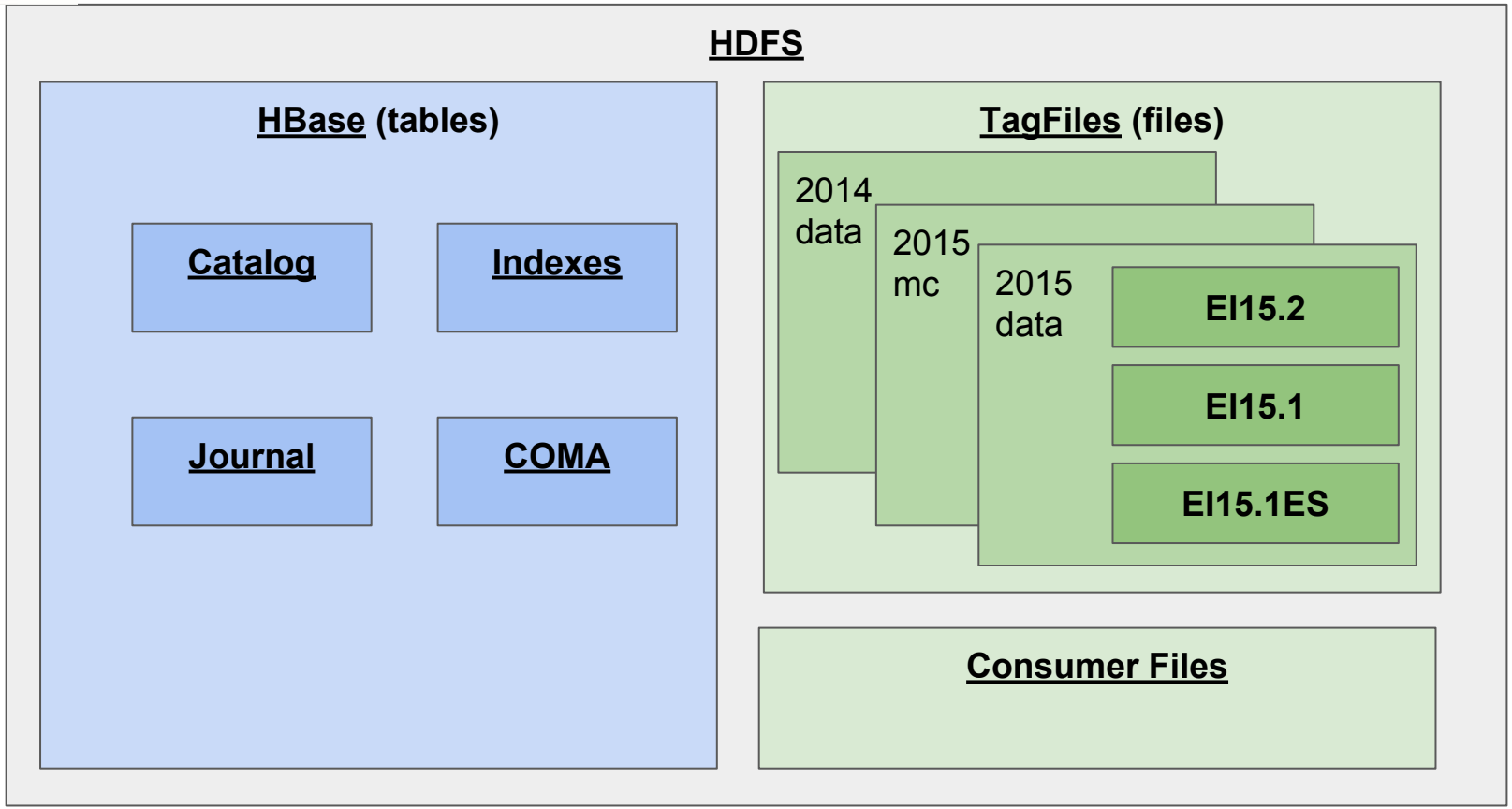
Import, Storage and Access

- Data
- Ecosystem
- Import
- Access
- Deployment
- Command Line
- Web Service
- Catalog
- Problems & Solutions
- Other Plans
- Summary

