The Extension of Data Analytics Architectures to support AI agents

Suresh Bysani

7th August 2025

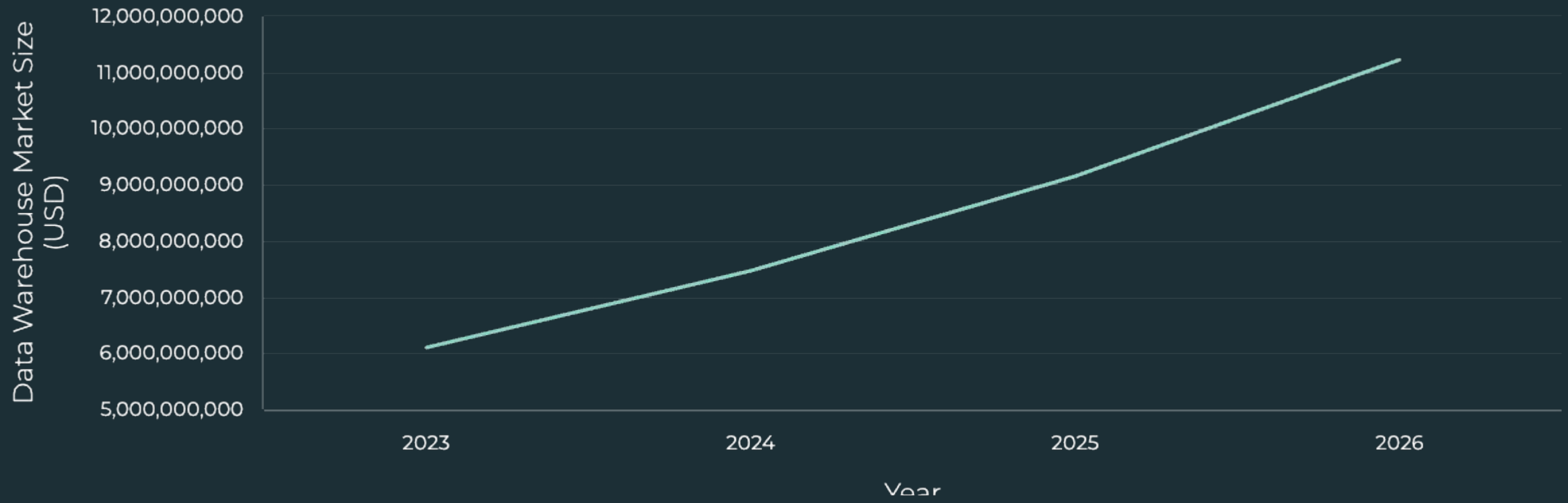
Suresh Bysani

Engineering Leader

Palo Alto Networks, Eightfofld.ai, Cohesity

SAAS AND AI Infrastructure

Increasing Analytics Spend



The data warehouse market is projected to grow rapidly at a CAGR of 22.5% from 2024-2032.

Why do companies need OLAP

Multidimensional Data Analysis

OLAP systems allow companies to analyze data from multiple perspectives, such as time, product, region, and customer, providing deeper insights into business performance.

Rapid Reporting and Dashboards

OLAP systems enable companies to quickly generate customized reports and interactive dashboards, empowering decision-makers with real-time data-driven insights.

