Datasheet

Kognitio on Apache® Hadoop™
Ultra-fast, high concurrency SQL engine

The Kognitio Analytical Platform is the world's fastest in-memory data analysis engine and when deployed on Hadoop it allows organizations to connect interactive, self-service analytical tools to Hadoop-based data for thousands of end users even when the data volumes are very large.

Hadoop and its associated software projects have provided organizations with a cost-effective way of storing and processing vast amounts of diverse data and a range of tools that allow that data to be captured analyzed and to some degree, consumed.

However, mass consumption of Hadoop based datasets by large numbers of users in a business context, is still difficult, as the tools available do not possess the performance levels or enterprise capabilities needed to reliably support hundreds of concurrent users.

What's required is a SQL engine that will allow a business to connect the tools that your wider business users prefer to use and which generate SQL, e.g. Tableau, to your Hadoop-based data and perform interactive train-of-thought querying for hundreds of concurrent users.

The engines that come as part of the main Hadoop distributions (e.g. Hive, Hive LLAP, Impala) are not as performant as more mature SQL engines like Kognitio. To read more about benchmark tests of the various SQL engines on offer, visit kognitio.com/blog/tag/tpc-ds/.

The architecture

"Kognitio's high-speed processing means that we can offer more and more real-time queries through our customer portal. For example, it allows us to offer very detailed targeting that otherwise wouldn't be possible in a reasonable amount of time."

Andrew Wood, Director of Advanced Analytics at Inmar
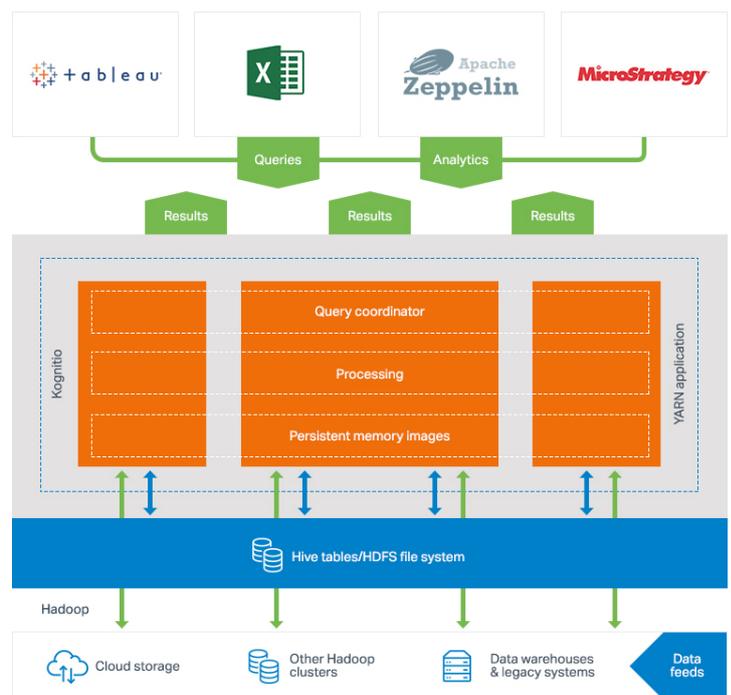
Kognitio's analytical platform on Hadoop

On Hadoop, Kognitio runs as a well-behaved YARN application, allowing Kognitio to share hardware infrastructure with other Hadoop applications.

It can support thousands of analytical queries per second from hundreds of concurrent sessions and can be scaled out to accommodate both larger data volumes and more query throughput.

In addition to the ultra-fast SQL Kognitio also has sophisticated NoSQL support that allows advanced analytical algorithms to be executed at scale and in almost any language eg R, Python.

How does Kognitio work on Hadoop?

✓ The tried and tested Kognitio MPP SQL database has been enhanced to allow it to run as a YARN application on Hadoop.

✓ Data is held in memory structures highly optimized for in-memory analysis – this is not a disk cache!

✓ The Kognitio architecture has been shared nothing for over 20 years and fits seamlessly into the Hadoop / YARN shared nothing paradigm to provide a platform that can be scaled across the largest of Hadoop clusters.

✓ Data can be explicitly pinned in memory - you know your expected usage patterns better than any optimizer.

✓ Kognitio's Massively Parallel Processing technology is used to localize processing and data into logical units known as ramstores to maximize efficiency.

✓ Sophisticated query planning allows queries on large data sets to be distributed to use all available CPU resource to process the query with maximum performance.

✓ Processor efficiency is further maximized by the use of sophisticated techniques to maximize processor instruction cache usage. These techniques include machine code generation and advanced query plan optimization.

✓ Granular queries that access smaller subsets of data are optimized by the query planner to access only the ramstores containing the data required to fulfil the request providing extremely high concurrency for these queries.

✓ The Kognitio database is a mature SQL implementation running all TPC-DS benchmark queries as well as providing functions to allow non standard SQL from other vendors to run unchanged.

✓ The Kognitio SQL extensions such as embedded R, Python or any other Linux process allow the easy integration of sophisticated statistical processing or specialized processing for IOT, telecoms or other complex data types.

✓ The wide variety of standard connectors and the ease of construction of custom connectors make connecting to any data source simple and transparent to the end user who sees the data as a standard table irrespective of whether it's in a Hadoop cluster, in an S3 bucket or retrieved by SSH from a remote file server.

We offer Kognitio on Hadoop completely free to use and offer paid support contracts to customers who wish to take out support for production environments. For more information visit kognitio.com/services-support/support-options/.

About Kognitio

For more than a generation Kognitio has been a pioneer in the development of scale-out, in-memory software for big data analytics.

Today the Kognitio software provides an ultra-fast, high concurrency SQL layer that allows modern data visualization tools to maintain interactive performance, even when the data volume is large and the user count high.

Kognitio is fully integrated with YARN on Hadoop or can be installed on standalone hardware infrastructure. In either case it can directly process data from a large variety of sources such as Hadoop HDFS (including Parquet and ORC format Hive tables), Amazon S3 or NAS.

The software also supports sophisticated NoSQL capabilities enabling scale-out advanced analytics alongside the ultra-fast, fully functional SQL.

To learn more visit kognitio.com

🌐 www.kognitio.com
f facebook.com/kognitio
🐦 twitter.com/kognitio
in linkedin.com/company/kognitio
▶ youtube.com/kognitio