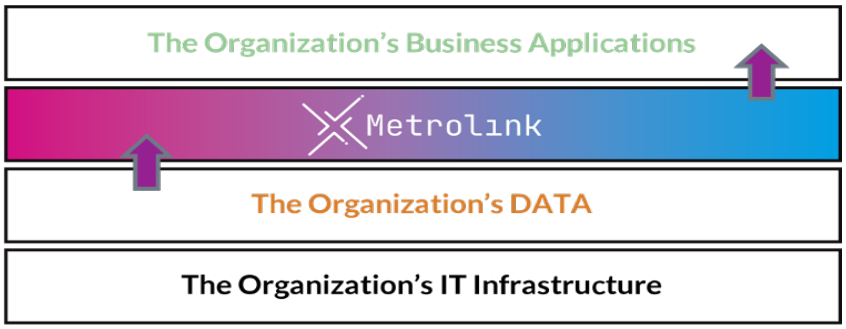
Datorios

Platform overview

Datorios offers a Data Transformation Framework that enriches the data engineer's toolkit, and saves them the burden of time-consuming tasks and the ones that are not a challenge anymore and gives them the infrastructure on which they can focus on **results** and complexity. Our ready-made data transformers and the ones that you will develop, run on a modern microservices infrastructure with autonomous scaling and dynamic optimization for unprecedented throughput and great latency. You get the latest technologies like Kafka, Kubernetes and more just waiting for you to operate.

Datorios is installed side-by-side with the organization's existing infrastructure. Once connected, Datorios gives immediate value without the need for any major changes.



Key features and capabilities:

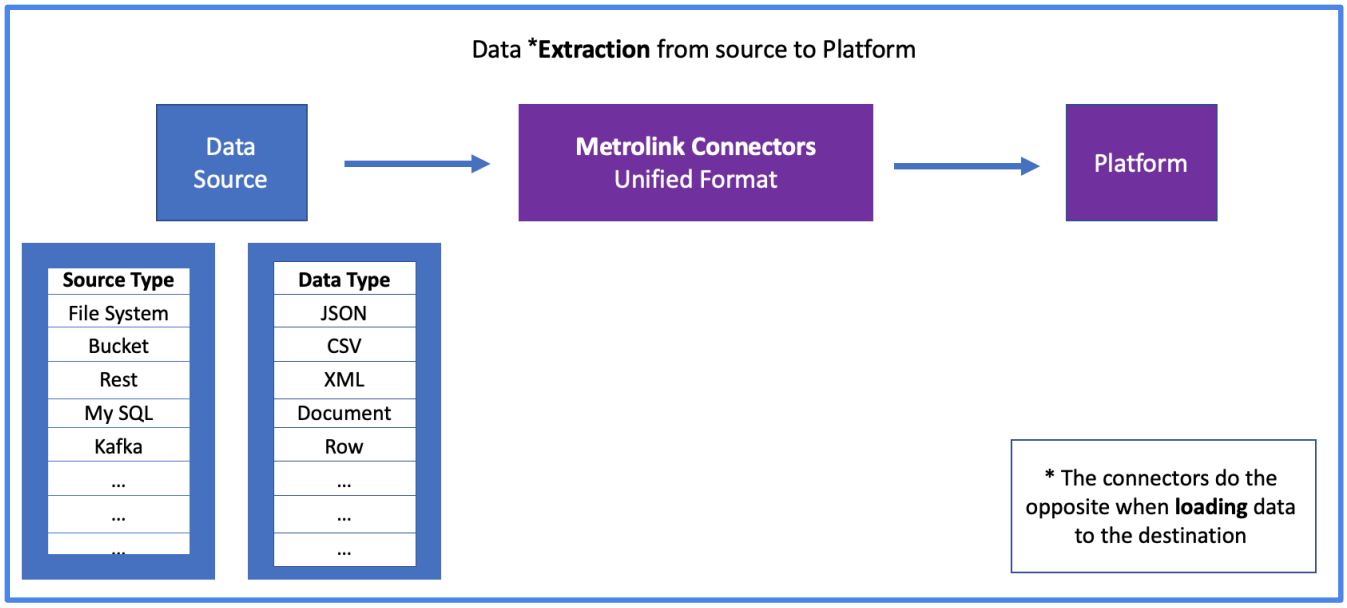
[1]Connectors - Decoupling data and applications

