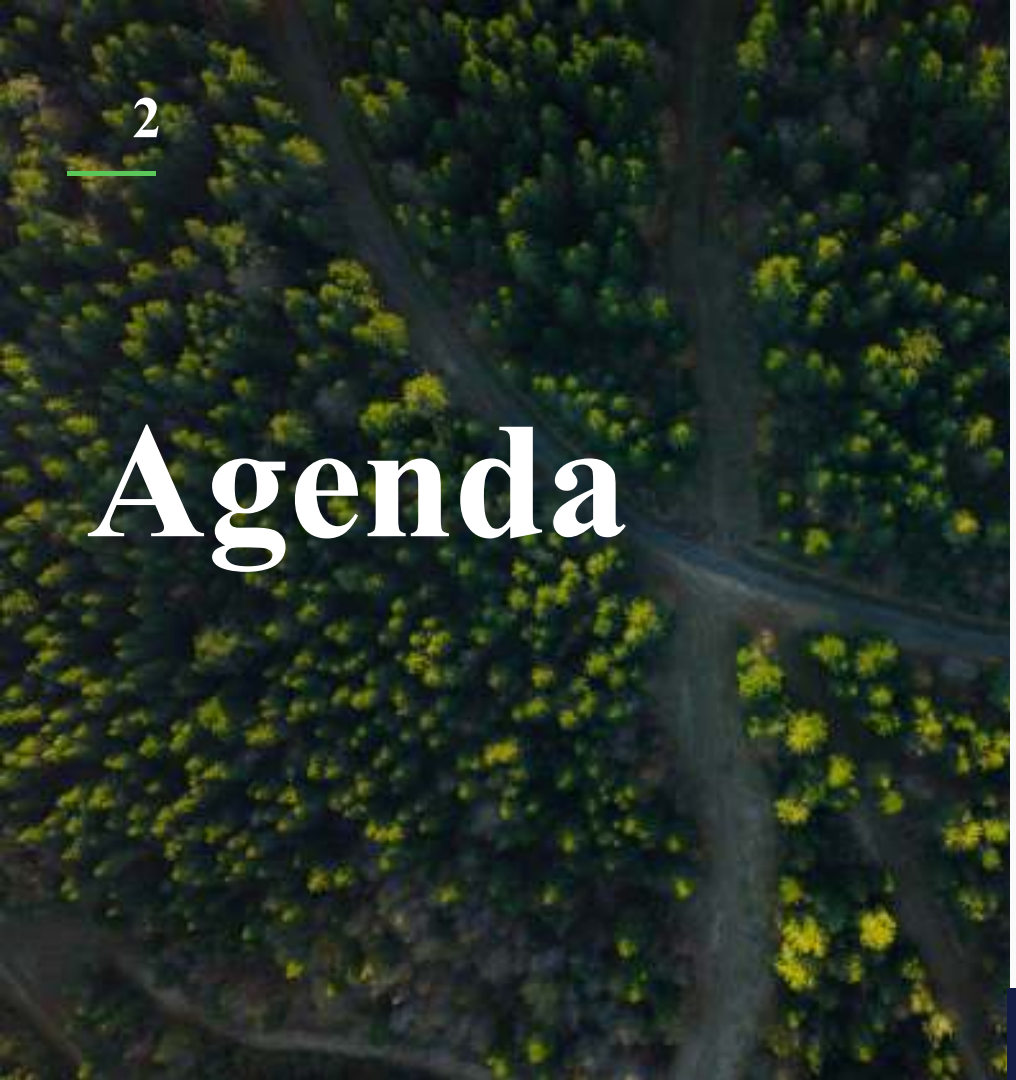
Five Ways Database Modernization Simplifies Your Data Life

Mike Boyarski, Senior Director Product Marketing



Agenda

- Quick Database Landscape Roundup
- Five (Customer) Ways to a Simpler Data Life
- An Intro to MemSQL

“Companies with **data-driven environments** have up to **50% higher market value** than other businesses.”

IDC

Database Limitations are Impacting Data-Driven Initiatives

Data demands

More people want data access, impacting database concurrency and scale

Analytic performance

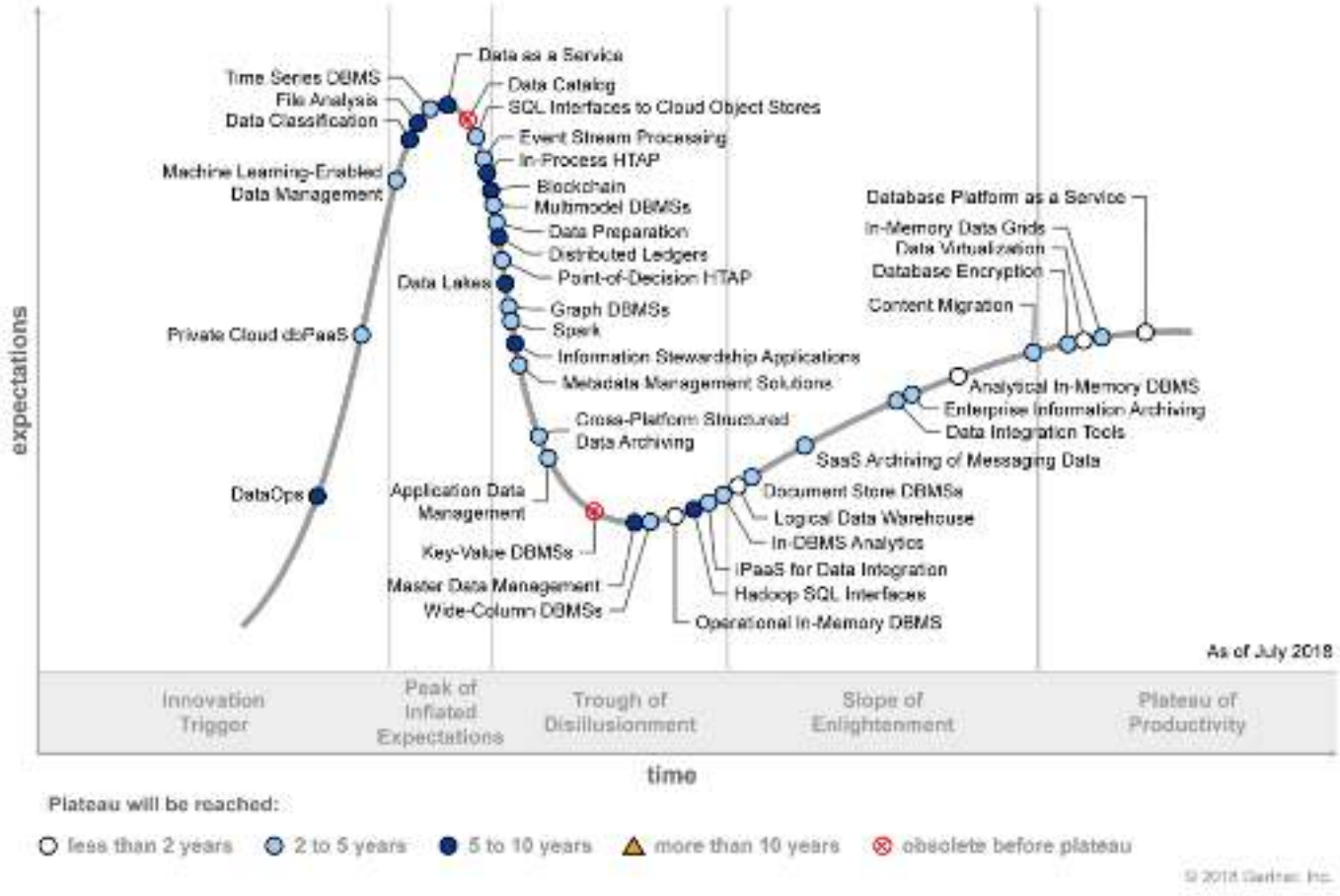
Faster, more frequent intra-day insights