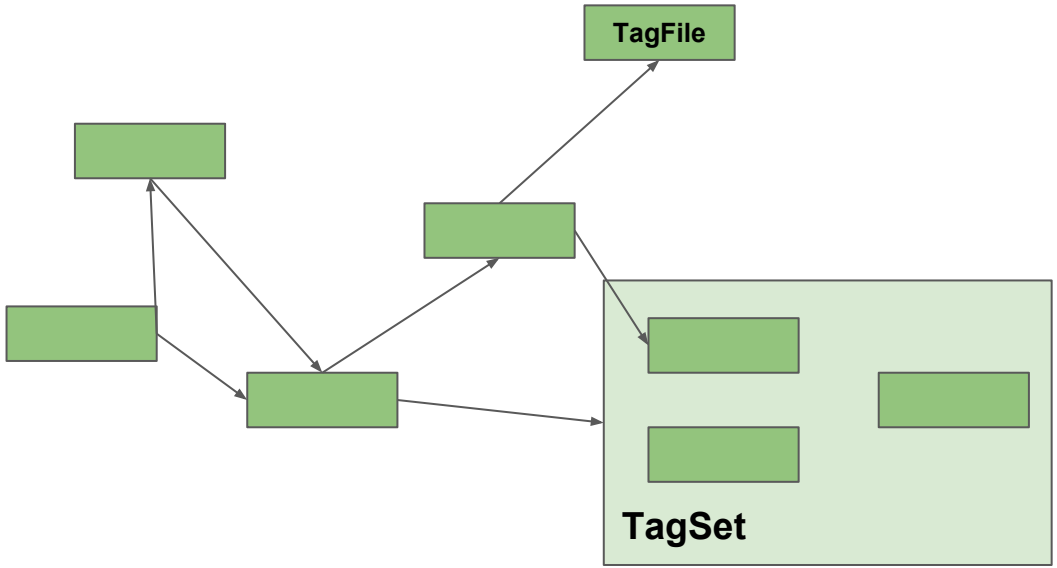
Julius Hrivnac, LAL Orsay
for
Event Index Technical Review, 24Nov2015

