Key Benefits

Operate at any scale.
Develop on a node,
deploy to a cluster.

Topology transparency.
Change your cluster,
not your code.

Flexible scalability.
Scale data management or
analytics independently.

Integrated cache.
Consistent, low latency
data operations.

Rich, optimized indexing.
Accelerate data read &
write operations.

Built-in data streaming.
Process data in real time.

High availability.
Operate on RDDs with
confidence, built-in
failover and recovery.

Millions of operations.
Enable web, IoT, and
mobile applications.

Strong integration.
Transparently use RDDs,
SQL and DStreams

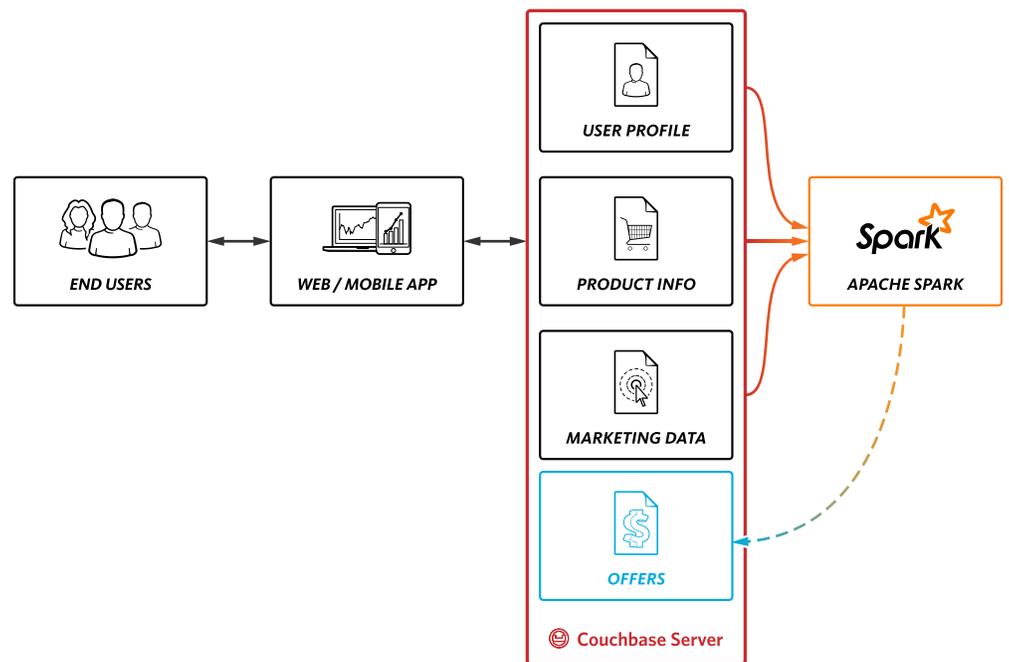**Data efficiency.**
Avoid cross-cluster
data transmission.

Schema Inference.
Automate schema
information generation.

Couchbase Spark Connector
In-memory Operational Analytics

Modern web, IoT and mobile applications generate tremendous volumes of data. The real value of the data is in the business decisions it enables and customer interactions that it can drive. Reducing the time it takes to go from raw data to usable information is crucial to leveraging its competitive value. Together, Couchbase and Apache Spark minimize the time-to-action that drives automated and user-driven applications.

Couchbase Server is an enterprise operational NoSQL database for modern web, mobile and IoT applications. It is a distributed, memory-first key-value and JSON document database with integrated caching, SQL-based querying and indexing, replication and data streaming. Couchbase Server is built for developing applications with agility, while operating at any scale.

Apache Spark is an open source, distributed, in-memory data processing platform built to provide low latency analytics. Apache Spark enables data analysts to easily build rich analytics, data processing and machine learning applications using general purpose, in-memory Resilient Distributed Datasets (RDDs). Apache Spark applications read and write RDDs to bring in the source data for processing and to write out the processed data results, interacting with a data store at both ends of the process.

Data + Analytics == Actionable Results

Why Couchbase and Spark?

The Couchbase Spark Connector provides best-of-breed technology integration, leveraging a common Develop with Agility and Operate at Any Scale architecture. It provides the most complete NoSQL and Spark integration available anywhere.

Couchbase and Spark can be scaled out together or separately, depending on application requirements. The Couchbase Spark Connector enables fast, scalable, reliable data access AND analytical data processing.

	Couchbase Spark Connector
RDDs Create and Persist	✅
DataFrames and Datasets Create and Persist	✅
Streams DStream input & output	✅
SQL Spark SQL, with optimized predicate pushdown	✅
Schema Automatic schema inference for JSON documents	✅
SQL Raw N1QL	✅
Minimize data shuffles Couchbase & Spark node location intelligence	✅
Optimized Indexing Use indexes to filter data, improve performance	✅
Languages Scala & Java APIs, Maven & SBT support	✅
Commercial Support With Couchbase Enterprise Edition	✅

The most complete Spark / NoSQL Integration available.

Typical Use Cases

Real time product recommendations. eCommerce applications suggest products to customers using proprietary, complex analytics that trigger recommendations, promotions or discounts.

Failure detection. IoT sensor-driven applications evaluate and predict component failures using complex machine learning algorithms and streaming analytics on real time data, which in turn can generate corrective actions, notifications and alerts.

Network intrusion detection. Network monitoring systems detect potential security threats using ETL and analytics over metadata and historical access logs, subsequently taking immediate protective action.

Fraud detection. Fraud detection applications compare questionable transactions to known statistical models using analytics and machine learning, and rating them.

Product and Customer 360. Comprehensive product and customer management tools aggregate data from multiple sources into a single unified repository, thereby providing a 360 degree view not previously available from the source repositories.

About Couchbase

Couchbase delivers the world's highest performing NoSQL distributed database platform. Developers around the world use the Couchbase platform to build enterprise web, mobile, and IoT applications that support massive data volumes in real time. The Couchbase platform includes Couchbase Server, Couchbase Lite - the first mobile NoSQL database, and Couchbase Sync Gateway. Couchbase is designed for global deployments, with configurable cross datacenter replication to increase data locality and availability. All Couchbase products are open source projects. Couchbase customers include industry leaders like AOL, AT&T, Bally's, Beats Music, BSkyB, Cisco, Comcast, Concur, Disney, eBay, KDDI, Nordstorm, Neiman Marcus, Orbitz, PayPal, Rakuten / Viber, Tencent, Verizon, Wells Fargo, Willis Group, as well as hundreds of other household names. Couchbase investors include Accel Partners, Adams Street Partners, Ignition Partners, Mayfield Fund, North Bridge Venture Partners, and West Summit.