Datorios serves as a mediation layer between the company's data and applications, using two kinds of connectors:

- Extractors - Connect the platform to any data sources and detach the data from its specific technological and format features into the unified format used by the platform. The triggering mechanisms of the connectors are batch-, event-based and CDC. SQL connectors can run a specific SQL query within the connector.
- Loaders -connect the Datorios platform to any data destination (databases, API, application, topics) by transforming the unified Datorios format into the desired data format with the technological features of the destination.

The result is the ability to gain straightforward multisource data correlations and route any data source, through any manipulation, to any data need.

[1] For the full list of our connectors, contact us



Transformers

Transformers are the heart of Metrolink's framework - we give you unprecedented transformation and correlation abilities with data in-rest, in-motion and together for data ingestion, migration, warehousing, and science.
A chain of Transformers, each one can be configured to the desired result, letting the user build the logic of “story” data towards the needed business result.
The Transformer configuration is as easy as building an Excel function or using regex, saving time, and building high visibility.
The stack of Transformers consists of pre-built generic ones and the ability to develop and deploy more transformers using our open SDK.
The pre-built Transformers include:

- **Filter** - Passes only events not meeting defined conditions.
- **Mapper -** Creates a new event structure out of an event. The mapper can be used to regulate keys names and values, build conditional values, insert metadata to the event. The mapper is usually used to regulate data at the beginning of the pipeline and before loading it to the destination with a needed schema.
- **Distributor** - The distributor is a multi-filter that can distribute events matching a predefined condition and feed sub-pipelines.
- **Aggregator** - Keeps data events stored in memory for correlation purposes and allows statistical evaluation of data events.
- **Correlator** - Correlates un-synced events, meeting defined conditions and joins them for further transformation along the pipeline. Correlations can be used for data enrichment, multi-source joining and state-based filtering.

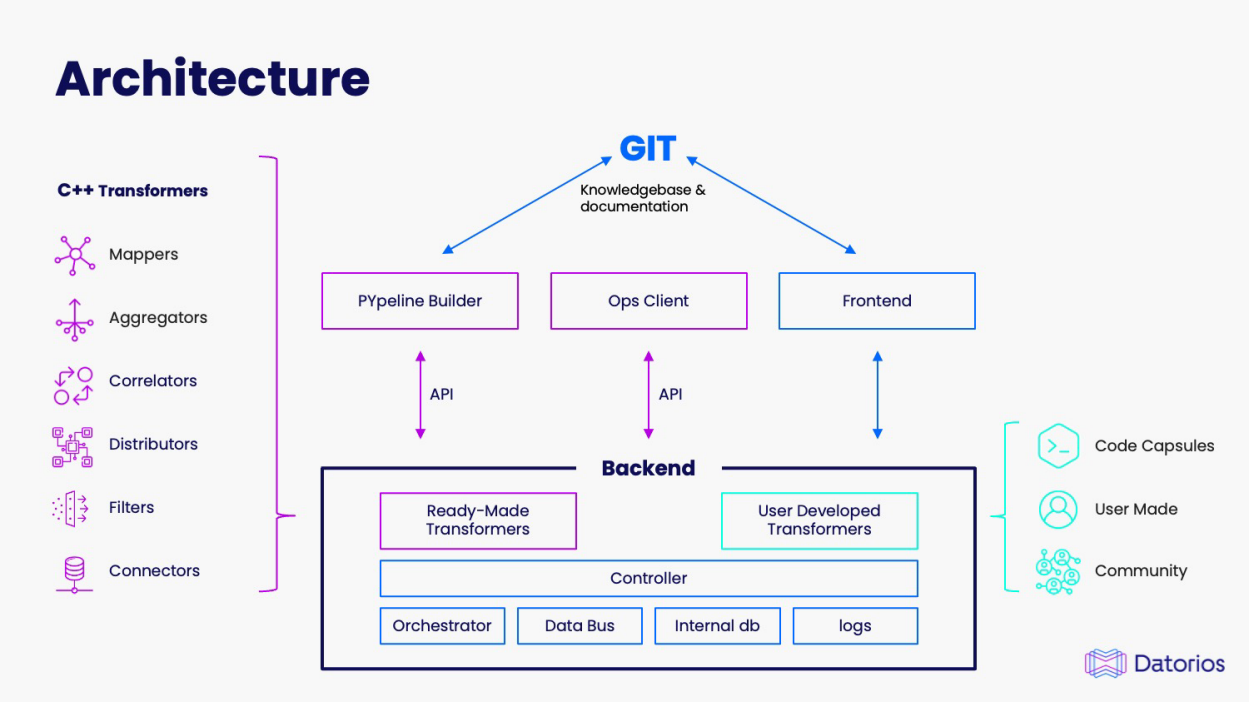
- **Enricher** - Event triggered encapsulated user code or scripts that creates a new event out of the current event. The enricher is also used for calling services outside the platform sending them the event's value and getting the results (legacy processing, high GPU machines, etc.). The enricher can be used for transforming data from unstructured to structured, with ML models embedding and other scripts' integrations along the pipeline.

