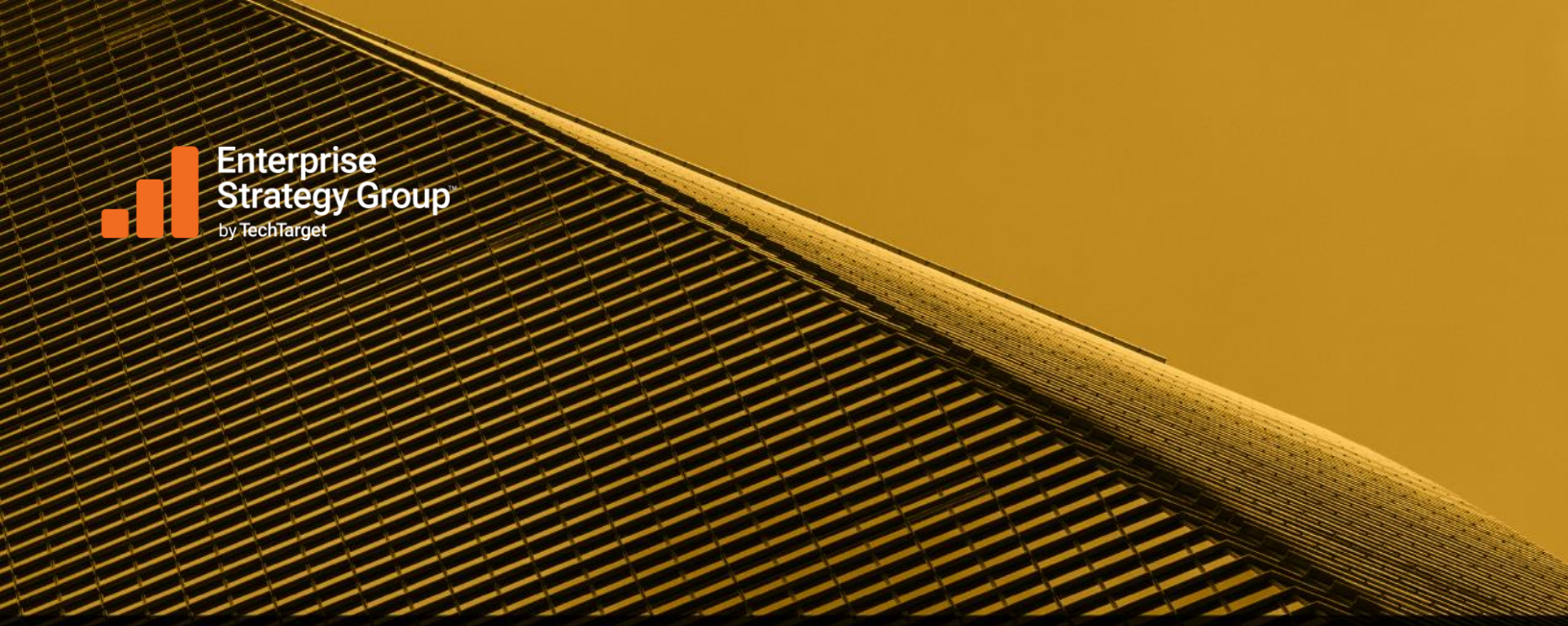


TECHNICAL VALIDATION

The Use of Couchbase for Developing and Supporting AI-powered Applications

By Alex Arcilla, Principal Analyst – Validation Services
Enterprise Strategy Group

January 2025

Contents

Introduction

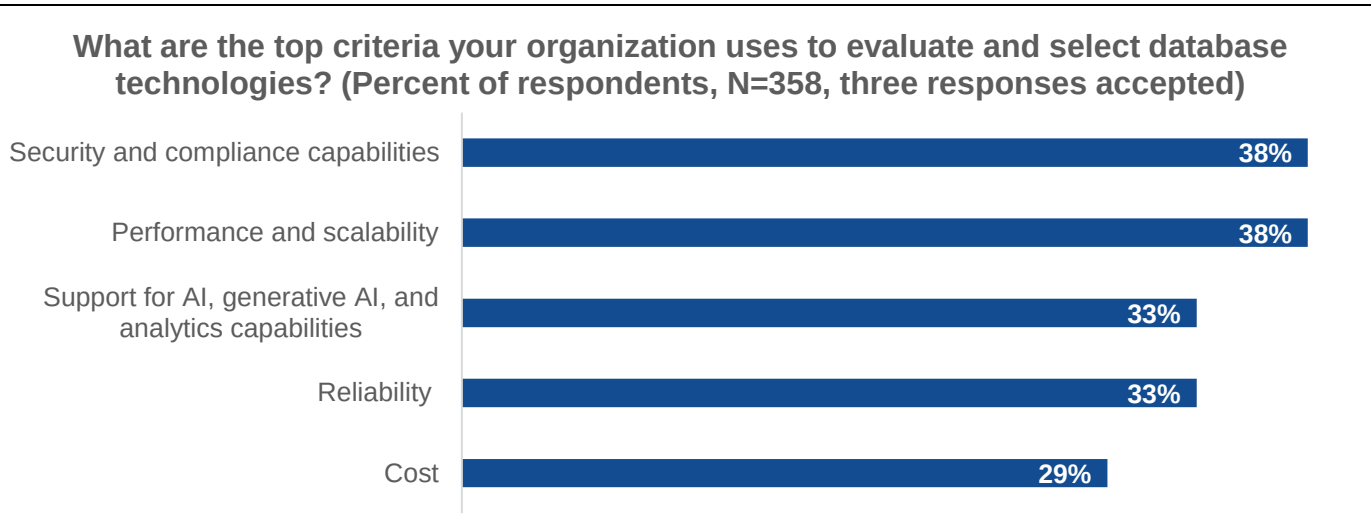
This Technical Validation from TechTarget's Enterprise Strategy Group documents our evaluation of Couchbase. We review how Couchbase can help organizations to develop and support applications that are increasingly relying on different data types that are distributed across multiple disparate environments and locations.

