Extract Server
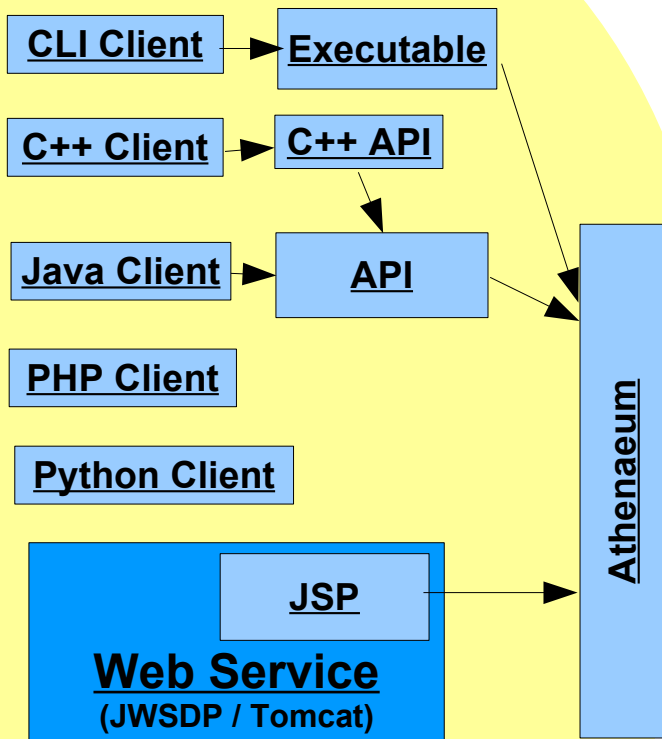
- Athenaeum Framework
- Extract Server Architecture
  - Manager
  - Worker
- Interfaces
  - http (accessed from ELSSI)
  - Web (Server Management)
  - CLI (direct user access, testing)
  - XML-RPC (internal)
- Distribution
  - How to install/update/configure/keep up
  - Where it already runs

Athenaeum framework
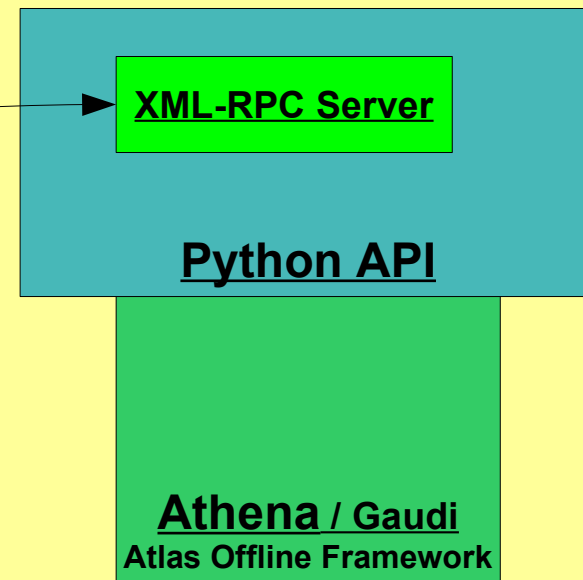
Remote Client to Atlas Offline Framework

Client(Manager)

- CLI Client → Executable
- C++ Client → C++ API
- Java Client → API
- PHP Client
- Python Client
- **Web Service** (JWSDP / Tomcat) — JSP

Athenaeum

Java / xMB (runs everywhere)

Python

XML

XML-RPC

Worker

XML-RPC Server

**Python API**

Athena / Gaudi Atlas Offline Framework

C++ / xGB (runs only on lxplus)