For the best performance, the transformers are written in C++ and the platform's controller arranges a set of transformers into optimal DPUs for best I/O. As a result, high throughput and resource optimization are achieved.

Using our open SDK, advanced users can develop and deploy their own transformers and connectors. The user-defined transformers building blocks are added to the pre-built stack, and benefit from the auto-scaling and monitoring of the platform.

Integration

Datorios is easily integrated to your current development environment and CI/CD procedures. The backend API allows the deployment of pipeline design (as declarative configuration), the operation and monitoring of the platform.

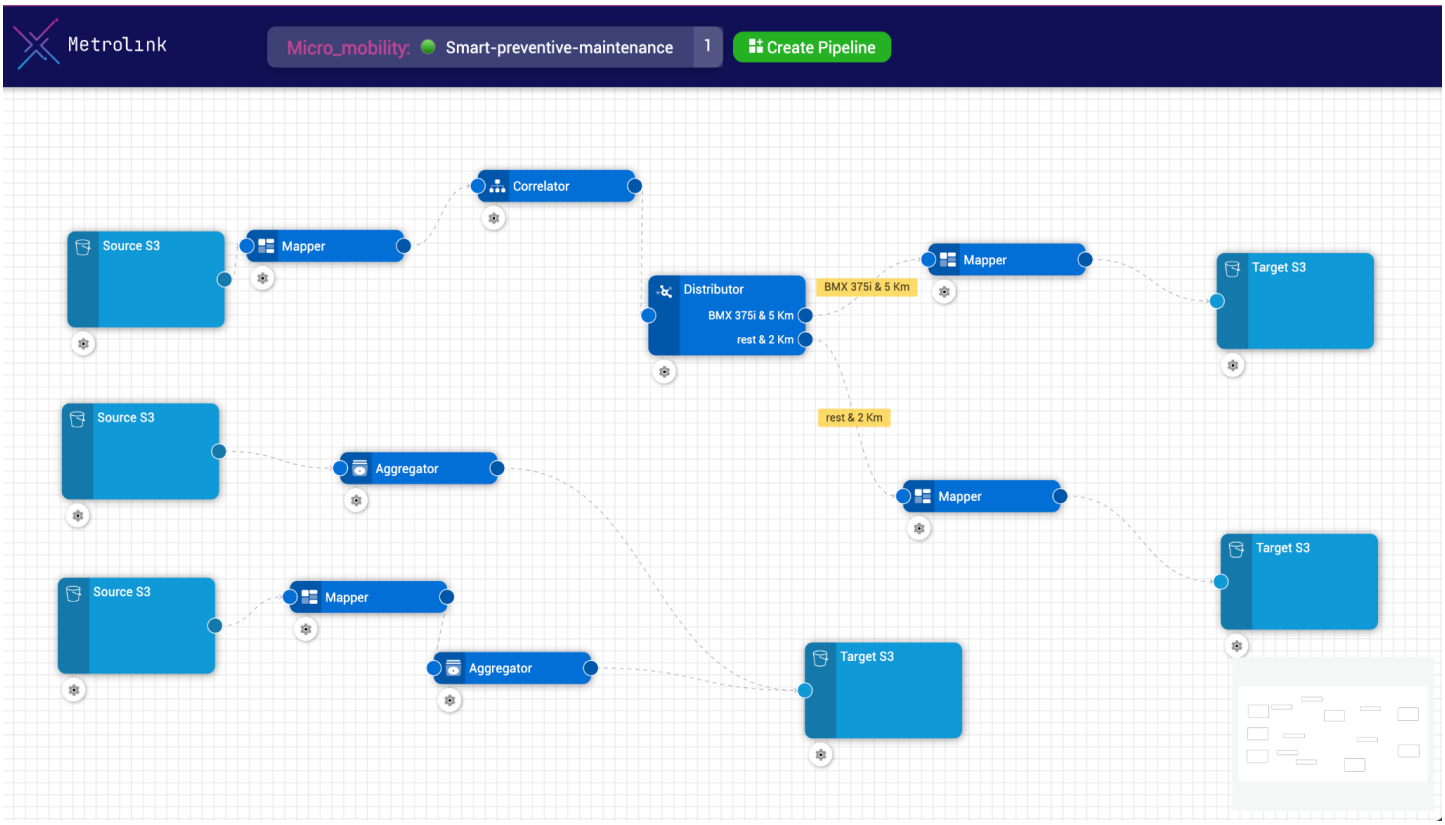


Visualization interface & Design tool:

Our frontend is a powerful visualization tool that lets you see in a glance, at any given time, a full view of the entire pipeline environment and its operational status including local and global metrics from the specific transformer to the pipeline and the entire resources consumption and status.

