

Oracle Big Data SQL: One fast query, on **all** your data.

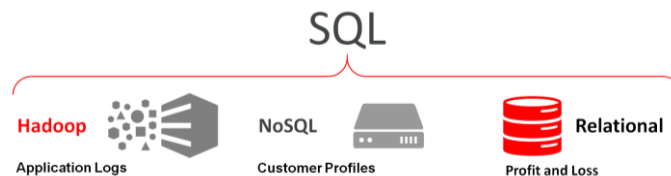
According to the McKinsey Global Institute, 140,000 to 190,000 skilled workers and 1.5 million more data-savvy managers are needed in the US alone to take full advantage of Big Data.

Oracle Big Data SQL helps you get value from big data today.

Leveraging your organizations existing skills, tools and governance to tame big data means you can realize value faster and at lower cost. Big Data SQL is the simplest way to analyze all your data: in Hadoop, NoSQL databases, and Oracle Database.

- Write one query that encompasses spans big data and databases, using the industry's richest SQL dialect.
- Add big data seamlessly into new reports and applications using existing SQL interfaces.
- Maximize performance and minimize data movement using Hadoop Smart Scan

Recent years have seen an exponential increase in the amount of data organizations produce and process. From the Internet of Things to troves of social and application log data, the reserves of data – the raw material of digital insight – grows by the hour. Coincident with this is an intensified pressure on modern, data-driven organizations to produce increased value from that data. To realize value from increasingly large and disparate datasets, big data technologies have risen to the challenge, providing organizations with the means to store and process more data than ever before.



Overcoming Big Data Barriers

While a host of new technologies have emerged to help process big data, both their rapid evolution and the inherently challenging task of application development for naively parallel distributed systems has led to a series of barriers to reliably producing value from big data. For those organizations that have found business cases likely to produce value to the business, significant barriers to actualizing this value exist: both in organizational skillsets and the integration of big data systems with existing enterprise information architectures.

To this end, there has been a resurgence of interest in the SQL language for managing and manipulating big data, particularly for manipulating data stored in the Hadoop ecosystem. Beyond its linguistic maturity, broad adoption, and deep integration with tools and applications, SQL is the natural language for working with data. The declarative, data-oriented nature of SQL requires that users describe the shape of the answers they seek from data, without having to deal with the complexities of how that data is accessed or processed. This enables SQL users to be much more productive, and focus their efforts on business problems, rather than complicated control flow.

Smart Scan for Hadoop and NoSQL

While big data can take up petabytes, the amount of data relevant to a given query can be orders of magnitude smaller. Hadoop Smart Scan makes sure only relevant data is analyzed.

- Data-local scans for fast, flexible reads
- Predicate evaluation and column projection ensure that only the rows and columns relevant to the query are transmitted
- Complex parsing of JSON and other semi-structured data happens data-local and on-the-fly

Given the inherently complicated nature of programming for large-scale distributed systems, it is little wonder that there is tremendous interest for SQL access to data in Hadoop. Many groups are working to provide SQL access to data stored in Hadoop. However, while these efforts alleviate some of the skills barriers to realizing value from big data, a fundamental problem remains with respect to integration of the big data technologies within the broader enterprise architecture.

Specifically, in a big data environment, data lives in many places. While increasingly more data is stored in Hadoop, most business-critical data is stored in a relational database. To that end, what is needed is not simply SQL access to data stored in Hadoop, but seamless SQL access to data stored in Hadoop, various NoSQL databases, and relational databases. Moreover, to truly integrate value from big data, the roles, policies and information governance which already exist must be extended to data residing in these new data stores.

Oracle Big Data SQL: One fast query, on **all** your data

Big Data SQL is Oracle's breakthrough approach to simplifying access and integration to big data sources. Oracle Big Data SQL provides the ability to query all data – in Hadoop, NoSQL datastores, or Oracle Database – in a single statement. Oracle Big Data SQL presents Hadoop and other sources as enhanced external tables. These tables are engineered to transparently map the external semantics of data access – horizontal parallelism, location, and schema – to Oracle internals. This ensures the best possible optimizations for access and native processing throughout. Oracle Big Data SQL allows users to

- Express their queries on all data using the world's richest SQL dialect
- Integrate big data quickly into reports or applications using existing interfaces
- Extend existing Oracle security and access control policies to data stored in Hadoop

Oracle Smart Scan for Hadoop and NoSQL: Maximum performance through minimized data movement

While big data may be massive, very often the amount of data that is relevant to a given query is smaller than the total data volume by an order of magnitude or more. This provides an opportunity for tremendous optimization in query performance. Smart Scan for Hadoop – based on Exadata Storage Servers Software – maximizes the performance of Oracle Big Data SQL by providing

- Data-local scanning: data is read and processed at the point of storage
- Predicate evaluation and projection: only relevant data is transmitted from Hadoop
- Complex parsing: data such as JSON and XML are processed locally at the source

CONTACT US

For more information about Oracle Big Data SQL, visit oracle.com/bigdata or call +1.800.ORACLE1 to speak to an Oracle representative.