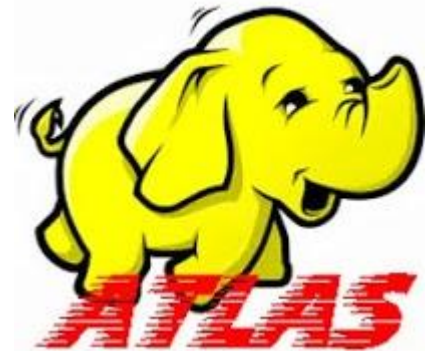


TagConvertor & ...



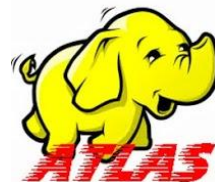
- Core
 - Index Types & Dereferencing
 - Inspector
 - Cron
 - Recipes
- CLI
 - Aux M/R
 - Remote CLI
- Web Service (Justin)
 - Remote Access
- Event Service
- Data Challenge ?

Also: Hadoop server has been upgraded (Rainer) and all following problems fixed.

latest version: 1.12.0

*J.Hrivnac, EI WS, 16Sep14, CERN
+ Justin, Fedor, Rainer, Andrea,...*

Core - Index Types & Dereferencing

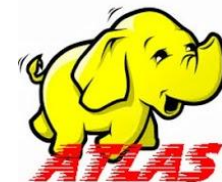


Andrea will give some more details

- There are three kinds of indexes:
 - **-output tags** => <new key>, <old key>, <other values>
 - writes complete output file (with all fields, but new keys = indexed values), this is the default
 - needs more disk space, but faster and simpler
 - it's in fact just re-arranging of data to fit searching pattern
 - **-output index** => <new key>, <foreign key>
 - writes only indexed values (which become new keys) + pointers to keys of satisfying events
 - needs less disk space, but reading requires dereferencing of pointers
 - when indexing on boolean selection (often in trigger search), we will not write indexes, but selection of satisfying entries (because new key would have just two values: true/false)
- Pointers dereferencing can be very slow
 - List of keys can contain overlapping intervals
 - Key searching has been optimised - before the search, the list of keys intervals is consolidated
 - It can still be quite slow when keys point into several files
- Note: Map files have internal indexes (to add confusion)
- Need more benchmarking to be able to use appropriate way in each case
- Looking at Map files extensions, ...

Each search is just a selection & transformation - created TagFile can be used in the same way as source TagFile. Full TagFiles or index TagFiles are treated in the same way (user doesn't see any difference). Collections of TagFiles (TagSets) can be used in the same way.

Core - Inspector



inspector [map|seq|txt|csv|chains] <path> [<min> <max>]

to see what is inside files in HDFS
(extended by Fedor)

\$ inspector csv EICache/prof/2014_7_28_4_33_40_318/part-r-00000

EF_2MUL1_I2j30_HV,515571.0

EF_2MUL1_I2j30_HV_cosmic,14195.0

EF_2MUL1_I2j30_HV_firstempty,69475.0

EF_2MUL1_I2j30_HV_unpaired_iso,4449.0

EF_2MUL1_I2j30_HV_unpaired_noniso,6605.0

\$ inspector chains EICache/prof/2014_7_28_4_33_40_318/part-r-00000 -1 3

EF_2MUL1_I2j30_HV,515571

EF_2MUL1_I2j30_HV_cosmic,14195

EF_2MUL1_I2j30_HV_firstempty,69475

\$ inspector seq EICache/prof/2014_7_28_4_33_40_318/part-r-00000 2 4

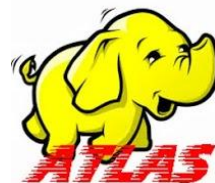
key = EF_2MUL1_I2j30_HV_firstempty

value = 69475.0

key = EF_2MUL1_I2j30_HV_unpaired_iso

value = 4449.0

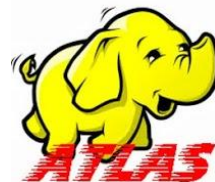
Core - Cron



Nightly cron:

- All standard tests (ant check-*)
- Catalog verification
- Catalog backup
- Results stored in aiatlas016:/data/TagConvertor-nightly and sent to me

Core - Recipes



Typical commands collected on the TagConvertor Home Page:

Convert CSV into MAP file

```
$ ant -Dtxt=/user/alvarof/miniprod_14Jul/FECE7697-...fb5e.csv -Dseq=ElHadoop/miniprod_14Jul/firstFile sequencify
$ ant -Dfn=ElHadoop/miniprod_14Jul/firstFile sort
$ ant -Dfn=ElHadoop/miniprod_14Jul/firstFile mapify
```

