
<p><b>CONNECTED</b></p> <p>Connect to all your data sources even if they are behind the firewall. Build data integration APIs.</p>	<p><b>EASY</b></p> <p>Select from multiple of built-in common data integration scenarios such as export, import, data replication, backup and many others.</p>	<p><b>COST EFFECTIVE</b></p> <p>Let us host and manage a solution for you or install it on your own hardware and have a full control.</p>	<p><b>SIMPLE</b></p> <p>Manage your own data integration workflow using simple drag and drop interface. Test and schedule flows with just a few mouse clicks. No coding, no complicated build-test-deploy process.</p>

## Etlworks Integrator – cloud data integration platform

WHITE PAPER

## Table of contents

- Overview ..... 2
- Connectors ..... 2
- Features ..... 2
  - Cloud-based solution ..... 2
  - Integrating heterogeneous data sources..... 3
  - On-demand, real-time and scheduled execution ..... 3
  - Automatic and manual mapping ..... 3
  - Change replication ..... 3
  - Web services ..... 3
  - Enterprise Service Bus..... 4
  - HL7 ..... 4
  - Support for online data warehouses ..... 5
  - Automating complex workflow..... 5
  - High level transformations..... 5
  - Executing SQL on any data source ..... 5
  - Data integrity ..... 5
  - Business Intelligence (BI) and data visualization ..... 6
- Benefits ..... 6
  - Connected ..... 6
  - Easy to use ..... 6
  - Cost effective ..... 6
  - Simple, yet powerful ..... 6
  - Working with data behind firewall ..... 7
- Performance ..... 8
- Security ..... 9
  - Perimeter protection ..... 9
  - Encryption during data transmission ..... 9
  - Encryption of credentials ..... 9
  - Encryption of files ..... 10
  - Encryption of the JWT tokens ..... 10
  - Protection for the API endpoints ..... 10

Application Data and Credentials .....	10
Monitoring .....	10
Disaster Recovery and high availability .....	10
Architecture .....	11
Deployment .....	11
Technology stack.....	12
Clients.....	12
Why Integrator?.....	12

## Overview

Etlworks Integrator is an all-in-one cloud platform for all your data integration projects, regardless of data location, format and volume. It is a completely online solution and does not require installation of any local software, except for a web browser.

## Connectors

You can access data in relational databases, files, web services, and social media web sites.

Databases	Data Exchange Formats	File Storage	Web Services and APIs	Other
MS SQL Server	CSV	Local files	any REST	HTTP
Oracle	Fixed length	FTP	any SOAP	SMTP
MySQL	XML	SFTP	Marketo	IMAP
PostgreSQL	XML Data Set	Amazon S3	Google Analytics	POP3
DB2	JSON	Google Cloud	Amazon MWS	MLLP
Informix	JSON Data Set	Storage	Facebook	JDBC
Sybase	HL7 2.x	Microsoft Azure	Twitter	
Snowflake	HL7 3.x	Storage	Journey	
SQLite	Excel XLS	WebDAV		
Any JDBC	Excel XLSX			
Redis	PDF			
	HTML			

If the connector you are looking for is not on the list, we can build it for you in no time.

## Features

### Cloud-based solution

Integrator is a completely online solution with cloud-based servers, allowing end-users to access it from any web browsers. Integrator does not require installation of any local software, except for a web browser.

## Integrating heterogeneous data sources

In Integrator, all data sources are treated equally. You can easily connect databases to files and web services, and vice versa. At the same time, Integrator supports all format-specific features. For example, it automatically generates PL/SQL code if the destination is an Oracle database or uses PostgreSQL specific syntax for MERGE (UPSERT).

## On-demand, real-time and scheduled execution

Flows created in Integrator can be scheduled or executed manually. Scheduled flows can be executed as often as every 30 seconds or triggered by an external event.

## Automatic and manual mapping

Integrator supports automatic and manual per-field mapping for all data sources, regardless of the format and location.

When automatic mapping is engaged (by default), it is assumed that the destination field names are the same as the source field names. It is also assumed that some of the source fields do not exist in the destination, and vice versa.