Trend Identification and Forecasting

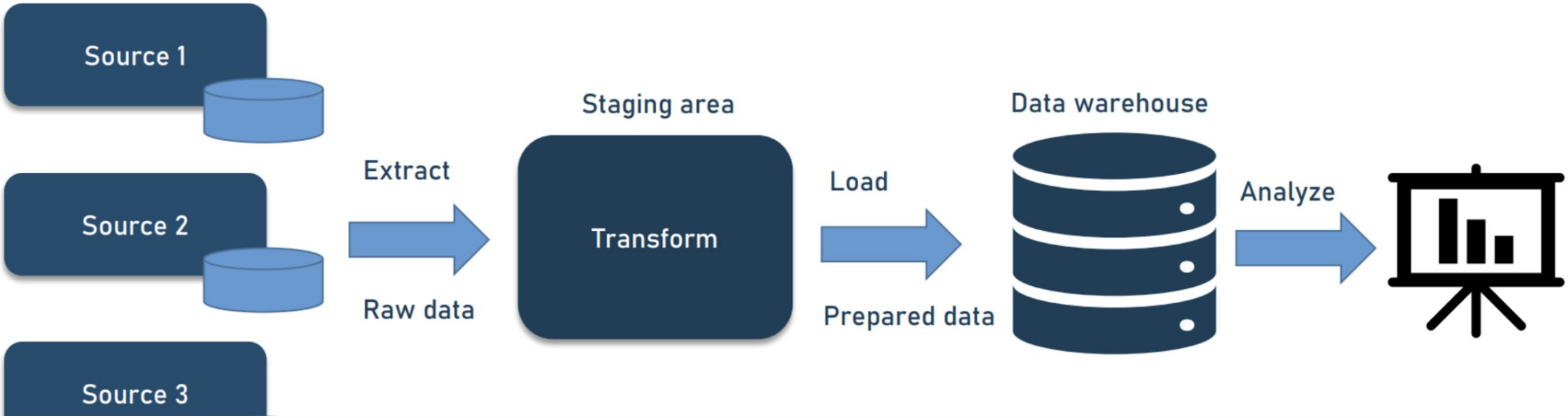
OLAP systems help companies identify patterns, trends, and anomalies in their data, allowing them to make more informed decisions and better predict future performance.

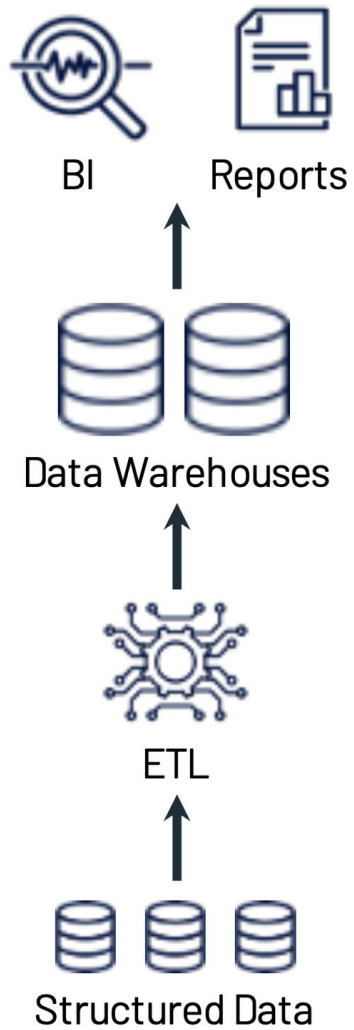
Improved Business Agility

OLAP systems provide companies with the flexibility to adapt to changing business requirements, enabling them to quickly access and analyze data from various sources.

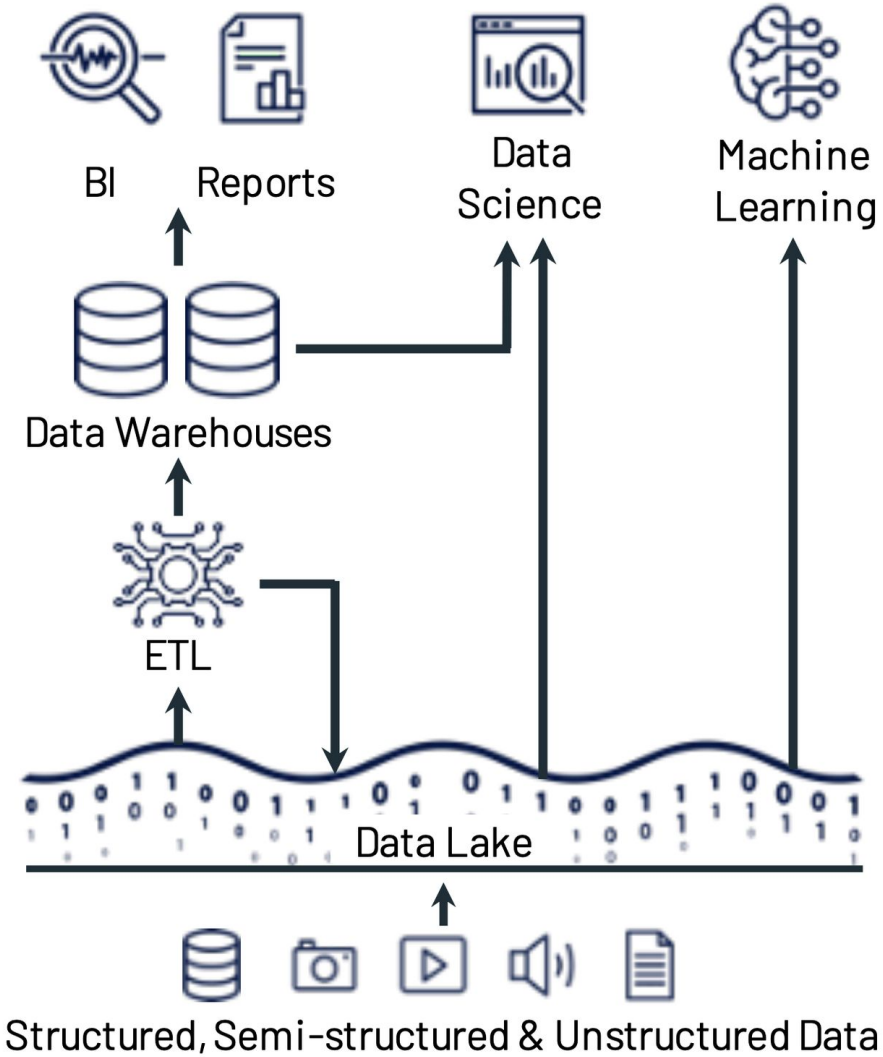
OLAP systems are a powerful tool for companies to gain deeper, more comprehensive insights into their business operations, enabling them to make more informed, data-driven decisions and stay competitive in their respective markets.