The frontend allows the probing of data after each transformation as part of the design or debugging process.

The frontend serves as design and/or configuration tool and can commit to pipeline versions as part of the development cycle. This allows a quick configuration process and the true collaborations with other data functions.



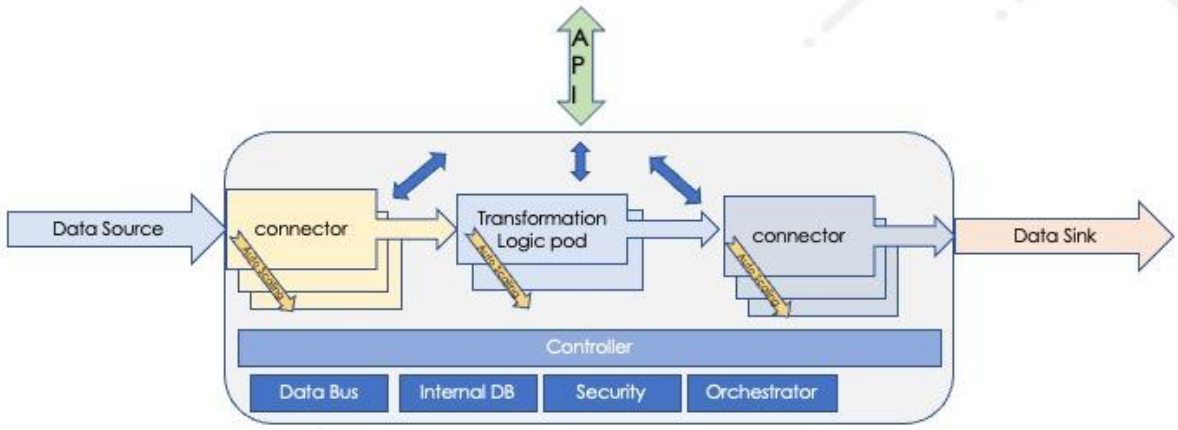
Streaming data processing and event-based pipelines:

Modern data use-cases have growing demands for dealing with streaming/IOT data. The huge amount of data from sensors and the motivation for low latency results produce an increasing need for data processing along the pipeline and the reduction of pipeline results to insights instead of raw data. Existing ETL solutions, built for batch processing, are far from optimal for dealing with data in motion.

