Twister® DATA FRAMEWORK

Twister Data Integrator

Data Sheet

Overview

Data integration and quality play an increasingly important role for organizations implementing business intelligence and analysis strategies. *Twister Data Integrator* enables these organizations to extract, transform, and load data from across the enterprise to create consistent, accurate information.

By linking legacy data and applications, web-based technologies, and third-party applications into an efficient pipeline that coordinates the data retrieval, enrichment, and delivery processes, Twister Data Integrator allows analytics users to exploit information enterprise-wide.

By using *Twister Data Integrator,* organizations can quickly process digital information and automatically connect an unlimited number of disparate data sources and applications without the need for manual processing or intervention.

Features

Data Sources and Targets

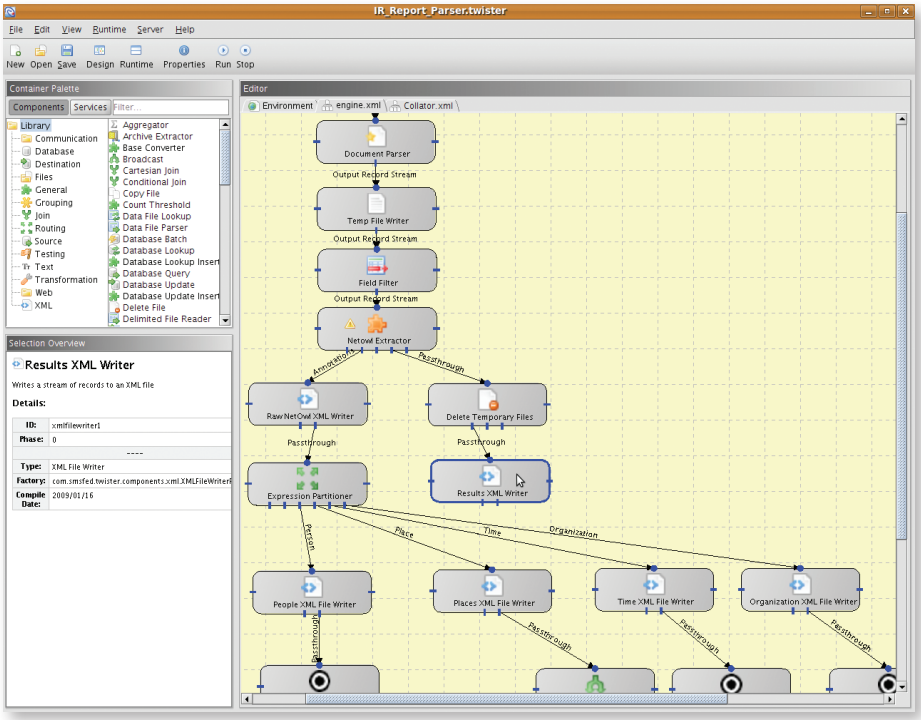
- Read and write to relational databases such as Oracle, SQL Server, Sybase, MySQL and Postgres.
- Read and write structured file formats such as CSV, fixed-width, spreadsheets and XML.
- Read and write unstructured document formats such as PDF, DOC and HTML or create custom text formats using templates.
- Exchange files over network connections using FTP, CIFS, SFTP, SCP, and HTTP
- Read and write to content repositories such as Endeca and MarkLogic.
- Read and write to XML web services
- Crawl websites and download internet resources such as HTML and RSS

Standard Data Manipulation Components

- File Manipulations: Search, Copy, Move, Delete, Checksum, Unzip, Split
- XML Manipulations: XPath, XSLT, Validation
- Text Manipulations: Regular Expressions, String parsing, Date and number parsing and formatting
- Routing: Partition, Gather, Join, Sort, De-duplicate, Intersection
- Other: Aggregate Functions, JavaScript, Lookup Files
- And more…

Ease of Use

- Intuitive visual programming interface with the power to define and change dataflow using drag, drop and configure.
- 90+ reusable data access and data manipulation components
- The same programming interface and components can be used to create bulk processing applications or transactional web services.
- Self-documenting dataflow diagrams are understandable by both technical and non-technical users.
- Construct and test complicated business integration applications much quicker than with custom code.



3rd Party Data Manipulation Components

- Choose between several of the leading 3rd party entity extractors for making use of unstructured data.
- Choose between several of the leading 3rd party gazetteers to add geographical information to your datasets.
- Choose between several of the leading 3rd party libraries for extracting metadata from imagery files.

Enterprise Scalability

- A dataflow can run on a single CPU on a single machine or on hundreds of CPUs on hundreds of machines.
- The multithreaded runtime takes advantage of multiple processors and cores for greater throughput.
- The platform makes it easy to partition data and processing across a cluster of machines allowing your dataflow throughput to scale up with the addition of new hardware.
- Data is streamed through the system which allows for immediate notifications and interactions with real-time systems.

Ease of Administration

- Web-based interface provides a central point for monitoring the status of jobs and the machines in the cluster.
- Controller provides a central location for installing configuration files and code which is then pushed to all machines in the cluster.
- Data flows and components are version controlled to allow for easy configuration management

Extensibility

- A Java API with documentation and sample code provides developers with the ability to add their own data access and data manipulation components.
- Custom components are completely self-described and can easily be shared with other users.
- An embedded JavaScript interpreter and function library provides scripting support within the dataflow.
- A dataflow can call out to command line applications to leverage existing tools and processes.

Platform Support

- Windows 2000
- Windows XP / 2003 Server
- Windows Vista
- Solaris / Solaris x86 / Open Solaris
- Red Hat Enterprise Linux
- Suse Linux
- Ubuntu Linux

SMSi
www.sms-fed.com
info@sms-fed.com