Data

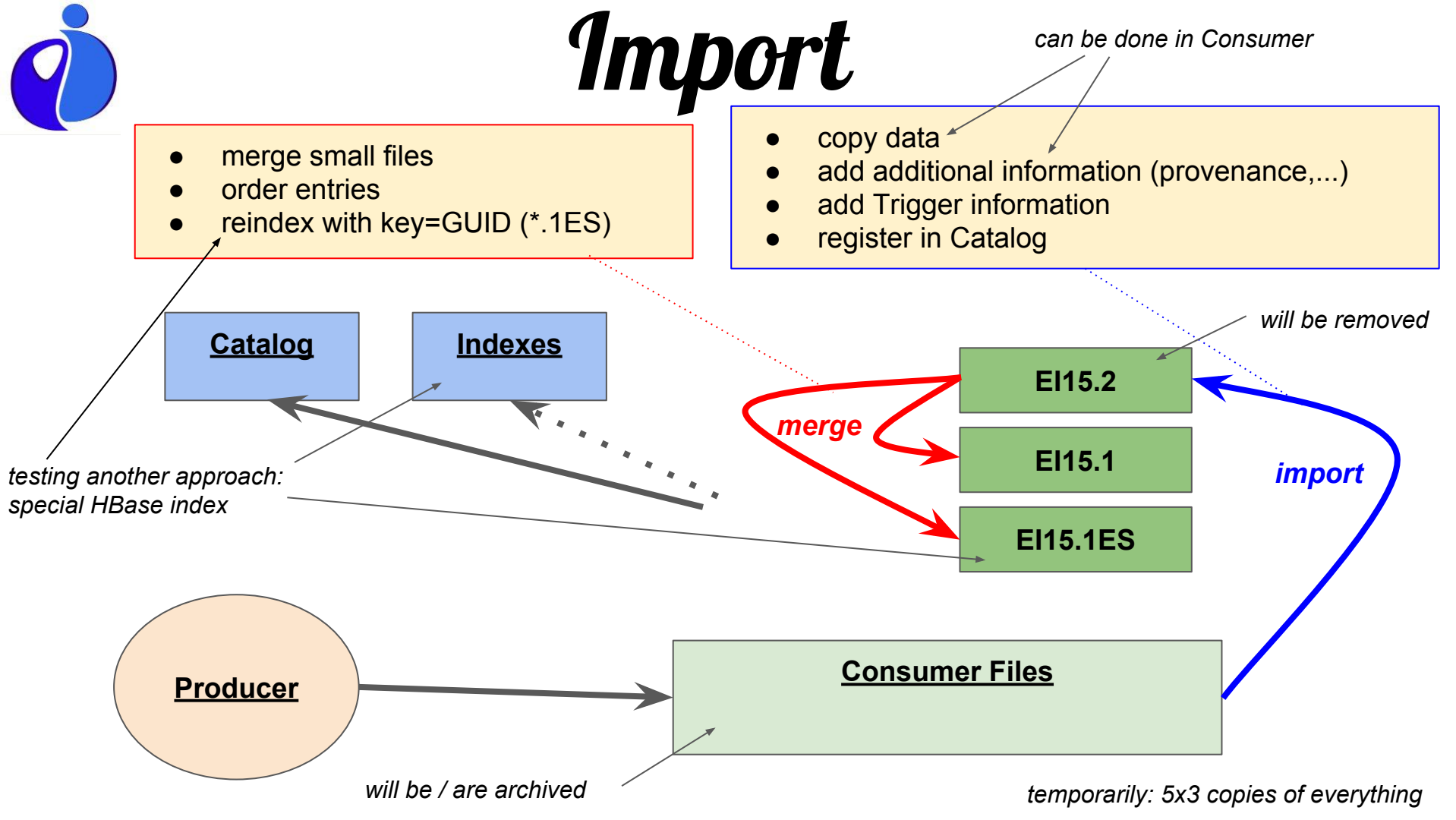

HDFS
HBase (tables)
Catalog
Indexes
Journal
COMA
TagFiles (files)
2014 data
2015 mc
2015 data
EI15.2
EI15.1
EI15.1ES
Consumer Files

Ecosystem

- TagFile = datase or anything derived from it
- TagFiles can be related
  - by origin
  - by collection (= TagSet)
  - via sequence
  - …
- each operation is transformation
  - writing new TagFiles to be re-used
    - to download result
    - to refine search
    - ...
  - output writing may be skipped
    - currently skipped in most cases