Rising data management costs

Costly upgrades and maintenance with existing technology

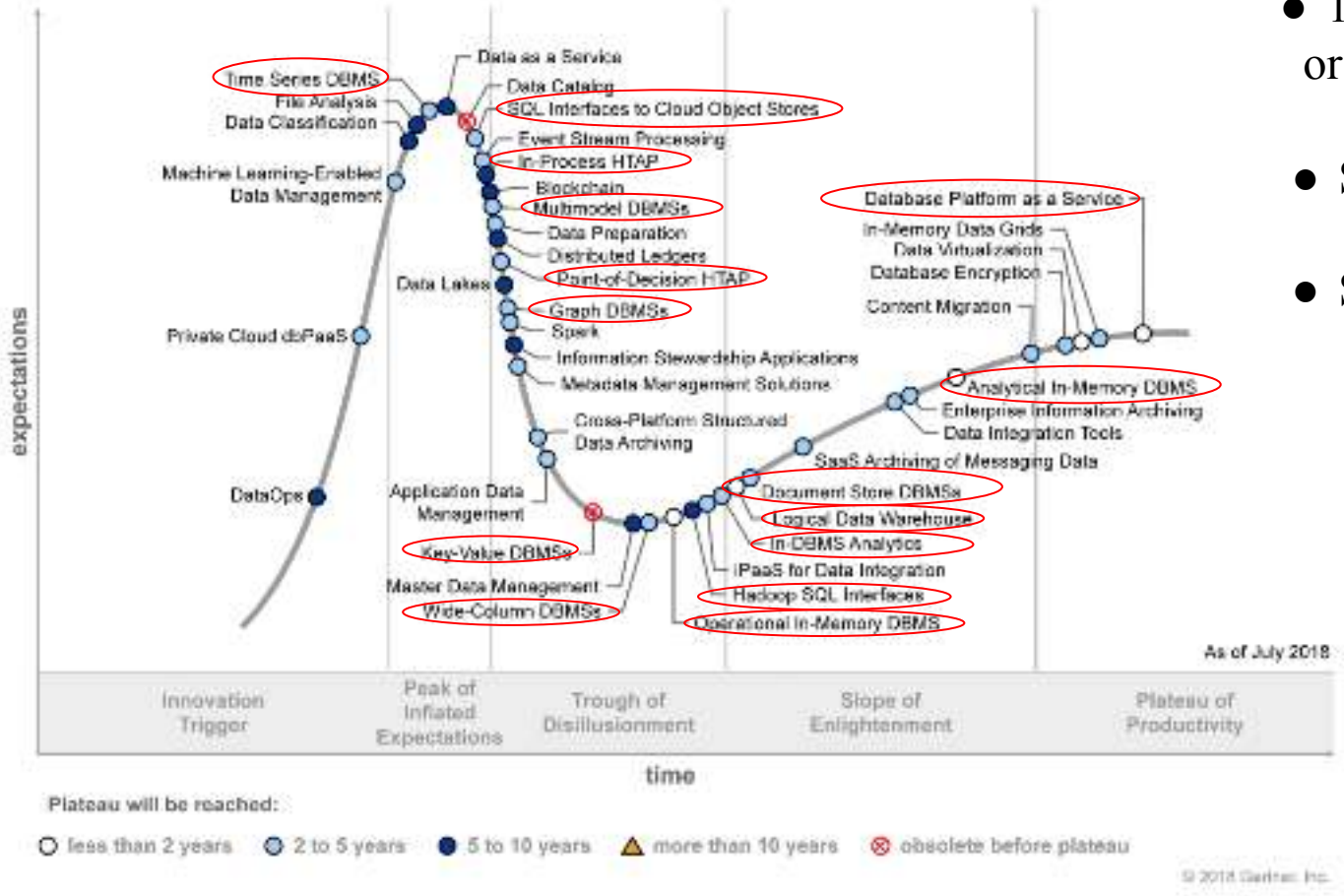
Database Landscape

Figure 1. Hype Cycle for Data Management, 2018



Source: Gartner (July 2018)

Figure 1. Hype Cycle for Data Management, 2018



- 15 Different Databases or DB Options
- Some are features
- Some are platforms

Source: Gartner (July 2018)

An Attempt to Simplify

Workload	Database Type
Transactions/Writes?	Operational
Queries/Reads?	Analytical
Both transactions and queries?	HTAP (next slide)

An Intro to HTAP

Hybrid Transaction/Analytical Processing

"HTAP architectures **remove the latency** between when a piece of **data is generated** and **when it is ready for analytics**"

"Delivering Digital Business Value Using Practical Hybrid Transactional/Analytical Processing", Analysts: Adam M. Ronthal, Roxane Edjlali

An Intro to HTAP

Hybrid Transaction/Analytical Processing

"HTAP architectures **remove the latency** between when a piece of **data is generated** and **when it is ready for analytics**"

"The traditional batch-oriented cycle of business intelligence insight is giving way to **new deployment architectures delivering real-time access to data and the expected insights** that it will provide."

"Delivering Digital Business Value Using Practical Hybrid Transactional/Analytical Processing", Analysts: Adam M. Ronthal, Roxane Edjlali