Background

Organizations deploying business-to-consumer applications seek to leverage data from multiple source points to provide personalized experiences and content. And these sources are not limited to traditional relational database management systems (RDBMS); organizations want to unlock value by combining, aggregating, and analyzing different types and formats of data—structured, unstructured, time-series, geospatial, relationship, vector, analytic, semantic, etc. Yet, recent research from Enterprise Strategy Group found that, when integrating data, organizations encounter multiple issues preventing them from gaining the insight desired. When asking about their top four challenges, 26% of respondents cited combining data from disparate silos/environments to be time-consuming, while another 25% cited combining data of different structures with a goal of analyzing together as another top challenge.[1]

As organizations increasingly build applications with data from disjointed sources simultaneously to provide differentiated customer experiences, having the right underlying database technology is key. This is true as more organizations are incorporating artificial intelligence technologies, such as generative AI (GenAI), into these applications to create additional competitive advantage. Indeed, this driver is affecting the criteria organizations are using to evaluate databases. Our research found that performance and scalability (38%); support for AI, GenAI, and analytics capabilities (33%); and cost (29%) are among their top five criteria (see Figure 1).[2]

Figure 1. Top Criteria for Evaluating and Selecting Database Technologies



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Organizations that want to leverage diverse data types cannot rely on traditional relational databases to organize the data while ensuring consistency, availability, and high data access performance. While relational databases maintain transactional data in a highly consistent manner, the underlying architecture supports features—

[1] Source: Enterprise Strategy Group Research Report: *The State of DataOps: Unleashing the Power of Data*, January 2024.
[2] Source: Enterprise Strategy Complete Survey Results, *Rethinking Database Requirements in the Age of AI*, November 2024.

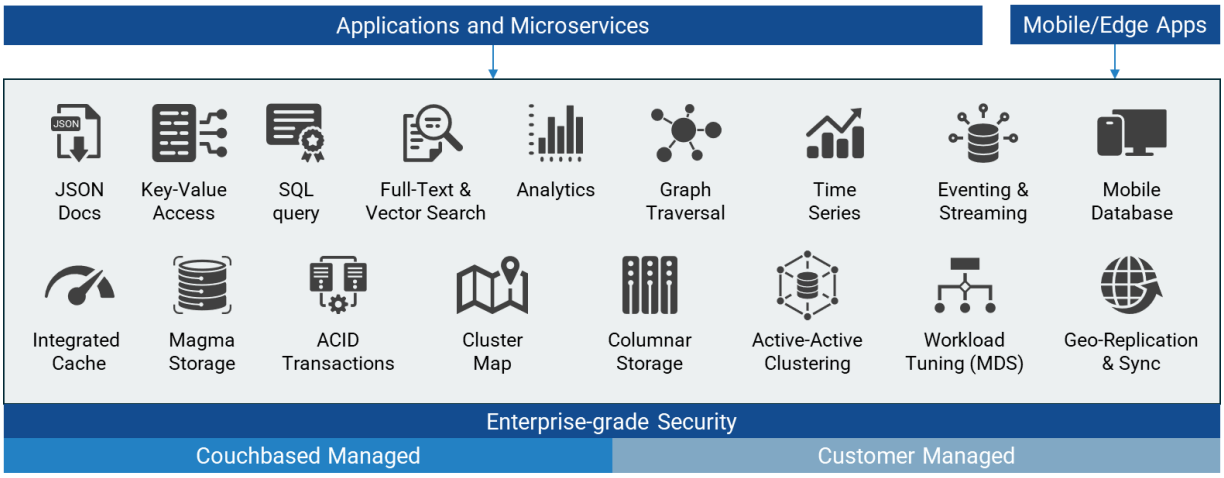
normalization of objects, adherence to fixed schema and data typing, single node transactional design—that are difficult to modify once deployed. Relational databases are not designed for other data types and formats. The architecture is also not designed to scale easily as applications and data become more distributed. Modifying how varied data types, aside from relational, are modeled consumes manual effort and requires more tools.

When faced with the task of combining different data types from multiple environments within their applications, organizations can invest in disparate database tools and technologies to achieve this goal. Yet, the investment in money and time to acquire the necessary skills to learn these tools and technologies quickly becomes prohibitive, especially as IT budgets remain constrained. Such challenges will be magnified as more organizations incorporate AI technologies into these applications.