Import

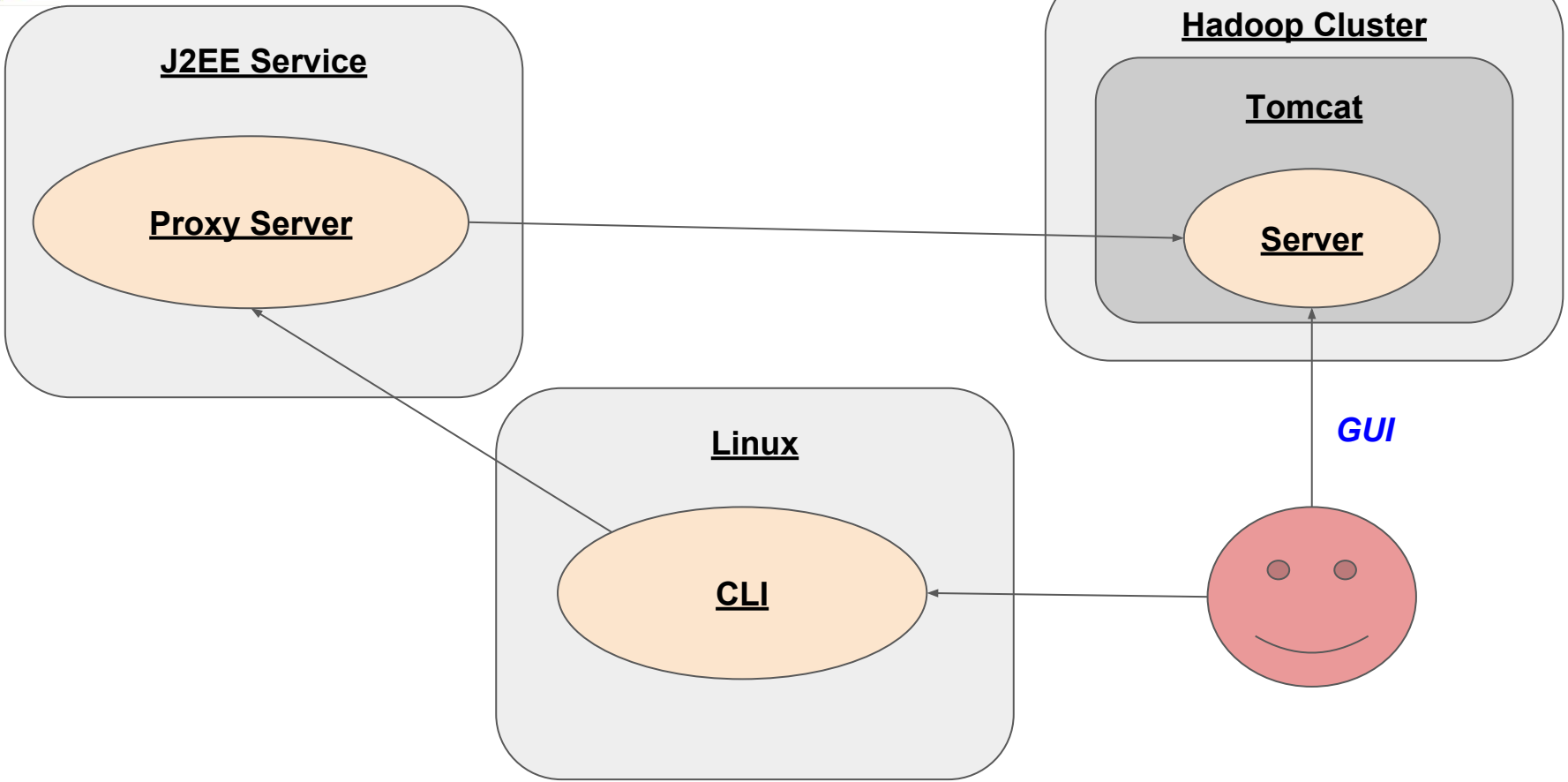

can be done in Consumer
merge small files
order entries
reindex with key=GUID (*.1ES)
copy data
add additional information (provenance,...)
add Trigger information
register in Catalog
Catalog
Indexes
will be removed
EI15.2
merge
EI15.1
import
EI15.1ES
testing another approach:
special HBase index
Producer
Consumer Files
will be / are archived
temporarily: 5x3 copies of everything

Import

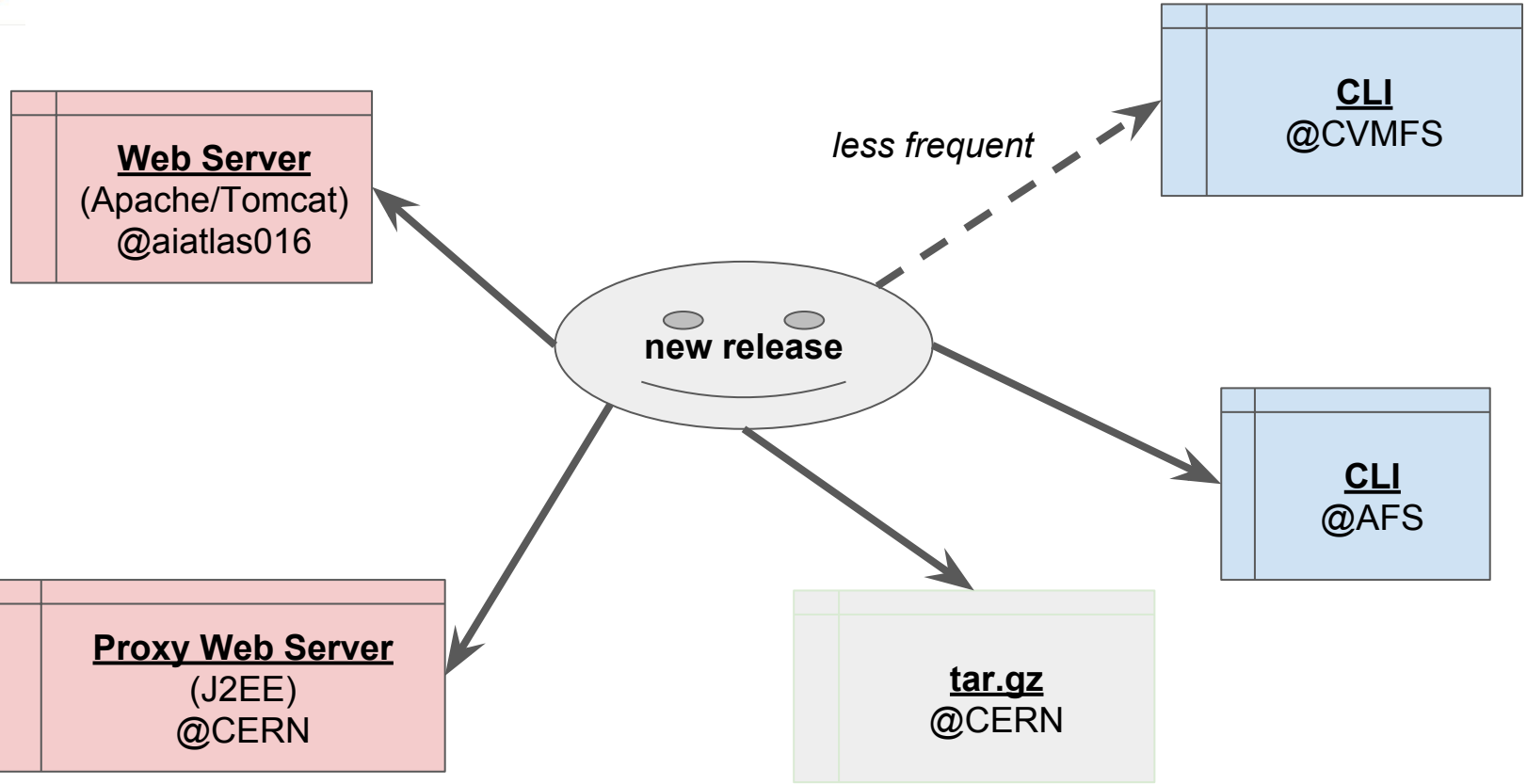
- Consumer records successfully consumed files in conf-files in AFS directory
  - may use Catalog to pass the information around
- Importer reads those files and use them to start two stage import
  - some of that can move to Consumer
- Verificator is running every three hours to find import inconsistencies and accumulate import statistics

