

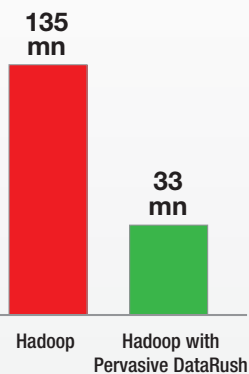
Pervasive DataRush for Apache Hadoop



For Faster, Greener, More Cost-Efficient Hadoop Runs

Hadoop Runs 4x Faster

Malstone B Benchmark,
0.5TB data,
10-node cluster on
Amazon EC2 HPC



Hadoop Delivers Big Data Scalability

With industry backers like Yahoo! and Facebook, it didn't take Hadoop long to become one of the core technologies used to wrangle Big Data. Hadoop and the Hadoop Distributed File System provide a reliable and scalable solution on inexpensive commodity servers. As Hadoop-based solutions began to migrate into production, the requirement to increase performance and reduce operational cost began to surface.

Pervasive DataRush™ combined with Hadoop can help meet this challenge.

Turbocharge Hadoop with Pervasive DataRush

Pervasive DataRush is a leader in delivering increased performance with less hardware. Adding

Pervasive DataRush to Hadoop can boost performance by more than 4x. What you do with that level of performance is up to you.

- Run your Hadoop jobs on 1/4 of the hardware—cheaper and greener
- Run more jobs and perform deeper analysis



*We've put Pervasive DataRush 5.0 to work inside wukong, our open source tool for executing Ruby tasks on Hadoop clusters, which we use to drive Infochimps' production jobs on Amazon's EC2. The result is exactly on-target: **improving computational efficiency means we can get the largest jobs done using fewer resources in the cloud, which translates to hard-dollar cost savings.***



Philip Kromer
CTO at Infochimps

Pervasive DataRush for Hadoop

Fast

- Run your MapReduce jobs on a quarter of the machines
- Perform deeper analysis in the same time
- Reduce capital and operational costs

Easy

- No changes to your HDFS or data
- Leverage existing Hadoop cluster
- Supports any Hadoop distribution

Flexible

- Integrate with your existing mapper and reduce functions
- Or, integrate directly with HDFS using dataflow as an alternative to the MapReduce model

Hadoop Integration

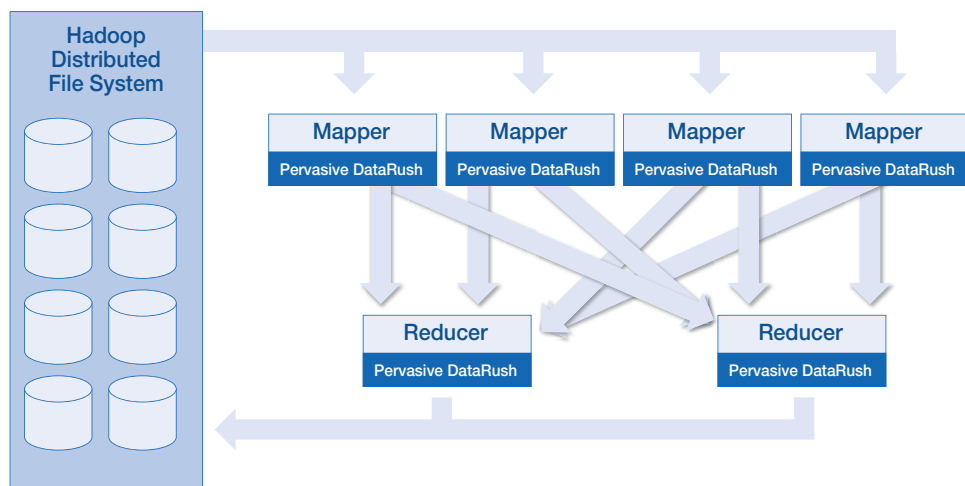
Pervasive DataRush integrates with Hadoop in several ways, depending on your needs.

If you already have an established Hadoop cluster and are using MapReduce, you can call Pervasive DataRush from your Map and Reduce functions, allowing each mapper and reducer to fully use the parallelism of multicore processors. You could see a 4x improvement in performance.

You can also integrate Pervasive DataRush with the data in HDFS, as an alternative to using the MapReduce model. The operators and convenience methods provided by the Pervasive DataRush-Hadoop library are optimized for HDFS. They take full advantage of Pervasive DataRush parallelism. Using this method, you could see a 13x improvement in performance.

Pervasive DataRush Engine

At the heart of every Pervasive DataRush for Hadoop application is the patented Pervasive DataRush engine. This embedded engine transforms the way an application handles large volumes of data. The Pervasive DataRush engine is designed to take full advantage of the disruptive shift to multicore processors. Designed to handle gigabytes to petabytes of data, the Pervasive DataRush engine scales with both the size and the dimensionality of the data. The embedded Pervasive DataRush engine delivers unprecedented performance to Hadoop users.



Why Pervasive DataRush for Hadoop?

Scalable. Performance dynamically scales at runtime with increased core counts and increased nodes.

High Throughput: Fast, deep analysis of large data sets with no limit on input data size.

Cost Efficient: Maximum performance from commodity multicore server clusters.

Easy to Implement: No highly specialized programming skills required. Programmable in any JVM language, including Java, Python, JRuby, Scala.

About Pervasive Software

For over two decades, Pervasive has helped thousands of customers around the world in every industry to manage, integrate and analyze their data. Pervasive DataRush is an embeddable software platform solution for data-intensive processing applications such as claims processing, risk analysis, fraud detection, bioinformatics and life sciences.

Contact Us

www.pervasivedatarush.com
1-866-980-RUSH (7874)
1-512-231-6000 (Main Office)

info@pervasivedatarush.com