Introduction:

This project is to crawl pages of CIA’ factbook, which contains data around 300 different fields for each country and zone on the earth.

The crawling was done by a python-based spider, scrapy. Raw-data is kept in format of json.

To insert this data to INM database, several python scripts have been used to build a proper schema and convert raw data to INM scripts.

I should admit that there is mismatch in every step. Thus, in each step, there is a little data lost. This part of data can be manually processed. But most of most part of data is kept and inserted.

Log in:

The database is on the server moons. And the port is 1989. My directory on server is “ouyang”

List of Directory: