

Data Processing Framework Design

Yutian Wang
Yihang Liu
Xupeng Shi

Data Processing Framework Design

Content

1. Description of Propose
2. Dataset Plugin
3. Data Processing Core Framework
4. Visualization Plugin

Description of Propose

Process and group the data in time order and show them with different styles.

Example:

- US population migration to the west
- Chinese people migration during spring festival
- COVID tested positive people in major cities starting in 2020
- Weather forecast (change of cloud and rain position over time...)
- Google office numbers over the world with time
- Support for republicans and democrats of the States over time

Dataset Plugin

The plugin is used to process all the input file data, like CSV, Excel or Twitter, to a consistent format that meet the input format requirements of core framework

Three steps:

1. Process to input data to transfer to a intermediate state
2. Transfer the intermediate state to the framework input state
3. Optional: Some extra manipulation of transformation due to different data

Data Processing Core

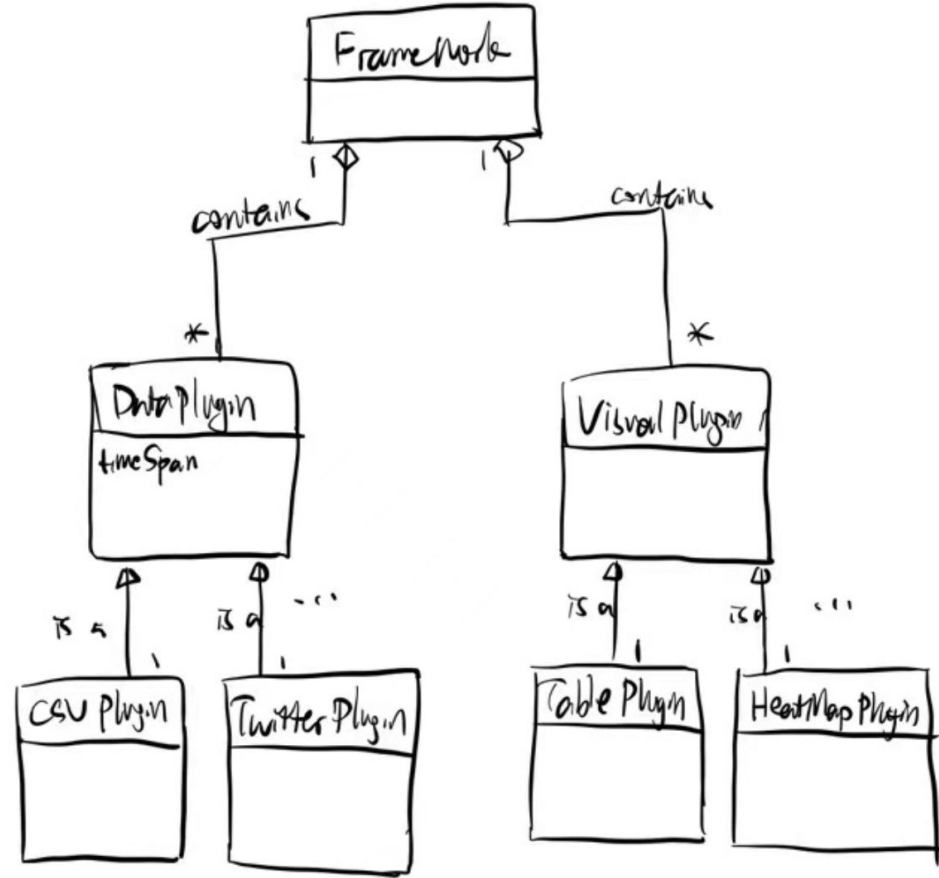
The jobs of this framework include

- sort the data in chronological order
- group the data with different keys

Core data structure:

1. List
2. HashMap

In the dataset plugins we have a specific data format. We used a list to store all data corresponding to the same



Visualization plugin

The plugin is to deal with the output from the core framework and transform them to different style for front-end render. We suppose to support three styles:

1. Geographic Map render
2. PieChart render
3. X-Y Map render

Three steps to process the data:

1. Transform the data from core framework to intermediate status.
2. Transform the intermediate data to render with different display methods.
3. Optional: Store some extra information due to different render formats.

THANK YOU