➢ *Athenaeum allows to access (remote) Athena Server.*
➢ *Any (Athena) Python script can be send directly to Athena from the Client.*
➢ *Results (usually in XML) are send back.*
➢ *Results can be stored on the server (on AFS in case of CERN).*
➢ *Special Python scripts are provided to present some Athena data.*
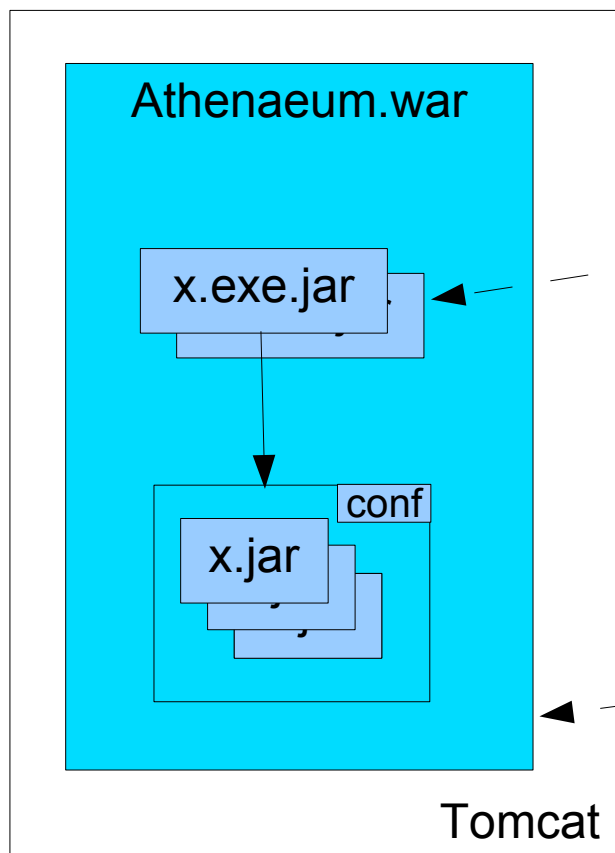➢ *Several Clients exist.*