Couchbase

Couchbase has been designed to be a distributed, multi-model, NoSQL, cloud-to-edge database with a flexible JSON data model that simplifies how data is stored, accessed, transacted, and updated when needed. This hybrid multipurpose database offers familiar relational and multi-model data access services so that organizations can supply data to a wide variety of operational and analytic applications. Couchbase consolidates multiple data access patterns (see top row of gray box in Figure 2) and performance optimizations (see bottom row of gray box in Figure 2) into a single platform that would otherwise require organizations to use and manage single-purpose databases simultaneously. This platform's integration eliminates the need for organizations to manage data models and consistency between multiple systems, learn different languages and vendor-specific application programming interfaces (APIs), and manage independent technologies.

Figure 2. Couchbase



Source: Couchbase and Enterprise Strategy Group, a division of TechTarget, Inc.

Used by numerous organizations today, Couchbase is delivered in two ways:

- **Couchbase Capella.** This database-as-a-service model is supported in Amazon Web Services, Microsoft Azure, and Google Cloud.
- **Couchbase Enterprise.** This model is consumed via subscription, enabling organizations to deploy Couchbase on-premises, in the public cloud, and at the edge.

The Couchbase architecture has been designed for:

- **Faster data access performance.** Through its memory-first design, Couchbase is designed to process, read, and write data in milliseconds and incorporates geo-replication and synchronization to ensure that applications

can access data when needed. The fast access performance is also supported with multidimensional scaling that enables a specific service (e.g., data, query, index, search, analytics, and eventing) to scale horizontally across nodes within a cluster. Scaling has zero impact, in terms of memory utilization, with other services. Together, these features deliver extremely low latency, resulting in faster application response times.

- **Versatility and flexibility.** Couchbase's flexible JSON model enables simpler and easier incorporation and normalization of new data, regardless of data type and format, without the need to redefine data schemas, formats, and relationships. Organizations can also add and modify a wide variety of data easily, without relying on third-party extract, transform, and load tools or vendor-specific APIs. Couchbase is also designed to support both desktop and mobile applications across on-premises, cloud, and edge environments.
- **Easier to learn and use.** Couchbase supports the easy-to-use query language SQL++ (SQL for JSON). Developers with SQL experience can query Couchbase using familiar syntax and behavior, with little time and effort spent on upskilling.
- **Less IT complexity.** Instead of using multiple database technologies and management tools, as well as proprietary development tools and APIs, to integrate data when accessed, organizations can instead leverage Couchbase's cost-effective integrated and simplified database design of multiple data access methods and data stores. Organizations incur lower capital, cloud, and operational expenses, as they do not need to deploy and operate multiple purpose-built databases. Also, because Couchbase is designed to auto scale compute resources, organizations do not need to manually add excess compute capacity to accommodate any anticipated increased use of data services. Existing compute capacity is used more efficiently, so overall costs decrease.
- **High availability.** To deliver consistent and low-latency application user experiences, Couchbase's geo-replication service automates replication across data centers to simplify data replication for global data distribution. To ensure that edge and mobile applications can still operate in lieu of reliable network connections, Couchbase's embedded local database and sync service provides local data processing, advanced offline-first syncing, and peer-to-peer syncing.

Couchbase is also designed to support mobile and edge applications so that data access performance is not compromised. Delivered as Couchbase Lite, this data can either be deployed locally as a fully searchable standalone database within mobile applications, or it can manage data sync automatically from cloud to edge when used with Sync Gateway and Couchbase Server. Using Couchbase's peer-to-peer synchronization service, Couchbase Lite clients can synchronize data without depending upon a central data repository in cases when, for example, mobile connectivity becomes unavailable.

For organizations looking to enhance their analytics applications with JSON data, Couchbase has incorporated Capella Columnar service into the platform. Organizations can now combine JSON data with enterprise data for enhancing real-time analytics. Once insights have been uncovered, this feedback, expressed via metrics, is fed back into the affected applications to enable a real-time feedback loop that drives immediate action when required. Insights can also be fed into AI prompts.

Couchbase has partnered with multiple AI frameworks, including Amazon Bedrock, LangChain, and LlamaIndex to provide complementary capabilities that enable organizations to speed up the development of context-aware AI-enabled applications. With LangChain, an open source framework for developing applications powered by large language models (LLMs), organizations have a common API interface for connecting with an extensive LLM library. Organizations using the LlamaIndex framework can leverage its tools to ingest, parse, index, and process enterprise data when building out context-aware LLM applications (e.g., retrieval-augmented generation, also known as RAG).

In addition to its integration within the AI ecosystem, Couchbase offers its own AI services to further boost developer productivity while decreasing overall IT environment complexity and operational and management expenses. These include the following:

- Vector search for retrieving relevant points within unstructured data based on vector embeddings. This is available for deployments in the cloud, on-premises, and on mobile devices in Couchbase Lite.
- Support for federated external models and hosting LLMs (open, embedding, guard, and prompt models) within Couchbase Capella.
- Local hosting of models from Mistral and Llama 3 projects that enable low-latency interactions between enterprise data and LLMs, without compromising data security.
- Vectorization service for automating data vectorization within Capella, as well as providing inference endpoints with external models from OpenAI, Amazon SageMaker, Google Gemini, Anthropic Claude, and others.
- Unstructured data services for preprocessing, chunking, and vectorization of unstructured data such as PDFs and images from enterprise data stores.
- Agent catalog service to help organizations accelerate development of callable and reusable AI agents via a central catalog.

