FUJITSU

shaping tomorrow with you

Japan MB's Reports

2015/11/16

Fujitsu, Ltd. Kotaro Noyama

Fujitsu Social Science Laboratory, Ltd. Chieko Hiramatsu

FUJITSU

RepOSS Collector’s status

Overview

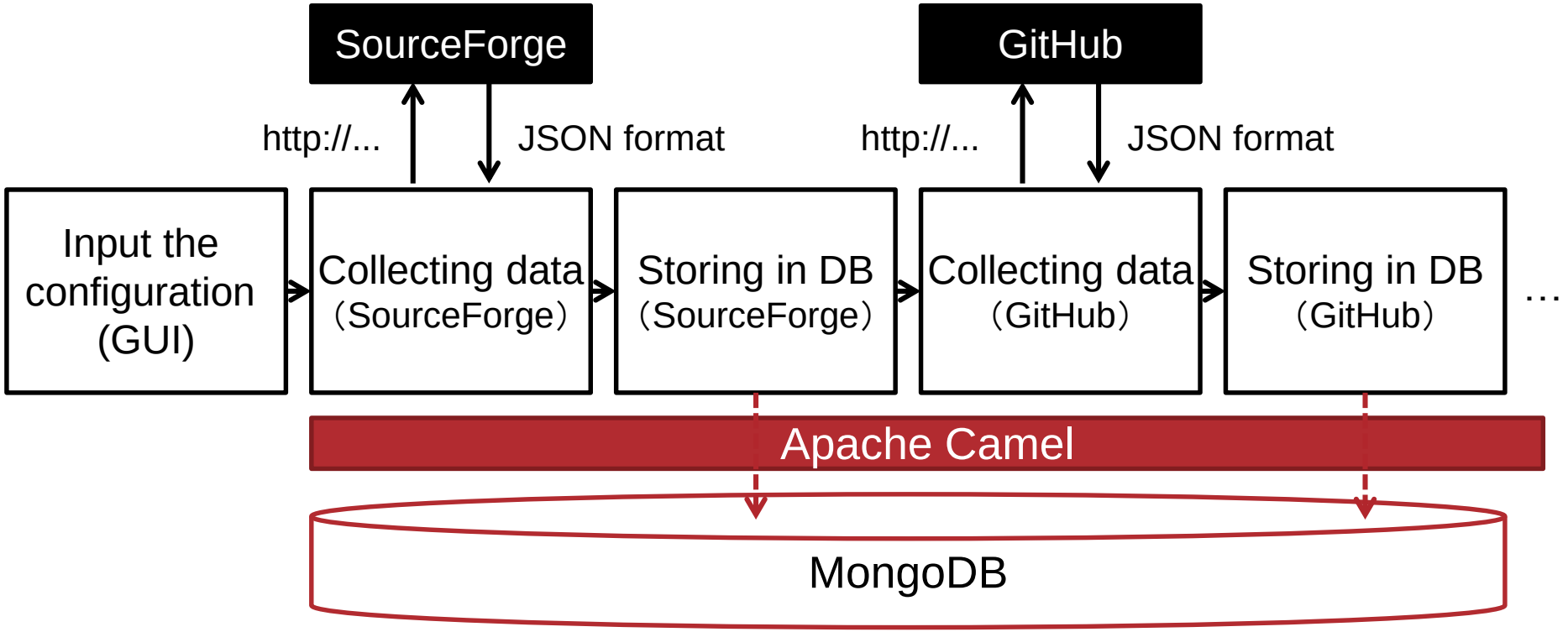
FUJITSU

- Data collecting tool for evaluation of OSS
  - Collecting information of projects from Web services using Web APIs
  - Storing the data into a database (MongoDB)
  - GUI for setting the configuration to collect data
    - Import from the configuration from Excel file
    - Registration / Revision of settings
    - Monitoring the status of progress with GUI

Image of RepOSS Collector

FUJITSU

- Collecting and Storing data by crawling Web services
  - Processing each of data by using Apache Camel
  - Storing the data in MongoBD with JSON format



Target data and implementation status

FUJITSU

Target Web services and information (in beta version)