Extract Server Architecture

Manager

Athenaeum.war

x.exe.jar

conf

x.jar

Tomcat
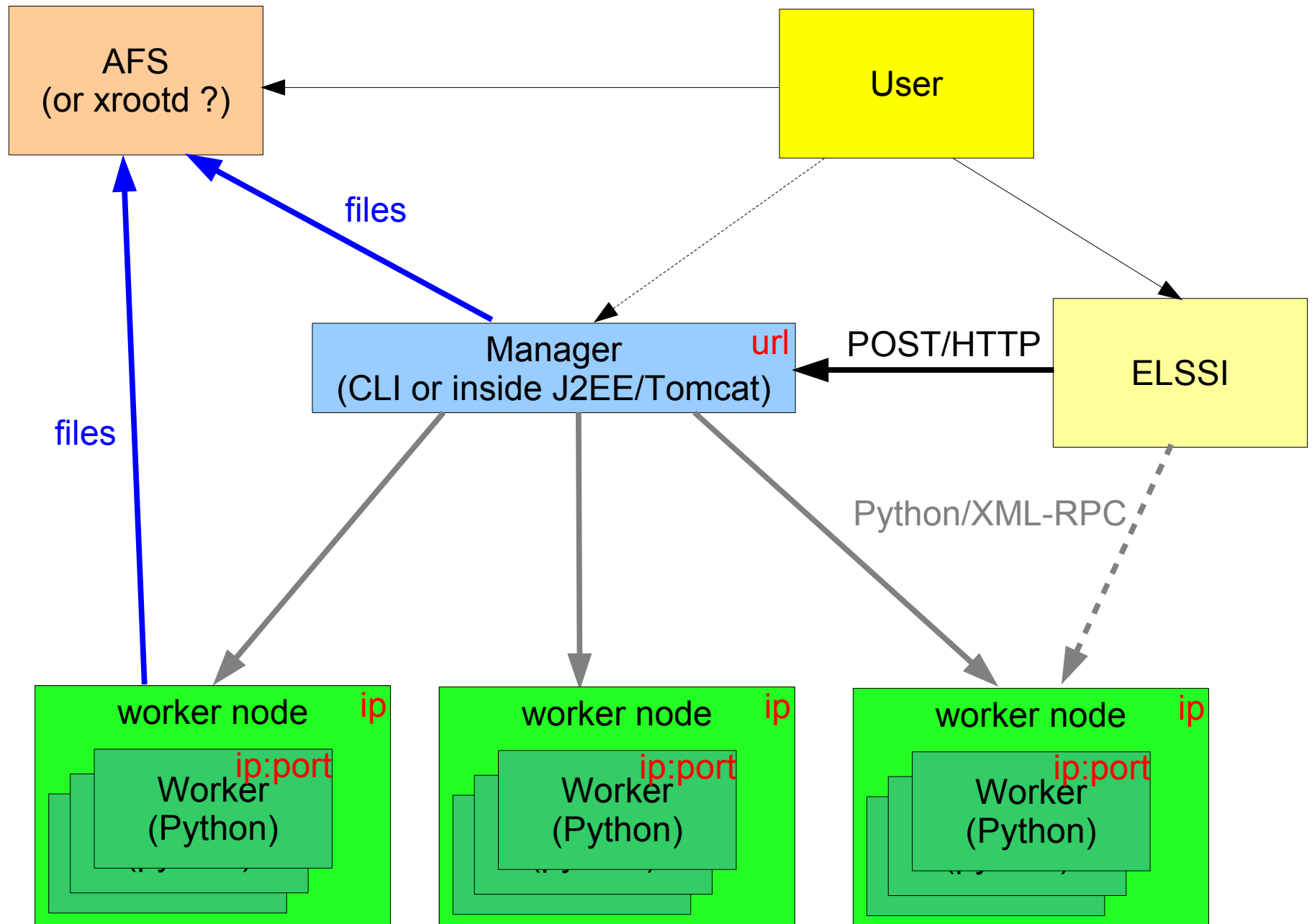
- x.exe.jar *files are directly executable*
- *they are aliased to commands* **athenaeum** *and* **extract**
- *those commands are used to*
  - *test*
  - *re-issue extraction (user gets a command line with all options in the mail, it is also available from the monitoring web page)*
- *all this can be used on any machine (any OS) as long as firewalls to workers are opened – just get Athenaeum-dist.tar.gz*

- Athenaeum.war *can be installed on any Tomcat, JWSDP or similar container – as long as you have proper rights*
- *the service can then be used*
  - *calling appropriate URL (like from ELSSI)*
  - *using simple Web GUI (for testing and server management)*

- *Manager is trivial to install and will work out-of-box everywhere*
  - *Known sides are treated in a preferential way by filling their characteristics into Athenaeum configuration (list of available worker nodes, email for monitoring/debugging,...)*
- *When inside Tomcat, it is isolated from the environment*
  - *=> it can't do any harm*
  - *=> it doesn't depend on local configuration (which is difficult to control in distributed environment)*
- *=> I try to put as much as possible functionality on the Manager*
- *Workers method calls pass often via Manager (even when they could go just within a Worker) because a Manager has global overview and control over whole node (load-balancing, monitoring,...)*
- *Manager uses Twitter to inform about problems (*atlastags *account)*

Worker
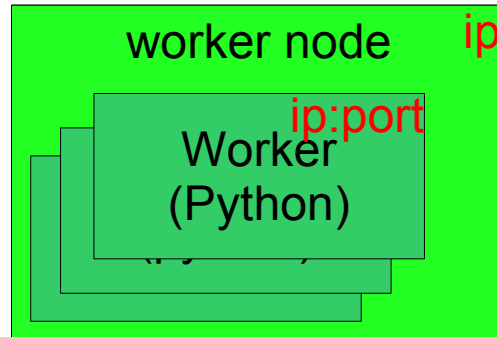
➢*Each worker node contains:*
- ➢ *Python implementation with Athenaeum-aware XML-RPC server*
- ➢ *A set of scripts to start/stop/restart/inspect running services (should be locally customized to handled needed services on designed ports and to fit in local environment)*
- ➢ *A cron job to monitor running servers and restart them if needed to fit in local environment)*
- ➢ *A set of testing scripts*
- ➢ *A directory for monitoring files*
- ➢ *A complete Athena able to run* CollAppend