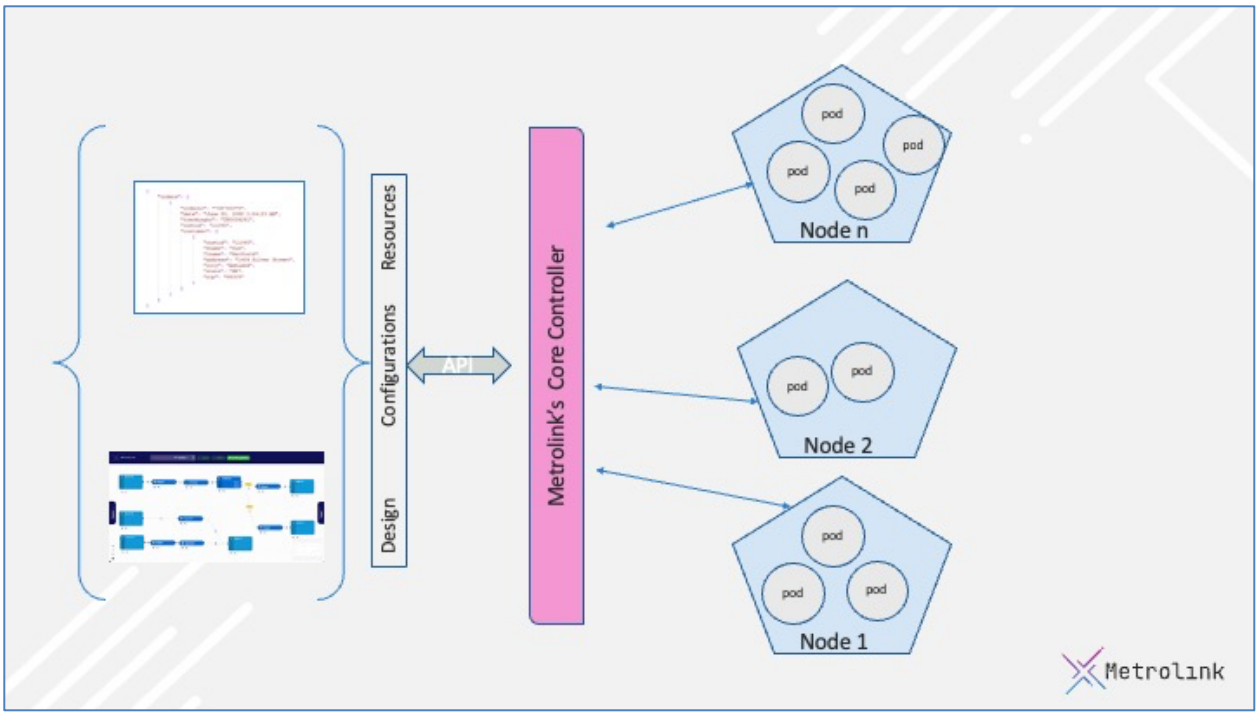
- Unlike most of the competitors we are streaming native
- Pipelines are triggered by events both from streaming events and query results, CDC and batch events
- Correlation between any number of sources from any kind of data (streams / data in rest)
- The basic installation can handle more than 250k events per second