Manual per-field mapping is performed using graphical drag & drop interface. Integrator can suggest default mapping, the user accept it or rename fields, disable fields, add fields, define rules to calculate field values, and much more.

## Change replication

Integrator supports change replication between all data sources using the high watermark concept. High watermark is the highest level of field value that can be used for change replication.

Examples:

- Change replication between two databases: the flow pulls changed records from the source database and updates the destination database.
- Incremental database backup: the flow pulls changed records from the source database and packages them into the "update" files.
- Get recently changed records from the API and load them into any supported destination: the flow pulls changed records from the supported API or web service and loads them into any supported destination, for example a data warehouse.

Change replication can be performed in [quasi] real time (as often as every 30 seconds) or on-demand, when the flow is triggered by a call to the Integrator's end-point.

Integrator natively supports MERGE (upsert) operations for the following databases: MS SQL Server, Oracle, DB2, PostgreSQL, MySQL, Informix, and Sybase.

## Web services

Integrator supports practically any REST or SOAP based web services using a generic HTTP connector. The following authentication methods are supported:

- Basic (user name and password)
- One-step OAUTH1

- Two-step OAUTH2
- API token
- Any combination of the aforementioned methods

In addition to the generic HTTP connector, Integrator includes a growing list of connectors to the third-party APIs:

- Facebook
- Twitter
- Marketo
- Google Analytics
- Journey
- Amazon MWS

In many cases, it is possible to develop a connector to the third-party API not currently on the list or to a service, which does not have a public API at all. For example, one of our clients developed a connector to Google Trends (which does not have a public API) using a combination of generic HTTP connectors and complex multi stage flows. The connector takes an input from the database, creates a quasi JSON payload, and makes a request to the Google Trends backend. It then scrubs the response and creates well-formatted JSON. All without writing a single line of code!

### Enterprise Service Bus

Integrator includes a full-blown enterprise service bus (ESB). The following scenarios are supported:

- Building external APIs, which can be used to query data behind the firewall and create data feeds.
- Building CRUD APIs.
- Executing flows on demand.

### HL7

HL7 is a combination of the data exchange formats and communication protocols used to connect to health care applications.

Integrator fully supports all existing HL7 formats and protocols, and is one of the best available HL7 interface engines.

In Integrator, the HL7 messages can be sent and received using one of the following connectors:

- HTTP connection - send and receive.
- HTTP Listener - receive.
- Cloud or file-based storage - send and receive.
- HL7 MLLP Sender - send
- HL7 HTTP Sender - send
- HL7 MLLP listener - receive

Supported HL7 formats:

- HL7 2.x
- HL7 FIHR

## Support for online data warehouses

Integrator includes several pre-built flows, optimized for the online data warehouse Snowflake:

- Change replication from any of the supported databases to Snowflake
- Direct file upload into Snowflake
- Upload data from the Web services and APIs into Snowflake

## Automating complex workflow

In Integrator, it is possible to build very complex data integration and automations flows using our drag and drop interface, as well as scripting languages, such as JavaScript.

### Some of the supported scenarios:

- Pipelines – flows conditionally executed in order
- Loops – flows executed in a loop, for example: until there are files in the folder, for all records in the database table, etc.
- On-demand execution – flows executed when certain events happen
- Parallel execution – multiple flows executed in parallel
- Validation and verification
- Using JavaScript to perform very complex logic and calculations

## High level transformations

Integrator supports very high-level transformations, without writing a single line of code:

- Filters
- Sorting
- De-duplication
- Extracting dimensions from the complex nested data objects
- Partitioning
- Merging
- JavaScript transformations
- XLST transformations

## Executing SQL on any data source

In addition to running any SQL with any of the supported databases, Integrator supports running SQL on any of the supported data sources, including files and web services. The built-in SQL engine supports filters (WHERE clause), grouping and aggregated functions (SUM, MAX, MIN, AVG, COUNT), unions and other set operations, and sorting (ORDER BY).