➢*Distribution is done via Atlas SVN*

➢*Manager can*
- ➢ *inspect running server (their configuration, history, status)*
- ➢ *restart running server*
- ➢ *clone running server*

➢ *All tasks run in independent threads, identified by unique pid*

Extract Server Architecture
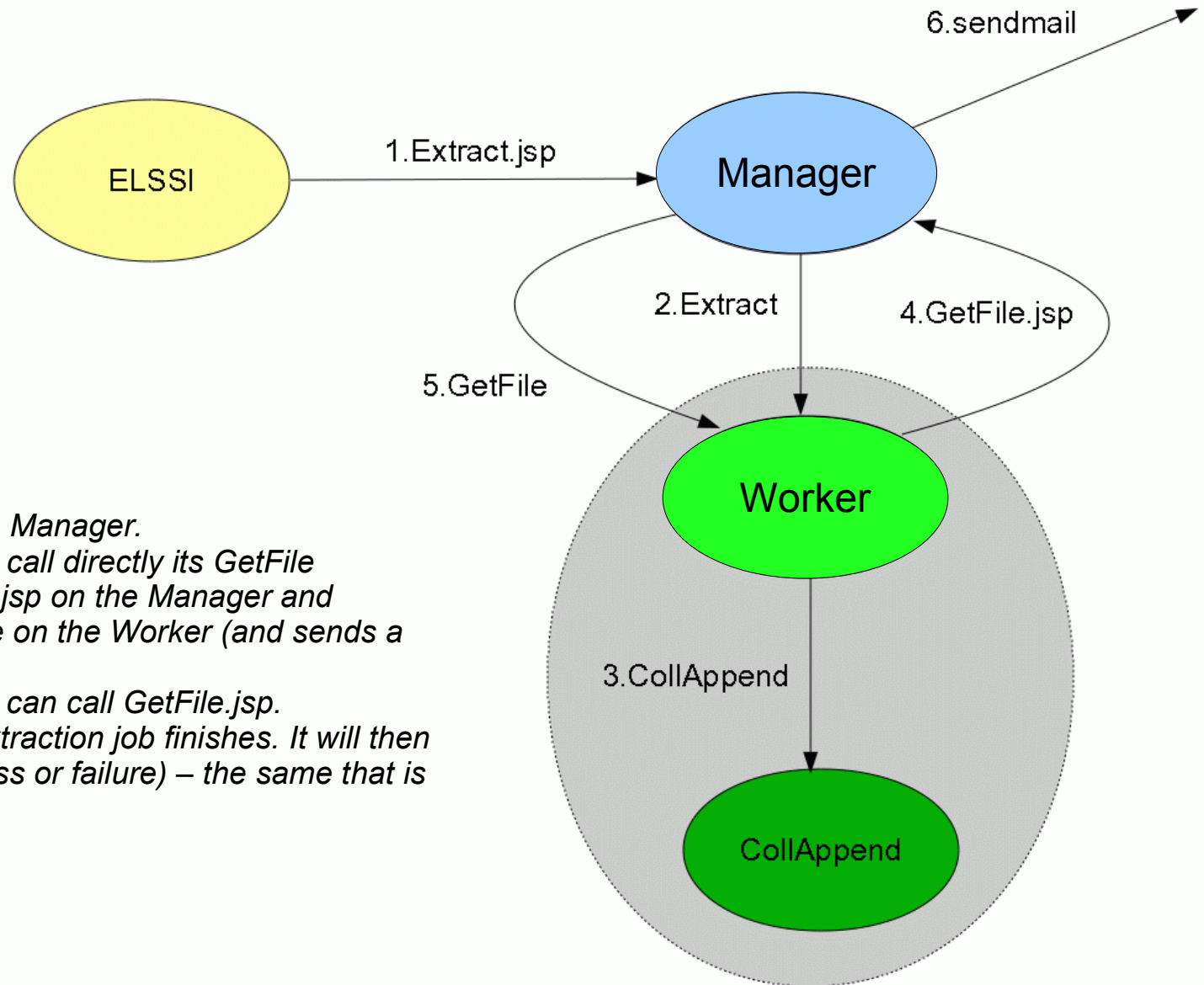
> All important actions pass via Manager.
>> This way, Worker doesn't call directly its GetFile method, but calls GetFile.jsp on the Manager and Manager then call GetFile on the Worker (and sends a notification mail).
> A user (directly or via ELSSI) can call GetFile.jsp. GetFile.jsp will loop until the extraction job finishes. It will then return the job summary (success or failure) – the same that is sends in the notification email.