An Intro to HTAP

Hybrid Transaction/Analytical Processing

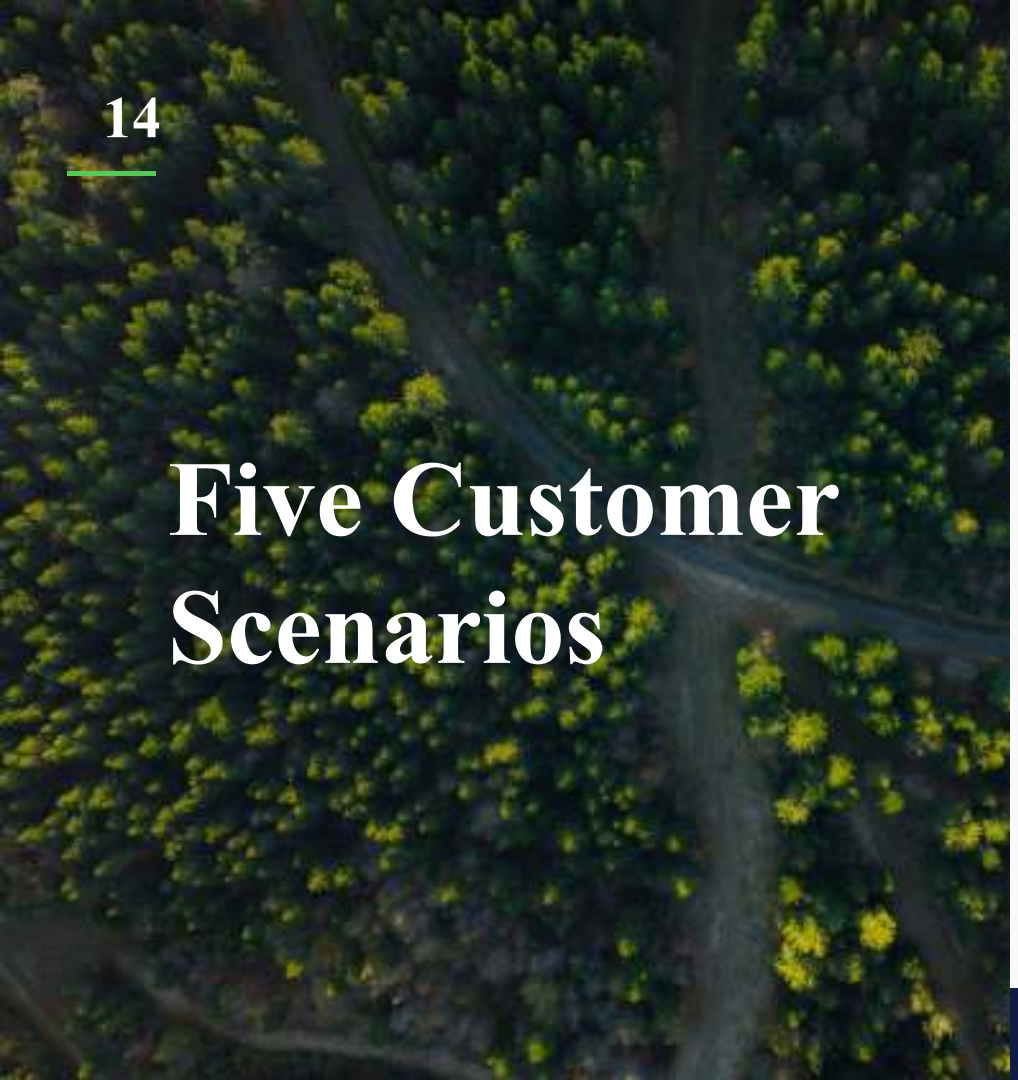
"HTAP architectures **remove the latency** between when a piece of **data is generated** and **when it is ready for analytics**"

"The traditional batch-oriented cycle of business intelligence insight is giving way to **new deployment architectures delivering real-time access to data and the expected insights** that it will provide."

"To support this demand, the **convergence of analytic and operational platforms is becoming more common**. "

"Delivering Digital Business Value Using Practical Hybrid Transactional/Analytical Processing", Analysts: Adam M. Ronthal, Roxane Edjlali

Workload	Database Type
Transactions?	Operational
Queries?	Analytical
Both transactions and queries?	HTAP (will explain next slide)
Lots of transactions per second? 10,000+ -> Millions	In-Memory Boosted Operational
Fast reads + ad-hoc? Milliseconds on 100Ms/Bs of rows	In-Memory boosted +Columnstore Analytical
Data Science/ML	Analytical or Data Lake



Five Customer Scenarios

Faster Event to Insight

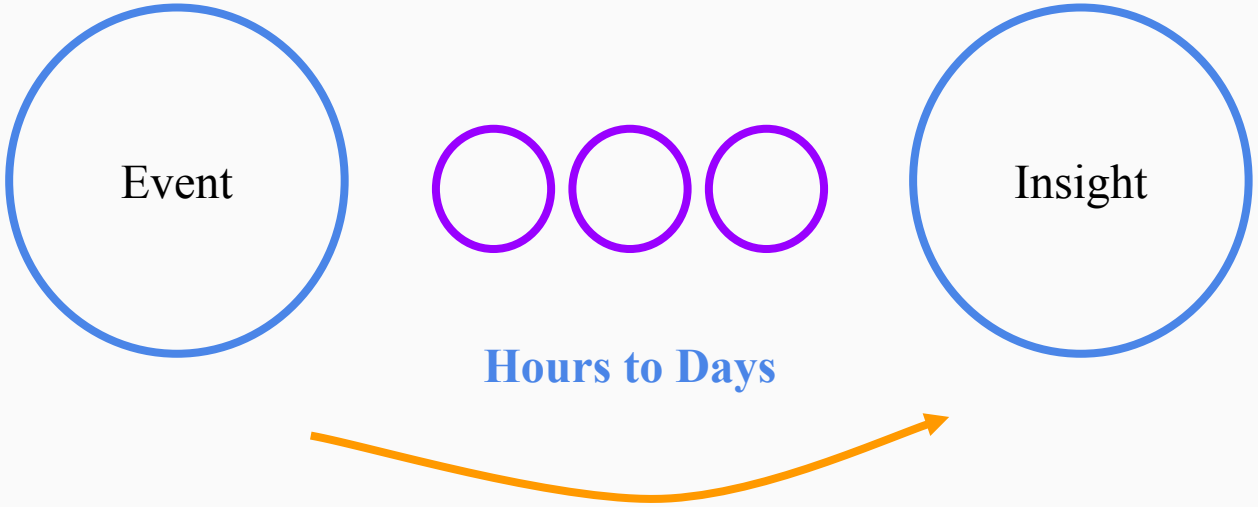
Growth in Concurrency

Cost Effective Performance

Accelerate Big Data

Deployment Flexibility

Event to Insight Delays



Event to Insight Delays

Common Challenges

- Protect performance limiting ad-hoc aggregates
- Load data during “non-peak” operations
- Writes slow down reads

Faster Event to Insight Example

Situation:
Oracle Financials database could only support once/day reporting resulting in **fraudulent and duplicate orders**

Industry: Energy

Faster Event to Insight Example

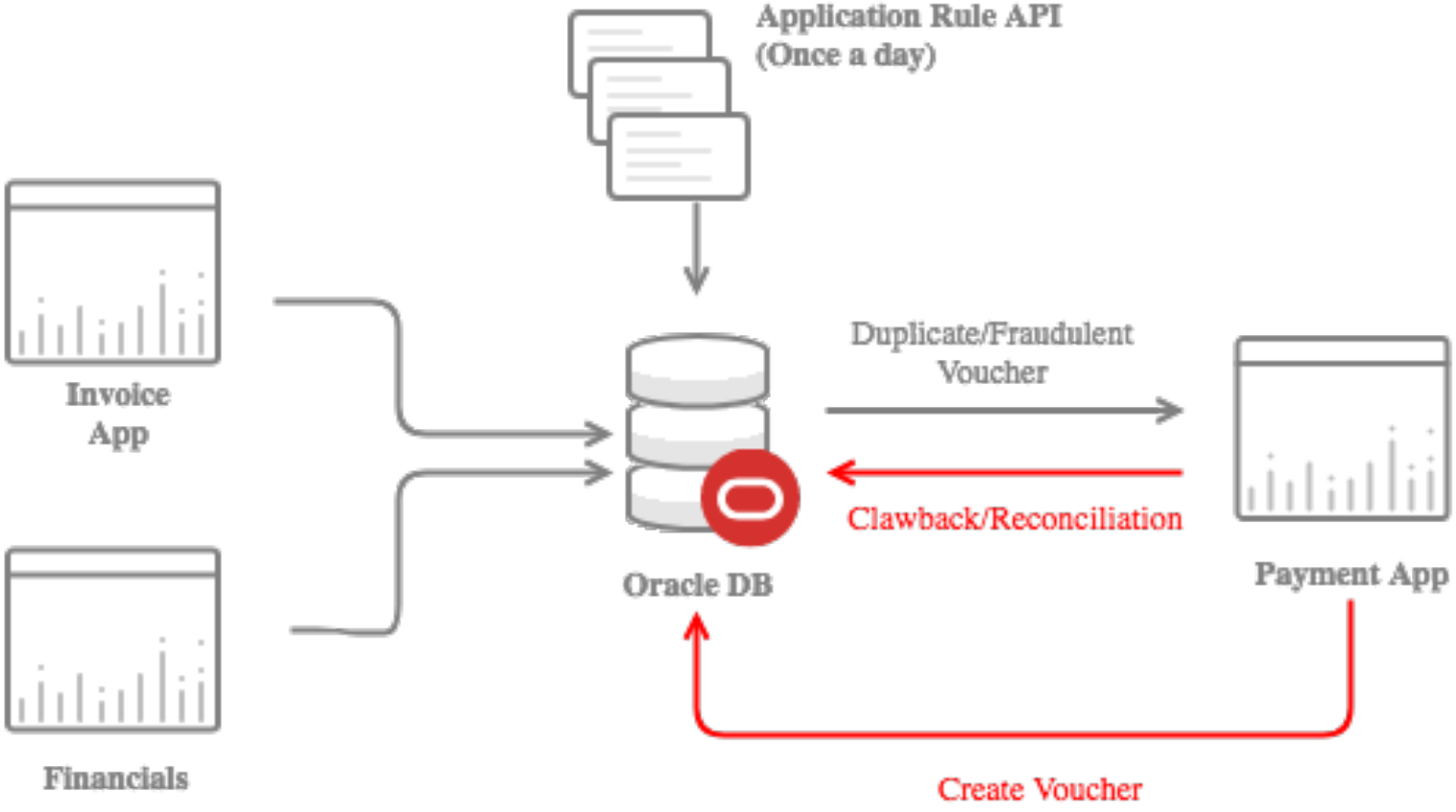
Situation:
Oracle Financials database could only support once/day reporting resulting in **fraudulent and duplicate orders**