EI15.2/data15_13TeV.00276161.debugrec_hlt.merge.DAOD_TOPQ3.g49_f618_m1480_p2411**/**
85D56A13-1175-824E-A2E5-223115B595F6_927f49e38bfa42aaa8c8cf831274d8f6_af2bd08d-5312-4e30-ab9f-f19047f903d0_6555935.G_2617928832.1
.....
EI15.1/data15_13TeV.00276161.debugrec_hlt.merge.DAOD_TOPQ3.g49_f618_m1480_p2411

Access


J2EE Service
Proxy Server
Hadoop Cluster
Tomcat
Server
Linux
CLI
GUI

Deployment


Web Server
(Apache/Tomcat)
@aiatlas016
less frequent
CLI
@CVMFS
new release
CLI
@AFS
Proxy Web Server
(J2EE)
@CERN
tar.gz
@CERN

Access (Search)

- Search executed in two steps:
  - (HBase) Catalog finds TagFiles to search from
    - speed depends on search criteria from < 1s (exect key based search) to 20s (full scan search)
  - Hadoop executes search in those files:
    - key-search (almost immediate) if searching directly on RunNumber-EventNumber (key of *.1), GUID (key of *.1ES)
    - Map/Reduce job if searching with other attributes (Trigger,...)
- Evaluating possibility to create small index HBase tables pointing directly to TagFiles
  - to replace *.1ES
  - looks good
- Access (local and remote) is tested nightly
- Remote access goes via proxy server to Tomcat @ Hadoop cluster to assure
  - Hadoop isolation from outside environment
  - fallback & load-balancing (not yet)

Command Line

- CLI commands available on the cluster and remotely from AFS, CVMFS and downloadable tar file:
- **catalog** to search and update Catalog
- **ei** to search data (uses catalog for first pass selection) gives very rich searching interface
  - allows to use any code (using TagFile fields) as a searching formula delivering
  - can deliver found entries (formula gives boolean)
  - can count results (formula gives number)
  - can use another formula to create result from found entries
- **inspector** to inspect TagFiles content
- **importer** to import new data from Consumer
- **verificator** to verify correctness of imported data:
  - finds problems due to faults during import
  - checks consistency of Catalog wrt data, internal consistency of Catalog, completeness of import,...
- **eventlookup** (wrapper over ei) to provide Even Lookup - like API
- **eventservice** (wrapper over ei) to provide API requested by Event Server
- other wrappers can be added
- Python clients are available
- Web GUI is available

Command Line Examples

catalog -query id:EI14.2.data14_cos.physics_Main.merge.AOD.f529.00248373'
catalog -count 'path:EI14.2/data14_cos.00248373.physics_Main.merge.AOD.f529' # prefix match
catalog -query 'exact path:EI14.2/data14_cos.00248373.physics_Main.merge.AOD.f529_m1359'
catalog -query 'runNumber:00248373 version:f594_m1435 dataType:AOD streamName:physics_Main'
catalog -add   'path:EICache/Test name:ATest'
catalog -query 'path:EICache/Test' -modify 'mytag:tag1'
catalog -delete 'path:EICache/Test'