Server Management

Athenaeum JSP @ CERN

Server URL: http://lxvm0341.cern.ch:10001

i Get Status of Standard Servers 🔡 Google

❓ Get Help

Generic Functions
🖳 Run Script on Server

Specific Functions
❄ Browse Cool DB (connect to any/recent database)
❄ Browse Cool DB (choose from standard databases)
📝 Extract Tag file

Information and Documentation
i Get Server Info
❓ Get Server Help
📇 See latest log
📇 See server running jobs
📇 See server accumulated statistics
📇 List parallel Servers

Server Management
U Restart Server
U Start new Server

Athenaeum JSP Global Help

1. Select *Athena Server URL* and *key* (key value is not filled in to prevent Robots from flooding the Server. Use the standard Atlas password as a key).
2. Select action to be performed on the Server:
   - **Run Sc...**
   - **Get Se...**
   - **Get Se...**
   - **Restart...** in the S... request...
   - **Start n...** impleme... sending...
   - **Family...** impleme...
   - **Browse...** DB usin... connection parameters. The most recently used parameters are filled in as default.
   - **Browse Cool DB (choose from standard databases)** allows to browse Cool DB using Athena Server. You can choose from the set of standard databases. The most recently used database is shown on the top of the page.
   - **Extract Tag file** allows to extract Root Tag file by consulting Tag SQL database..
   - **Get Status of Standard Servers** tests all Standard Server.
   - **Get Help** gives this help.
   - **See latest log** gives the latest server log.
   - **See server running jobs** gives status of the running and finished tasks. All servers (running on all ports) are shown.
   - **See server accumulated statistics** gives statistics about finished tasks. All equivalent servers are shown.
3. Some Actions will require more input on the right frame. In such case, fill in the

---

Demo: http://cern.ch/Athenaeum

The same tasks can be performed by athenaeum *command*

- Family of equivalent parallel Servers suported.
- Server statistics is available.
- Tag Extraction is available as a standard function, including in the Web Service.

