



Node Packs

Lavastorm Analytics Library

Copyright

© 2016 LAVASTORM ANALYTICS, INC. ALL RIGHTS RESERVED. THIS DOCUMENT OR PARTS HEREOF MAY NOT BE REPRODUCED OR DISTRIBUTED IN ANY FORM WITHOUT THE WRITTEN PERMISSION OF LAVASTORM ANALYTICS, INC.

Apache Hive, Hive are trademarks of The Apache Software Foundation.

Apache Spark, Spark, Apache, and the Spark logo are trademarks of The Apache Software Foundation.

Microsoft and SharePoint are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

MongoDB and Mongo are registered trademarks of MongoDB, Inc

Qlik®, Qlik Tech®, QlikView® and the Qlik Tech logos are trademarks or registered trademarks of Qlik Tech International AB.

Salesforce, SALESFORCE.COM and others are trademarks of salesforce.com, inc. and are used here with permission.

Tableau and Tableau logo are registered trademarks of Tableau Software, Inc.

TIBCO® Enterprise Runtime for R are either registered trademarks or trademarks of TIBCO Software Inc. and/or its subsidiaries in the United States and/or other countries.

The names of actual companies and product mentioned herein may be trademarks or registered trademarks of their respective owners.

Confidentiality

This document contains confidential information that is proprietary to Lavastorm. The original recipient of this document may duplicate this document in whole or in part for internal business purposes only, provided that this entire notice appears in all copies. The recipient agrees to make every effort to prevent the unauthorized use, distribution or disclosure of the proprietary information contained in this document.

Disclaimer

No representation, warranty or understanding is made or given by this document or the information contained within it, and no representation is made that the information contained in this document is complete, up to date or accurate. In no event shall LAVASTORM ANALYTICS, INC. be liable for damages or losses of any kind in connection with, or arising from its use, whether or not LAVASTORM ANALYTICS, INC. was made aware of the possibility of such damage or loss.

Contact Details

For product demonstrations, enhancement requests or technical questions regarding the use of any Lavastorm product, contact us as follows:

HQ Address: Lavastorm, 321 Summer Street, 3rd Floor, Boston, MA 02210 USA

Telephone: +1 617 345 5422 ext. 244

Fax: +1 617 345 5475

Email: info@lavastorm.com

Internet: www.lavastorm.com

Node Name	Node Description	Availability	Compatibility		
			5.1	6.0	6.1
Data Converter	Converts the data type of each field based on the discovered data type.	Standard	NA	NA	6.1.0.0
Spark Agg	Aggregates data using a Spark cluster.	Standard	NA	NA	6.1.0.0
Spark Data Viewer	Previews data from Spark.	Standard	NA	NA	6.1.0.0
Spark Download	Downloads data from a Spark cluster.	Standard	NA	NA	6.1.0.0
Spark End Session	Ends a Spark session.	Standard	NA	NA	6.1.0.0
Spark Filter	Filters data using a Spark cluster.	Standard	NA	NA	6.1.0.0
Spark Java Submit	Provides an interface with a Spark cluster to execute a Spark application.	Standard	NA	NA	6.1.0.0
Spark Join	Joins data using a Spark cluster.	Standard	NA	NA	6.1.0.0
Spark Sample	Samples data using a Spark cluster.	Standard	NA	NA	6.1.0.0
Spark Sort	Sorts data using a Spark cluster.	Standard	NA	NA	6.1.0.0
Spark Start Session	Starts a Spark session.	Standard	NA	NA	6.1.0.0
Spark SQL Query	Provides an interface with a Spark cluster to execute a Spark SQL query.	Standard	NA	NA	6.1.0.0
Spark Upload	Uploads data to a Spark cluster.	Standard	NA	NA	6.1.0.0
HDFS Directory List	Lists the contents of a HDFS directory using the WebHDFS API.	Standard	NA	6.0.0.0	6.1.0.0
Predict Quantile Regression	Uses a Quantile Regression model to predict the value of a dependent variable based on the specified values of the independent variables.	Premium	NA	6.0.0.0	6.1.0.0
Output to	Ouputs LAE data to the Qlik QVX file	Standard	NA	6.0.0.0	6.1.0.0

