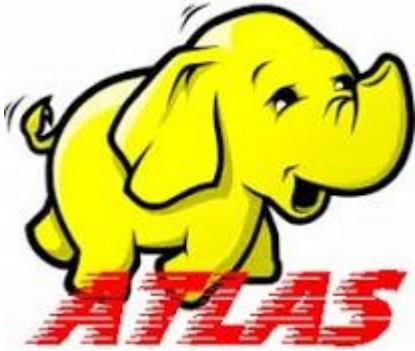
TagConvertor & ...

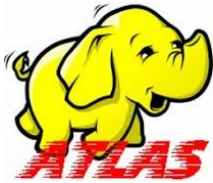


- CLI
  - Options
  - EICache
- Web Service
  - Authentication
- Event Service
- Import

latest version: 1.7.0

J.Hrivnac, EI WS, 3Jun14, CERN
+ Justin, Fedor, Rainer, Andrea,...

CLI - Options



-mr/-scan:
Can contain any formula giving boolean.

-count:
Doesn't create list of results but counts a sum of formulas evaluated for each result.
Can contain any formula giving double.

-filter:
Specifies, what should be given to user.
Can contain a list of variables or any formula giving String (starting with *String* keyword)..

-index:
Specifies index key for output (=> sorting).
Can contain any formula giving String.
-output option removed.

ant/setup.sh defines aliases for common tasks:
catalog
ei
inspector

All formulas can access tag variables and special functions ("closures" in TagFile), many trigger functions provided by Fedor..

Variables should be properly casted,
eg.: *EventTimeNanoSec* is *long* => to be used for sorting like: **-index 'String.valueOf(EventTimeNanoSec)'**

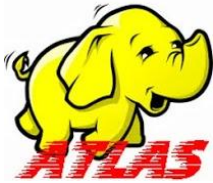
examples:

-count '1'
counts number of results (doesn't write results to file => faster)

-mr 'hasGuid("1234567890")' -filter 'String token()' -index 'oid0'
implements Event Service using Event Index API, can be wrapped in more user-friendly call

CLI - ElCache

Strategy:
Hadoop is fast in processing & converting
=> when a new selection is needed,
make it !



Search results are stored (cached) as standard TagFiles.
They can then be re-used for next search. (When run from Web Service or as atlevind id.)

Example:

"Bigger is Better"

ei -query path:ElHadoop/data11_7TeV.periodK.physics_Muons.PhysCont.AOD.repro09_v01 -mr 'true'
writes new TagFile ElCache/2014_4_27_4_34_21_561 which contains tags from all searched files
(because -query path:... works on path substring and -mr 'true' selects everything).
This takes 20' (mostly taken by sorting and writing results into output TagFile).

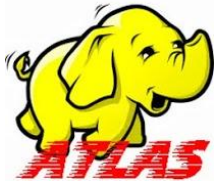
ei -query path:ElCache/2014_4_27_4_34_21_561...
will then search in that new TagFile.

"Query Spaces"

Cached files are stored with their genealogy
=> that will allow automatic selection of the best match for selections
(the closest bigger TagFile).

Note:
-mr 'true' -index 'abc'
will re-arrange TagFile so that it is indexed by 'abc' variable
=> -key query searches on 'abc' variable and gives immediate answer

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

three sub-services:

- Catalog
- Event Index: full interface
- Event Service: without decoration, usable from wget/curl

closer integration with Hadoop (Justin):

- progress bar
- links to log,...
- links to results (to bookmark),...

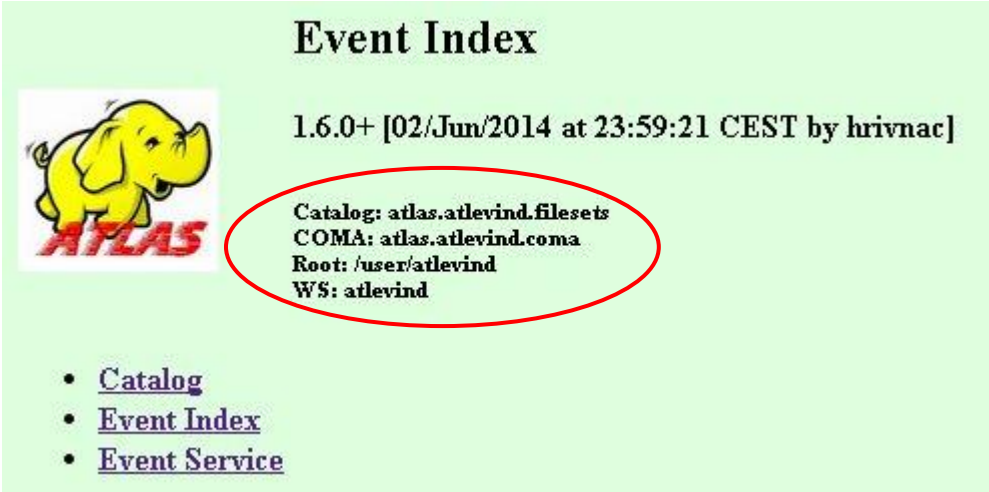
Web Service - Authentication

Overlap of several access control systems:

- Hadoop
- HBase
- Tomcat
- Kerberos

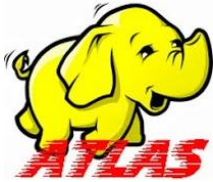
Solved :-) mainly thanks to Rainer.

Now all subsystem run under service account **atlevind** (= **atl**as**ev**ent**ind**ex).
Another account can be specified in *build.properties* (for procedural & testing reasons),
it can't be changed at run-time.



Event Service

(Event Index Service for Event Service)



Ad-hoc prototype (of functionality & api) was replaced with special call to standard Event Index service (should be a bit faster):

ei -query ElHadoop/testtagfiles -mr 'hasGuid("1234567890")' -filter 'String token()' -index 'oid0'

function called in search step, can be more complex if needed

function called in presentation step, can ask for anything else too

results ordering definition

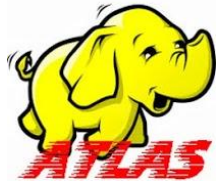
In Web Service, details are hidden behind interface:

https://aiatlas016:443/ElHadoop/ES.jsp?query=ElHadoop/testtagfiles&guid=1234567890&show=5

If not fast enough, index with guid key will be created:
-mr 'true' -index 'guid' to create new TagFile (with *RunNumber_EventNumber* key replaced with *guid* key).
-key '1234567890' -index 'oid0' to perform fast search on new TagFile based on *guid* key, ordered by *oid0*.
(Could also generate pure index TagFile with references to original TagFile.)

Only txt report format supported, other formats can be added if needed.

Import



data11_7TeV.periodK.physics_Muons.PhysCont.AOD.repro09_v01
imported, but some entries are missing because source was changing during import.
It will be re-imported.

For each new import, list of “final” files (directories) should be specified.

<u>import:</u>

1. convert *csv* into *seq*
2. sort by key
3. convert *seq* into *map* (= create index file)
4. insert into Catalog
5. move files and Catalog entries into *atlevind* account

So files are copied several times to have better control over procedure => quite slow
steps will be merged to make it faster.

Or *DataCollection* can create proper map file.

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
- JavaDoc:
  - http://cern.ch/hrivnac/Activities/Packages/TagConvertor/JavaDoc
- SourceDoc:
  - http://cern.ch/hrivnac/Activities/Packages/TagConvertor/Src
- SVN:
  - svn+ssh://svn.cern.ch/reps/atlasoff/Database/TAGHadoop/TagConvertor
- Ant targets:
  - ant -p