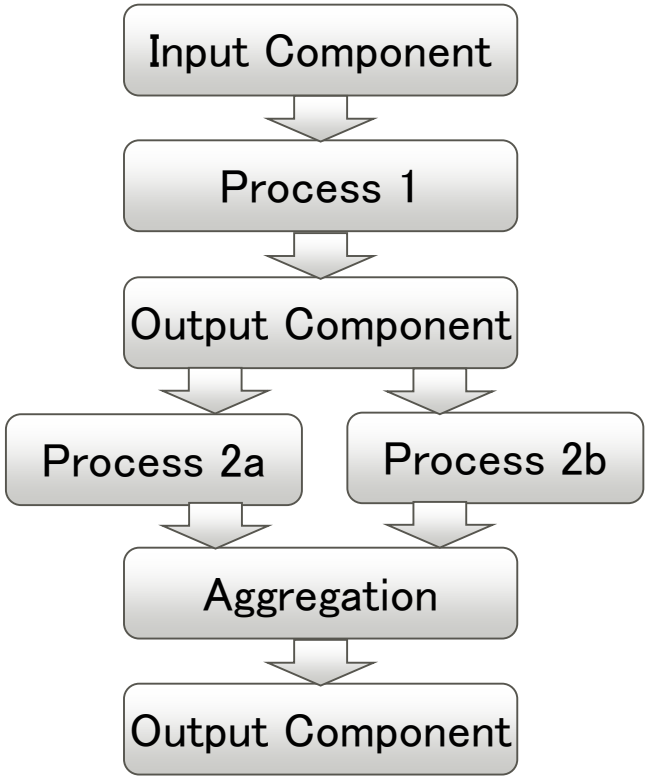
target	collecting information (example)	Impl.
SourceForge	Basic information of the project	Done
GitHub	Basic information of the project	Done
Bugzilla	Bug information	Done
JIRA	Bag information	Done
SlideShare	Slides about the target OSS	Done
Google Custom Search	The number of search results	Done
Google Trend	Transition of the number of search results	Done
Amazon	Books about the target OSS	Not yet
Wikipedia	Information of the project on Wikipedia	Done
CVE	Vulnerability information	Done
Twitter	Tweets with the keywords about the target OSS	Done
Mail archive site	Flow in the mailing list	Done
Package information	Version of the package in major Linux distributions	Not yet

(Appendix) Apache Camel

FUJITSU

- Rule-based routing engine
  - Defining the routing rule, how to input / process / output the data for automating data processing

