MarkLogic®

MarkLogic-Hadoop Deployment Accelerator

The MarkLogic-Hadoop Deployment Accelerator jump-starts projects that integrate MarkLogic® Enterprise NoSQL database and Hadoop*. We advise on technical direction, deliver architecture and design documents, and provide hands-on implementation support to get an initial prototype up and running.

Whether you have Hadoop now and would like to leverage MarkLogic for real-time data access and analytics, or you already have MarkLogic and would like to augment your system with low-cost storage and efficient batch processing using Hadoop, MarkLogic can help.

Overview

MarkLogic Consulting has been building large-scale, real-time systems for unstructured and heterogeneous data sets for over a decade. We have real-world experience delivering value to end users, have best practices for these systems, and know the pitfalls involved.

The Hadoop accelerator brings MarkLogic's extensive experience to bear specifically on combined Hadoop-MarkLogic deployments. MarkLogic fits into the Hadoop ecosystem much like Pig, Hive, HBase or other technologies. MapReduce works directly on MarkLogic data and MarkLogic forests can be stored directly in HDFS. The MarkLogic Content Pump moves data between storage tiers in MarkLogic or Hadoop using massively-parallel processes that give the speed and horizontal scalability needed for large deployments.

As part of the Hadoop Accelerator, MarkLogic consultants will analyze the range of data formats and types you have, how they need to be used, and which components can best process them. We provide design documents, configure MarkLogic tools and technologies, and build working prototypes to get you started.

Service Details

The activities in the MarkLogic-Hadoop Deployment Accelerator vary according to your needs and how you want to combine MarkLogic and Hadoop.

Adding MarkLogic to an Existing Hadoop Ecosystem

If you already have Hadoop, you probably now need to make all the information actionable and valuable to users. Hadoop and MarkLogic both deal with "any data in any format" – which is wonderful and powerful, but also presents challenges due to the variation and structure of the data. Having all the data in one place does not immediately solve the problems inherent in heterogeneous or unstructured data, much less make it available in real-time, or in a dynamic way.

![MarkLogic logo]

Fortunately, MarkLogic Consulting Services has a decade of experience making varied, unstructured data immediately useful with the same agility you have seen when adding new data into Hadoop/HDFS. MarkLogic can extract data from HDFS, store data directly into HDFS and use Hadoop Map/Reduce on MarkLogic partitions in or outside of HDFS. The Hadoop Accelerator engagement is designed to navigate these and other options – not as a technical exercise, but to solve real problems and bring our Big Data experience to bear on your organization's specific challenges.

Adding Hadoop to an Existing MarkLogic Deployment

If you already have a MarkLogic cluster up and running, we will advise on how to use Hadoop to make existing MarkLogic functionality work better, or at lower cost, using Hadoop. MarkLogic deployments typically focus on real-time data access and analytics, rather than staging, ETL, and batch – appropriate Hadoop use can augment MarkLogic in these areas, reducing your TCO and increasing agility.

- To reduce storage costs, MarkLogic deployments can store data on Hadoop HDFS which provides lower-cost storage than a SAN

- Hadoop MapReduce jobs can run directly on MarkLogic data, in massively parallel ways without extensive programming

- Older, less-valuable data can be moved from MarkLogic to a separate Hadoop cluster for slower batch processing, while retaining the full power of MarkLogic on your most valuable data

The MarkLogic-Hadoop Deployment Accelerator advises on these and other technology choices to move data to the right place in the right format, use the right tools, and deliver valuable information to users.

MarkLogic-Hadoop Deployment Approach

Design and Architecture

The deployment accelerator includes delivery of design and architecture documents that specify data flows, transformations and processing models. The MarkLogic team will investigate the data available, where it comes from and how it needs to be used, and use this information to develop and deliver a design that works for your enterprise.

Setup and Configuration

After delivering, discussing, and agreeing on the design approach with your team, MarkLogic Consultants will identify key setup and configuration tasks that need to be performed. This may include setting up MarkLogic Server, configuring it to talk to Hadoop or store data directly in HDFS, and deploying the MarkLogic Content Pump to move data.

Implementation

To fully jump-start your Hadoop and MarkLogic deployment, we work with your developers to build initial data flows and processes and expose data as services or information access applications. This hands-on work will empower and train your own development staff and leave behind working code that can be extended by your team.

While our design and architecture activities recommend directions for the overall Hadoop enterprise, our implementation efforts focus on MarkLogic technologies.

Getting Started

Our skilled MarkLogic consultants are ready to review your specific needs for MarkLogic and Hadoop and accelerate your deployment. Whether you have a system in place now or are still choosing your technical direction, this accelerator package will open up new opportunities and jump start your effort.

About MarkLogic

For more than a decade, MarkLogic has delivered a powerful, agile and trusted Enterprise NoSQL database platform that enables organizations to turn all data into valuable and actionable information. Organizations around the world rely on MarkLogic's enterprise-grade technology to power the new generation of information applications. MarkLogic is headquartered in Silicon Valley with offices in Chicago, Frankfurt, London, Munich, New York, Paris, Singapore, Stockholm, Tokyo, Utrecht, and Washington D.C. For more information, please visit www.marklogic.com.

MARKLOGIC CORPORATION
999 Skyway Road, Suite 200 San Carlos, CA 94070
+1 650 655 2300 | +1 877 992 8885 | www.marklogic.com | sales@marklogic.com

MarkLogic®