Industry: Energy

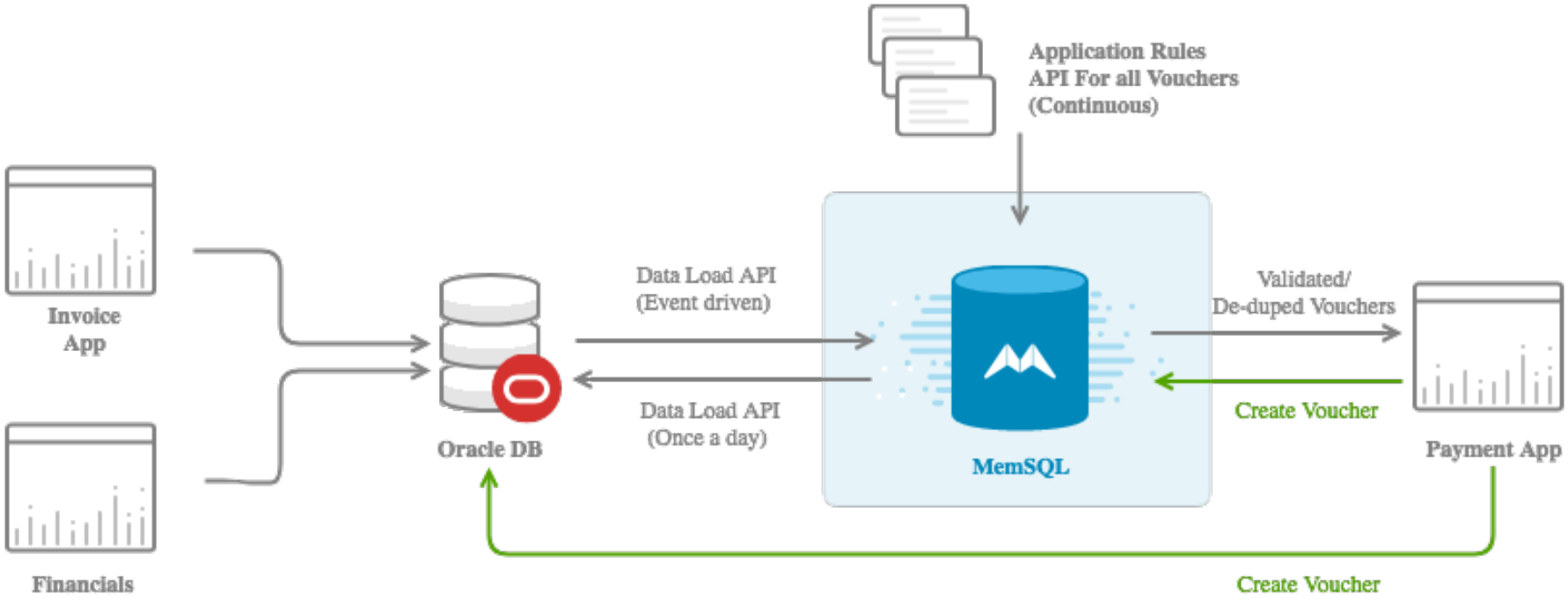
Solution:
Real-time synchronization from Oracle to MemSQL with change data capture enabling continuous reporting

