AstroLabNet

Catalog & FileServer

Two (significant) missing pieces:

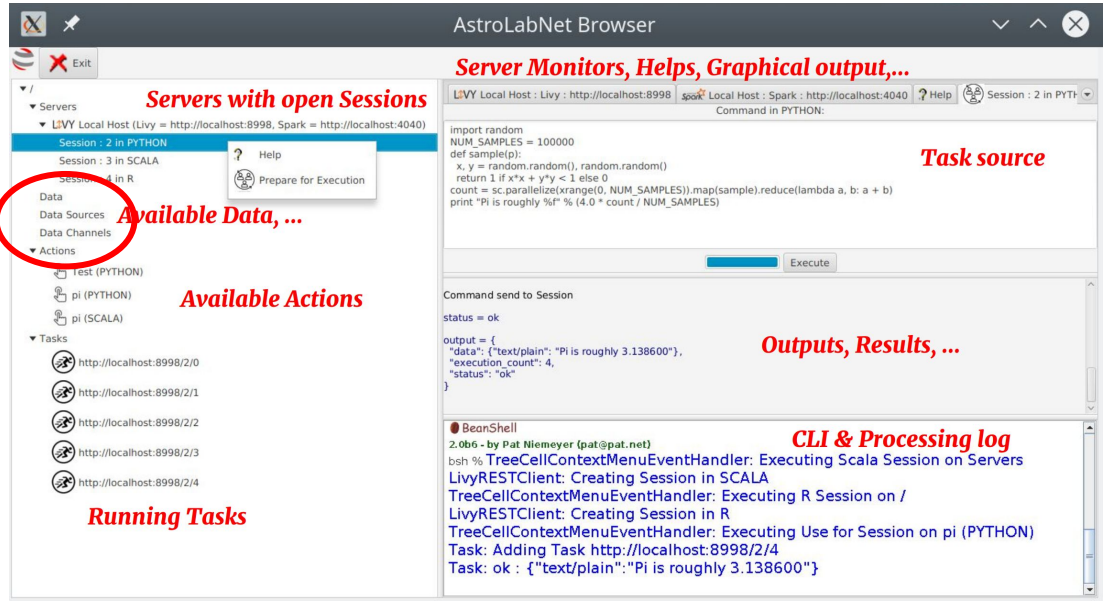
➔ To know, what is available, and where - **Catalog**.

➔ To get it / copy it / access it remotely - **File Server**.

Julius Hrivnac
26/3/19, LAL, Orsay

Archipel

- Catalog + File Server will allow full remote access to resources
- Once ***AstroLab Archipel*** is in place:
  - Global map of all available data
  - Job execution strategies
    - Sending jobs to data
    - Accessing data remotely
    - Bringing required data locally
  - Scheduled data replication



Catalog

- Technology:
  - **HBase**
    - Well integrated in Apache HDFS/Spark environment
    - Flexible
    - Scalable and fast (but may need tuning)
    - Simple to use
- Content:
  - Alarms (or alarms-like)
  - Physical addresses
  - Free attributes/tags
    - Pre-defined
    - User-defined
  - Archipel topology
  - Journal

HBase

- Data arranged in Column Families
- Each Cell can have several versions and expiration period
- HBase native language is Ruby
  - But other interfaces available
- RESTfull service used by AstroLabNet
  - Can move to more direct connection if performance problems
- Currently three tables (it's useful to have tables with version):
  - `astrolabnet.topology.1`
  - `astrolabnet.catalog.1`
  - `astrolabnet.journal.1`

Topology

- Client connects to one (known) server and transitively acquires the full network

```
# Topology
##########
# key: name
# i = info
#   name
#   location
# d = description
#   spark
#   livy
#   hbase
#   xrootd
# r = reference
# c = comments
#   comment
# a = attributes
#############################################
create 'astrolabnet.topology.1', 'i', 'd', 'r', 'c', 'a'
```