## Data integrity

In Integrator, it is possible to build complex validation logic using JavaScript. The validators can be used to:

- Reject entire datasets
- Reject single rows
- Generate exceptions so that the flow execution halts completely

## Business Intelligence (BI) and data visualization

Integrator is a powerful BI tool. It is possible to visualize and query data in any of the supported formats:

- Databases
- Files
- Web services
- Key-value storage

### Supported functionality:

- Browse database metadata: databases, schemas, tables, views, columns
- Browse files: folders, files, columns
- Browse web services: endpoints, columns
- View data in a grid
- Execute any SQL on any supported data source
- Execute multiple SQL statements at the same time
- Execute DDL and DML SQL statements
- History of the executed SQL statements
- View data in the raw format (JSON, XML, etc.)
- Export any data set into any of the supported formats and download the file on your local machine
- Upload local files into any connection (local and remote) using drag and drop

## Benefits

### Connected

You can connect to all your APIs and data sources, even if they are behind the firewall. Build data integration APIs. View and explore data and metadata right in the browser.

### Easy to use

Select from multiple of built-in common data integration scenarios such as export, import, data replication, backup and many others. No need to deploy anything or learn a new programming language.

### Cost effective

Let us host and manage a solution for you or install it on your own hardware and have full control.

### Simple, yet powerful

Manage your own data integration workflow using a simple drag and drop interface. Automate very complex data integration tasks with just a few clicks of the mouse.

## Working with data behind firewall

Q. MY DATABASE IS IN A CORPORATE NETWORK, BEHIND THE FIREWALL. HOW CAN INTEGRATOR ACCESS IT FROM THE CLOUD?

A. We install and configure a zero-maintenance background service called INTEGRATION AGENT right in your network, which talks to your corporate database and to the Integrator on the cloud.

Q. HOW DOES THE INTEGRATION AGENT COMMUNICATE WITH A DATABASES AND APIS IN A CORPORATE NETWORK?

A. Integration Agent works behind the corporate firewall, so it communicates with databases and APIs directly. It uses JDBC to talk to the databases and HTTP to get and push data to REST and SOAP APIs.

Q. HOW DOES THE INTEGRATION AGENT, WORKING BEHIND THE CORPORATE FIREWALL, COMMUNICATE WITH INTEGRATOR ON A CLOUD?

A. Integration Agent can push data to the secure cloud file storage such as Amazon S3, which is accessible by Integrator, or it can push data directly to the Integrator REST API.