Qlikview / Sense	format, allowing users to work with the output data within the QlikView application and Qlik Sense application.				
Quantile Regression	Models data using quantile regression allowing identification of data trends for conditional quantiles of a response variable distribution.	Premium	NA	6.0.0.0	6.1.0.0
Lavastorm Metadata for Salesforce	Retrieves metadata in Salesforce.com objects via the Salesforce.com REST API.	Standard	NA	6.0.0.0	6.1.0.0
Lavastorm Update for Salesforce	Inserts, updates, upserts or deletes records in Salesforce.com objects via the Salesforce.com REST API.	Standard	NA	6.0.0.0	6.1.0.0
Lavastorm Get for Salesforce	Queries Salesforce.com Objects via Salesforce Object Query Language (SOQL) over the Salesforce.com REST API.	Standard	NA	6.0.0.0	6.1.0.0
Linear Regression Diagnostics	Generates diagnostic plots from a linear regression model.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Logistic Regression Diagnostics	Generates diagnostic plots from a logistic regression model.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Decision Forest	Models data using the Random Forest model allowing identification of data trends using an 'ensemble' of decision tree models.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Predict Decision Forest	Predicts the value or classification for a dependent variable in a Decision Forest (Random Forest) model based on the value of the independent variables.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Time Series Forecast	Uses the Holt-Winters method to forecast a time series that can optionally contain a trend and seasonal variations.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Time Series Diagnostics	Generates plots of univariate time series data.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Market Basket	Generates Frequent Itemsets and	Premium	5.1.0.0	6.0.0.0	6.1.0.0

Analysis	Association Rules from transactional data.				
Market Basket Miner	Extracts association rules from an association rule model.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
HDFS Download	Downloads file(s) from a specified HDFS server using the WebHDFS API.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Hierarchical Clustering	Classifies data into a specified number of clusters. The data are partitioned into a hierarchy of sub-groups by agglomerative clustering of the observations such that the distance between observations is minimized at each step.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
K-Means Advisor	Recommends the best number of clusters to use when performing a K-Means analysis of the data.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
K-Means Clustering	Classifies data into a specified number of clusters. The data are partitioned into K groups such that distance from the points to the assigned cluster centers is minimized.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Linear Regression	Models data using linear regression allowing identification of data trends.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Logistic Regression	Models data using logistic regression allowing identification of data trends	Premium	5.1.0.0	6.0.0.0	6.1.0.0
R Library Package Download	Downloads the primary 'Comprehensive R Archive Network' (CRAN) R packages that are prerequisites for the correct operation of the nodes within the Statistical and Predictive Analytics Node Pack.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Predict Linear Regression	Uses a linear regression model to predict of the value of a dependent variable based on the specified values of the independent variables	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Predict Logistic Regression	Uses a logistic regression model to predict the probability of a successful outcome of a dependent variable based on the specified values of the independent variables.	Premium	5.1.0.0	6.0.0.0	6.1.0.0

HDFS Upload	Uploads file(s) to a specified HDFS server using the WebHDFS API.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Upload for SharePoint 2010	Uploads file(s) to a specified SharePoint 2010 server using the SharePoint 2010 SOAP API.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Calculate Workdays	Calculates the number of whole days between two dates. Specified weekend days and holidays are not included in the total.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Lavastorm Power R	Executes R script on the embedded TIBCO® Enterprise Runtime for R engine using LAE input data and returning any output back to LAE.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Download for SharePoint 2010	Downloads file(s) from a specified SharePoint 2010 server using the SharePoint 2010 SOAP API.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Folder List for SharePoint 2010	Reads a specified folder within a SharePoint 2010 server using the SharePoint 2010 SOAP API and produces a list of files within that folder that meet the specified criteria.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Send Email+	Sends a Unicode-encoded email or group of emails, with optional attachments, to designated recipients.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Output for TIBCO Spotfire	Outputs LAE data to the TIBCO Spotfire STDF file format, allowing users to work with the output data within the TIBCO Spotfire application.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Hive Join	Orchestrates the execution of a single specified join within an Apache Hadoop Hive environment, producing the result set returned by the Hive environment at its output. Uses an ODBC connection to the Hive environment.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Hive Metadata Query	Queries an Apache Hadoop Hive environment to retrieve schema data. Uses an ODBC connection to the Hive environment.	Standard	5.1.0.0	6.0.0.0	6.1.0.0

Hive Sample	Generates a sample of the data in a specified Apache Hadoop Hive table. Uses an ODBC connection to the Hive environment.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Folder List for SharePoint	Reads a specified folder within a SharePoint server using the SharePoint 2013 API and produces a list of files within that folder that meet the specified criteria.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Download for SharePoint	Downloads file(s) from a specified SharePoint server using the SharePoint 2013 API.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Upload for SharePoint	Uploads file(s) to a specified SharePoint server using the SharePoint 2013 API.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Lavastorm Query for Salesforce	Enables the Lavastorm Analytics Engine to integrate with Salesforce using the Salesforce REST API to query Salesforce Objects.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Lavastorm Transact for Salesforce	Enables the Lavastorm Analytics Engine to process Create, Read Update and Delete (CRUD) operations on Salesforce Objects using the Salesforce REST API.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Net Present Value	Implements a financial function to calculate the Net Present Value (NPV) of a series of regular future cash flows using a specified discount rate.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Get MongoDB Data	Converts JSON data output from MongoDB queries to BRD data.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Query for MongoDB	Performs queries on a MongoDB database and outputs the results as JSON documents.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Update for MongoDB	Performs updates and deletions to documents within a MongoDB database.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
JSON Data	Reads and converts JSON data from an input data source into tabular BRD data.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
XML Data	Reads and converts XML data from an	Standard	5.1.0.0	6.0.0.0	6.1.0.0

