# System Documentation - Team 2

26 Aprile 2020

## Team Members

- Elia Vicentini
- Luca Landolfo
- Simone Gandini
- Adrian Munteanu
- Enrico Marocchio

## Contents

1) Introduction 2) Data collection and usage 3) Installation on a computer 4) How ETL-Support works 5) System Maintenance

## Revision History

| Version | Date | Name | Description |
| --- | --- | --- | --- |
| 1 | 26/04/2020 | Elia Vicentini | Initial document |

## 1) Introduction

The team behind ETL has the task of providing a script to manage, modify, update and provide a backup plan, for the data obtained from another project in our working environment, MarconiTT, a web application. This script has the task of speeding up, supporting and updating the data needed throughout the web app process. This document will provide some information about ETL-Support Program.

## 2) Data Collection and usage

The purpose of the ETL-Support program is to download data from one server, process it and then upload it to another server.

There are **2 servers** from which ETL-Support downloads data: - **edu-x04:** There is a direct connection to a mysql database - **edu-x08:** There is a connection via API at the following address: http://edu-x08:8090/etl/getData

The downloaded data is saved locally in the directory where the ETL-Support program is installed and started. The files will have the corresponding names depending on the server from which the download was made: - **database-direct.csv** (msqli database) - **database-api.csv** (api connection)

After data processing, the data is loaded into an SQL Server database on the **EDU-SYS01** machine.

**(All servers listed here are located within the institute and made accessible only from within)**

## 3) Installation on a computer

**Requirements**

- OS: Windows
- Python 3.7
- Python libraries:
    - PyMySQL
    - pyodbc
    - requests
    - tqdm

- argparse
- Git

**Installation**

- Clone the following repository in a directory: https://github.com/IlVice26/ETL-Support-2019.git
- Start **exportData.py** via cmd (Windows)

**Warning:** The program **may have conflicts** with pre-installed Python libraries. It is highly recommended to install the above libraries and **start** the program via **Python's "Virtual Environment"**

# 4) How ETL-Support works

To start the ETL-Support program you need to enter some parameters that are visible by writing: - exportData.py -h

The complete syntax is as follows: - exportdata.py [-h] --mode MODE [--noupload] [--mysql MYSQL SERVER] [--sql SQL SERVER] [-v]

| Arguments | Description |
|---|---|
| -h, --help | show this help message and exit |
| --mode MODE | program start mode: **api** or **direct** (required) |
| --noupload | the program will not load data on SQLServer |
| --mysql MYSQL SERVER | selection of the MySQL server present in the conf.json file |
| --sql SQL SERVER | selection of the SQL server present in the conf.json file |
| -v | it provides additional details as to what the program is doing |

# 5) System Maintenance

## 5.1) Python Code

No great experience of the Python programming language is required for code maintenance. The code is accessible via the link: https://github.com/IlVice26/ETL-Support-2019 **(The repository is private due to application access data security reasons)**