- Support for JAS3 client has been stopped.
- Access to server log files is available.
- voatlas16 server has replaced lxgate01 server
- athenaeum command is available on CERN AFS.
- SQLite databases (readable from CERN AFS) are supported.
- When you get org.apache.xmlrpc.XmlRpcException: I/O error while communicating with HTTP server: Connection refused exception, it means (almost certainly) that Athena has crashed. It that case, try another server (port) or wait for server restart (which should happen within an hour).
- JAS3 with integrated Athenaeum client is available with one-click instalation procedure using WebStart. The only requirements is the correctly installed Java (1.5 or 1.6) on the local machine.
  [⚡ Launch]
- JAS3 Plugin can access Athenaeum Servers behind firewalls via SSH tunnels.
- Tag extraction servers are supported..
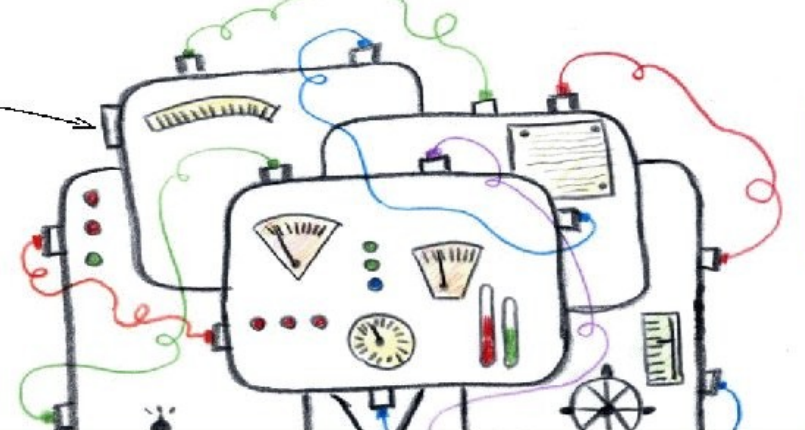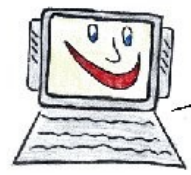
---

progress bar not (yet) started   0%   STOP

CLI

```
$ source /afs/cern.ch/sw/lcg/external/Java/bin/setup.sh
```

```
$ athenaeum
AthenaClient <url>:<port> <key> <task> [<options>] # executes <task>
AthenaClient <task>                          # shows   <task>
available tasks: Cool Dummy Event Extract Fork Info Restart Help Log Statistics Family
```

```
$ extract
extract or java -jar AthenaeumExtract.exe.jar
      -manager [CERN|CHICAGO|BNL], default = CERN
      -python <Python options file>
      -url <worker ip:port>, default = http://lxvm0341.cern.ch:10001
      -key <insider key>
      -execution [extract|skim|prun], default = extract
      -output <output Root file>, default = test_<random>.root
      -query <sql query>
      -collname <collection name>
      -lumi <luminosity>, default = Unknown
      -release <release>, default = Atlas,takeFromEnv
      -conn <connection string>
      -target <target directory>, default = .
      -atts <requested attributes>, default = RunNumber,EventNumber
      -proxyname <skimming proxy>
      -stream_type <type of stream>
      -athena_jo <athena JO>
      -user_jo <user JO>
      -outputdata_type <type of output data>
      -utility <POOL Run utility name>, default = CollAppend
      -params <xml parameters file for utility>
      -email <notification email>
      -d, debug
```

How to install/update/configure/keep up

- *Manager:*
  - *Tell me*
    - *email of a human monitor to be notified about problems*
    - *available smtp server (otherwise gmail is used)*
    - *ip:port of all workers*
    - *local directory and URL to store config files to make them available to others*
  - *Deploy Athenaeum.war on local Tomcat container (re-do when new version becomes available)*
  - *Untar Athenaeum-dist.tar.gz on local filespace (re-do when new version becomes available)*
- *Worker:*
  - *Install Atlas software*
    - *including Database/TagPoolServices*
  - *Modify scripts/*.sh to reflect local configuration and desired servers*
  - *Submit monitoring cron scripts/cron.sh*
- *Make sure firewalls are opened*
- *Manage using Manager Web Service*
- *Read notification/error emails (if configured)*

<u>Where it already runs</u>

➢ *@CERN*
 ➢ *Manager*
   ➢ *Web Service on central J2EE server (very well managed): http://cern.ch/Athenaeum*
   ➢ *CLI on AFS (memoryless):*
     ➢ source /afs/cern.ch/sw/lcg/external/Java/bin/setup.sh
     ➢ extract .... # to perform extraction/skimming/prun
     ➢ athenaeum ... # to perform a management task
 ➢ *Workers (Extract, Skim, PRUN) on lxvm0341 (dev) and voatlas18 (prod)*
➢ *@Chicago – installed and died*
➢ *@BNL*