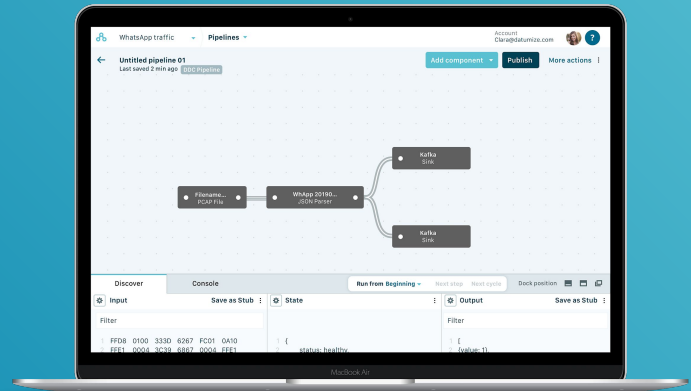


Datumize Data Collector (DDC)

Powerful data collection engine for sophisticated data sources.



DDC Version 5

Reliable real-time data collection for Dark Data sources

Datumize Data Collector (DDC) is a lightweight, high-performance, streaming enterprise data integration software focused on dark data collection from sophisticated and complex sources that most vendors don't cope with.

DDC allows companies to integrate data from various sources, such as in-transit network transactions, Internet of Things (IoT), mobility and industrial protocols, and enrich their customer, business and operational intelligence with the resulting new data insights.

Product features:

- **Modular:** every functionality is a pluggable component.
- **Non-intrusive:** adaptive data gathering to access data sources either unobtrusively (network traffic sniffing and Deep Packet Inspection) or in a polling (request-response) fashion.
- **Real-time:** data is processed in streaming mode with minimal latency.
- **Remote Management:** installation and upgrades automated with no downtimes and reliable configuration.
- **Graphical Management:** intuitive GUI for system configuration.
- **Edge Computing:** lightweight agent runs on minimal hardware, allowing acquired data to be enhanced, filtered or joined with other sources prior to delivering it for further processing or storage.
- **Highly Scalable:** designed with multithreading and concurrency for best performance in the most demanding environments, yet can also run on single-board computers.
- **Throughput:** of up to hundreds of millions of transactions per day.
- **Security and Privacy:** data integration always happens within your premises to ensure security and privacy best practices and laws.

Benefits



Collect and leverage Dark Data - data not stored and therefore not used.



Expand your data by tapping into your organization's Dark Data.



Manage large volumes of sophisticated data for real-time ingestion.



Discover new business and operational insights.

Complete DDC documentation, extra resources and testing sandbox available at:

<https://docs.datumize.tech>

Datumize Data Collector (DDC)

Powerful data collection engine for sophisticated data sources.

Sample Use Cases

From Network Transactions to Real Demand insights

DDC network traffic connectors leverage network sniffing and deep packet inspection (DPI) techniques to capture live network traffic and crunch several protocols such as HTTP, SOAP or industrial protocols like OPC-DA.

That allows gathering transactional data from an availability web service integration and transform it into valuable real demand and customer insights. If the returned product availability is empty, you are missing a sale.

From Wi-Fi data to motion intelligence insights

API traffic can be both polled or sniffed with DDC. The sniffing feature is particularly handy for the IoT and Industrial domains where you find many closed and proprietary data sources.

Wi-Fi and Mobility devices are usually polled by using SNMP or RTLS APIs. That allows, for example, the collection of very low-level metrics from an existing Wi-Fi infrastructure in a logistics warehouse or a hotel, and transforming it into valuable location-based insights and motion intelligence (position, movement, distance), without the need for costly infrastructure investments.

Product Specifications

General Specifications

- Operating system: Runs on Linux or Windows (Java runtime 8)
- Environment: from Raspberry Pi to high-end multi-core cloud servers
- Configuration: intuitive GUI tool (Datumize Zentral)
- Deployment: Automated deployment via encrypted connection to destination hardware or manual installation for highly secure, disconnected environments.
- Fully programmable edge computing support via Python or Groovy

Data Sources

- Network PCAP files
- USB, REST and HTTP
- Socket
- SNMP and RTLS APIs
- Industrial: OPC-DA, Modbus, Fieldbus.
- Relational: MySQL, Postgres, Oracle, JDBC.
- NoSQL: MongoDB, Elastic, Redis.
- Files: text, CSV, JSON.

Data Processors

- IP/TCP/UDP assembler: OSI Level 4, timeouts, client/server, payload & headers extract, missing packets drift.
- Specialized HTTP assembler.
- SQL database traffic: MySQL.
- Mobility assembler: trilateration, fingerprint.
- AI assembly: path finding.
- Advanced filtering.
- Enrichment: lookup tables.
- XML/JSON payload parser, supports malformed/missing content (data drift).
- Cook processor (script) for edge processing.

Data Sinks

- File storage: binary/text/xml/json, partitioning.
- API/REST endpoints.
- Big data storage: AWS S3, Hadoop, Google.
- Database: JDBC (SQL)
- Message brokers: Kafka, JMS, MQ.

Full list of connectors can be found at docs.datumize.tech