ETL PIPELINE

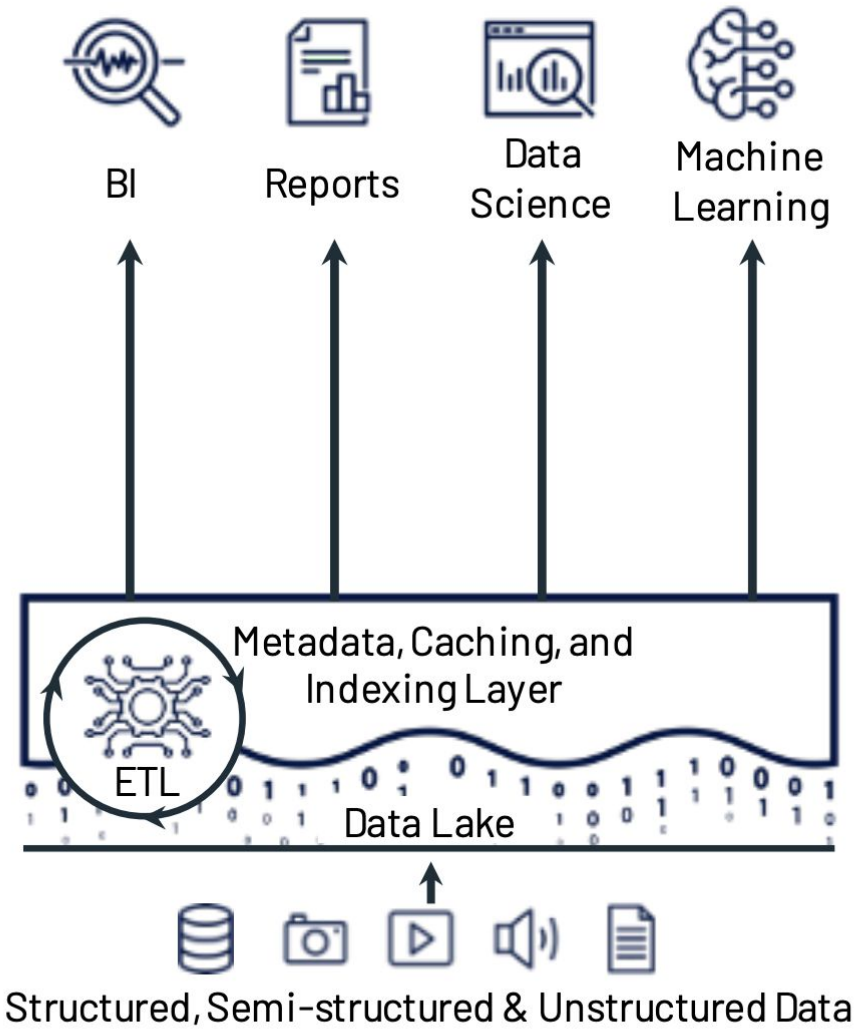




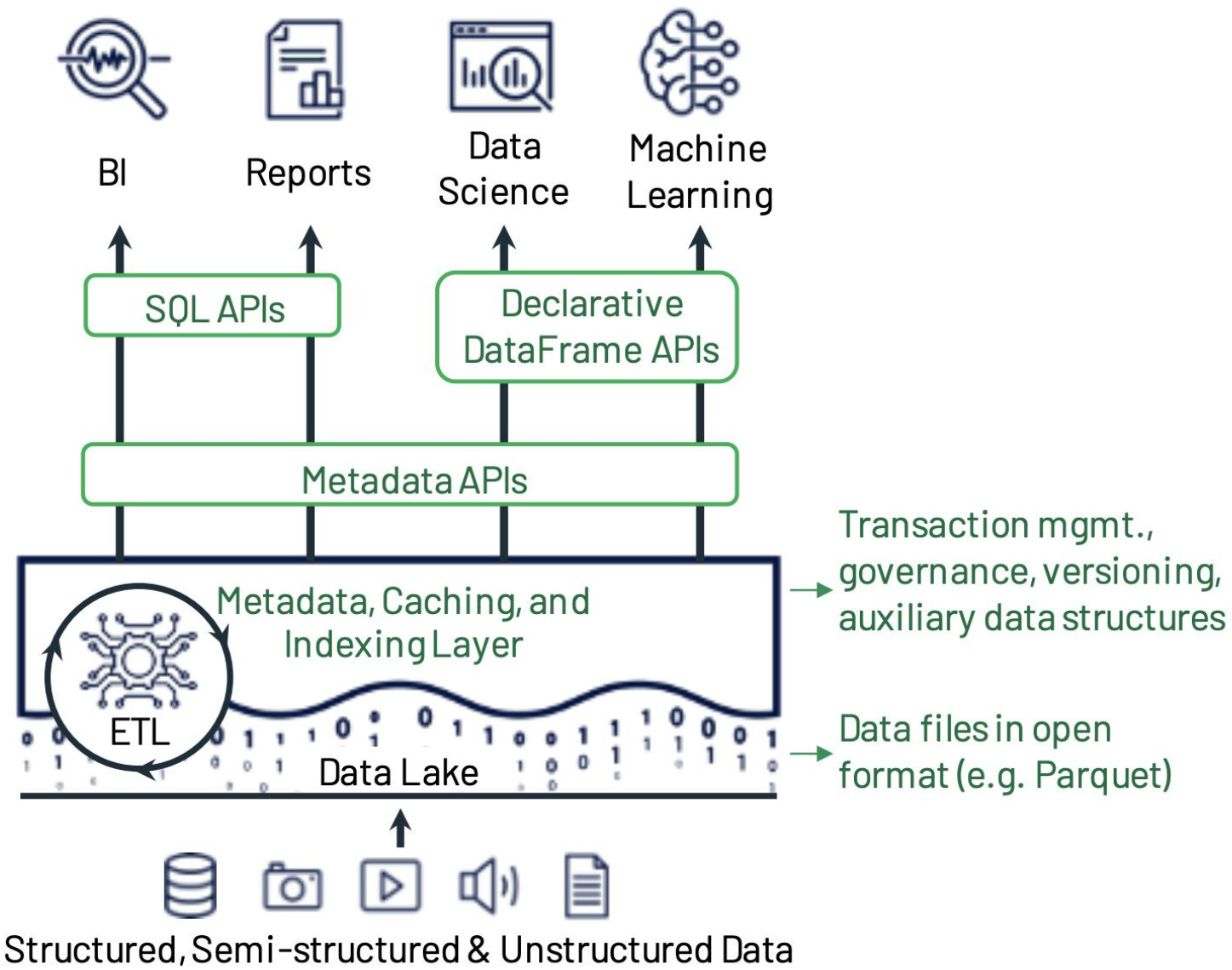
(a) First-generation platforms.



(b) Current two-tier architectures.



(c) Lakehouse platforms.


BI
Reports
Data Science
Machine Learning
SQL APIs
Declarative DataFrame APIs
Metadata APIs
Metadata, Caching, and Indexing Layer
ETL
Data Lake
Transaction mgmt., governance, versioning, auxiliary data structures
Data files in open format (e.g. Parquet)
Structured, Semi-structured & Unstructured Data

Lakehouses

Lakehouse Architecture

Lakehouse architecture integrates the data warehouse and data lake concepts, providing a unified platform for analytics and machine learning.

