

SAP HANA Vora: A Distributed Computing Platform for Enterprise Data Lakes

Christian Sengstock¹, Christian Mathis²

Abstract

Businesses are increasingly leveraging the power of Big Data to improve their services and products. We call the infrastructure to process and manage the heterogenous kinds of data their “data lakes”. Data lakes are used to store and process massive streams of sensor data, service data, collected or generated media, archived enterprise data, and massive transactional databases, among others. Such infrastructures are often realized by Hadoop clusters and low-cost persistence layers, such as S3 or SWIFT data stores.

SAP HANA Vora is a distributed computing platform that sits on top of Data Lakes and was developed to build a basis layer for upcoming Big Data applications in the enterprise. It provides high-performance in-memory data processing and management capabilities, is easily extensible by new computing engines, extends the existing Big Data software stack, and integrates with the existing enterprise IT by design. We present an architectural overview of the system.

¹ SAP SE, 69190 Walldorf, c.sengstock@sap.com

² SAP SE, 69190 Walldorf, christian.mathis@sap.com