THE ULTIMATE ANALYST GUIDE TO PREPARING DATA FOR IBM WATSON ANALYTICS

ANALYZE MORE DATA, FASTER!
Since the advent of self-service analytics platforms such as IBM Watson Analytics, data scientists, data analysts and even ordinary business users can turn their data into intelligence that drives decision-making and delivers business value. As a result, a world of new opportunities, once hidden beneath a debris of datasets, has emerged.

However, to yield maximum ROI from Watson Analytics’ capabilities, organizations must feed their data into it in a structured, analysis-ready format. The process of collecting, manipulating and blending data from a variety of sources and preparing it for analysis is known as data preparation. Research has found that analysts spend up to 80 percent of their time manually preparing data, which doesn’t leave many hours in the day for analysis.

80% Research has found that analysts spend up to 80 percent of their time manually preparing data, which doesn’t leave many hours in the day for analysis.

To solve this challenge, **Datawatch and IBM teamed up** to deliver better and faster data access and self-service data prep to IBM Watson Analytics. Datawatch Monarch helps data analysts and business users prep less and analyze more. Users can easily and rapidly access, manipulate and blend data from virtually any source and prepare it for data analysis. And Datawatch Monarch is available for purchase directly from IBM for a one-stop solution.
“Datawatch has increased the use of information throughout the organization. Before this, it was inconceivable to bring data from different areas into one data set that could be used for security, marketing, audits and other business line decisions that would directly impact the bottom line. The relationship with IBM Watson Analytics extends the value by delivering powerful predictive analytics capabilities using simple natural language queries.”

The Fauquier Bank
BEFORE YOU CAN

Answer business questions using your own words to explore and visualize data

Take confident action on insights and see what influences outcomes

Tell a compelling story through a dashboard filled with insight

Be confident in the data you are using for analysis regardless of its origin

YOU MUST FIRST

Unlock more data including "dark data" that often goes unanalyzed

Prepare and transform raw data for data discovery and analytics

Automate and share for transparency and reuse

Optimize data for Watson Analytics for better, faster insight
Acquire data from anywhere

Users can drastically improve the completeness and trust in their visualizations by quickly acquiring and preparing data locked away in previously inaccessible formats— including invoices, sales reports, balance sheets, customer lists, inventory, logs, etc. – along with traditional data sources. With Datawatch, even novice business users can easily access the data they need for analysis, without relying on IT.

- PDF and text reports, web pages, JSON and log files, telemetry data, print spools, EDI, etc.
- Databases such as SQL Server and applications like Salesforce
- Big data sources such as IBM Cloudant and Hadoop
- Content management system repositories, including IBM Manager OnDemand

With the ability to retrieve and use not only the right data, but all of the data required to get the whole story, users can focus on performing analysis that will result in timely, more informed business decisions and better operational processes.
Prepare your data faster

Acquire
Drag & drop/point & click access to virtually any data source or type and instantly unlock and extract data into rows and columns. Access all major big data and relational sources through a library of pre-built connectors. And, automatically extract data trapped in semi-structured documents like PDFs or machine data, log and text files, JSON and more. You can also connect directly to an enterprise content management (ECM) system to use existing reports/documents as a data source.

Preview
Visually navigate data and metadata before you work with it so you can better understand what you have and what you need to do. Datawatch Monarch shows column name, data type and count of unique values as well as statistics around rows, null values, blanks, word counts and length along with histograms highlighting the distribution of values for each field.

Clean
In Datawatch Monarch, data preparation is simple and intuitive. Drag and drop files or documents onto the prep canvas, then apply more than 100 pre-built functions to clean, format, manipulate and even mask sensitive data with mouse clicks – not scripting! Each change is recorded and visible to show how it was prepared, providing more transparency around manipulation and promoting reuse of models.

Combine
For most analyses, you will need to integrate multiple, disparate sources into a single dataset to get a more complete picture of your data. In the past, this would have been a painful, time-consuming process as you tried to determine what data to merge.

Datawatch does the hard part for you. In addition to being able to simply choose which columns you want to combine, Fuzzy Matching is available when there is not a perfect match between two sets of data. And when you really are not sure what or how to combine datasets, Monarch automatically recommends the highest match rate using powerful algorithms.
Think about how much time you would save if you could automate data prep routines wherever possible.

With Datawatch Monarch, data preparation operations are tracked and recorded in the workspace for easy audit and repeatability. You can create a reusable load plan that can then be automated based on a scheduled time or when new data becomes available. An intuitive visual process flow interface will allow you to deliver specific data sets to different users and roles or to other systems like data warehouses or multiple relational sources.

You will likely want to share frequently used data sources or automated data preparation routines with others. This will not only instill confidence around secure access, but also support consistency in the data used to make important business decisions and result in higher levels of efficiency and productivity.
Speed and agility are the common benefits that result from self-service data prep, but they shouldn’t come at the expense and risk of strong governance.

While most organizations have well-defined strategies for governing data that lives in managed systems, such as enterprise applications or data warehouses, analysts are also pulling data from non-managed sources, like CSV or text extracts from transactional systems, personal spreadsheets, third-party reports or other semi-structured content. Without proper governance in place, this can create big headaches around version control, data breaches, reconciliation, auditing...the list goes on.

At the enterprise level, it’s critical to bridge the gap between the ease-of-use and agility that business users demand and the scalability, automation and governance required by IT.

Comply with corporate, industry or regulatory requirements

**DATA RETENTION.**
At a minimum, you must employ document version control for consistency. To meet regulatory or business requirements, relevant source data and documents should also be archived.

**DATA MASKING.**
Sensitive data, such as account numbers and social security numbers, can be hidden or obfuscated so that it is still usable for analytics but unreadable by unauthorized users.

**DATA LINEAGE.**
With a repository for the underlying source content, you should be able to provide complete data lineage and drill down into any source document. This capability will be critical when you need to audit or reconcile data.

**ROLE-BASED ACCESS.**
If you segment prepared data sets based on user roles, you’ll ensure that the right subset of data is delivered to authorized users.
Once you are done preparing it, you can export optimized data directly into Watson Analytics in its native format. Datawatch Monarch pre-configures your clean data for easy upload and ensures it is readily consumable for further exploration and advanced analytics. You will be immediately productive and become more efficient in creating visualizations, answering questions and sharing insights.

You can also save data to Excel, CSV, Access or other analytic tools, including IBM Cognos Analytics.
The powerful combination of Datawatch's intuitive self-service data prep technology and IBM’s powerful cognitive computing solution empowers everyday business users to spend their time on high-value analytic activities, rather than getting bogged down in data prep tasks.

Users can bring in data that otherwise may have never been utilized, analyze it and look for patterns that could provide meaningful insight for their business. It allows the data to reveal new discoveries.

And being able to derive new insights every day that have the potential to innovate and transform operational processes is an invaluable business benefit.

Want to learn more?

Watch
Datawatch Monarch and IBM Watson Analytics Three-Minute Demo

Read
Top 10 Ways Data Preparation Helps IBM Watson Analytics Users

Try
Download a 30 day trial of Datawatch Monarch for IBM Watson ANalytics