Routing rule



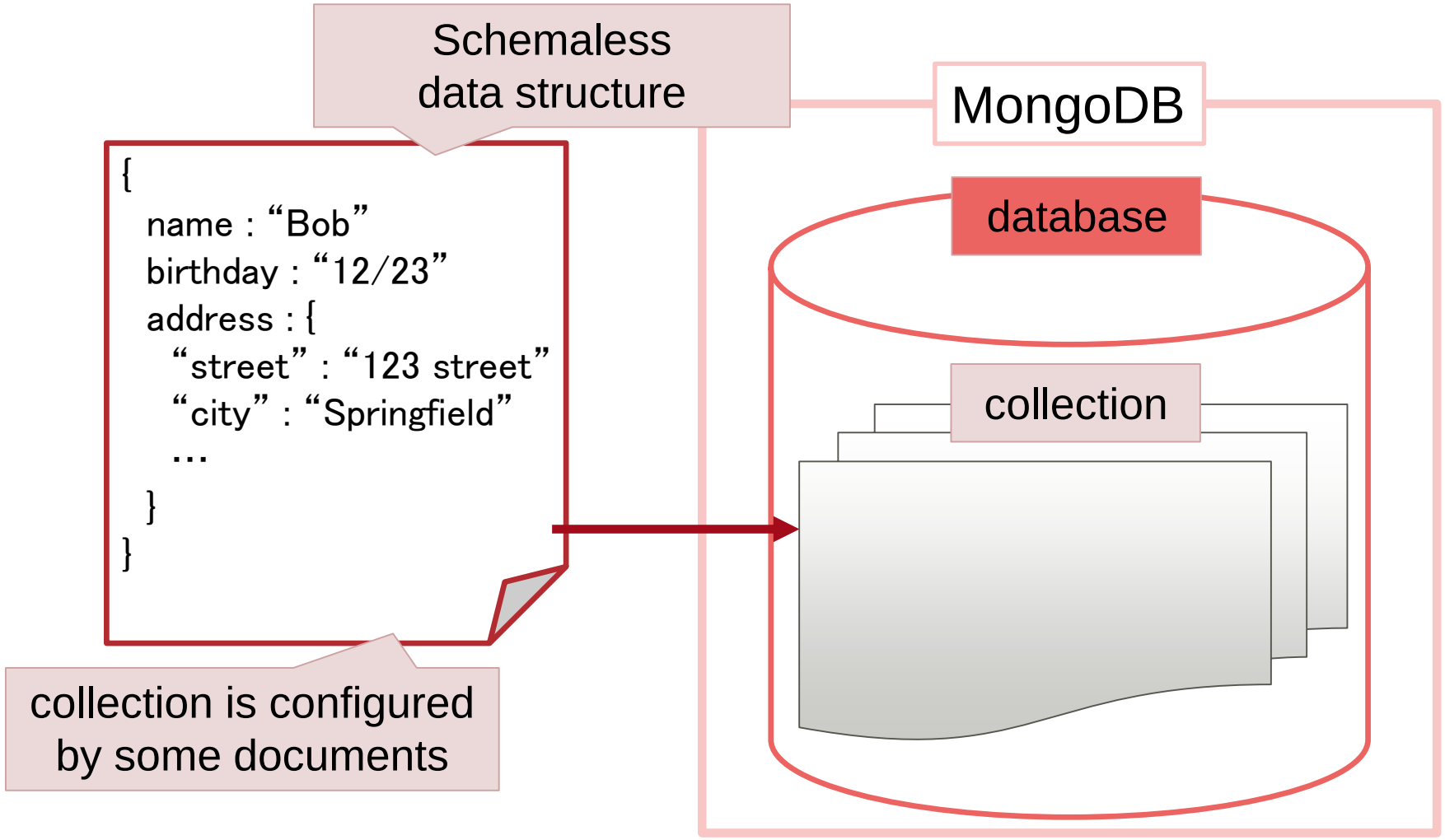
Receive / import data (file, HTTP, DB, etc…)
Data conversion
Send / Restore data (file, HTTP, DB, etc…)
⋮