Result: Cost savings

Previous Architecture

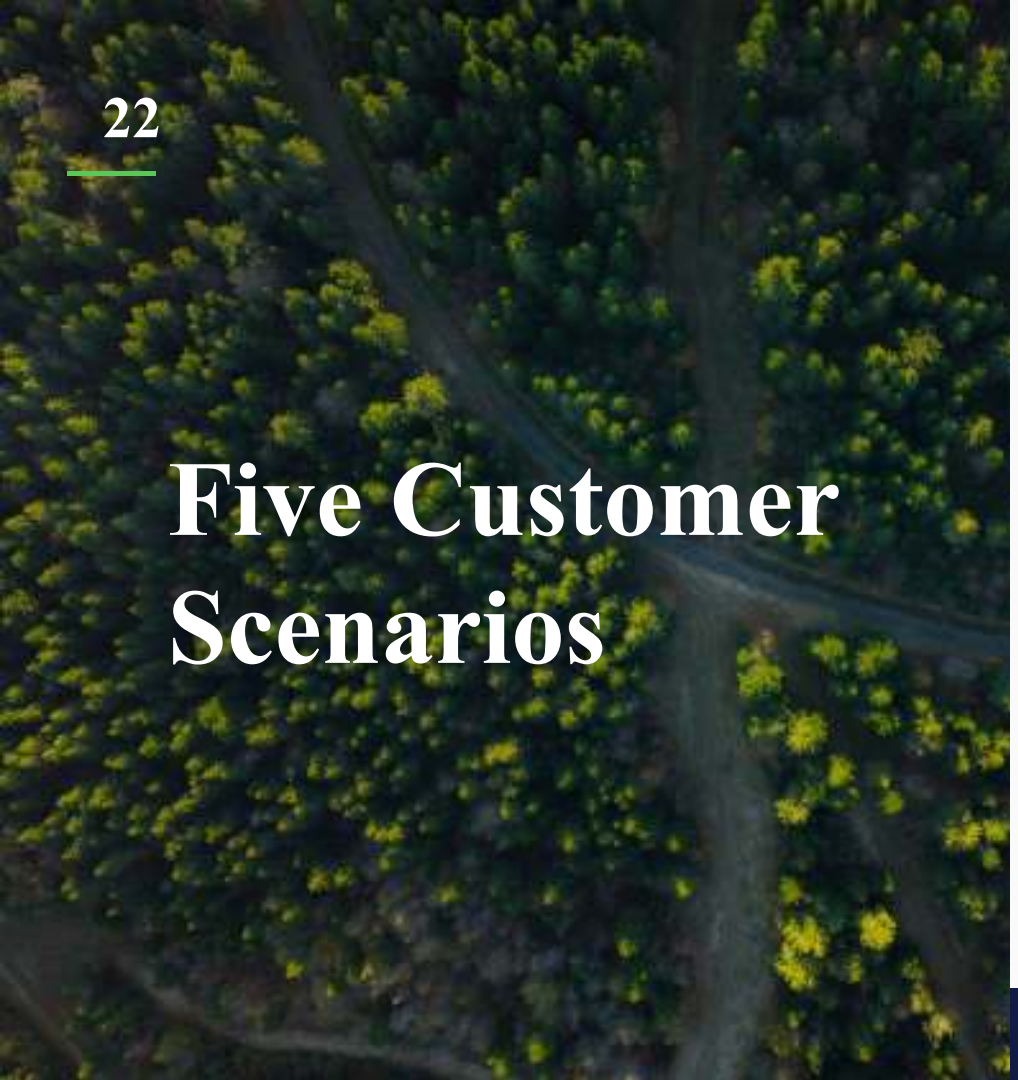


Modernized Architecture



DB Modernization for Faster Event to Insight

- Lock-free ingestion
- Distributed for faster parallel loading
- Scalable durability for reliability and accuracy
- Relational SQL for BI tool compatibility



Five Customer Scenarios

Faster Event to Insight

Growth in Concurrency

Cost Effective Performance

Accelerate Big Data

Deployment Flexibility

“Wait in Line” Analytics

Common Challenges

- Under-powered database can’t support data and user growth
- Too many non-standard ad-hoc queries
- Transactions impact queries

High Query Concurrency Example

Situation:
Operational dashboards couldn’t support demand of roughly 250k queries across 1,000+ users

Customer: UBER

High Query Concurrency Example

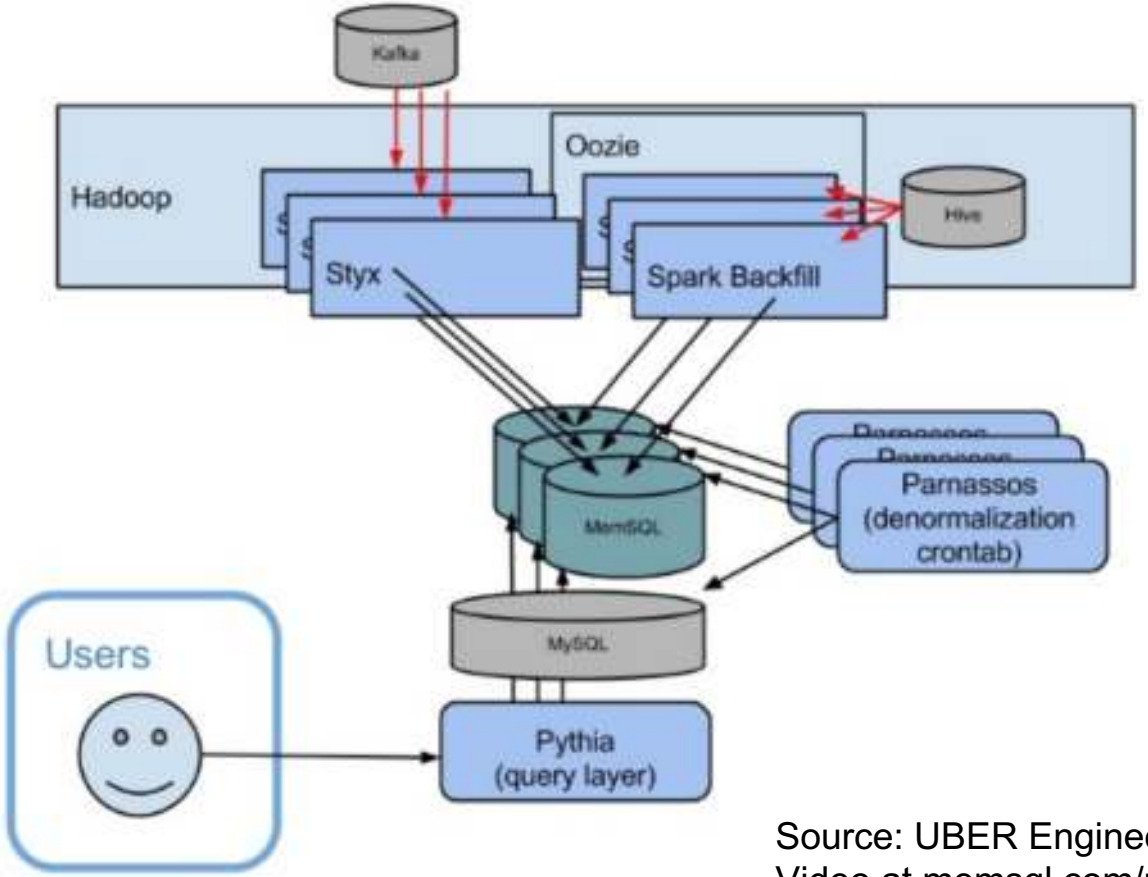
Situation:
Operational dashboards couldn't support demand of roughly 250k queries across 1,000+ users

Customer: UBER

Solution:
MemSQL delivers sub-second query response for 1,000s of users while supporting continuous writes