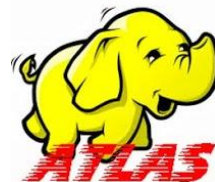
Prepare ES TagFile

```
$ ei -query path:ElHadoop/EVNT.738098._000003 -mr 'true' -index 'guid=db0'
$ ei -query path:ElHadoop/XXX -mr 'true' -index 'guid=Sref0'
created: ElCache/atlevind/2014_7_11_11_7_24_425 213s (10s)
```

Merge Set of TagFiles

```
$ ei -query path:ElHadoop/data11_7TeV.periodK.physics_Muons.PhysCont.AOD.repro09_v01 -mr 'true'
created: ElCache/atlevind/2014_7_7_12_28_22_181 1692s (296s)
```

CQ1 - Aux M/R



- A user can provide his version of Hadoop Mapper and Reducer for **-mr** task
- TagConverter provides interface to needed data and Catalog
- Used by Fedor for Trigger analyses

ei ... -aux "net.hep.atlas.Database.EIHadoop.Accessor.Aux.TrigStat;myparams"

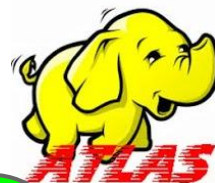
This call will use

net.hep.atlas.Database.EIHadoop.Accessor.Aux.TrigStatMapper

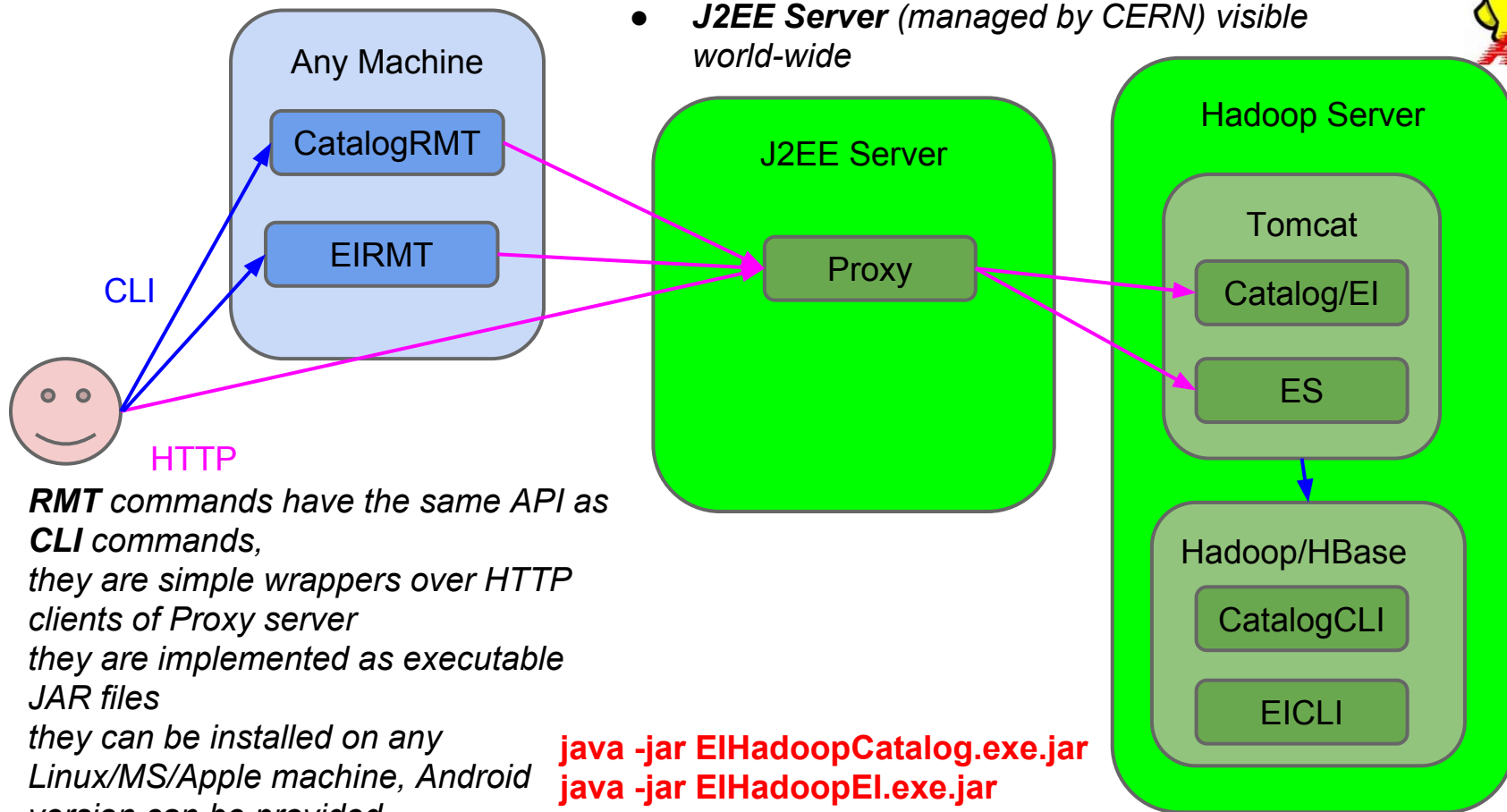
net.hep.atlas.Database.EIHadoop.Accessor.Aux.TrigStatReducer

and pass them myparams

CQ1 - Remote CQ1



- **Hadoop Server** only visible inside CERN domain
- **J2EE Server** (managed by CERN) visible world-wide

