panoply blog   Hot Topics   News & Events   Search Blog

Data Engineering

Top 9 MongoDB ETL Tools

Panoply

Smart Data Warehouse

Free for 21 days.
No credit card required.

Watch Demo

By Yetunde Abass | May 7, 2019

The ETL process is as old as the digital collection of data itself. Getting information from a database and into an understandable form is arguably the most crucial step in the process of extracting intelligible insights from the data. One problem many companies face when they begin to collect and store data in various places is reconciling data from traditional relational (SQL) and NoSQL databases.

MongoDB is one of the most popular NoSQL databases used today, so we compiled a list of the top MongoDB ETL tools to extract data out of a MongoDB database. Our list is a mixture of free/open source tools and along with paid options--some of which have a free "community" version. Check out the offerings below. You'll be sure to find a tool that works for you and your team.

MongoSyphon

MongoSyphon is lightweight, open source ETL tool that transforms data into documents in JSON or XML format. It can also do the reverse, sending documents directly into

MongoDB, differing from other ETL tools that try to create relational structures. MongoSyphon is a good tool for folks with intermediate to advanced knowledge of SQL, as there is no GUI. This tool also assumes that the user has an intimate understanding of the structure of the source data. On its Github page the author notes that the program is in its early stages and still needs work. However, the MongoSyphon is being actively managed, with its latest update published in April 2019, as of this writing.

Transporter

Transporter is an open source tool created by Compose that transfers data between different databases. Transporter connects to different databases through the use of adaptors, which can be configured into a data source and/or sink. The MongoDB adaptor has dual functionality and can read from or write to MongoDB databases. The adaptor communicates between each database by converting data to JSON documents. Users are also able to make changes to data during the transfer process through the use of transformers, which are JavaScript functions able to manipulate data.

Krawler

Krawler is an open source ETL tool created by geospatial consultants, Kalisio. The program's purpose is to extract geographic and geospatial data and convert it into more readable forms. Krawler aims to reduce the time it takes to download and analyze information about every point on the planet (the authors go into detail about this process here). MongoDB is not the only data source compatible with Krawler. However, the authors provide extensive documentation specific to ETL with MongoDB.
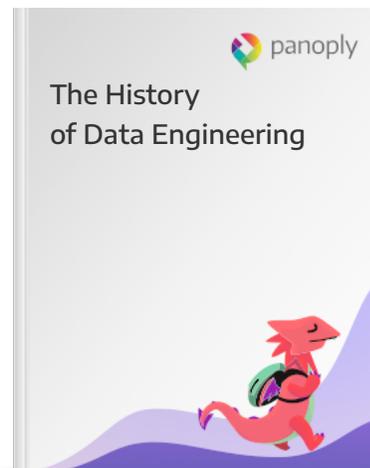
Download Now

Panoply

Panoply is a bit different from most ETL tools because the all-in-one data platform combines the ETL process with a managed cloud data warehouse. This means that importing data from MongoDB and over 100 other sources, just involves a few clicks–without having to define data warehouse schema in advance. Panoply also handles all the scaling and query performance optimization automatically making it easy for data users of all levels to access your MongoDB data easily with SQL. Check out the full list of MongoDB integrations on our website.

Stitch

Stitch is a cloud-based ETL service designed for developers. Stitch differs from many other ETL tools in that it can connect to thousands of SaaS-based data sources through its Import API. Stitch integrates with a variety of data sources and data analysis tools including Chartio and Looker. MongoDB is one of many database integrations that Stitch supports.

Talend Big Data Open Studio

For more than a decade, Talend has been a popular creator of open source big data and ETL tools. The organization's Big Data Open Studio is the community version of its Big Data Platform. Its popularity stems from its easy-to-use drag and drop UI, prebuilt components, and integration with multiple databases including MongoDB. Although there is no company support for the community version of Big Data Open Studio, Talend's users are very active in support forums.

Pentaho

Hitachi Ventara offers paid and open source ETL tools through its Pentaho platform. The Pentaho Big Data Analytics platform aims to be a one-stop data analysis and BI shop, serving as a data extraction and visualization tool. Pentaho integrates with NoSQL databases like MongoDB and has detailed documentation about this particular integration. Pentaho is particularly useful for organizations analyzing IoT data, as Hitachi Ventara has specialty in this area.

SYNC

SYNC is an open source tool designed for data migration between MongoDB and traditional relational databases. Although the tool's author only tested the tool with Oracle and MySQL, they suggest that it should work for any SQL database. SYNC has graphical interface making it easy for users to create mapping and joins between databases. Other features include email notifications with summaries about successful migrations and failures.

YelpDatasetETL

The YelpDatasetETL project has a very narrow, but incredibly useful purpose--moving data from MongoDB to an Elasticsearch index. What's with the name? The authors chose to use Yelp's publicly-available dataset to perform their series of transformations. The first step involves converting Mongo's binary JSON (BSON) format to JSON notation that conforms to Elasticsearch specifications. The second transformation is applied to text fields so that a sentiment analyzer can be applied. This could be a useful ETL tool for anyone trying to analyze large troves of social media data.

Get a free consultation with a data architect to see how to build a data warehouse in minutes.

Request Demo

Read more in:

Data Engineering   MongoDB

Also Check Out

MySQL

MariaDB Vs MySQL In 2019:
Compatibility, Performance, And
Syntax

Data Engineering

Connecting A Jupyter Notebook
With PostgreSQL For Python Data
Analysis

Data Engineering

MongoDB And BI: Frie

Work smarter, better, and faster with wee

tips and how-tos.

Subscribe to our newsletter

Platform	Company	Support	Learn	Get in touch
Product	About	Documentation	Data Warehouse Guide	Get a Demo
Integrations	Partners	Community Forum	Content Library	Contact Sales
Customers	Careers	Status Page	Panoply Blog	Contact Support
Pricing	Panoply Swag		News & Events	Contact Us

panoply

No preview

2

Learn More

Live Webinar May 30:   Surviving and Thriving as a Data Team of One