

# Actian DataFlow

High speed data extraction, transformation, analysis and loading

## Access Hadoop Data

Process data natively on Hadoop, 10x faster than MapReduce

Analyze all your data. Push analytics to the data and eliminate loss of granularity

Collect, prepare, optimize and analyze data fast for predictive analytics

Integrate analytics seamlessly into your enterprise

## Code free workflows

Design, test and deploy workflows in minutes without coding

## Capture More Data Value Faster

Actian DataFlow provides a visual end-to-end solution for data preparation, analytics development and execution on Windows, Linux and Hadoop. An embedded data flow engine delivers unmatched auto-scaling and parallelism, processing data natively on Hadoop over 10 times faster than MapReduce. It's an ideal combination of features and performance for everyone from business analysts to data scientists.

### Improve Accuracy – Use All of Your Data

If your software is limiting the amount of data you can analyze, and you've been settling for 2% samples or relying on best-guess scenarios, it is time to try Actian DataFlow. You can run analytics directly on Hadoop clusters. Push the compute to the data source for optimum performance and eliminate loss of granularity from too much aggregation. Process more data, find patterns and outliers, and get fast, accurate answers.

### Add Context – Use All of Your Data Sources

It's hard to know your customers and anticipate their needs when the social media data is in Hadoop, account information is in a data warehouse, purchasing history is in a data lake, and the support call history is in an application. Actian helps you bring all this data together and gain contextual depth in your analysis. Whether your customer is an advocate or about to switch to a competitor, whether your new ad campaign is red-hot or not, analytics can deliver insights that give you a competitive advantage.

### Be Agile – Shorten the Time from Uncertainty to Action

Spend less time preparing your data in Hadoop and more time analyzing it to uncover new insights. Actian DataFlow, in partnership with KNIME, lets you design workflows visually without MapReduce coding. In minutes, you can create workflows that collect, prepare and analyze data, from all of your data sources, and output it to your visualization tool of choice.

- **Connect Fast** – parallelized access to Hadoop, warehouses, log files ...
- **Prepare Fast** – parallelized data profile, de-dupe, join, sort, aggregate ...
- **Optimize Fast** – automatically parallelized workloads on any hardware
- **Analyze Fast** – built-in drag and drop parallel operators for basic

# Empower your business users with analytics

## Design, Test and Deploy Without Coding

Build end-to-end analytics workflows in minutes. Actian provides a generous set of pre-built data preparation and analytics operators that you can drag, drop and configure easily to meet your needs.

- Gain proficiency quickly with no need for Hadoop or MapReduce expertise.
- Test, refine and re-test models rapidly, iterate as needed and get the most accurate results.
- Deploy analytics models locally or on a cluster. There's no need to wait for IT.
- Change you can deploy today.

## Ensure Scalability and Future-Proof your Work

Actian DataFlow automatically detects and utilizes all available cores and nodes at runtime. This gives unmatched data-crunching performance on inexpensive industry standard cluster hardware, using as much as 90% of hardware power without tuning. It also eases the transition from test to production environments. Execution moves seamlessly from desktop to server to cluster, without the need to modify code or re-design models.

## Integrate New Models Easily into Your Architecture

Actian DataFlow designs and executes on virtually any Hadoop platform, connecting to virtually any data source, and cooperates with virtually any other software. It's a good corporate citizen that integrates data analytics with your enterprise architecture.

- Executes on nearly all Hadoop distributions - YARN certified
- Win, Linux and various UNIX (anything with a JVM)
- Reads and writes PMML, works with R, SAS, SPSS, etc.
- Reads and writes HDFS, HBase, RDBS, log files, etc.
- Works with static data, streaming data, or both at once
- Performs advanced analytics right on the Hadoop cluster
- Fills in missing values, sorts, aggregates, joins and transforms