Metadata Layer Significance

The metadata layer is the crux of the lakehouse architecture, as it manages the data schema, partitioning, and other metadata to enable efficient querying and processing.

File Formats

Common file formats used in lakehouse architectures include Apache Iceberg, Hudi, and Delta Lake, which provide features like ACID transactions, time travel, and schema evolution.

The lakehouse architecture, with its emphasis on the metadata layer and use of advanced file formats, offers a powerful and flexible data platform for modern analytical and machine learning workloads.

In Product and Conversational Analytics

Increased Demand of In product Use Cases

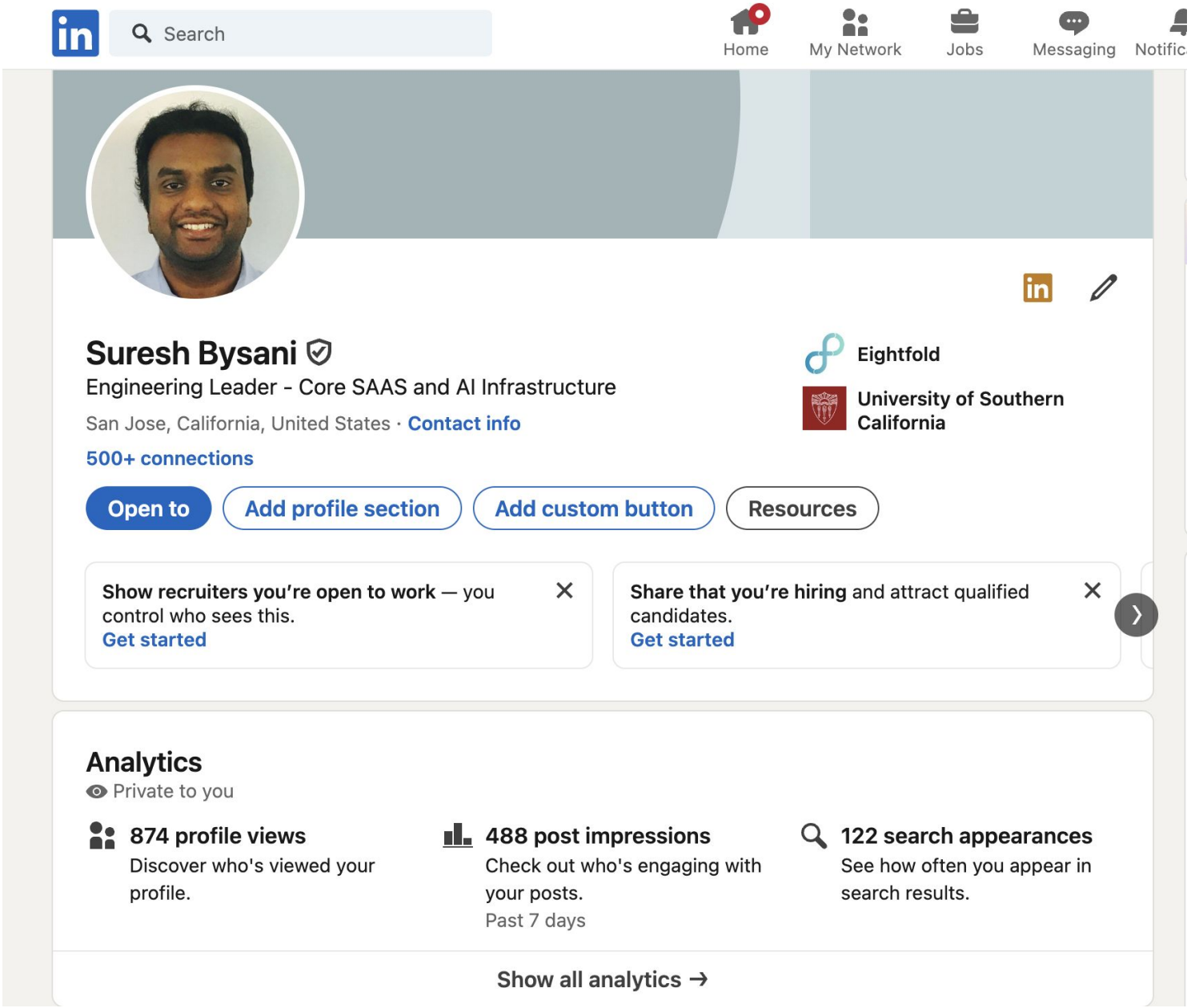
Customers need valuable analytics Directly In Products

AI agents have high data needs and high QPS needs

AI agents will make lot of analytics Queries.