	input data source into tabular BRD data.				
FTP Get	Downloads files from a server using the FTP, FTPS, or SFTP protocol.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
FTP Put	Uploads files to a server using the FTP, FTPS, or SFTP protocol.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
ERP Connector	Provides integration with the SAP platform using the BAPI collection of interfaces, allowing users to acquire data from and publish data to the system.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Extract ERP Metadata	Extracts metadata for available BAPIs from an SAP system.	Premium	5.1.0.0	6.0.0.0	6.1.0.0
Input Raw	Acquires a file and produces an output field containing the raw file contents.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Input Static	Produces an output field containing raw data entered manually by the user.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Output QlikView	Outputs LAE data to the QlikView QVX file format, allowing users to work with the output data within the QlikView application.	Public	5.1.0.0	6.0.0.0	6.1.0.0
Output Tableau	Outputs LAE data to the Tableau TDE file format, allowing users to work with the output data within the Tableau application.	Public	5.1.0.0	6.0.0.0	6.1.0.0
Histogram+	Combines the Histogram and InFlow Bar nodes to create a histogram of the data and a bar chart rendering of it.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Deep X-Ref	Extends the X-Ref node to compare the key fields similar to the X-Ref, but for each record from both inputs that match key fields, performs a field-by-field comparison using all other (non-key) fields. This can be useful when two data sources are expected to be exactly equal in all common (overlapped) fields.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Change Metadata	Changes a field's name and data type within the source data; allows the fields to be reordered, and fields to be duplicated	Standard	5.1.0.0	6.0.0.0	6.1.0.0

as well.

Pivot Table	Creates a cross tabulation of data, summaries of two criteria, similar to spread sheet pivot table functionality.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Quick Stats	Executes a number of statistical functions over numeric data in the input records: sum, min, max, average, count, range and null count.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Encrypt Fields	Takes an input BRD file and encrypts fields of the user's choosing using the triple DES algorithm and SHA-1 password hashing technique.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Decrypt Fields	Takes an input BRD file with fields encrypted by the Encrypt Fields node.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Get MetaData	Produces an output describing the metadata on the input given, by producing a record for each field in the input, describing the metadata for that field.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
X-Ref MetaData	Performs a cross-reference against the metadata of two inputs to find the common fields and types among the input data sets.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Output Raw	Takes data from a provided field, and writes it to a data file, dealing with internal encoding as appropriate.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Output Static	Takes any text that is entered into the text box and writes those contents directly to a file, without interpretation.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Convert Tabular XML	Transforms a field containing XML tabular data into the native BRD format used by LAE nodes.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
HTTP	Provides an interface for making URL requests to HTTP servers and retrieving response data.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Remove Duplicates	Removes duplicate records from the input set.	Standard	5.1.0.0	6.0.0.0	6.1.0.0

Tabular XML File	Transforms an input XML file which contains tabular data into the native BRD format used by LAE nodes.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
URL Query Builder	Creates a list of URLs which can be used for HTTP requests.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Fuzzy Join	Compares data sets using keys which don't exactly match, but which are logically correlated, using fuzzy matching techniques.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Fuzzy X-Ref	Extends Fuzzy Join to provide not only the matching sets by also the orphaned populations from each side of the match.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Band by Strata	Organizes data into "bands" based upon common characteristics. Given an increment value, assigns a band number to each value based upon distance, in number of increments, from zero.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Band by Deviation	Organizes data into "bands" based upon common characteristics. Calculates the mean and standard deviation of a data set, and then assigns a band number based upon each value's count of deviations from the mean.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Percent Rank	Calculates the percentile for each value relative to the rest of the data set.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Interval Inspection	Identifies gaps in sequenced data sets; verifies that the values in each row are increasing according to a specific pattern defined by a sequence interval.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Threshold Test	Identifies records which, when a running aggregate is calculated, exceed a designated threshold.	Standard	5.1.0.0	6.0.0.0	6.1.0.0
Dimensional Analysis	Calculates statistics for a multi-dimensional dataset, which may help a user to identify interesting patterns or anomalies in data which suggest areas of further analysis.	Standard	5.1.0.0	6.0.0.0	6.1.0.0

END OF DOCUMENT