Enterprise Strategy Group Technical Validation

Enterprise Strategy Group validated how organizations can benefit from the Couchbase architecture by evaluating how this multimodel database supports easier data access and high performance for business-to-consumer applications, regardless of where they are deployed.

Higher Performance to Achieve Faster Application Response Times

Of all the factors that can affect application response times (e.g., database location), database performance plays a significant role. Couchbase has been designed to deliver fast data access performance. This becomes especially important when more organizations incorporate AI into their applications, as massive amounts of data must be accessed, captured, and validated first before each response is generated.

Enterprise Strategy Group Testing

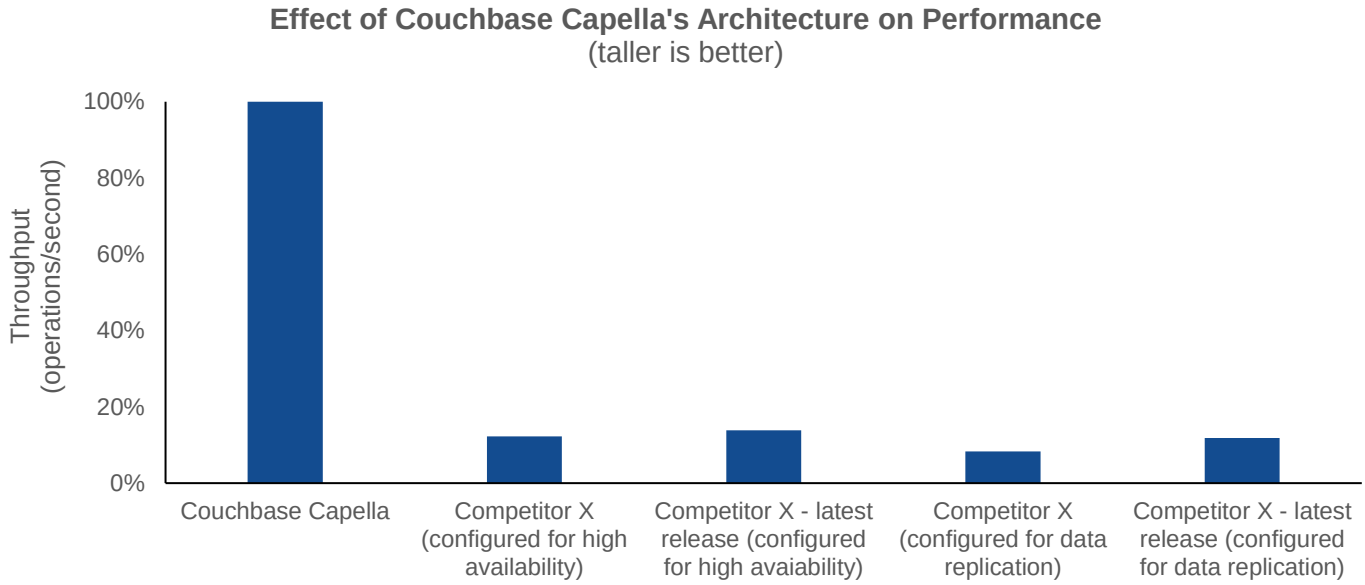
Enterprise Strategy Group evaluated test results comparing the performance of Couchbase against a leading competitor. We utilized the Yahoo Cloud Serving Benchmark (YCSB), an open source specification and program suite for evaluating retrieval and maintenance capabilities of computer programs, to measure performance. YCSB has been commonly used to measure the relative performance of NoSQL databases.

These tests compared Couchbase Capella with two of the most recent stable versions of the competitive offering.[3] Tests were conducted using three-node clusters from Couchbase and the competitive offering deployed in the AWS US-East-1 region. (All node configuration details are in the Appendix.) Testing employed YCSB Benchmark Test A using the 50% read/50% read-modify-write workload.

We evaluated how Couchbase Capella's architecture affects performance—specifically how the architecture achieves high availability (via allowing reads and writes to and from all nodes) and automatically distributes data across nodes (via automatic sharding). Clusters of the competitive offering were configured to closely mimic these two architectural features. Results shown in Figure 3 revealed that, in both test scenarios, Couchbase Capella's performance consistently exceeded that of both releases of the competitive offering.

[3] The competitor announced the newest version shortly before the release of the report and specifically touted performance improvements. Tests included this version to ensure that performance differences were noted appropriately.

Figure 3. How Couchbase Capella's Architecture Affects Performance



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Given these results, Enterprise Strategy Group noted how Couchbase Capella's performance can aid organizations in decreasing operational costs. Organizations can leverage the faster data access performance to actually decrease the overall number of clusters provisioned and managed. Multidimensional scaling further decreases operational costs, as utilization of existing resources can be maximized without adding nodes. This becomes especially important as organizations deploy AI-enabled applications that use continuously increasing amounts of data via LLMs.

