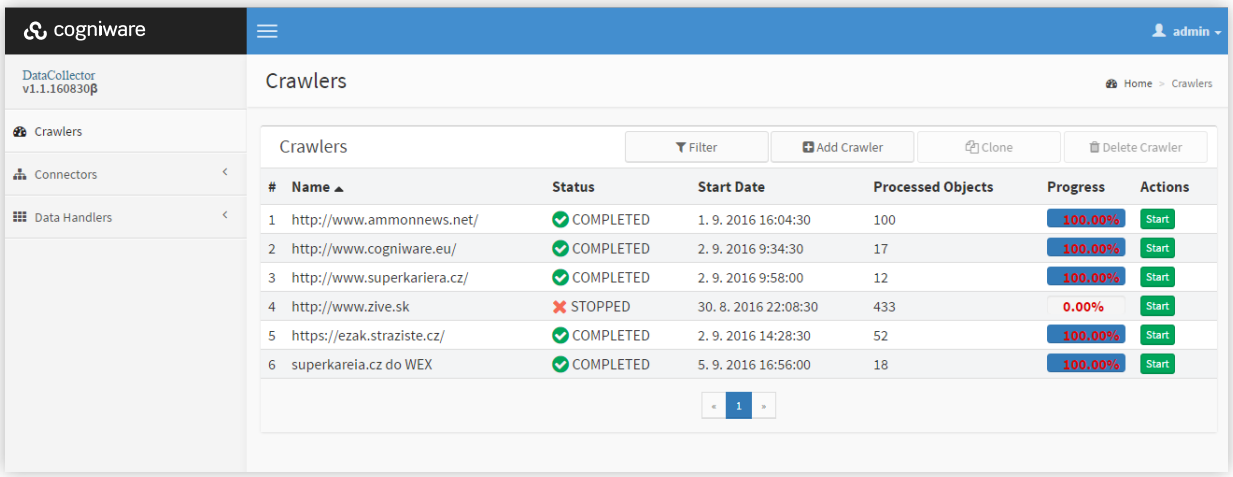
cogniware

Cogniware Data Collector

The ultimate solution for data aquisition from online and offline sources

PRODUCT SHEET

Cogniware Data Collector (CWDC) is an enterprise, robust and scalable system for collecting information from social networks, websites, forums, e-mails, existing local data sources and various other sources for further analysis in specialized analytics software systems, such as IBM Watson Explorer Content Analytics.



OVERVIEW & MAIN FEATURES

Cogniware Data Collector consists of various components such as the Core, User Interface, Connectors, Data Handlers and API.

CWDC Core	Job management, scheduling, load balancing, crawler failure recovery
CWDC User Interface	User interface for CWDC management
CWDC Connectors	CWDC component fetching data from source system, e.g. social media
CWDC Data Handlers	CWDC component sending collected, normalized and cleansed data to target system
CWDC Crawler	CWDC Job definition consisting of one Connector configurations mapped to one Data Handler configuration
CWDC Open Framework & API	Open Framework and API for custom connectors and data handlers development