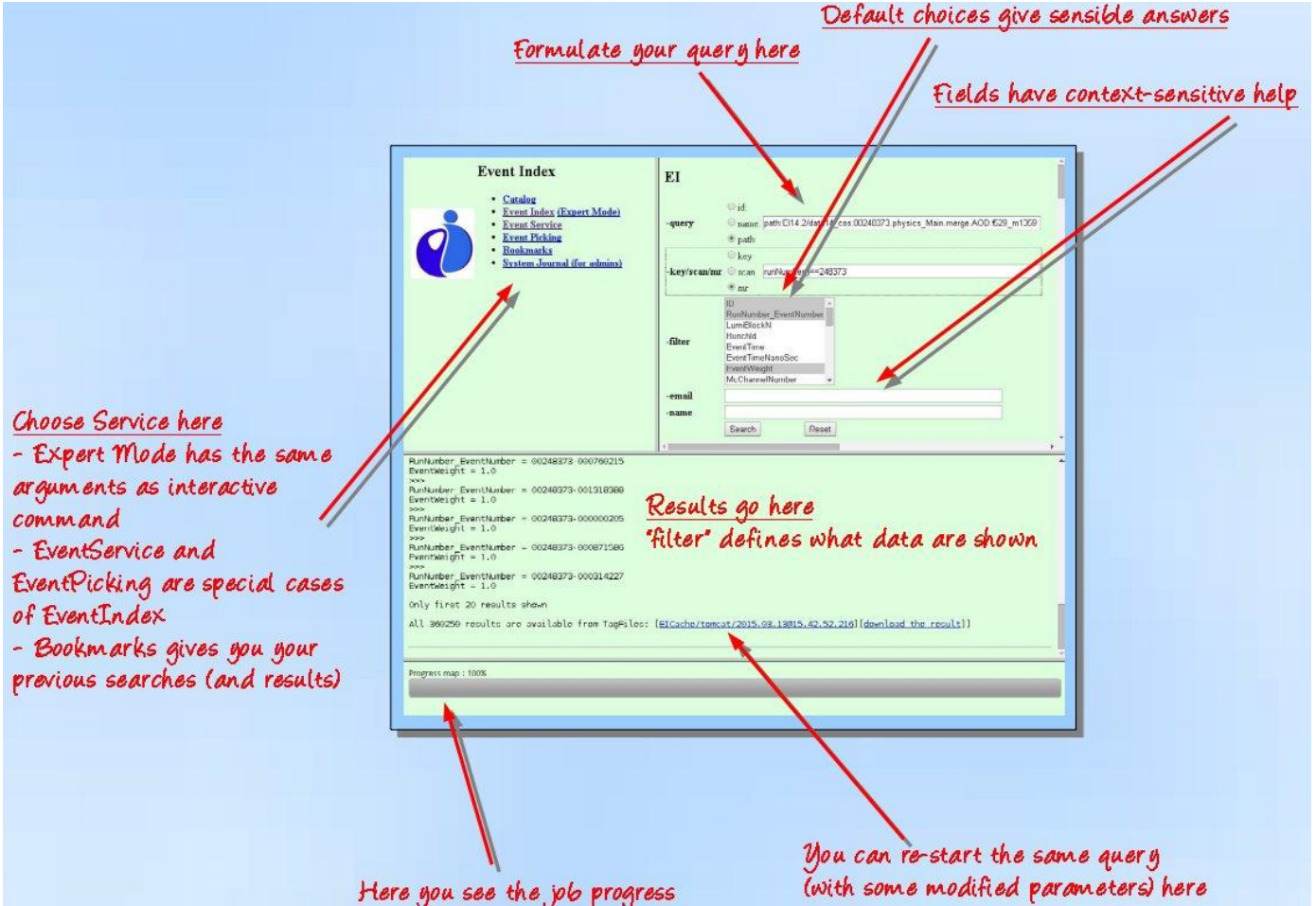
Catalog query

ei -query **$qrx** -show 2 -key '00248373-000983650..00248373-000983656'
ei -query $qry -key '00248373-000983650..00248373-000983656' -scan 'hasGuid("EC2886B0-519E-704B-8F3A-92CF124E3D5E")'
ei -query $qry -key '00248373-000983650..00248373-000983656' -scan 'hasGuid("EC2886B0-519E-704B-8F3A-92CF124E3D5E")' -count '1'
ei -query $qry -key '00248373-000983650..00248373-000983656' -scan 'hasGuid("EC2886B0-519E-704B-8F3A-92CF124E3D5E")' -count 'EventWeight*EventWeight'
ei -query $qry -key '00248373-000983650' -output index -index 'id=String.valueOf(BunchId)'
ei -query $qry -key '00248373-000983650..00248373-000983656' -output index -index 'id=String.valueOf(BunchId)' -scan 'hasGuid("...")'
ei -query $qry -mr 'BunchId==2390 && LumiBlockN==70'
ei -query $qry -mr 'runNumber()==248373' -filter 'BunchId RunNumber_EventNumber'
ei -query $qry -mr 'true' -count 'EventWeight'
ei -query $qry -mr 'BunchId==2390' -output index -index 'id=String.valueOf(BunchId)'
ei -query $qry -mr 'hasGuid("EC2886B0-519E-704B-8F3A-92CF124E3D5E")' -filter 'String token()' -index 'oid0'
ei -query $qry -eventlist eventlist.txt -filter 'clid0'
ei -query $qry -show 2 -mr 'true' -extent 'NewField=String.valueOf(BunchId*BunchId)'
ei -query $qry -show 2 -mr 'true' -update 'NewField=String.valueOf(2*NewField)'

