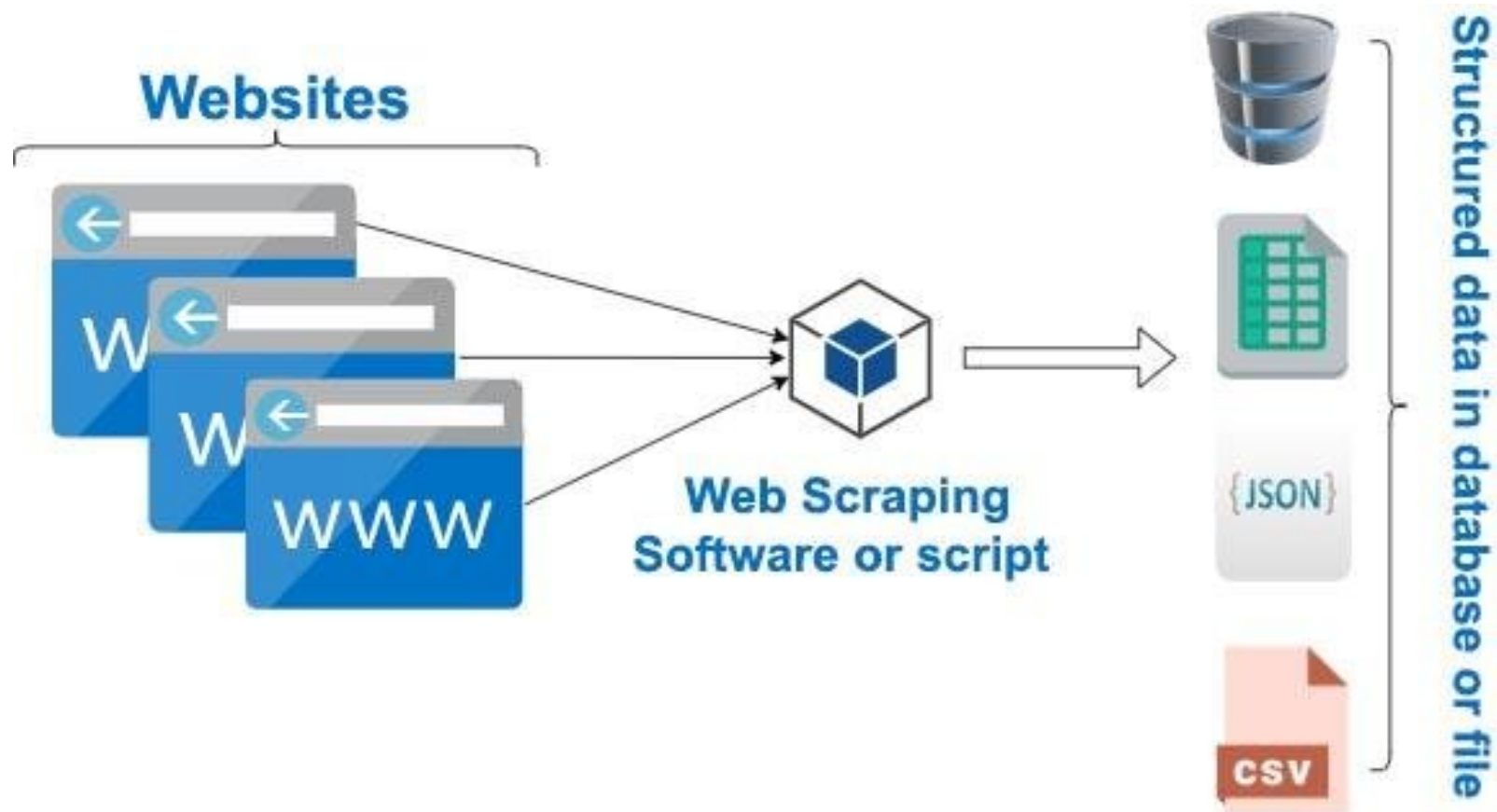# Python web scraping

Objective

Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites. Web scraping software may directly access the World Wide Web using the Hypertext Transfer Protocol or a web browser.

Benefits :

- Efficient data collection. Data scraping allows for the automated and efficient collection of large amounts of data from the internet. ...
- Variety of uses. ...
- Access to otherwise difficult-to-obtain data. ...
- Cost-effective. ...
- Automation. ...
- Real-time data.

Websites

W

W

WWW

**Web Scraping
Software or script**

{JSON}

CSV

Structured data in database or file

Data Validation and Data Transformation :

Name Validation - Data Validation is used to validate the data. checking the accuracy and quality of source data before training a new model version.

Number of Columns – Validation of number of columns present in the files.

Name of Columns - The name of the columns is validated and should be the same as given in the schema file.

Data type of columns - The data type of columns is given in the schema file. It is validated when we insert the files.

Null values in columns - If any of the columns in a file have all the values as NULL or missing, we discard such a file

Storing:

Storing the data in form of data frame in CSV file

Thank you