External library / Original implementation are available on the processes

(Appendix) MongoDB

FUJITSU

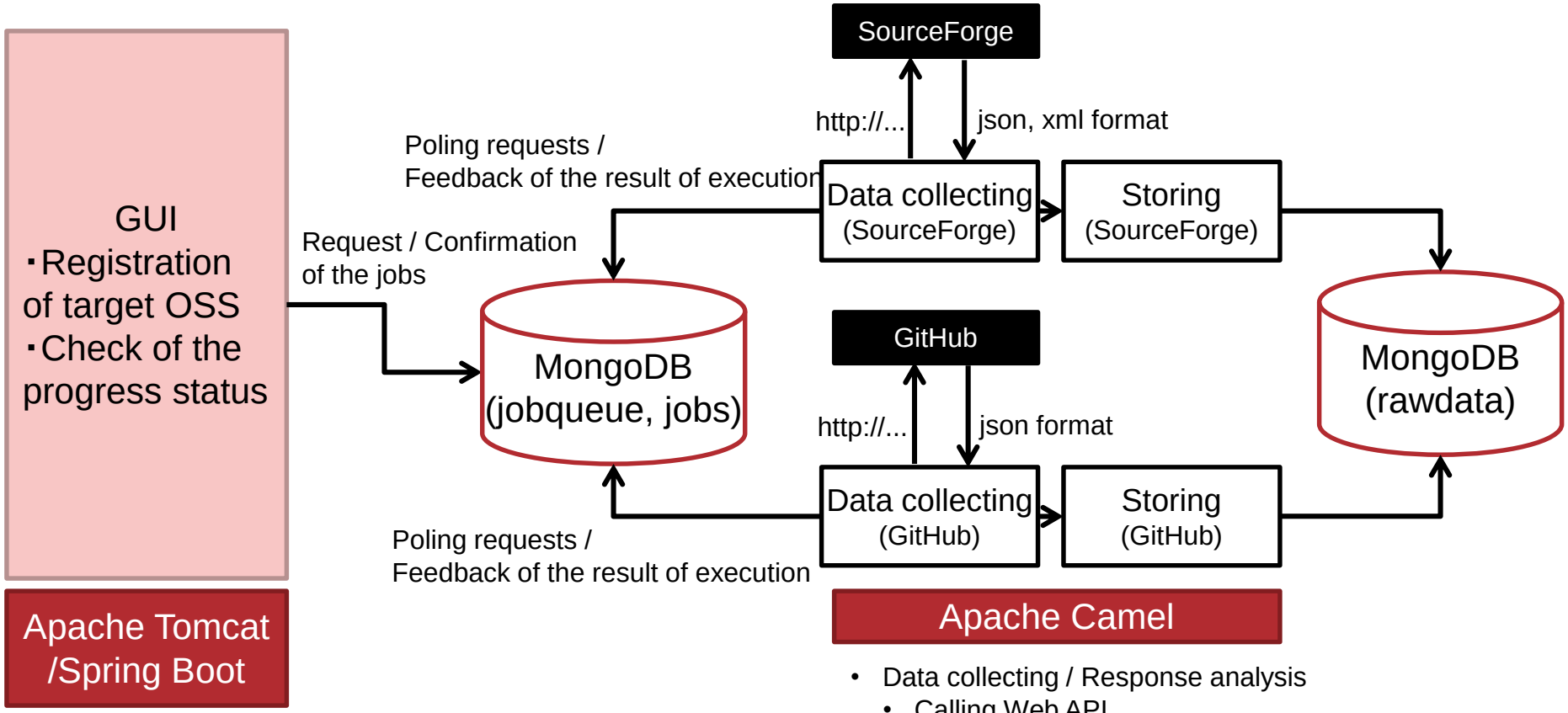
- No SQL, document-oriented database
- Using JSON format to store the data



Detail of RepOSS Collector

FUJITSU

Detailed image of processing in RepOSS Collector



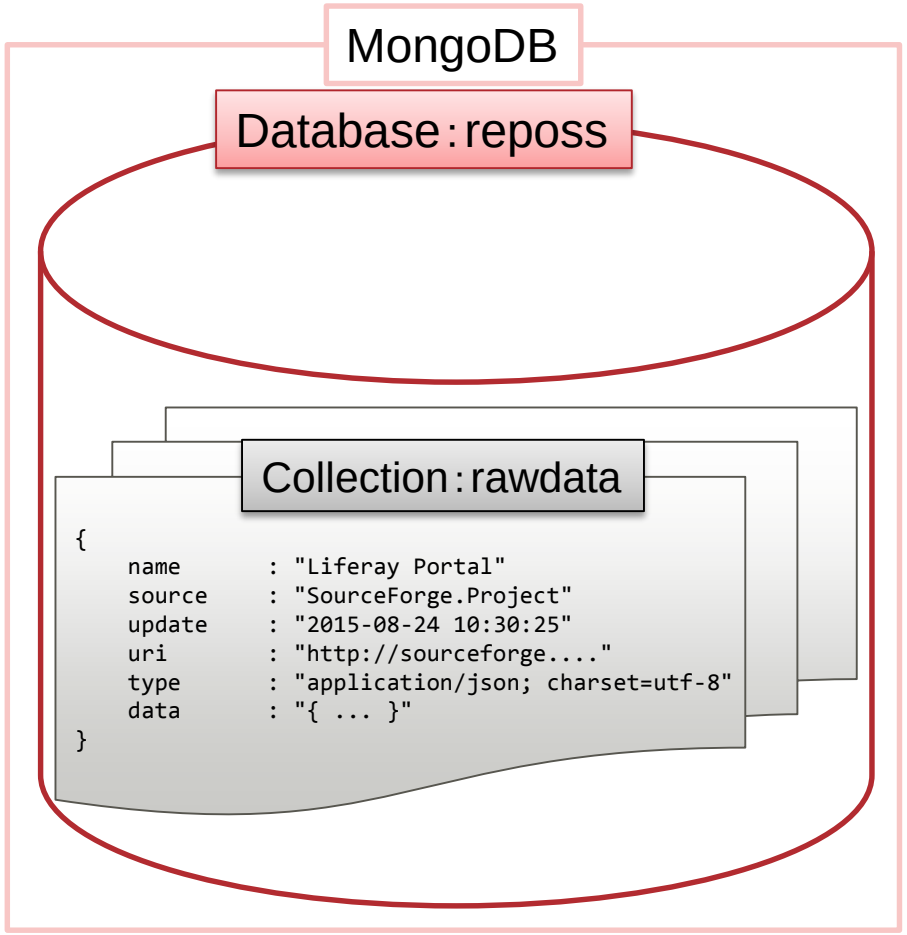
- Registration of target OSS
- Import of the list of target OSS and shelf registration (from Excel Book manipulated by Apache POI)
- Request of the jobs (to jobqueue)
- Check of the progress status

