

SAP Big Data Warehouse on Pure Data Platform



This post was originally posted on this [site](#). It is from 2017.

Digital companies have imperial views of how their business is run. They realize data is no longer for back-end offices. It is how customers experience today's products and services. Initiatives around data and analytics are fast becoming a priority in the boardroom and as necessary means to achieve desired outcomes. *Traditional* data from transactional systems combined with data from emerging technologies such as Machine Learning provide us both the capability to respond quickly and appropriately in today's fast-pace business environments.

In the last blog post, I described the concept of a "Big Data Warehouse" where HANA and Hadoop can live together in harmony on the same platform and open new possibilities to enterprises. This is important because traditional databases and architectures are not aimed at being able to flexibly master today's flood of data and data-sources.

Nowadays, systems must store, process and analyze very large volumes ranging from hundreds to thousands of petabytes. With Hadoop, the advantage of the distributed file system extends to processing large amounts of data. It can scale-up, almost limitlessly and cost effectively. Hadoop is able to asses unstructured or semi-structured data. When combined with the most up-to-date business data in-memory (HANA), HANA and Hadoop stored on Pure Storage become well-matched to create a versatile system that can serve all types of data needs. HANA delivers the high-performance database, and Hadoop delivers the mass-data ecosystem, together they provides a full range of data needs while keeping costs low. The

potential is limitless.

How Pure Storage can simplify architectures and offer a more affordable SAP Big Data Warehouse

To process large quantities of data, modern databases use high-caliber hardware components and employ in-memory processing such as HANA. HANA is an extremely high-performance technology, but it is also not cheap. For the digital transformation, a certain flexibility is needed. Here, the flexibility needed is to store large quantities of data over the short-term without overstepping budget limits and thus raising the pressure on performance.

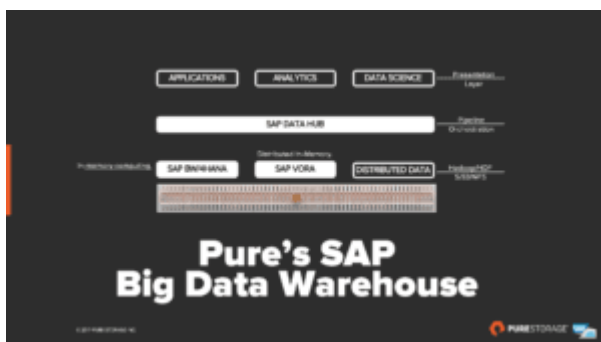
Enter Pure Storage, a data platform that was created for precisely this purpose. In this concept, I will refer to hosting BW/4HANA In-Memory platform along with Hadoop and potentially even including SAP Vora and SAP Data Hub, the newly announced Big Data orchestration tool, all on the same data platform.



Big Data Warehouses have been around for a while. Big Data storage and processing environments can not only complement traditional business warehouse systems but also save customers cost and give them a market advantage in very competitive markets.

With Pure's FlashBlade, you can:

1. Easily process semi-structured and unstructured data, such as photos, videos or text. at incredibly high speeds.
2. The platform offers extremely high-performance data processing storage for fine data.
3. In addition to this concept, the same platform can be utilized in many, many traditional and new data technologies. The more you use Pure, the higher your savings.



A standard Big Data Warehouse setup usually has the following storage layers:

- An Ingestion Layer: to collect data from many sources, e.g. IoT devices in the edge.

- A Processing Layer: for distributed processing of large and/or many files,
- An Analytics Layer: This is where BW/4HANA comes in this concept, providing structured data to end users.



With Pure, you will be able to simplify and consolidate the architecture by utilizing it for all three tiers' storage layer. This tight integration between BW/4HANA, Big Data, SAP Vora and SAP Data Hub means workflows work more in harmony. Data movement between these different systems is optimized and aligned for performance, making pipelines part of your Business Warehouse strategy and virtually part of your process chain.

Last but not least, you have so much more freedom utilizing data-tiering in FlashBlade and leverage it for archiving and colder data. This concept can translate to incredible time and cost savings, and architecture consolidation. This puts you in control of your data and enables you to leverage it in the most profitable way.

"Any powerful idea is absolutely fascinating and absolutely useless until we choose to use it." – Richard Back, One.

The ability to share data between technologies is going to make a big difference for companies going forward and the need for a data platform such as Pure's will become a necessity. Technology changes human behavior and how we collect information. Companies must follow the data and leverage it to keep up with their customers' and shareholders' demands. As the famous writer and pilot wrote Richard Back wrote, "Any powerful idea is absolutely fascinating and absolutely useless until we choose to use it." Data and analytics will help companies understand where they need to put their dollars to find growth opportunities. High-performance and diverse types of data is a huge accelerator to that.