Result: Real-Time Dashboards

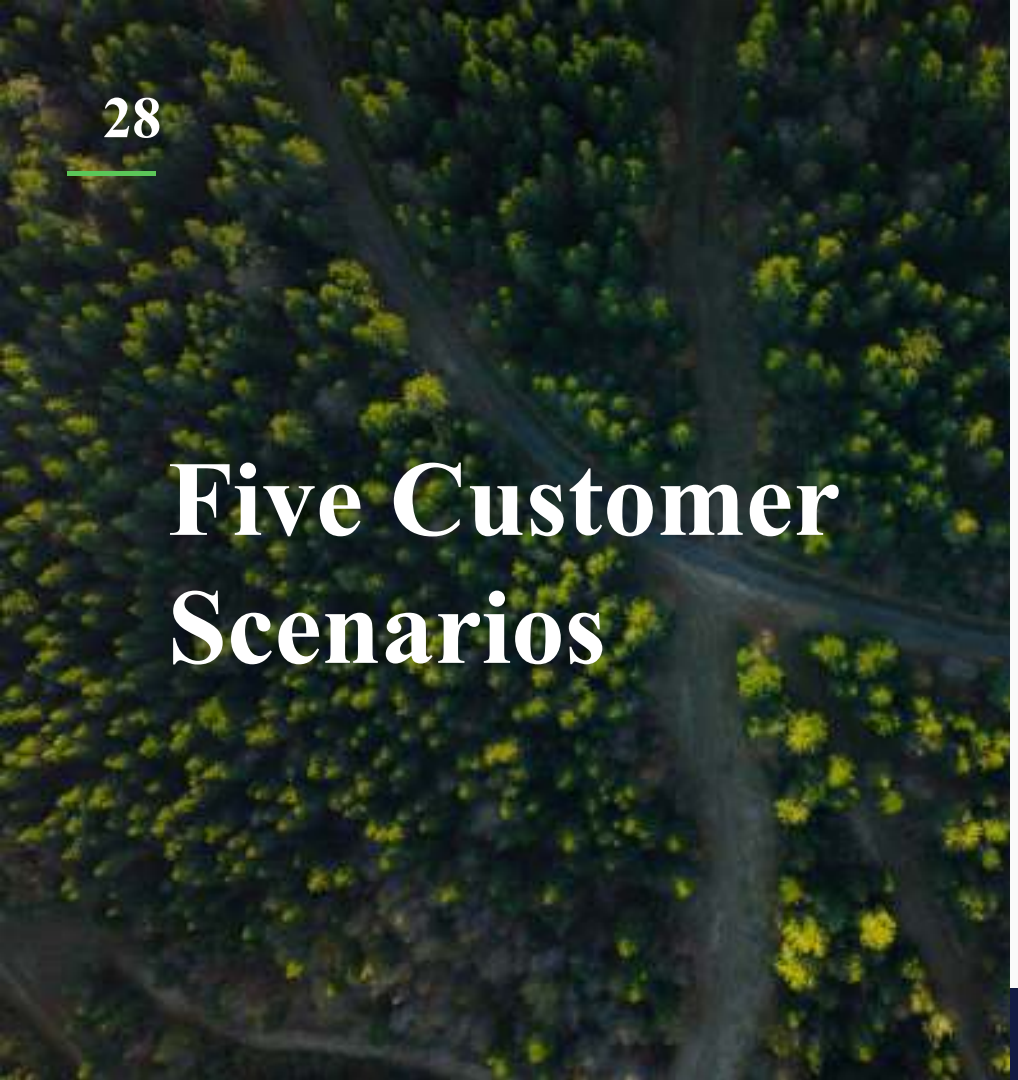
Architecture



Source: UBER Engineering
Video at memsql.com/uber

DB Modernization for Concurrency Growth

- Distributed architecture simplifies scale demands
- Query compilation for fast continuous queries
- In-memory optimized tables for continuous ingestion



Five Customer Scenarios

Faster Event to Insight

Growth in Concurrency

Cost Effective Performance

Accelerate Big Data

Deployment Flexibility

Costly Performance

Common Challenges

- Specialized HW required for growing data
- Costly DB options and accelerators
- Professional service tuning

Costly Performance Example

Situation:
Degrading query and ingestion performance couldn't deliver daily billing requirement resulting in lost revenue

Industry: High Tech

Costly Performance Example

Situation:
Degrading query and ingestion performance couldn't deliver daily billing requirement resulting in lost revenue

Industry: High Tech

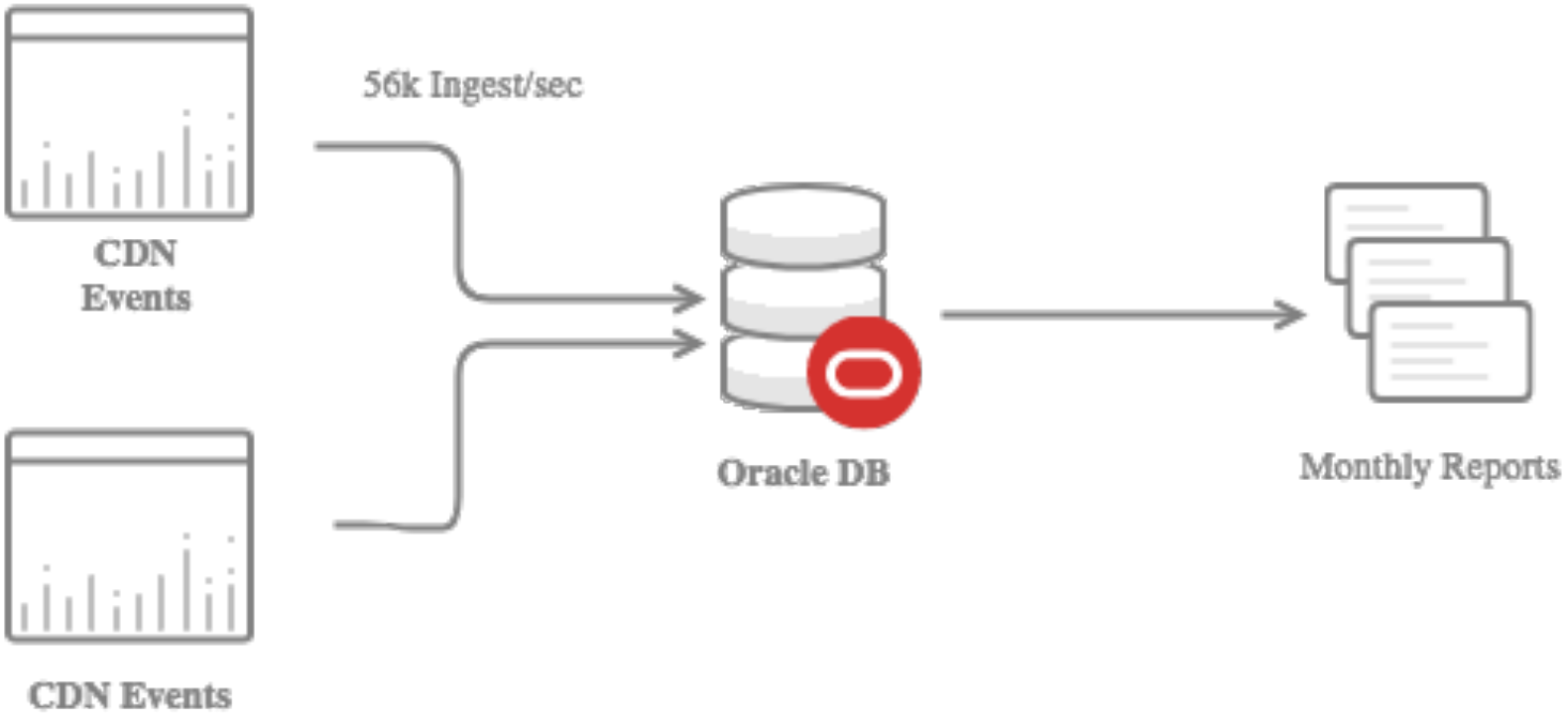
Solution:
Accelerated ingestion from **56k/s** to **6M/s**

