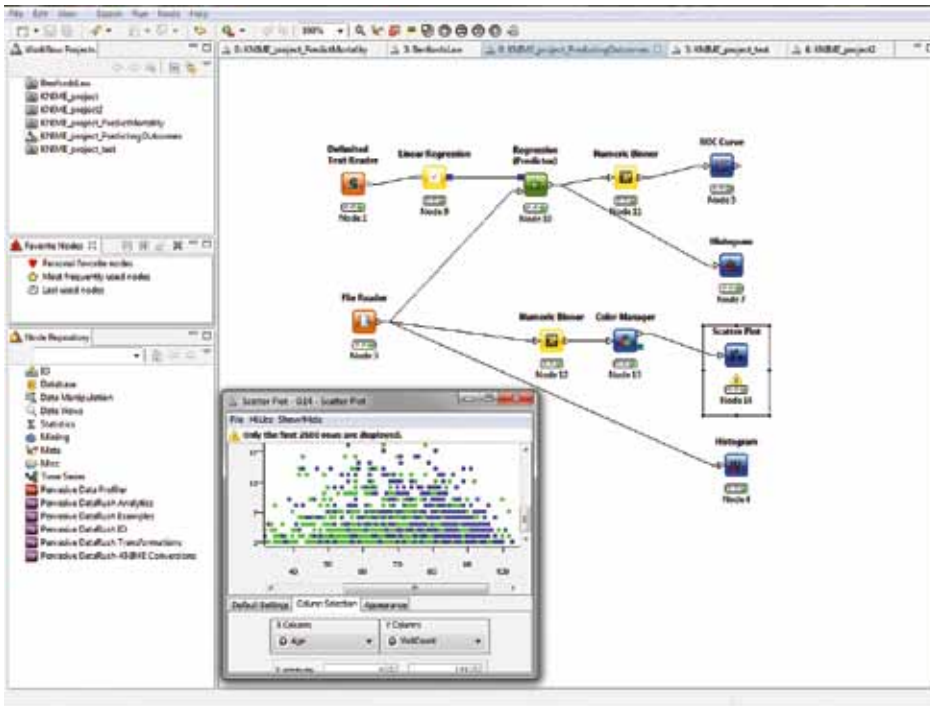


Pervasive DataRush for KNIME

A high-throughput, scalable, visual data mining solution



Pervasive DataRush™ for KNIME Benefits:

- **Big Data Mining.** Easily and quickly analyze gigabytes to terabytes of data.
- **Visual.** Build your models and visualize the results from an intuitive visual interface.
- **Scalable.** Dynamically scale your data mining performance by increasing the number of cores.
- **High-Throughput.** Parallel dataflow technology enables fast, deep analysis of large data sets with no limit to input size.
- **Cost Efficient.** Get the highest possible performance from commodity multicore desktops and servers.

KNIME Data Exploration

KNIME is a comprehensive open-source data integration and analytics platform. Thousands of professionals use KNIME in industries such as pharmaceuticals to perform analytics on their data. Users can easily assemble individual “nodes” into a complete workflow. KNIME already incorporates more than 100 processing nodes for data manipulation and mining as well as plug-ins that run R-scripts, giving users access to a vast library of statistical routines. Visualization of results is provided through interactive scatter plots, parallel coordinates, and more.

Analyzing Big Data Sets

The size and complexity of data is exploding. Data mining and analytics professionals are being asked to process hundreds of gigabytes and even terabytes of data to uncover actionable knowledge. They are straining to achieve timely knowledge when faced with this ever-growing data sets.

Fortunately, this doesn’t mean you have to resort to using an expensive cluster, or run your models overnight, or even sample the data. KNIME’s unique ability to handle big data combines nicely with Pervasive DataRush, enabling you to scale up as data volumes grow.

Pervasive and KNIME

Pervasive now brings the performance and scalability of Pervasive DataRush to the KNIME platform. Complex workflows created in KNIME can now process much larger data sets in a fraction of the time.

With Pervasive DataRush for KNIME, users have access to a library of scalable, high-throughput Pervasive DataRush nodes to tackle big data analysis. Pervasive DataRush leverages the parallel processing power of today’s multicore servers to deliver a high throughput data mining solution that runs on a single server.

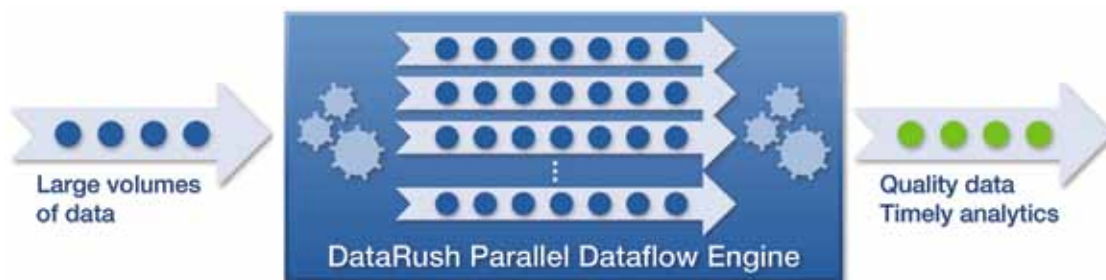
Algorithm	KNIME Runtime	Pervasive DataRush for KNIME Runtime	Performance Increase
Naïve Bayes Numerical Predictor	4136 secs	17 secs	243X
Naïve Bayes Categorical Predictor	4732 secs	65 secs	72X

Pervasive DataRush Parallel Dataflow Engine

At the heart of every Pervasive DataRush-enabled application, including Pervasive DataRush for KNIME, is the Pervasive DataRush Parallel Dataflow 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 enabling performance on a single server that you thought was only possible on a large cluster. 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 Parallel Dataflow Engine delivers unprecedented performance to KNIME users faced with mining very large datasets.



Pervasive Software invites you to learn more or request a free trial on our website at www.PervasiveDataRush.com or email info@pervasivedatarush.com.

Pervasive DataRush for KNIME Nodes include:

- Data Profiling
- Analytics
 - Association Rules
 - FP-growth
 - Classifiers
 - Decision Tree
 - K-Nearest Neighbors
 - Naïve Bayes
 - Clustering
 - Recommender
 - K-Means
 - SVM
 - Neural Networks
 - Feature Selection
 - PCA
 - Fraud Detection
 - Benford’s Law
 - Regression
 - Linear Regression
 - Logistic Regression
 - Multivariate Polynomial Regression
 - Polynomial Regression
- Input/Output
 - PMML Reader
 - Delimited Text I/O
 - Sparse Data I/O
- Transformations
 - Join
 - Lookup
 - Regular Expression
 - Group
 - Cluster Duplicates
 - Discover Duplicates
 - Filter Rows
 - Select Rows
 - Missing Value
 - Overlay Fields
 - Run Script
 - Trim Whitespace

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 a parallel dataflow platform enabling data-intensive applications such as claims processing, risk analysis, fraud detection, data mining, predictive analytics, bioinformatics and life sciences.

Contact Us

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

info@pervasivedatarush.com