 # wrappers over ei:
eventlookup -e '0263962 000000623,00263962 000000511' -s physics_MinBias -p m1420
eventlookup -f events.txt -s physics_MinBias -p m1420
eventservice -guid 00020A71-86E3-EA44-B62D-62F19E303574

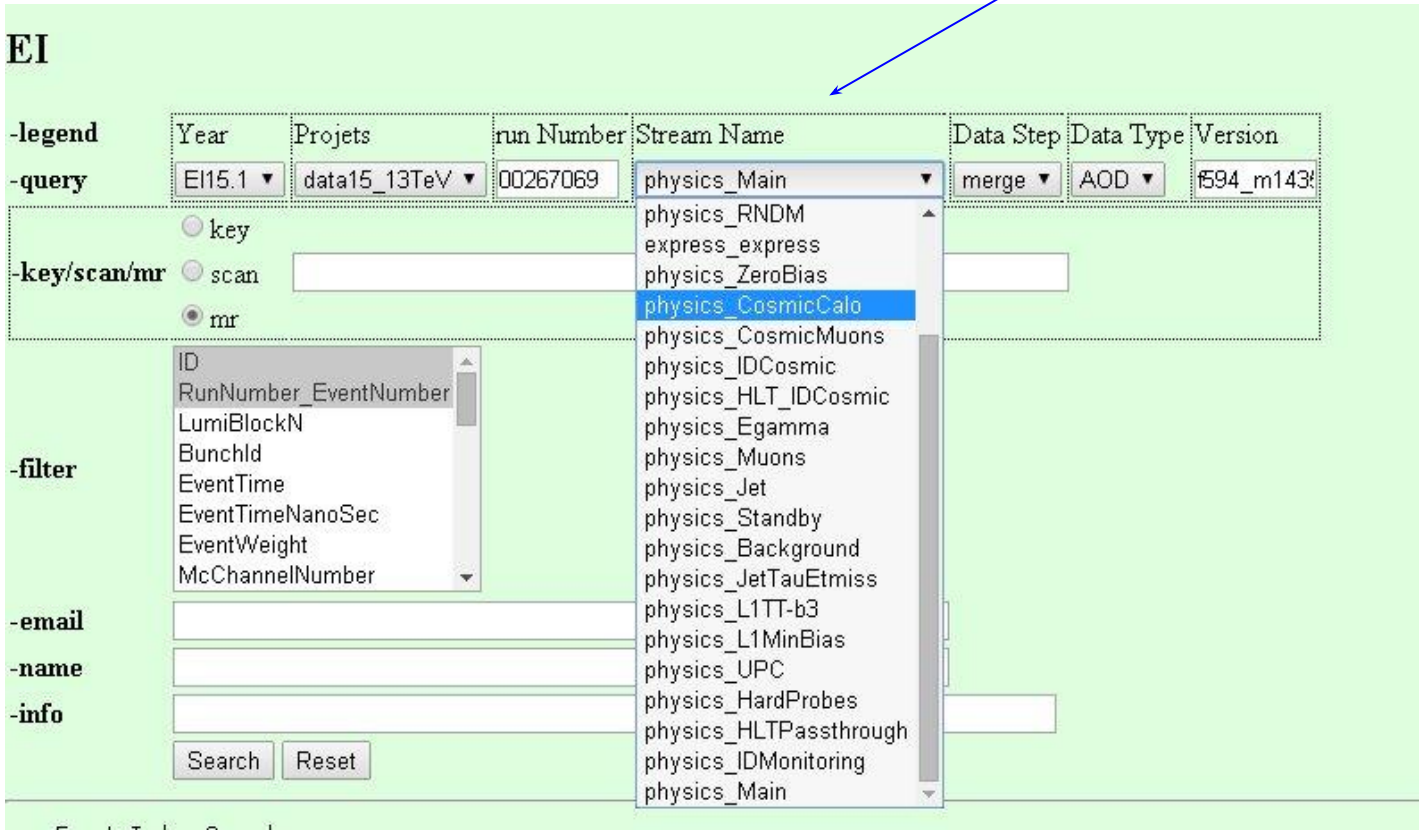
*all **ei** commands write result into standard TagFiles,*
they can be re-used later
(as results or source for another search)