Why This Matters

As organizations ramp up their capabilities to deploy AI-enabled applications, the choice of database becomes key, as performance becomes a critical consideration. In fact, Enterprise Strategy Group found that 92% of respondents agreed or strongly agreed that database performance is critical to the success of their GenAI projects.[4]

Enterprise Strategy Group validated that Couchbase does indeed deliver higher performance than a leading competitor. After reviewing the test methods and results defined by YCSB, an open source specification and program suite for evaluating database performance, we found that Couchbase exhibited as much as 12x the performance of the competitor.

Simplified Data Modeling and Schema Migrations

Ensuring that data is consumable and up to date is no easy feat when dealing with multiple data stores and databases. Combining multiple data types from a wide range of data sources is manually complex, as multiple tool sets and APIs are typically needed to present data in a common and usable format. Furthermore, updates and changes to data are not necessarily reflected immediately, as experienced traditionally in an RDBMS. With

[4] Source: Enterprise Strategy Complete Survey Results, *Rethinking Database Requirements in the Age of AI*, November 2024.

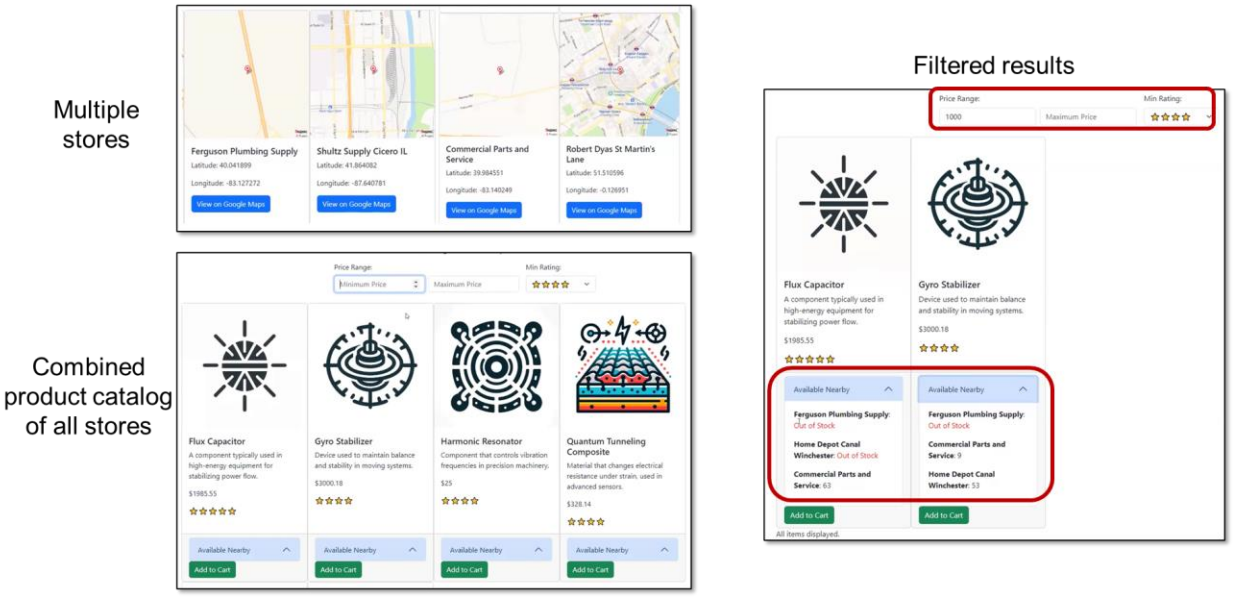
Couchbase's support of versatile JSON documents and schema, combined, updated, and modified data from any source for use by transactional and analytics applications is greatly simplified.

Enterprise Strategy Group Testing

Enterprise Strategy Group observed how different data types were combined to power an e-commerce application. Specifically, we saw how Couchbase can support a SQL++ hybrid search for data, combining multiple search methods within a single query.

Our application was designed to browse product catalogs from multiple stores (see left of Figure 4). From the combined catalog, we queried products with a minimum price of $1,000 and a minimum rating of four stars. Results displayed the stores carrying products fitting the criteria, along with the available inventory.

Figure 4. Web Application Enabling Product Search of Multiple Stores



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

This search was completed using a single SQL++ query, as shown in Figure 5. To browse the combined product catalog spanning multiple stores, the query combined a geospatial search to locate the nearest stores to the location of the user and a standard SQL search for selecting the nearby stores with the queried product, based on the price and rating parameters.

Since the application was supported with Couchbase, multiple and separate queries using language compatible with the different data stores (containing transactional and geospatial data) did not have to be written, as SQL++ was the common language used. The different data types did not require normalization so that they could be combined to produce the search result.

