



# SQL Data Skipping

BigDataStack Software Component developed by IBM

## Description

SQL Data Skipping is part of the “Data as a Service” BigDataStack offering. It is one of the major success of BigDataStack project since this technology has been currently included within IBM portfolio, specifically in four commercialized products.

## Features

According to today’s best practices, cloud compute and storage services should be deployed and managed independently. This means that potentially huge datasets need to be shipped from the storage service to the compute service to analyse the data. This is problematic even when they are connected by a fast network and highly exacerbated when connected across the WAN e.g. in hybrid cloud scenarios. To address this, minimizing the amount of data sent across the network is critical to achieve good performance and low cost. Data skipping is a technique which achieves this for SQL analytics on structured data.

## Areas of Application

Big Data Analytics.



# SQL Data Skipping

## Market trends & opportunities

All products/services/projects with which data is retrieved from Object Storage through SQL queries are part of the potential market. Within the “Innovation Radar project”<sup>22</sup>, the EC identified the Market Creation Potential of the innovation as addressing the needs of existing markets. Additionally, IBM is on its way to Open Source the Data Skipping component, which will help to gain global visibility and foster adoption beyond IBM.

## Customer benefits

- 1) Better performance.
- 2) Lower price due to reduced data read.

## Technological novelty

Data skipping techniques are novel in the industry, in addition, envisioned synergy with data layout technology may lead to further advances in big data analytics. IBM solution is the most complete and advanced, and hence it has been selected by the EU “innovation radar” as a key innovation. SQL Data Skipping has also been identified as “Tech Ready”.

**TRL level: 9**