSI SQL ↔ Spark SQL ↔ Other SQL

Pro-tip: Use a SQL translator

Pitfall warning: Don't assume it will translate your queries



# Step 5: Benchmarking and performance tuning

## Query Engine

- Number of Workers
- Machine type
- Memory per machine
- CPU per machine
- Cluster architecture
- ...

## Storage

- Clustering
- Partitioning
- Compaction
- Locality
- ...

## Query

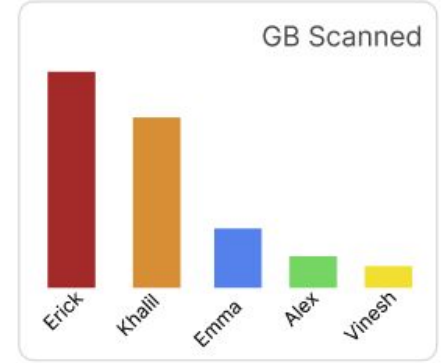
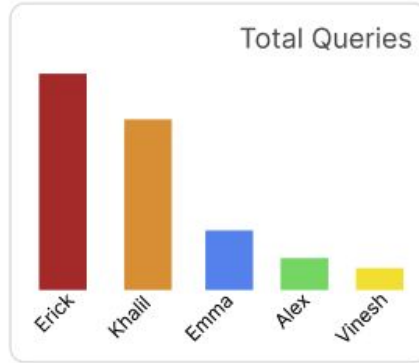
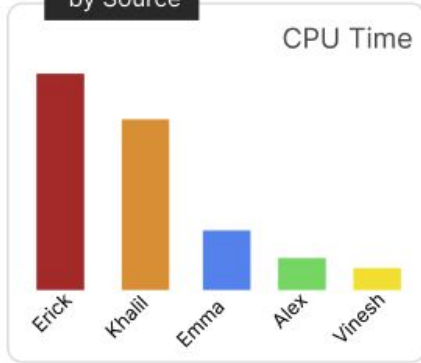
- Join Order
- Partition Key
- ...



# Step 6: Resource management and cost attribution

Usage by User ^

by Table  
by Source



## Step 7: Establishing new habits



# Conclusion

- Data & workload selection
- Table Registration and Metadata Management
- Securing your data
- Establishing workload parity
- Benchmarking and engine tuning
- Resource management and cost attribution
- Establishing new habits

# THANK YOU

erick@inquery.app

www.inquery-data.com

Linkedin: <https://www.linkedin.com/in/erick-enriquez/>