Q. IS IT POSSIBLE TO LOAD DATA INTO MY CORPORATE DATABASE OR API USING THE INTEGRATION AGENT?

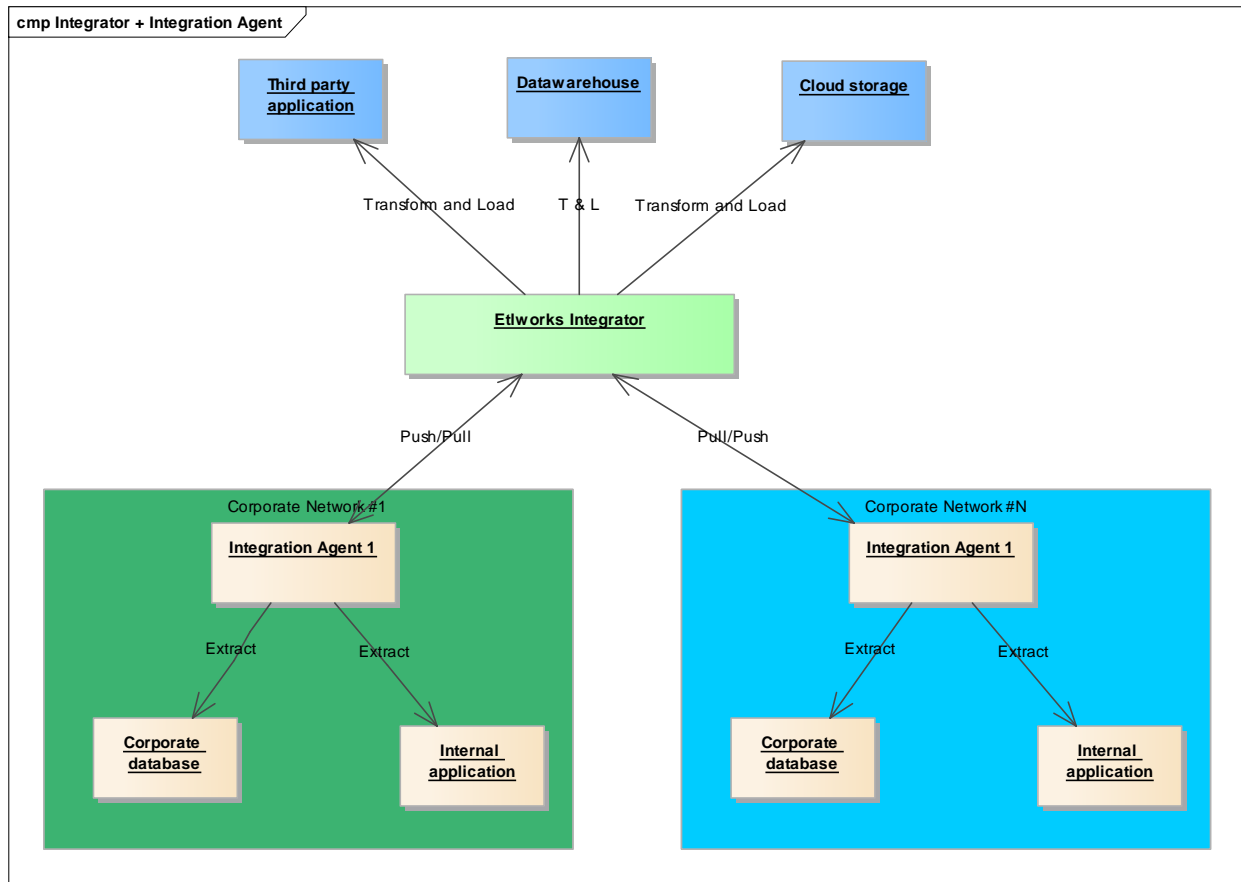
A. Integration Agent can load data into the database or API just as well as extract it.

Q. WHY DO I NEED A CLOUD-BASED INTEGRATOR IF I HAVE THE INTEGRATION AGENT WORKING BEHIND MY CORPORATE FIREWALL?

A. Integration Agent is limited to simple change replication, extract and load scenarios. Integrator and Integration Agent work together to support very complex data integration flows with transformations, mapping and logic.



The following figure shows how Integrator works together with Integration Agent:



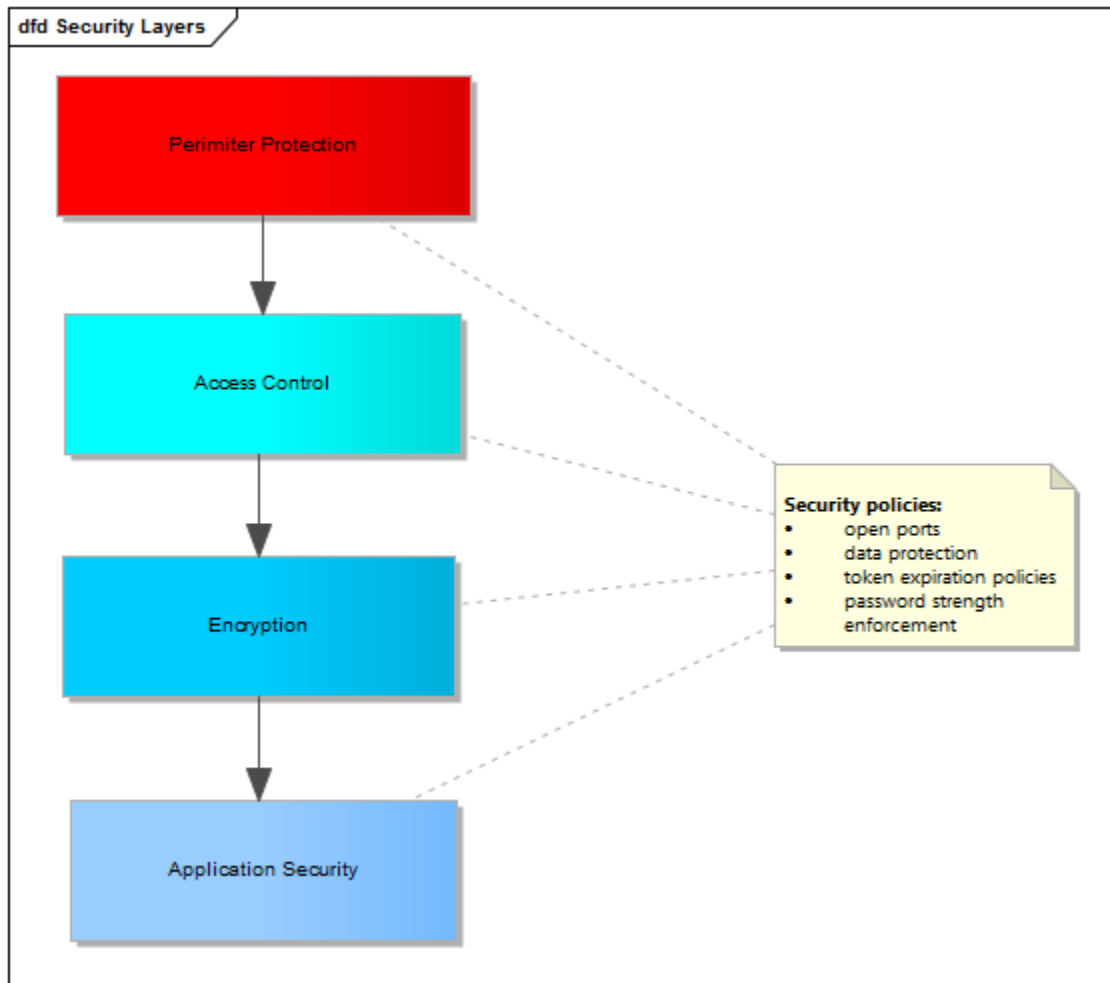
## Performance

Integrator is built with performance in mind. If needed, the back-end can be deployed to the multiple servers behind the load balancer so it can scale to permit billions of concurrent operations to run at the same time. Performance monitoring is also built-in, so we can easily spot and fix bottlenecks.

To improve performance of the ETL operations, we created a unique streaming engine, which allows Integrator to transform practically unlimited datasets from the source to destination, while maintaining a very small memory footprint.

## Security

We take security of our customers' data very seriously. We have implemented stringent safeguards to ensure the security of your data.



### Perimeter protection

We install and configure the following elements of the perimeter protection infrastructure:

- system firewall
- reverse proxy server
- load balancer

### Encryption during data transmission

We encrypt all sensitive information transmitted to and stored on our servers.

### Encryption of credentials

In Integrator, all passwords, access keys, and other credentials are encrypted using a strong encryption algorithm with a 512-bit private key.

## Encryption of files

In Integrator, you can configure flows to store all files in the encrypted archive. This policy is not enforced.

## Encryption of the JWT tokens

In Integrator, JWT tokens are hashed using strong encryption algorithm with 512-bit private key.

## Protection for the API endpoints

All API endpoints in Integrator, including those that are private and protected by the short-lived JWT tokens.

## Application Data and Credentials

Our data protection policy is very simple – typically, we don't have access to your data at all unless you opt-in to store it on our servers.

### Exceptions:

- passwords and other secure credentials - we always encrypt them.
- data files which you decide to store on our servers - you have an option to store data in our password protected archive files.

### Policies:

- we never send passwords and other secure credentials to the web browser, so there is no way to see them anywhere in the Integrator.
- we monitor and regularly install security patches for all elements of the infrastructure, including the core operating system.

## Monitoring

Our monitoring suite is a combination of third-party services and home-grown solutions, based on industry standard elk stack. It includes the following elements:

- uptime monitoring using uptimerobot.com
- health dashboard
- log collection.

In case of downtime or any critical problem in the infrastructure elements, we are notified within a few minutes. Our support team is able to jump right on the problem and fix it without disturbing the customer's operations.