100x faster queries

⅓ the cost of incumbent solution

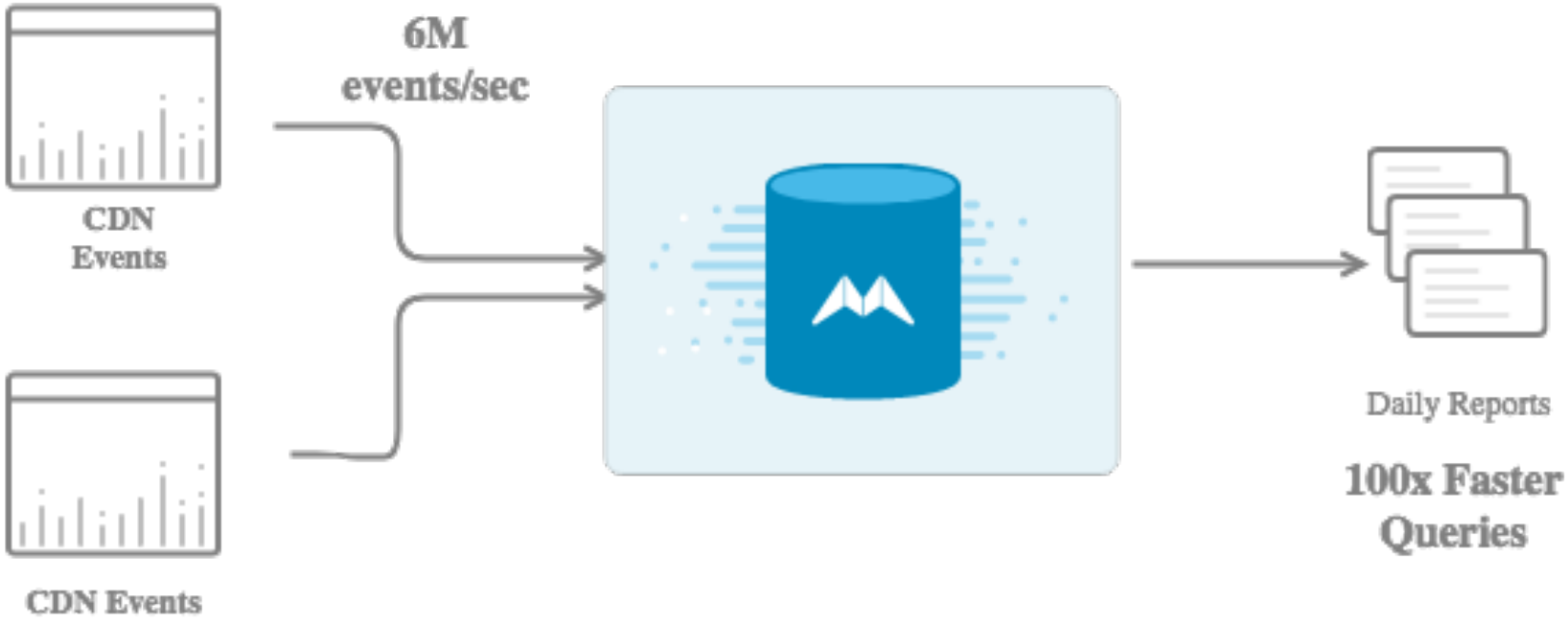
Result: Added revenue

Before Architecture



Source: Akamai

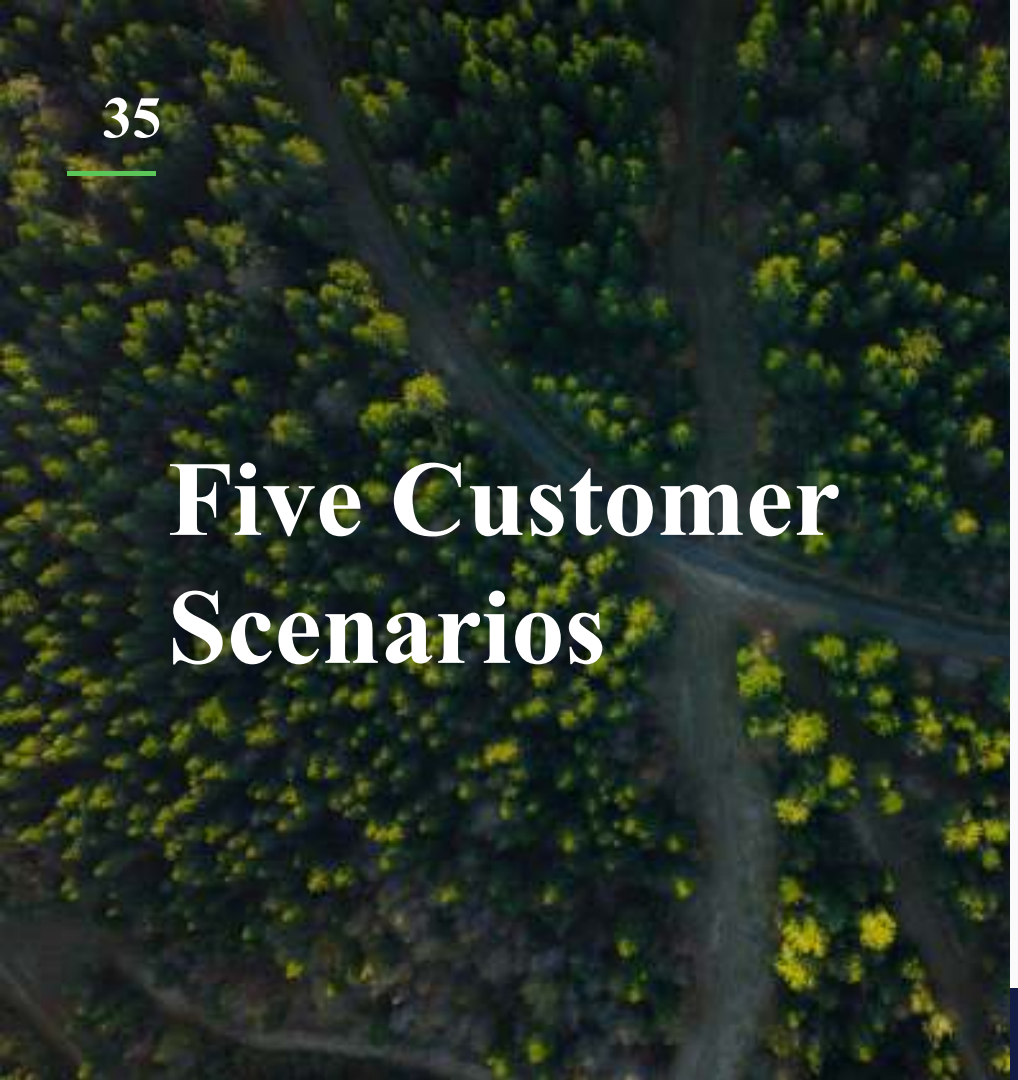
After Architecture



Source: Akamai

DB Modernization for Cost Efficient Performance

- Optimized for industry standard hardware
  - Query vectorization - CPU acceleration
  - Scale-out architecture
- Columnar data compression - 5-10x
- Memory optimization at no additional license cost



Five Customer Scenarios

Faster Event to Insight

Growth in Concurrency

Cost Effective Performance

Accelerated Big Data

Deployment Flexibility

Accelerate Big Data

Common Challenges

- Slow queries
- Existing BI tools not compatible
- Flexible data structure hard to understand/use

Accelerate Big Data Example

Situation:
Queries taking hours to days to complete limiting use of analytics for decisions

Customer: Pandora

Accelerate Big Data Example

Situation:
Queries taking hours to days to complete limiting use of analytics for decisions