- Data collecting / Response analysis
  - Calling Web API
  - Scraping
- Storing collected data
- Feedback of the result of execution
- Throttling (for flow limitation)

Example of the data collection in MongoDB

FUJITSU

- Collection
  - rawdata
- Items (data type)
  - name (string)
    - name of software
  - source (string)
    - source of information
  - update (string)
    - date of request from Console
      - yyyy-MM-dd HH:mm:ss
  - type (string)
    - value of the Content-Type header*
  - data (string)
    - collected information
      - JSON with response body, or made by partial information on body



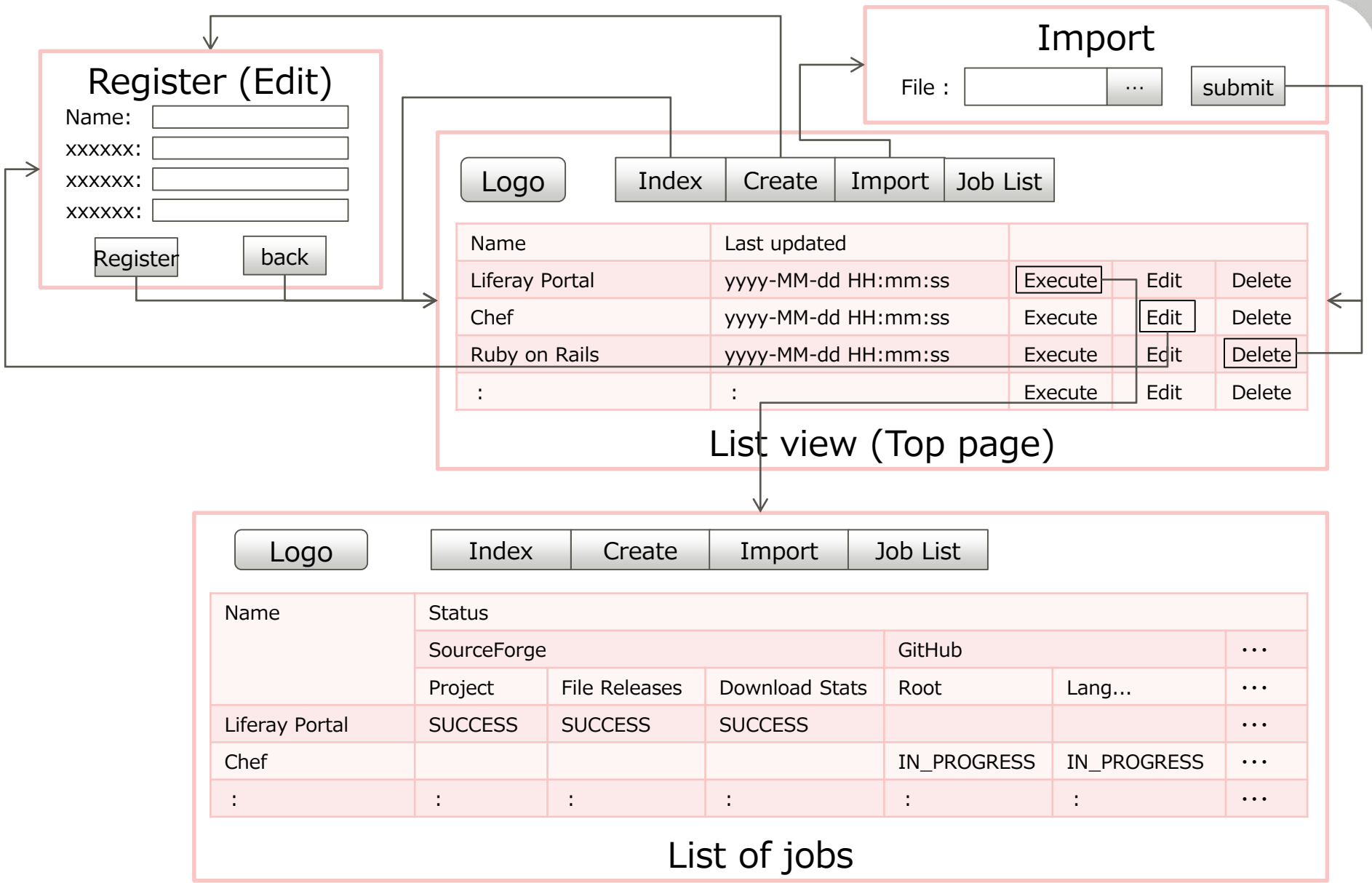
* necessary only if 'data' is raw response body