Table creation (in Ruby)

```
hbase(main):005:0> scan 'astrolabnet.topology.1'
ROW                                                  COLUMN+CELL
LAL                                                  column=c:comment, timestamp=1553168510756, value=Institute server
LAL                                                  column=d:hbase, timestamp=1553168510738, value=http://134.158.74.54:8080
LAL                                                  column=d:livy, timestamp=1553168510712, value=http://vm-75222.lal.in2p3.fr:21111
LAL                                                  column=d:spark, timestamp=1553168510703, value=http://vm-75222.lal.in2p3.fr:20001
LAL                                                  column=i:location, timestamp=1553168510689, value=Orsay
LAL                                                  column=i:name, timestamp=1553168510677, value=LAL
Local Host                                           column=c:comment, timestamp=1553168510654, value=Default server
Local Host                                           column=d:hbase, timestamp=1553168510633, value=http://localhost:8080
Local Host                                           column=d:livy, timestamp=1553168510592, value=http://localhost:4040
Local Host                                           column=d:spark, timestamp=1553168510567, value=http://localhost:8998
Local Host                                           column=i:location, timestamp=1553168510542, value=here
Local Host                                           column=i:name, timestamp=1553168510526, value=Local Host
2 row(s)
Took 0.0241 seconds
```

Catalog

- Populated automatically from Alerts
  - Schema converted almost automatically

Avro alert

```
{"schemavsn": "3.2",
 "publisher": "ZTF (www.ztf.caltech.edu)",
 "objectId": "ZTF18abyouwp",
 "candid": 697251920015010010,
 "candidate": {"jd": 2458451.7519213,
               "fid": 1,
               "pid": 697251920015,
               "diffmaglim": 19.647123,
...
```

```
hbase(main):001:0> scan 'astrolabnet.catalog.1'
ROW                                   COLUMN+CELL
ZTF18abyouwp                          column=d:fileName, timestamp=1553516397668, value=candid697251920015010010_pid697251920015_targ_sci.fits.gz
ZTF18abyouwp                          column=d:objectId, timestamp=1553516396357, value=ZTF18abyouwp
ZTF18abyouwp                          column=d:publisher, timestamp=1553516396346, value=ZTF (www.ztf.caltech.edu)
ZTF18abyouwp                          column=d:schemavsn, timestamp=1553516396331, value=3.2
ZTF18abyouwp                          column=d:type, timestamp=1553516397655, value=cutoutScience
ZTF18abyouwp                          column=r:candid, timestamp=1553516396321, value=697251920015010010
ZTF18abyouwp                          column=r:prv_candid_1, timestamp=1553516396665, value=ZTF18abyouwp_1
ZTF18abyouwp                          column=r:prv_candid_2, timestamp=1553516396785, value=ZTF18abyouwp_2
ZTF18abyouwp                          column=r:prv_candid_3, timestamp=1553516396889, value=ZTF18abyouwp_3
ZTF18abyouwp                          column=r:prv_candid_4, timestamp=1553516396996, value=ZTF18abyouwp_4
ZTF18abyouwp                          column=r:prv_candid_5, timestamp=1553516397104, value=ZTF18abyouwp_5
ZTF18abyouwp                          column=r:prv_candid_6, timestamp=1553516397185, value=ZTF18abyouwp_6
ZTF18abyouwp                          column=r:prv_candid_7, timestamp=1553516397318, value=ZTF18abyouwp_7
ZTF18abyouwp                          column=r:prv_candid_8, timestamp=1553516397446, value=ZTF18abyouwp_8
ZTF18abyouwp                          column=r:prv_candid_9, timestamp=1553516397535, value=ZTF18abyouwp_9
697251920015010010                    column=d:candid, timestamp=1553516396426, value=697251920015010010
697251920015010010                    column=d:dec, timestamp=1553516396456, value=-26.9677112
697251920015010010                    column=d:decnr, timestamp=1553516396515, value=-26.967766
697251920015010010                    column=d:fid, timestamp=1553516396391, value=1
697251920015010010                    column=d:isdiffpos, timestamp=1553516396435, value=f
...
```

