

Pervasive TurboRush for Apache Hive



Increase Hive Performance on Half the Hardware with No Additional Development

Apache 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 reliable and scalable storage on inexpensive commodity servers.

While the MapReduce model can prove challenging to work with, the resulting scalability was often worth the effort.

Apache Hive Delivers *Ease of Use*



Hive was created to allow Hadoop users to express queries in a familiar, high-level language without having to directly write MapReduce programs. The Hive Query Language (HiveQL) was born and Hive was now delivering on the promise of a data warehousing solution built on top of Hadoop for easy summarization, querying, and analysis of large datasets stored in Hadoop. Hive and Hadoop continue to excel in scalability, handling thousands of nodes and petabytes of data. It was now time to turn to the challenge of improving performance.

Pervasive TurboRush for Hive Delivers *Performance*



Pervasive DataRush™ is a leader in delivering increased performance with less hardware. And that is exactly what we do for your Hive solution. Simply add Pervasive TurboRush™ for Hive, and your Hive queries will run faster AND require fewer resources.

- No learning curve – just drop it in
- Preserve your HiveQL development investment
- Reduce the number of machines in your cluster – cheaper and greener
- Increase workload capacity – run more queries, perform deeper analysis
- Benefit from Hadoop's scalable and robust distributed file system

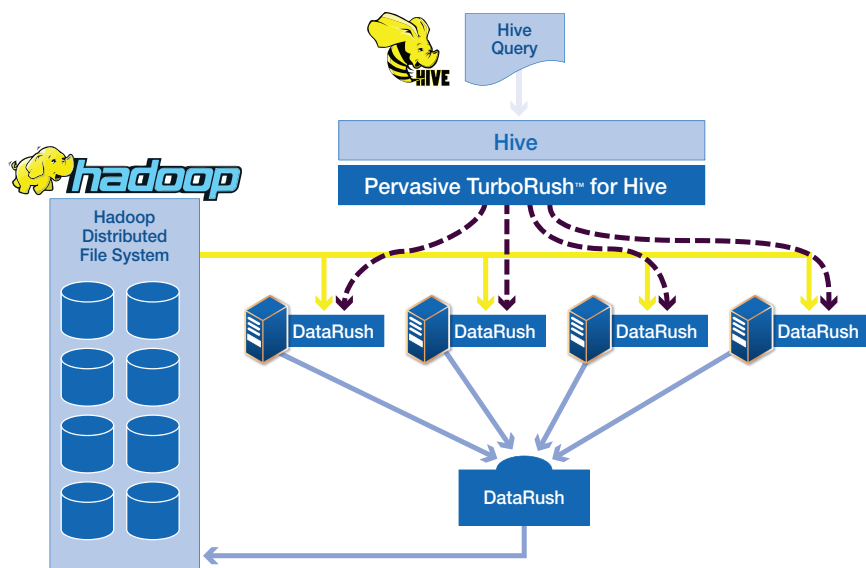
Pervasive TurboRush for Hive

Fast

- Run the same queries on half the machines
- Run more queries in the same time
- Reduce capital and operational costs

Easy

- No changes to your HiveQL code
- No changes to your HDFS or data
- Supports any Hadoop distribution
- Simple installation and configuration



Other Big Data Solutions from Pervasive Software

Pervasive DataRush for Hadoop

- Make MapReduce 4x faster

Pervasive DataRush for Analytics

- Perform deep analytics 20x faster

Pervasive DataRush for Data Preparation

- Remove performance bottlenecks in your cleansing and preparation of large data sets

Pervasive DataRush for KNIME

- Design custom analytics with a graphical composer

Pervasive DataMatcher™

- Quickly find duplicate data in large data sets

How It Works

In normal operation, when Apache Hive receives an SQL query, it parses the query, creates a query execution plan as a series of MapReduce tasks, and runs the plan. Pervasive TurboRush allows Hive to generate an execution plan using **dataflow graphs** as an alternative to MapReduce. It then executes these graphs using Pervasive DataRush distributed across the machines of the cluster. Once installed, it is completely transparent to the user. You simply type one command to enable Hive to use dataflow. This can result in substantial performance gains, due to the efficiencies of using a dataflow paradigm.

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.

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

info@persivedatarush.com