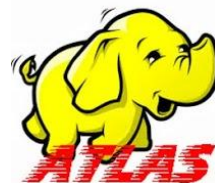


- **RMT** commands have the same API as **CLI** commands,
- they are simple wrappers over HTTP clients of Proxy server
- they are implemented as executable JAR files
- they can be installed on any Linux/MS/Apple machine, Android version can be provided

```
java -jar ElHadoopCatalog.exe.jar
java -jar ElHadoopEI.exe.jar
```

Web Service

version, when and by whom build



Event Index



1.12.0+ [12/sept./2014 at 14:41:21 CEST by rryan]

Catalog: atlas.atdevind.filesets
COMA: atlas.atdevind.coma
Root: /user/atdevind
WS: atdevind
Manager: yuam@lul.in2p3.fr

- [Catalog](#)
- [Event Index](#)
- [Event Service](#)
- [Bookmarks](#)

[Atlas Event Index in Hadoop](#)

EI

-query

-key

-scan

-mr

-filter

-index

-limit

-show

-count

-output

```
RunNumber_EventNumber = 189184-10000510
ID = 705564
>>>
RunNumber_EventNumber = 189184-10000547
ID = 715316
>>>
RunNumber_EventNumber = 189184-10000549
ID = 695814
>>>
RunNumber_EventNumber = 189184-1000055
ID = 30785
>>>
RunNumber_EventNumber = 189184-10000564
ID = 702259
>>>
RunNumber_EventNumber = 189184-10000565
ID = 695887

Only first 20 results shown

All 2211125 results are available from TagFiles: [EICache/tomcat/2014_8_16_9_38_45_343]
```

Progress map : 100%

run parameters

services

(bookmarked results)

parameters

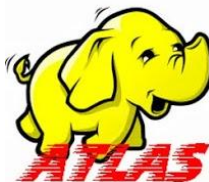
results

progress bar

link to complete results

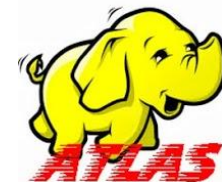
- to start finer search
- (to get results)

Web Service - Remote Access



- As Graphical Web Service is interactive, it needs latency-free access to Hadoop => Proxy server wouldn't be a good solution
- Opening the Tomcat @ Hadoop to outside World would be probably considered a security risk by IT
- Will use SSH tunnel
 - (Probably) can provide J2EE Web Service, which will create SSH tunnel for user

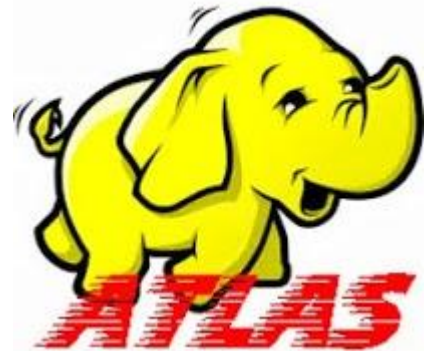
Data Challenge ?



As all pieces are in principle present, we should try to run the whole system (with expected speed):

- data collection, transport and assemblage
- data conversion into Map format
- data placement in HDFS
- registering into Catalog
- creation of Event Server structures (indexes)
- optimisation (indexes,...)

Links



- Home:
 - <http://cern.ch/hrivnac/Activities/Packages/TagConvertor>
- Web Service:
 - <https://aiatlas016.cern.ch/EIHadoop>
 - <https://aiatlas016.cern.ch/hrivnac-EIHadoop>
 - <https://aiatlas016.cern.ch/ryuan-EIHadoop>
 - `ssh -L 8443:aiatlas016.cern.ch:443 id@lxplus.cern.ch` to access it outside CERN
 - <https://j2eeps.cern.ch/atlas-project-SQLTuple/ES.jsp?guid=<guid>&show=<show>> for Event Service
 - `java -jar EIHadoopCatalog.exe.jar` for remote Catalog CLI
 - `java -jar EIHadoopEI.exe.jar` for remote Event Index CLI
- JavaDoc:
 - <http://cern.ch/hrivnac/Activities/Packages/TagConvertor/JavaDoc>
- SourceDoc:
 - <http://cern.ch/hrivnac/Activities/Packages/TagConvertor/Src>
- SVN:
 - `svn+ssh://svn.cern.ch/repos/atlasoff/Database/TAGHadoop/TagConvertor`
- Ant targets:
 - `ant -p` to get help on available targets