AI agents will extract ton of data from Data Warehouses. This will increase QPS needs.

In Product Analytics



How to Scale Data Warehouses for Agentic QPS

Seperate Storage and Compute

Once they are seperated, compute becomes stateless entity.

Predicate Push Down

Filter Less amount of data from S3. IO is expensive.

Cache As much as you can

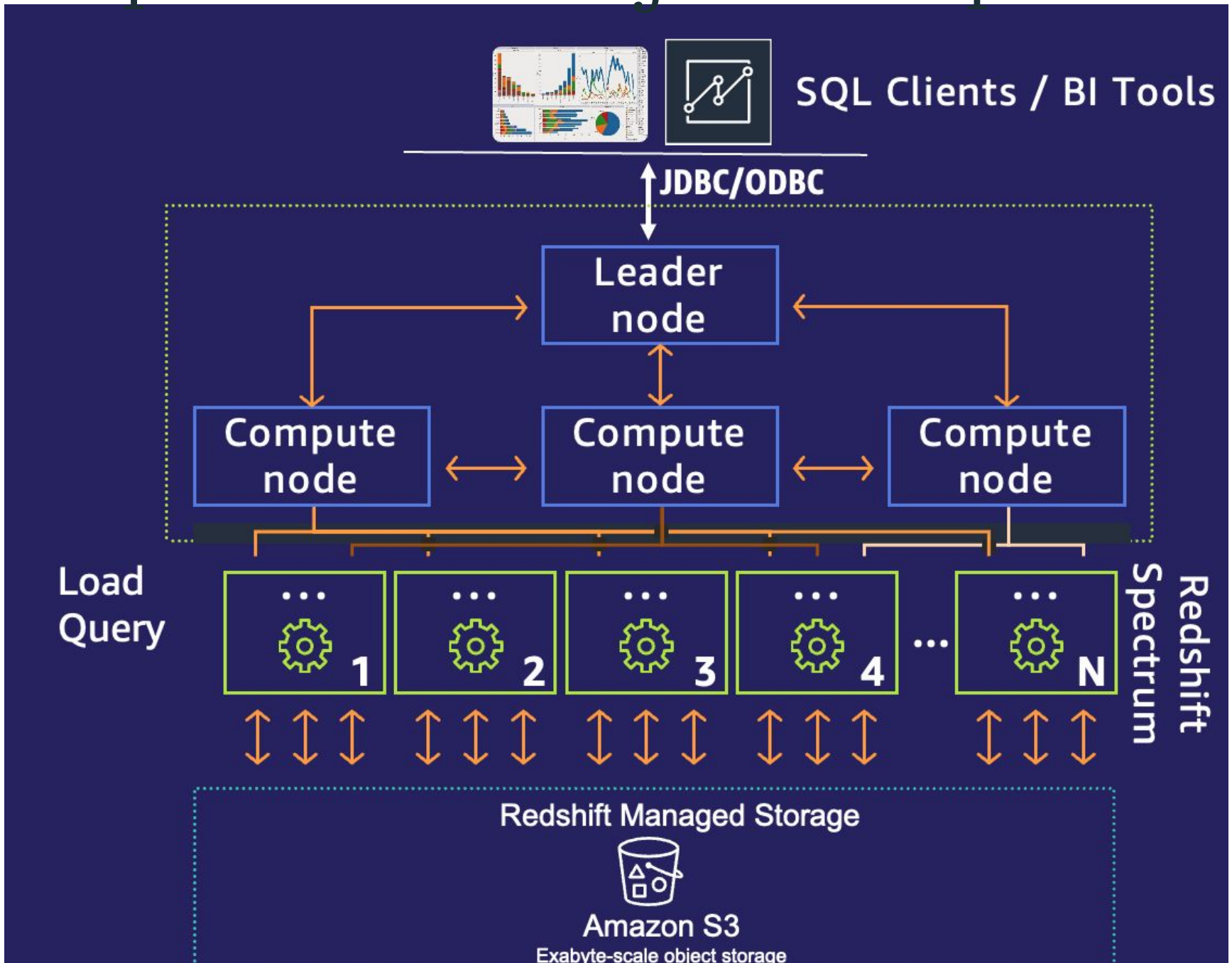
Store frequently used data on EBS

Do not have single leader bottlenecks

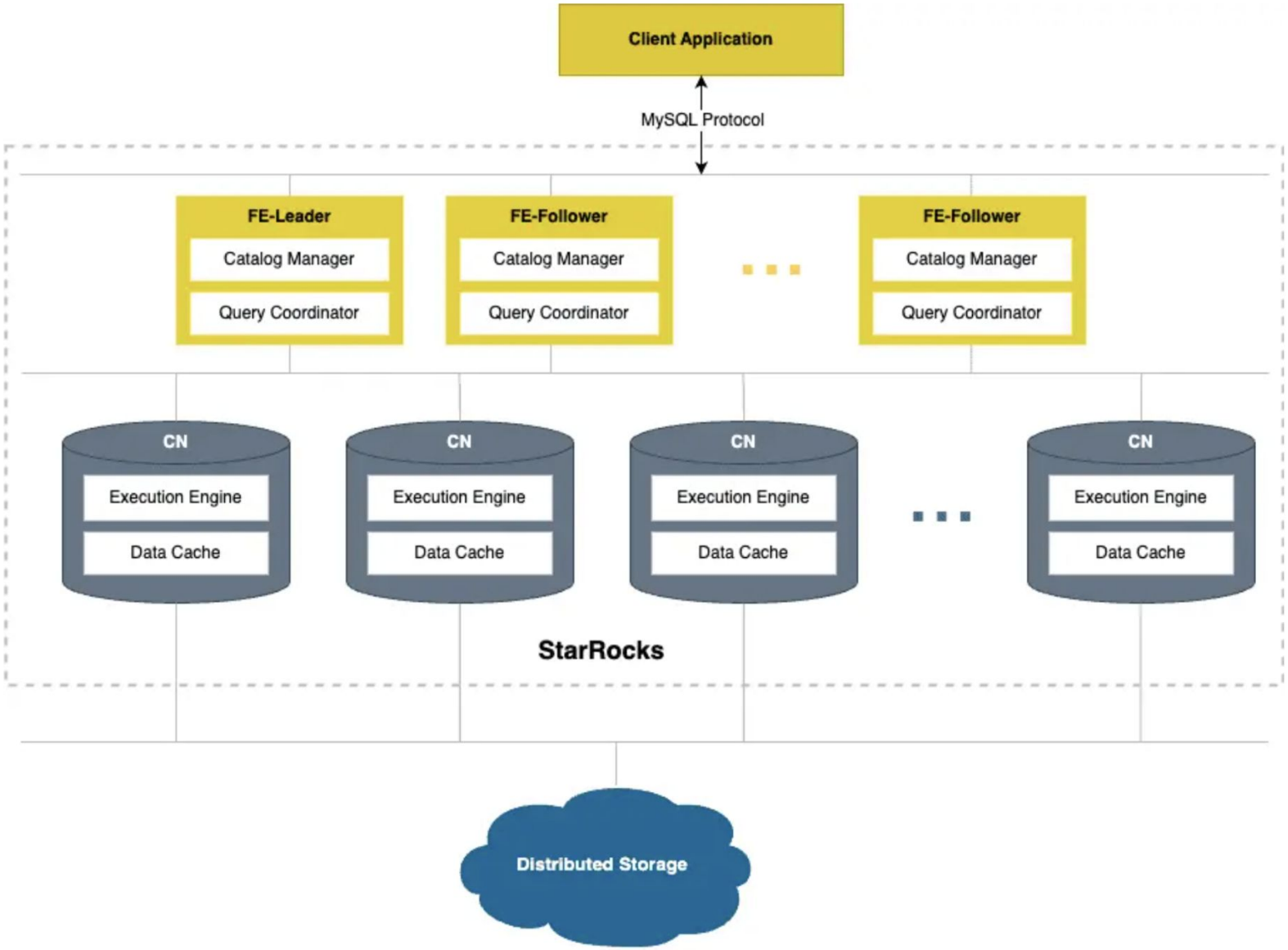
Single Leaders(masters) will bottleneck number of queries.

How can we scale Data Warehouses for AI agents and In Product Analytics

Separation of Storage and Compute - Leader



Additional EBS Cache



Thank You