Figure 5. Single Query Combining Standard SQL and Geospatial Search

Browse catalog (paged)

Geospatial search

```
WITH closestStores AS (
    SELECT x.name, META(x).id AS id
    FROM whatisthis._default.Stores x
    WHERE SEARCH(x, {
     "fields": ["*"],
     "query" : {
       "location" : {
         "lat" : 39.8666212,
         "lon" : -83.0683414
       },
       "distance" : "15mi",
       "field" : "geo"
     },
     "sort": [
      {
       "by": "geo_distance",
       "field": "geo",
       "unit": "mi",
       "location" : {
         "lat" : 39.8666212,
         "lon" : -83.0683414
```

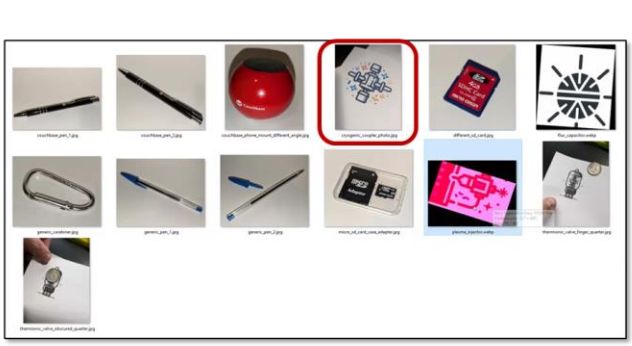
Standard SQL search

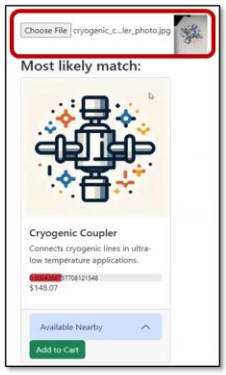
```
SELECT t1.name, t1.`desc`, t1.image, t1.price, t1.rating,
    ARRAY_AGG({
        "storeName": store1.name,
        "quantity": stock1.numInStock
    }) AS Stock
FROM whatisthis._default.Items AS t1
LEFT JOIN whatisthis._default.Stock AS stock1
    ON SPLIT(META(stock1).id, "::")[1] = META(t1).id
JOIN closestStores AS store1
    ON META(stock1).id LIKE store1.id || '%'
WHERE 1==1
     AND t1.price >= 1000            /* min price filter (if any) */
             /* max price filter (if any) */
     AND t1.rating >= 4          /* min rating filter (if any) */
GROUP BY t1.name, t1.`desc`, t1.image, t1.price, t1.rating
ORDER BY t1.name
LIMIT 6
OFFSET 0
```

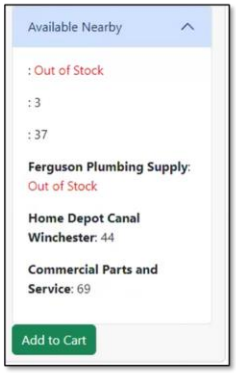
Source: Enterprise Strategy Group, a division of TechTarget, Inc.

We then observed how Couchbase enables vector searches, such as an image search. An example of this type of search is looking for an item without knowing what exactly it is. From a local computer, a picture was uploaded to the product search application (see top left of Figure 6).

Figure 5. Web Application Enabling Product Search of Multiple Stores







SQL++ code for vector search

```
WHERE SEARCH(t1,                    /* vector search using image embedding */
  {
    "fields": ["*"],
    "query": {
      "match_none": ""
    },
    "knn": [
      {
        "k": 4,
        "field": "imageVector",
        "vector": [ 0.93066406,0.1352539,-0.92626953,-1.7304688,2.1621094,0.39990234,0.5175781,-0.6777344,0.6640625,-2.6113281,2.746
      }
    ]
  }
)
GROUP BY t1.name, t1.`desc`, t1.image, t1.price, SEARCH_SCORE(t1)
ORDER BY score DESC
```

Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Once the picture was uploaded, we saw that the application found the most likely match. Scores were also generated to help produce the most likely matches.

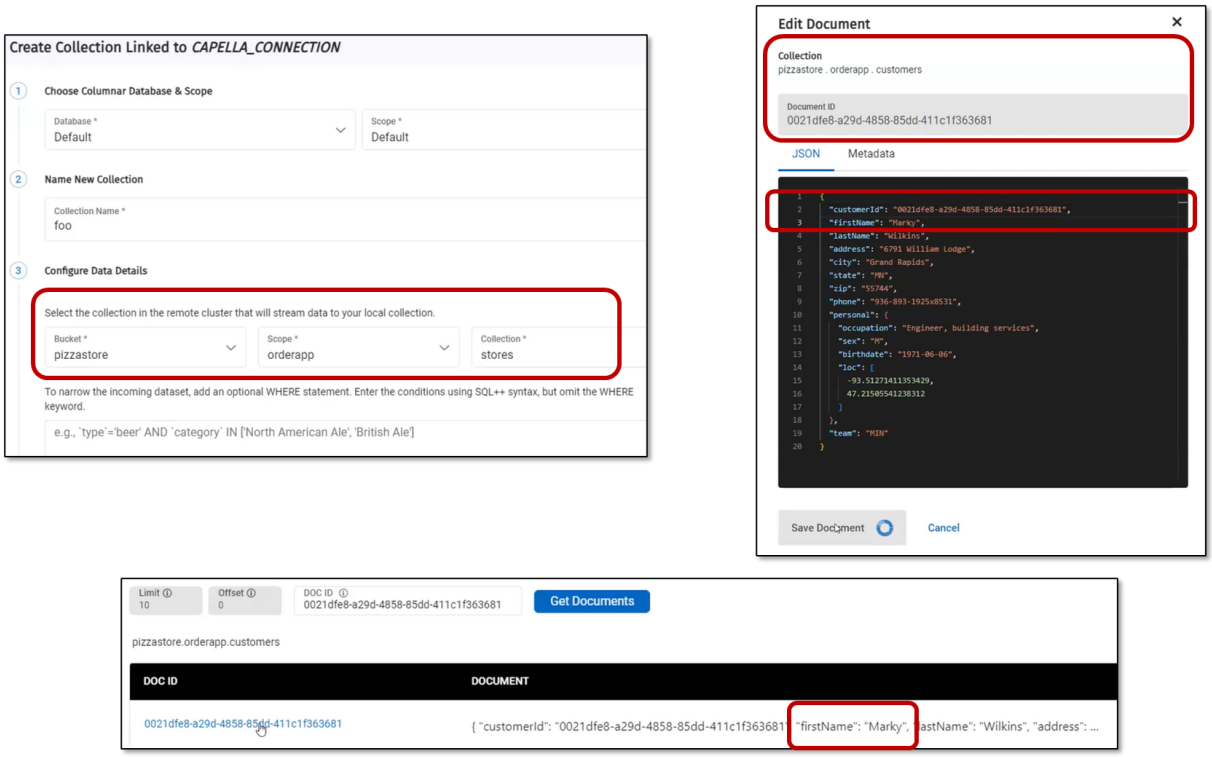
To provide a more complete product search experience, additional SQL++ language was added to the previous geospatial and standard queries (Figure 5) to locate nearby stores that sold the item and have available inventory (see bottom of Figure 6).