Functions of GUI

FUJITSU

- RepOSS Collector has the following GUI functions.
  - List view
    - Showing the all registered software information
  - Input / Edit of the configuration
    - Input / edit the configuration to collect data from Web services
  - Import of the configuration
    - Import of the configuration data from Excel Book (Excel2007/2010/2013)
  - Request for collecting data
    - Requesting the collecting tool to start collecting data
  - Monitoring the status
    - Confirmation of the progress (success / processing / error)

Views and transitions (sample)



Document

- Preparing two types of documents
  - Installation Manual
  - Introduction of RepOSS Collector

FUJITSU CONFIDENTIAL

FUJITSU

RepOSS Collector
Installation Manual

Ver. 0.1
2015/10

FUJISU LIMITED
Software Technologies Unit
Software Development Technologies Division
Center of OSS Technology Excellence

FUJITSU
shaping tomorrow with you

Automated data collecting tool for maturity evaluation of OSS

2015/10/8
FUJITSU LIMITED
Center of OSS Technology Excellence

Action items

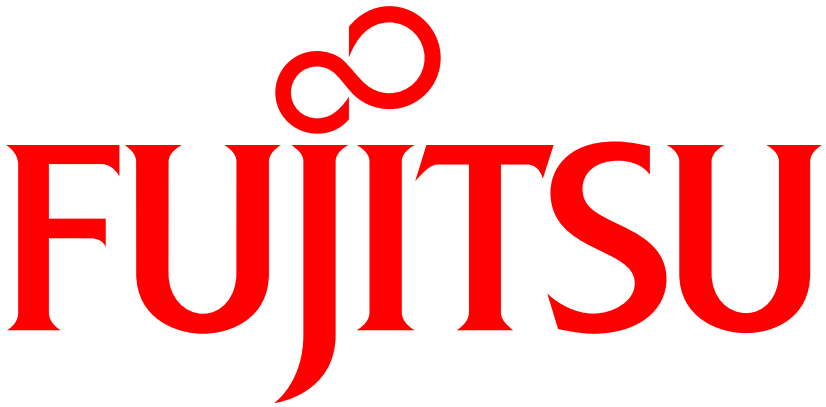
FUJITSU

- A2015101302 Japan MB will provide the information of GitHub before next F2F meeting.
- https://github.com/neaosspf-wg3/reposs-collector

Action items

FUJITSU

- A2015101303 Japan MB will provide the list of Open Source Software used in RepOSS Collector before next F2F meeting.
- Apache Camel (Apache License 2.0) only
- But the following software is necessary elsewhere to operate it.
  - OpenJDK 1.8
  - MongoDB 2.6.x
  - Apache Maven 3.x



shaping tomorrow with you