

Comparing SAP HANA's capabilities with other open-source platforms and solutions

András Béleczi, Bálint Molnár, Zoltán Vincellér and Attila Kiss

Eötvös Loránd University, Budapest, Hungary,
bearaa@inf.elte.hu, molnarba@inf.elte.hu, vzoli@inf.elte.hu,
kiss@inf.elte.hu

Abstract. There is a bunch of open-source tools and technologies for processing large amount of data or organizing them in uncommon data models (like graphs, hypergraphs, or other flexible data-structures). These solutions mostly focus on distributed computations and use distributed solutions. To improve a business performance to support its already well done information system requires a lot of research among these technologies and just by choosing the proper softwares the work is still incomplete: the next big issue is to integrate everything with the company's base information system. This part is one of the most complex and time-consuming part of these kind of projects.

SAP grants a possible solution to avoid the above mentioned problems: its in-memory HANA database has been extended with a lot of modules and functions to help businesses to reach their targets without big effort since it is working out of box. These functions cover multiple domains: text-mining, data-mining, graph-database, supporting web-services and it also can be used, as an application server. Also it is capable to cooperate with Apache Hadoop or Spark.

To check, how versatile this system is we compare its performance and results with other open-source software which have already proved their worth (Spark, Neo4J, Python Pandas).

Keywords: SAP HANA, text-mining, data-mining, graph representation