Customer: Pandora

Solution:
MemSQL delivers sub-second queries on 100s of billions of rows

Result: Frequent use of analytics

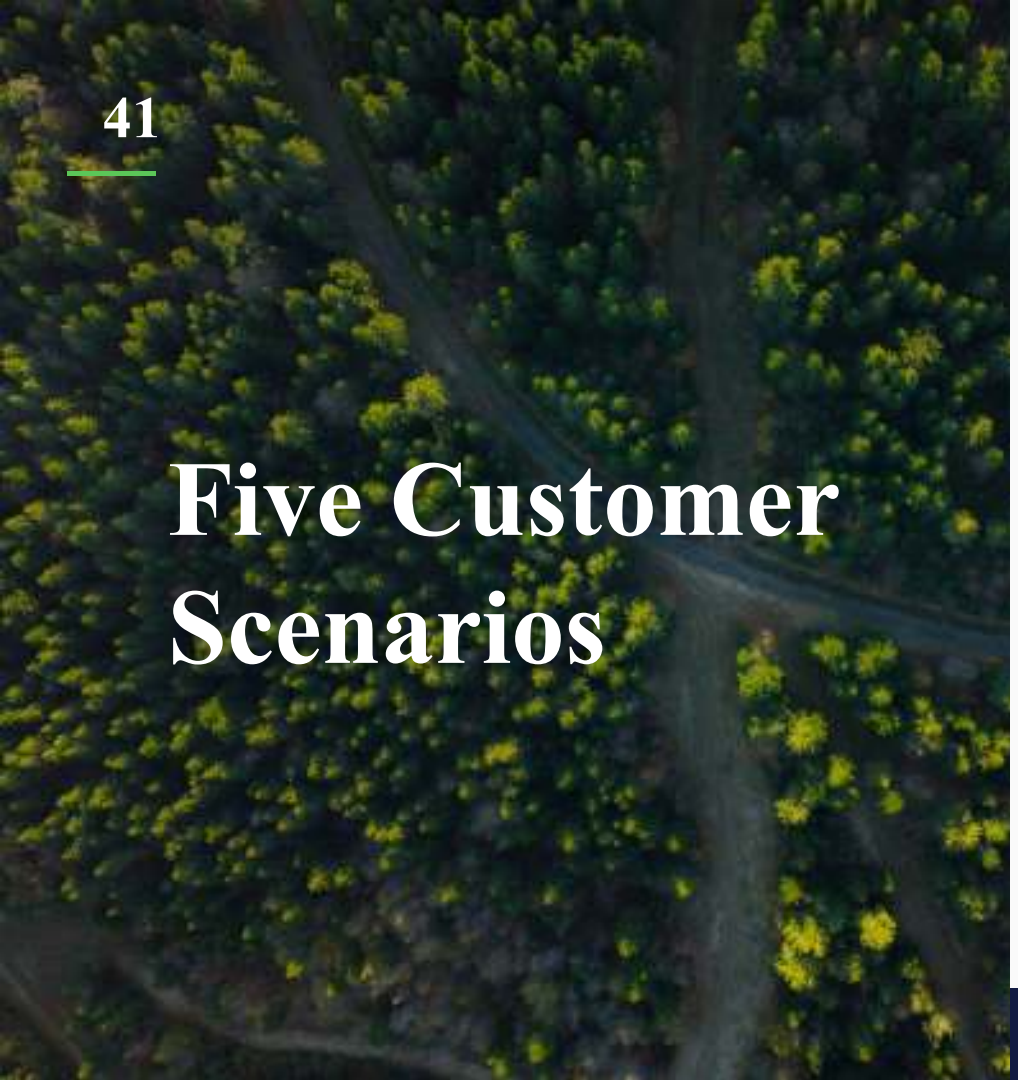
Kaan Erdener Follow
I'm a software engineer at Pandora Media
Aug 13 · 18 min read

engineering.pandora.com

Using MemSQL to Query Hundreds of Billions of Rows in a Dashboard

DB Modernization for Faster Big Data

- Lock-free architecture for fast ingestion
- Disk-based compute for cost effective storage
- Connectivity for bulk and stream loading (ie. Kafka, Spark, HDFS, S3, Azure Blob)
- Columnar table format eliminates pre-aggregations and background queries



Five Customer Scenarios

Faster Event to Insight

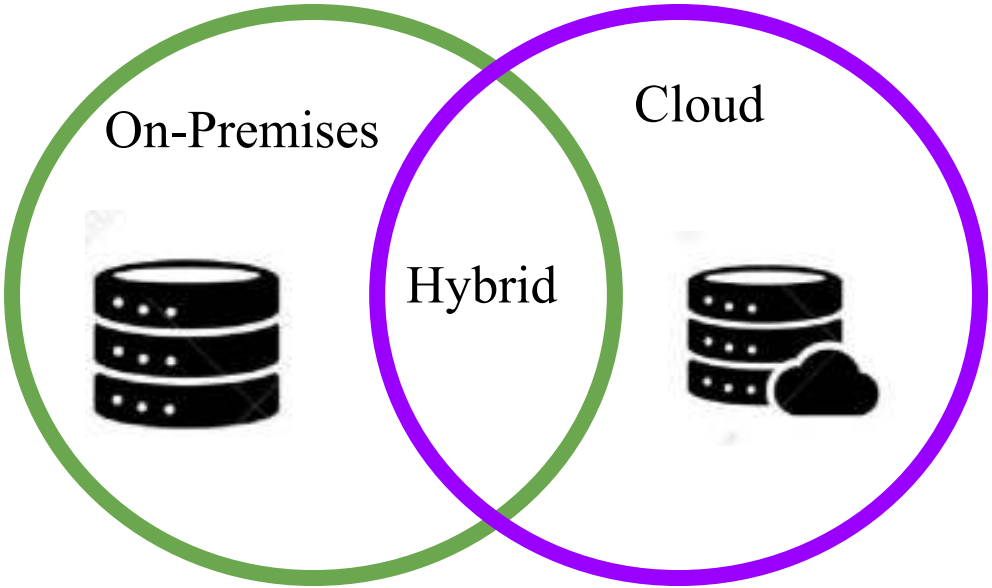
Growth in Concurrency

Cost Effective Performance

Accelerated Big Data

Deployment Flexibility

Deployment Inflexibility



Common Challenges

- Database runs on one cloud
- Cloud and On-Premises product versions are different

Deployment Flexibility Example

Situation:
Cloud application deployed on Azure and Equinix for customer security requirements, requiring multi-cloud database

Customer: Kollective

Deployment Flexibility Example

Situation:
Cloud application deployed on Azure and Equinix for customer security requirements, requiring multi-cloud database

Customer: Kollective

Solution:
MemSQL used on Azure and Equinix, achieving breakthrough performance for real-time dashboards

Result: Improved customer experience

What is MemSQL

The Modern Database for Today's Performance

Fast Loading

Stream data
On-the-fly transformation
Multiple sources

Low Latency Queries

Scalable ANSI SQL
Petabyte scale
Columnar compression

Scalable User Access

Scale-out for performance
Converged transactions and analytics
Multi-threaded processing

MemSQL: The No-Limits Database

High Speed Ingest
Fast bulk load or stream data with real-time pipelines

Memory Optimized Tables
Ultra-low latency for transactions and analytics

Disk Optimized Tables
Petabyte scale analytics with compression and performance

Ecosystem

Real-Time Data

Messaging and Transforms

Data Inputs

MEMSQL

Relational Key-Value Document Geospatial

High Speed Ingest | **Memory Optimized Rowstore** | **Disk Optimized Columnstore**

Real-Time Applications

BI Dashboards

Bare Metal, Virtual Machines, Containers

On-Premises, Cloud, As a Service

Fast Data Ingestion

- Stream ingestion
- Fast parallel bulk loading
- Built-in `Create Pipeline`
- Transactional Consistency
- Exactly-Once Semantics
- Native integrations with Kafka, AWS S3, Azure Blob, HDFS

Instant Insights

- Scalable ANSI SQL
- Full ACID capabilities
- Support for JSON, Geospatial, and Full-Text Search
- Fast Query Vectorization and Compilation
- Extensibility with Stored Procedures, UDFs, UDAs

Thank You

MEMSQL