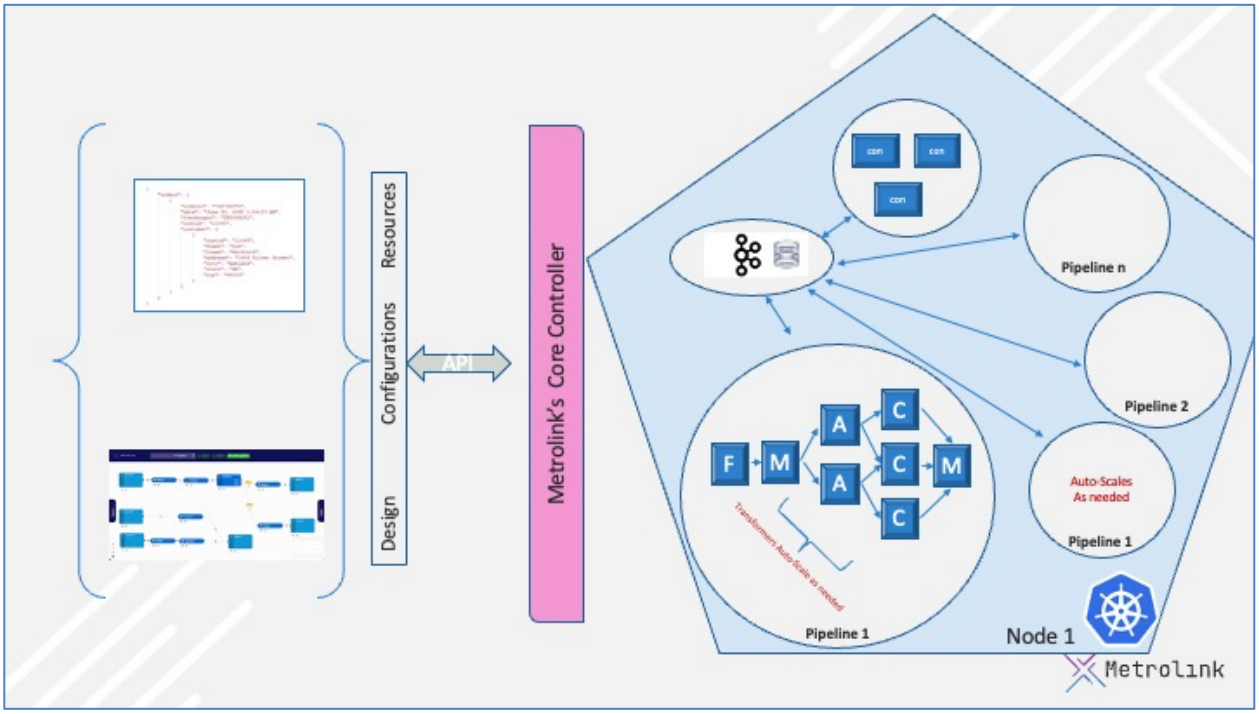
Autonomous scaling:
After deployment, each part of the pipeline is continuously monitored. The platform's controller identifies performance bottlenecks along the pipeline and automatically scales (and de-scales) DPUs for throughput optimization. Resources are distributed according to load, throughputs, and pipeline prioritization. Under the hood we have a unique implementation of DPUs & Kafka over K8s.

Auto-scaling allows for efficient usage of the overall computing power. The overall load and usage are constantly monitored and give the system administrator an alert where more computing power is needed.



Metrolink





Trust and reliability:

The controller orchestrates the data flow and ensures pipeline reliability. Moreover, when internal microservices fail, the controller maintains data integrity by ensuring that each event is processed once and only once.

Statistics and monitoring:

Datorios is highly verbose by design, uniquely designed to give high visibility with minimal influence on overall pipeline performance., Layers of statistics and alerts are implemented on top of the collected information. Monitoring is available within the embedded dashboard or by exporting raw data to an external application.

Privacy:

Metrolink.ai is deployed on the client's infrastructure (cloud/on-prem) and inherits the specific client's security policies.
Metrolink.ai is designed for security from day one, meeting the highest standards in the industry, with a state-of-the-art, embedded integrity mechanism to prevent pipeline abuse.

Installation requirements and process:

Metrolink's platform is available in both cloud (EC2) and on-prem installation. The installation requirements are:

- **Operating System requirements**
  Now Ubuntu 20.0.4 is the only operating system supported by Datorios software

- **Hardware requirements**
  A minimum of 3 servers that will run the software should meet the following hardware requirements:
  - **Minimum hardware requirements:**
    CPU: 8 Cores Memory: 16 GB storage: 500GB
  - **Recommended hardware requirements:**
    CPU: 16 Cores Memory: 64 GB storage: 2TB

The installation package is sent to the user obtaining all that is needed, no internet connection required.
Installing the platform takes only a few minutes. The Datorios technician support team is always available for assistance, if necessary.
Once Datorios is installed by the administrator, the following actions need to take place:

- Set up the user management roles following the company's policy
- Configure the sources and targets by setting up the connectors

Contact: contact@datorios.com / cs@datorios.com