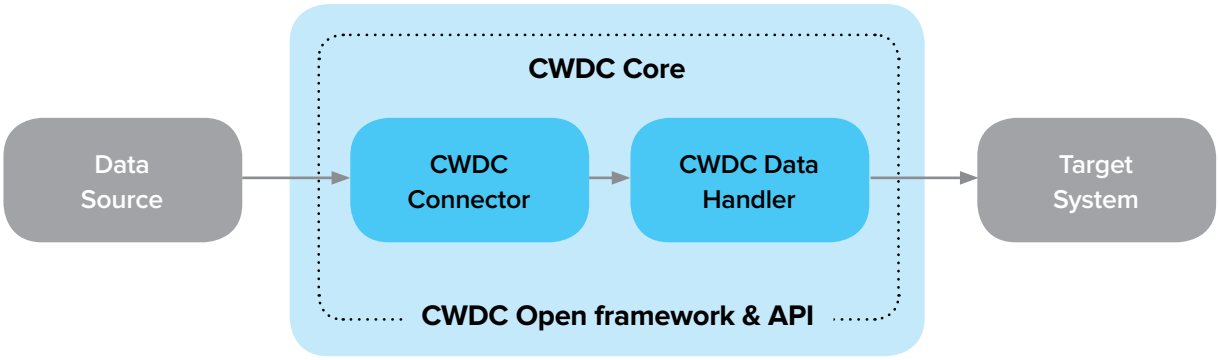
Core

Cogniware Data Collector Core is the heart of the system managing Connectors and Data Handlers by providing an extensible enterprise level configuration mechanism. Configurations such as Connector connection credentials, Crawler query configuration, scheduling information and Data Handler connection information are stored in the product built-in database.
The framework allows **multiple parallel crawling sessions** for different queries, where jobs can be defined as full crawl jobs (for initial crawling) or incremental jobs (when scheduled or run manually).

This approach enables collecting of structured and unstructured data such as social network posts, comments, user profiles or web pages, as separate entities while taking into account the relations between them. These relations are also stored as separate objects and have similar structure to the mapping table in relational database.

User Interface

Cogniware Data Collector User Interface is a HTML 5 user interface application, which interacts with the configuration database using REST API of the Cogniware Data Collector Core. It provides functionality for the system management using Administrator and Analyst roles to improve and define responsibilities and security settings



Cogniware Data Collector Architecture

Connectors and Data Handlers

Cogniware Data Collector enables collection of publically available data from social networks API, RSS channels, mailboxes, websites and offline sources, and stores the collected data on filesystem, in database or in IBM Watson Explorer Content Analytics. The output recorded in the filesystem is in JSON or XML format and the database output can be arranged into structured tables. Cogniware Data Collector internally stores information about the crawler's current state in case a failure of the crawler occurs. When a crawler encounters an error, the crawler shall be run again and the information about its state is restored from the database and the crawler will resume from the point where the error occurred.

The following online sources are currently supported:

Facebook | Twitter | Google+ | YouTube | Vimeo

Tumblr | Pinterest | Instagram | RSS | E-mail | Websites

The following output systems are currently supported:

IBM Watson Explorer Content Analytics | Filesystem | Database

Cogniware Data Collector contains the functionality of an API key recycling.

CWDC Social media connectors for Facebook and Twitter are able to recycle defined API keys to overcome the limitation with regards to the number of calls per one set. When a crawler encounters an API limit, it automatically switches to the next pair and when no API key is available, the crawler will wait for a short period of time before attempting to establish a connection again.

API

Cogniware Data Collector Core provides an API for developers to implement their own source system of Connectors (for data crawling) and their own destination system Data Handlers (to store data to). These Connectors and Data Handlers may transform the source data to enable an improved analysis of the crawled content.

USE CASES

Evaluation of Brand Perception by Customers

A marketing department can proactively monitor posts and comments on web forums and social networks, and thereby improve communication with clients, as well as manage the handling of potential threats to eliminate possible media pressure with regards to the product or brand.

Thanks to the implemented solution customer gains insights into:

- The public perception of the brand and whether it is generally positive or negative and in what context
- What are the main topics / keywords mentioned in connection to the brand or its products
- To what extent and in what connection are the company's products compared with products direct competitors

Cogniware Data Collector is vital part of the overall solution for data acquisition.

Monitoring Objects Related to Security Topics

Thanks to the broad range of CWDC Connectors, security organizations can continuously monitor and asses suspicious online activities (comments, posts, events), trace the surce of these activities and uncover the intentions of individual actors. This data can be valuable extension to the existing human intelligence databases.

cogniware

U Habrovky 247/11
140 00 Praha 4
Czech Republic

VAT ID: CZ02892081
E-mail: info@cogniware.com
Web: www.cogniware.com