



Data Store and Query System

- *Import*
- *Web Service*
- *Cleaning*

latest version: **1.17.4**

*J. Hrivnac, Atlas SWWS, 22Sep15, CERN
+ Justin, Fedor, Rainer, Andrea,...*

Import Status



- Import speed up
 - now importing 50k TagFiles per week
 - very unequal performance
 - total imported so far: 1390k TagFiles
 - about 1% of failures
 - was 5k TagFiles per week
 - thanks to
 - code optimisation
 - faster write
 - in case of HBase timeout: wait 10 min and then re-try (20 times)
 - less HBase timeouts (but not zero, even after 20 re-tries)
 - more imports in parallel
 - sometimes many HDFS warnings, slowing import down
 - some failed imports leave behind wrong or corrupted files, which should be corrected
 - could probably speed up by another factor 5 without big problem
- VALIDATED coming even faster
 - a big jump recently
 - now 7k EI + 9k MC conf files waiting to be imported
 - *1 EI conf file = cca 70 TagFiles + 1 MC conf file = cca 500 TagFiles*
=> cca 490k EI + 4500k MC TagFiles waiting to be imported
 - *just very approximative numbers, someone can do precise statistics and include in monitor:*
aiatlas016:/data/TagConvertor-statistics

Import Speed Up



- Import does five things:
 - 1) changes format of TagFiles (adds fields)
 - 2) puts some information from dir-name into files (so that they are searchable)
 - data15_13TeV.00267358.physics_CosmicCalo.recon.AOD.r6832
=>
project=String streamName=String prodStep=String dataType=String version=String
 - 3) adds Trigger information
 - L1trigChainsTAV=String L1trigChainsTAP=String L1trigChainsTBP=String L2trigChainsPH=String
L2trigChainsPT=String L2trigChainsRS=String EFtrigChainsPH=String EFtrigChainsPT=String EFtrigChainsRS=String
 - 4) updates Catalog
 - 5) merges into *.1 TagFiles
- first two tasks could be done already on the Consumer
 - so would write TagFiles in final format, just with missing Trigger information
- then import can be done in two stages:
 - 1) just simple copy + Catalog update + merge:
 - fast and gives all functionality (ES, EL,...) except Trigger search
 - 2) Trigger update:
 - should re-write TagFiles (could be done on merged files)
 - takes time
 - can be done lazily
 - not available for 2015 yet anyway

Web Service



- IT asked to open aiatlas016 Web Service world-wide
 - to give easy interactive access
 - to speed up and simplify CLI access
- IT has done security evaluation
 - found no application vulnerability so far
 - found many system & configuration (Tomcat) vulnerabilities
 - our Tomcat is official (puppet) installation, so strange that it has vulnerabilities
 - IT replied that it has only checked Tomcat version, and that probably everything is ok because security patch has been applied
 - still one conf vulnerability should be fixed (SSL?)

Cleaning



- Too many small files on HDFS
- Many are just intermediate phases of import:
 - *TOBEDELETED Consumer files
 - already imported Consumer files (imported into *.2 files)
 - *.2 files (already merged into *.1 files)
- Don't want to completely remove those files as they were expensive to create and may be useful in case of a problem
- Will archive them
 - tar.gz them in HDFS (that will already cut the number)
 - move them somewhere else, if necessary (where ?)
 - (maybe) we can delete *TOBEDELETED files ?
- That will remove > 90% of files and free 50% of space (if moved out)
- is there anything like “**hadoop fs tar ...**” ?