Global Data Platform Architecture

Global Data Platform collects information from social and other providers, applies some transformation and makes the new data available for other consumers and services.

The general view of the architecture is:

# Connectors

Connectors are used to collect data from different data sources and store them in the data lake. A simple flow overview is the following:

## Requirements

### Configuration

1. Users should be able to configure the connector from the portal application.
2. Configurations should be stored in a way they are accessible to connectors.

### Security

1. Data at transit should be secured and encrypted (SSL)
2. Any credentials and authentication information should be encrypted before stored.

### Throughput

1. Connectors should be designed to handle high volume data and spikes. The connector should specify high throughput thresholds.

### Fault Tolerant

1. Data loss is not acceptable, therefore connector architecture should be fault tolerant.

## Twitter Connector

# Development

## Dependency management

For dependency management pip is going to be used.

## Unit Testing

Unit tests are written using pytest.