Finally, Enterprise Strategy Group reviewed how quickly organizations can update, aggregate, and use data immediately with Couchbase Columnar. We used a database supporting an application to order pizza and saw how we could change customer data simply by modifying the SQL++ code associated with the JSON document.

In this example, we chose a JSON document associated with a customer record. The record was located in the "pizzastore" bucket, containing the "stores" collection within the "orderapp" scope. (In RDBMS terms, the bucket is the database, the scope is the schema, and the collection is the table.) After selecting the customer record according to the JSON document ID, we navigated to the related SQL++ code and changed the "firstName" field from "Mark" to "Marky." Once the change was saved, we navigated back to the JSON document and found that the change was recorded immediately (see Figure 7).

We noted that no additional SQL-like statements were needed to update or modify existing data within the database, saving both time and resources. In addition, any new transactions or analytics conducted on this new data would be reflected immediately, reducing any business errors.

Figure 7. Updating Existing Customer Record in Real Time



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Why This Matters

Combining multiple data types from a wide variety of sources is no small feat. In fact, Enterprise Strategy Group research found that 33% of respondents cited the difficulty in integrating data from diverse sources and diverse data as one of their top challenges in their data management environment.[5] The use of multiple tools and APIs to combine data into a consumable format, as well as updating and modifying data, becomes too complex and time-consuming, especially as organizations seek to make more data accessible to its applications.

Enterprise Strategy Group validated that Couchbase can help organizations to greatly simplify how data is combined, updated, and modified to ensure that the latest data is consumable and available for use by transactional and analytics applications. We observed how the use of JSON documents enables multiple searches, across different data types, within a single SQL++ query to deliver search results that can enhance the end-user experience. We also saw that updating and modifying existing data with JSON documents and SQL++ can be accomplished immediately, thus enabling business applications to deliver relevant and timely results based on the most up-to-date data.

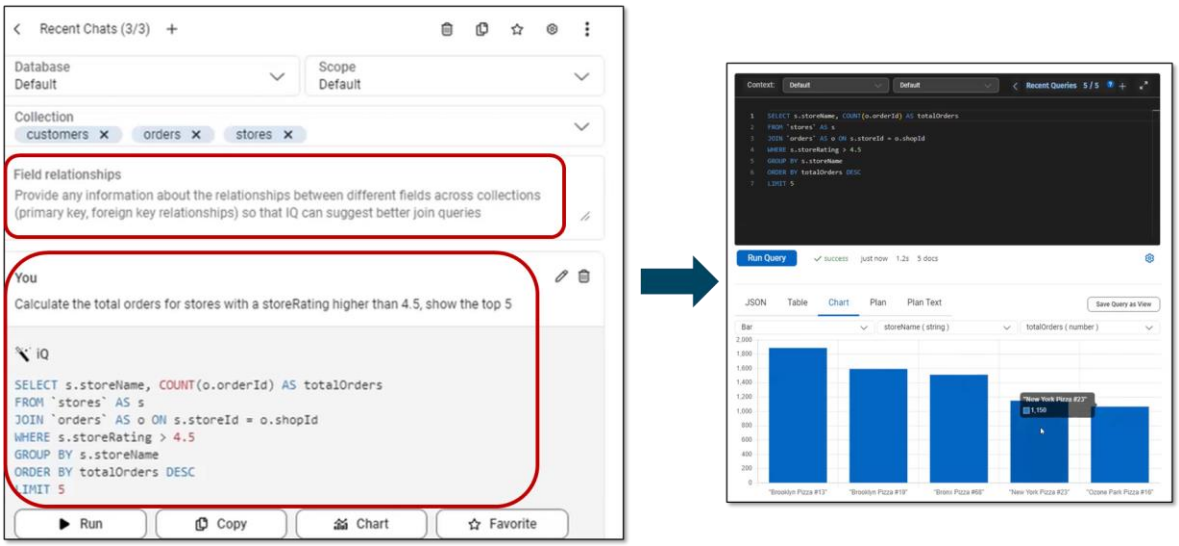
Increased Coding Efficiency With Capella IQ

To support the development of complex queries (as seen in previous sections), Couchbase offers Capella IQ, an AI-enabled coding assistant. Capella IQ can especially help organizations close any existing skill gaps when using SQL++.