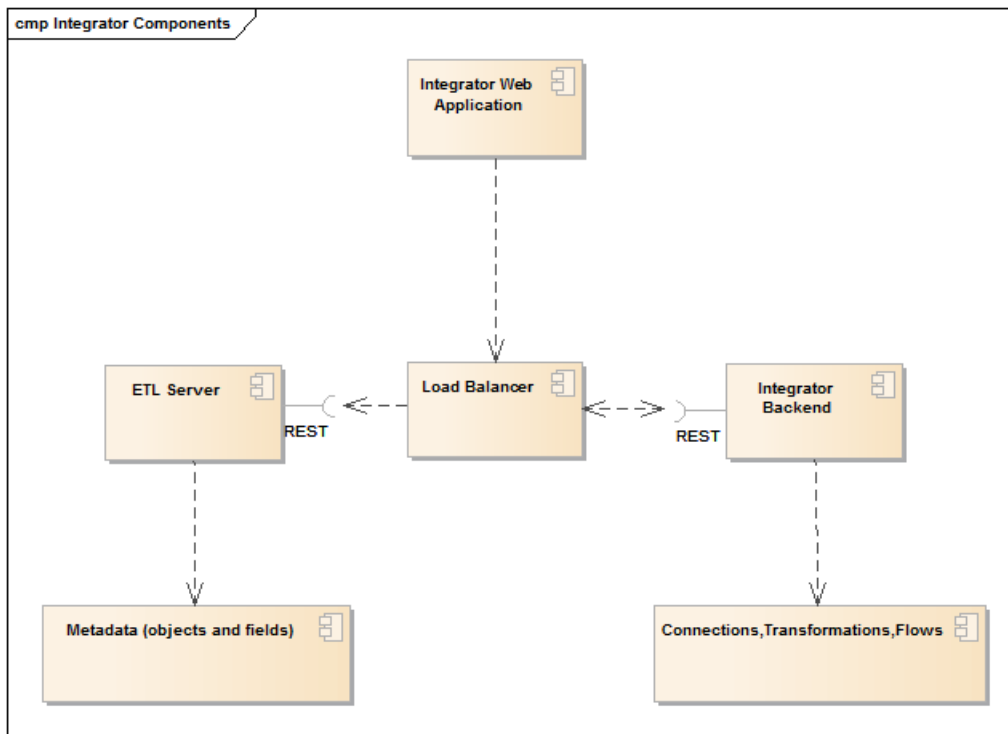
## Disaster Recovery and high availability

Our disaster recovery plan includes the following elements:

- daily database backups are transferred to the encrypted Amazon S3 bucket.
- High Availability (HA), zero downtime deployment option is available.

## Architecture

The following figure shows Integrator components:



User can access the Integrator via the web based interface.

Web based interface communicates with back end using REST web services.

Integrator back end calls (via the REST interface) independently deployed ETL Server which handles all data integration tasks.

Back end components can be deployed to the multiple nodes in a cluster, behind load balancer, which allows Integrator to scale horizontally and support a very high volume of transactions.

Flows, connections and formats, as well as users and run time information are stored in the Integrator database.

Cached objects, such as table and view names, fields, file names, API endpoints, etc. are stored in the metadata store.

## Deployment

Integrator can be deployed to one or multiple servers behind the load balancer on the cloud. We typically use Amazon Web Services (AWS) or Microsoft Azure as our cloud providers. If needed Integrator can also be deployed on-premise.

We use deployment and provisioning automation to simplify upgrades to new versions. If needed, zero downtime deployment process can be implemented.

## Technology stack

**Back-end:** Java, Tomcat

**Front-end:** EmberJS, Bootstrap CSS

**Database:** PostgreSQL

**Key-value storage:** Redis

## Clients

Here is a selected list of our customers:

- Data Systems Integration Group
- Transform Digital
- Cargoclic
- transport.nsw.gov.au
- Savino Del Bene
- Artis
- Sage
- Cerebral Palsy Alliance
- Bank of New Zealand
- Marketo
- Acer USA

## Why Integrator?

Integrator is THE most powerful, yet cost effective, cloud data integration tool on the market. We work directly with a customer, without layers of bureaucracy, to ensure that our system meets your needs. We implement features requested by the customers in a matter of weeks, if not days. Our support is a best in class. Connect to everything with us!