

Hadoop Integration Service – Optimal linking of large quantities of data

Take advantage of your large quantities of unstructured data with our Hadoop integration

Key characteristics

- Optimal integration of Hadoop into EXASolution – enables extremely fast linking of data warehouse with Hadoop data and native interlocking with SQL
- Simple inclusion of Hadoop MapReduce jobs or Hive queries in SQL – allows for flexible distribution of analysis jobs between Hadoop and EXASolution
- Direct parallel transfer from and to Hadoop Distributed File System (HDFS) using SQL
- New functions that boost compatibility, productivity, and manageability, guaranteeing a high-performance data transfer between Hadoop and EXASolution

More and more, during decision-making processes, companies have to evaluate Big Data that is available to them only in an unstructured format.

Data scientists are thus faced with the challenge of interlocking this unstructured data with data from a data warehouse – for example, with customer profiles or financial data. Yet the resulting blend of structured and unstructured data makes it virtually impossible to evaluate large quantities of data rapidly and reliably. In addition, data mining, Big Data, and new applications are driving the development of new analytical capabilities.

With the Hadoop Integration Service, large quantities of unstructured data from Hadoop can be linked with your EXASOL data warehouse in EXASolution. Moreover, new functions with regards to compatibility, productivity, and manageability complement the functionality in EXASolution and Hadoop. Take advantage of the analytical opportunities provided by the seamless integration of your structured and unstructured data. Interlock your data from social media, call centre activities, or clickstreams with your data warehouse to create a valuable analysis source.

Benefit from Hadoop integration

The Hadoop Integration Service is the optimal solution for communication between EXASolution and Hadoop to combine the strength of both worlds. Access is always transparent, and EXASolution's special MapReduce capabilities enable you to flexibly allocate and distribute the analysis work to be performed. Thus, Hadoop can be assigned tasks without influencing the extremely high-performance data analyses in EXASolution running in parallel. Besides the MapReduce integration, EXASolution offers interfaces for HDFS and Hive. The EXALoader allows direct and parallel reading and writing from and to HDFS. Furthermore, EXASolution allows you to run Hive queries embedded in a regular SQL SELECT statement. The native interlocking with SQL and the fastest possible linking of your data warehouse with Hadoop data delivers an entirely new capability for your analyses and evaluations.

We developed the Hadoop Integration Service working closely with our customers, and many of them are already using it to combine the strengths of EXASolution and Hadoop (petabytes of unstructured data vs. high-speed analytical Business Intelligence).