File Server (next)

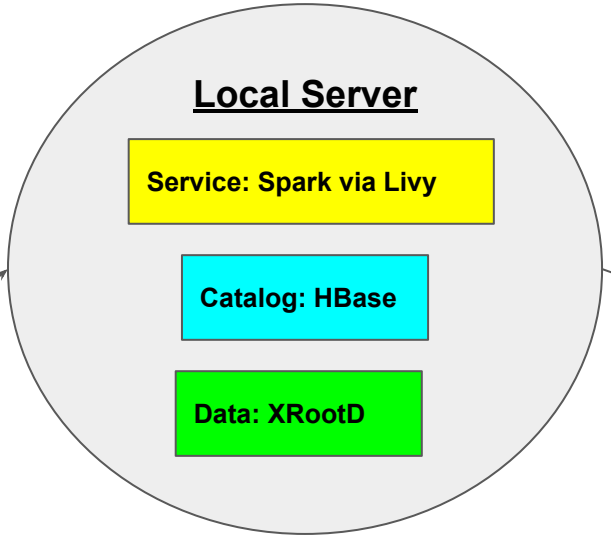
- **Hadoop-XRootD-Connector**
  - Well accepted technology (also in LSST)
  - Developed and maintained in CERN
    - Used mainly for backup (remote access to HDFS)

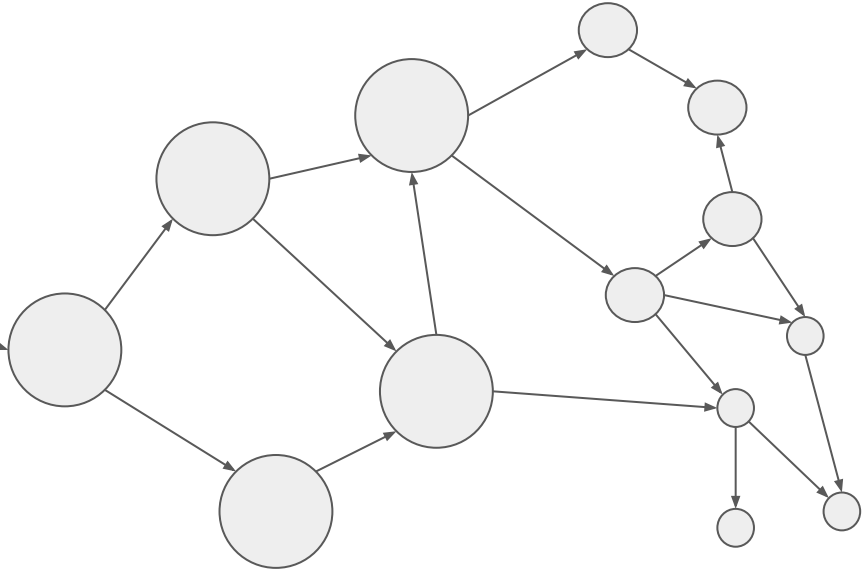
GIT: https://gitlab.cern.ch/db/hadoop-xrootd

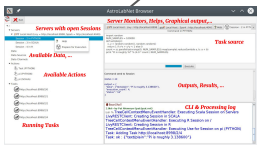
RPM: https://koji.cern.ch/kojifiles/packages/hadoop-xrootd/1.0.0/4.el7.cern/x86_64/hadoop-xrootd-1.0.0-4.el7.cern.x86_64.rpm

Final Aim:

A Job will find a convenient Server & Data itself







- Home:
  - https://hrivnac.web.cern.ch/hrivnac/Activities/Packages/AstroLabNet
- Source:
  - https://github.com/hrivnac/AstroLabNet.git