Enterprise Strategy Group Testing

In the Capella IQ module, Enterprise Strategy Group entered a query using plain text. The query was intended to calculate the total orders for stores with a store rating higher than 4.5, then show the top five results (see left of Figure 8). We also selected the database and collections on which the query would run. Results were generated as desired (see right of Figure 8).

Figure 8. Generating SQL++ Query With Capella IQ



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

[5] Source: Enterprise Strategy Group Research Report: *The State of DataOps: Unleashing the Power of Data*, January 2024.

With Capella IQ, we saw how this could help in bringing end users up to speed on the SQL++ language, should they still be learning how to write queries. Capella IQ could also assist in writing complex queries conducting multiple search types (e.g., transactional, geospatial, vector, time-series, etc.), minimizing the overall time for deploying applications dependent on those queries.

Why This Matters

Enterprise Strategy Group research found that the most-cited database investment priority is to integrate AI and machine learning capabilities.[6] Not only can database technologies integrate AI to support AI-enabled applications once deployed, but they can also help increase overall productivity in the development of these applications.

Enterprise Strategy Group validated that Couchbase Capella IQ can help organizations in developing SQL++ queries while helping end users to increase their existing skill sets. We observed how we could enter complex queries in plain text, which was then converted into SQL++ language that could run immediately and produce the desired result.

[6] Source: Enterprise Strategy Complete Survey Results, *Rethinking Database Requirements in the Age of AI*, November 2024.

Conclusion

Organizations continuously attempt to reach and retain customers via engaging customer experiences delivered through applications. With the emergence of AI, the opportunity to deliver a better and more timely experience is within reach. Delivering these experiences requires the combination, search, and analysis of volumes of data, spanning multiple data types located within multiple, yet siloed, databases and data stores. Furthermore, using purpose-built databases to access this data becomes a bottleneck. The tools and workflows required to access that diverse data and then combine it into a consumable format that can be easily augmented, updated, or modified lead to unnecessary and costly complexity. Such complexity delays how quickly applications can access the data required to deliver timely and relevant responses demanded by end users.

Couchbase has been designed to address the complexity challenges that organizations face when needing to manage and combine multiple data types into a consumable and easily accessible format for transactional, analytical, mobile, (agentic) AI, and analytics applications. The hybrid database consolidates multiple data access patterns and services into a single platform, helping organizations to reduce complexity and the associated costs of investing and operating multiple purpose-built databases simultaneously. To support the development of AI-enabled applications, Couchbase has been adding multiple features (e.g., vector search, AI services, support for local and federated LLMs) as well as forging partnerships within the broader AI ecosystem. With this extended support, organizations can increase developer productivity, reduce operating costs, and decrease time to market.

Enterprise Strategy Group validated that Couchbase can help organizations deliver highly responsive applications while decreasing overall operational and management complexity and costs within their IT environment. Specifically, we found that Couchbase:

- Exceeds the performance of a leading competitor by as much as 1,200%.
- Simplifies how data is combined and accessed via the use of JSON documents and a common query language (SQL++), eliminating the need to use vendor-specific tools and APIs to integrate data into an easily consumable format.
- Facilitates developer productivity and upleveling of skill sets via Capella IQ, Couchbase's AI coding assistant.

Enterprise Strategy Group research revealed that 63% of organizations are already using hybrid databases, with another 24% planning to invest in this database type over the next 12-24 months.[7] As organizations ramp up their deployment of AI-enabled applications, the need to combine, access, and manage larger amounts of diverse data, without negatively affecting performance or injecting yet more complexity into the IT environment (and subsequently incurring more cost), will become more challenging. Based on our evaluation, we strongly suggest looking closely at Couchbase to address these challenges effectively.

[7] Ibid.

Appendix

Table 1. Configuration Details for Clusters Used in Performance Testing

Couchbase Capella	Competitor X (high availability)	Competitor X – latest release (high availability)	Competitor X (data replication)	Competitor X – latest release (data replication)
Three nodes	Three nodes	Three nodes	12 nodes	Three nodes
Eight vCPU/node	Eight vCPU/node	Eight vCPU/node	Eight vCPU/node	Eight vCPU/node
32 GB RAM/node	32 GB RAM/node	32 GB RAM/node	32 GB RAM/node	32 GB RAM/node
VPC peering	VPC peering	VPC peering	VPC peering	VPC peering
256 GB data set	256 GB data set	256 GB data set	256 GB data set	256 GB data set
6000 IOPS	6000 IOPS	6000 IOPS	6000 IOPS	6000 IOPS
	Replica set	Replica set	Three shards	One shard

Source: Couchbase and Enterprise Strategy Group, a division of TechTarget, Inc.

About Enterprise Strategy Group
TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

contact@esg-global.com
www.esg-global.com