Web Service


Default choices give sensible answers
Formulate your query here
Fields have context-sensitive help
Event Index
Catalog
Event Index (Expert Mode)
Event Service
Event Picking
Bookmarks
System Journal (for admins)
EI
-query
-key/scan/mr
-filter
-email
-name
Search
Reset
Choose Service here
- Expert Mode has the same arguments as interactive command
- EventService and EventPicking are special cases of EventIndex
- Bookmarks gives you your previous searches (and results)
Results go here
"filter" defines what data are shown
Only first 20 results shown
Here you see the job progress
You can re-start the same query (with some modified parameters) here

Web Service

context sensitive menu



Catalog

- Catalog schema is **very flexible**, the only strictly defined field is key:
  <generation>.<project>.<streamName>.<prodStep>.<dataType>.<version>.<runNumber>
  (EI15.1.data15_13TeV.physics_Main.merge.DAOD_HIGG3D2.f618_m1480_p2411.00276147)
- Searching on the key is very fast
- All key components are available as individual columns to allow search like:
  -query “prefix id:EI15.1.data15_13Tev.Physics_Main runNumber:00276147”
- There are three column families;
  - **d**escription of the TagFiles
  - **r**elations to other TagFiles
    - to create TagSet = collection of TagFiles
    - to record TagFile genealogy (most operations are transformations)
    - …
  - **a**ttributes can be added and modified by any (authorised) process **[very useful]**
    - to set status (incomplete, deleted,...)
    - to tag
    - to add information
    - ....

Problems & Solutions

(so far)

- HBase access timeouts
- Slow Catalog
- Slow Import
- Slow Access

- learning how to use new technology, it's strong & weak points
  - sometimes developing several prototypes to test
- first concentrating on functionality, then (now) on speed
  - still a lot of debugging code present (log, journal, mails,...)
- new requirements

HBase access timeout

(=> missing & wrong data)

- serious problem last Summer
- too frequent timeouts of HBase access => Catalog unusable
- consequence: almost no import during Summer
  - big data backlog
  - data imported with errors
    - will fix them once backlog absorbed (or on request)
  - presence of many small files (which should have been imported and removed) hurts Hadoop performance and makes upgrade difficult
- solution:
  - unreasonably big HBase tables removed and HBase reconfigured
- “many small files” problem being solved by archiving files into HAR archives and importing from them

Slow Catalog

- almost all request required full-scan, which resulted in > 20s latency
- cause: Catalog key was a random string
- solution: Catalog key with navigational information
  - search on Catalog key
  - search on fields from catalog key is very fast (because it allows to define row range to search on)
  - migration of Catalog done, most access code already migrated

Slow Import

- import slow in processing backlog data from Summer (HBase timeout problems)
  - but able to cope with data coming from Consumer
- causes:
  - inefficient Catalog
  - slow sorting of imported files
  - one import server @Wigner (10x slower)
- solution:
  - Catalog update (done)
  - new import server @CERN (in place)

import speed went up a lot last two weeks thanks to code optimisation and new Catalog

- at this moment, there are several import queues:
  - 2015 runs, running backwards (FILO)
  - priority, on request
  - older runs (lower priority)
  - re-import of wrongly imported data
- MC stopped till all data imported

- **16/11:**
  - **11757 El15 TagFiles imported, 7055 waiting for import**
- **23/11:**
  - **17853 El15 TagFiles imported, 5150 waiting for import**
- **so imported 6000 TagFiles in a week, backlog went down by ¼**
- **+ 9000 MC15 TagFiles waiting, 3000 imported**
- **+ x100 TagFiles to be re-imported**

Slow Access (Search)

- causes:
  - inefficient Catalog
  - double Web Service re-direction
  - Apache/Tomcat/J2EE stuck (rare, but happens)
- solution:
  - Catalog update (done)
  - simplification of Web Access
    - fallback server
    - Tomcat @Hadoop ww visible
  - special purpose indexes
- originally, we wanted to merge all EI15.1 into one EI15.0 TagFiles
  - that would make (some) searches very fast
  - but merging (10xTB) failed (sorting)

Generic API and Event Lookup made much faster following Catalog update

ei -query 'id:EI15.1.data15_13TeV.express_express.merge.AOD runNumber:00280464' -key '00280464-00001567911 00280464-00001571727 00280464-00001584942 00280464-00001590701 00280464-00001607520 00280464-00001609308 00280464-00001609485 00280464-00001612678' -filter 'String guids()'

 # takes 11s @lxplus, 7s @hadoop; searching for just one event takes 4s @hadoop

Other Plans

- Proposed speed up improvements (feasibility to be evaluated):
  - multithreaded importer
  - importing and merging in memory (for small files)
- Testing of new file formats (instead of Map files)
- Web Service:
  - currently have only one Web Server for both DEV and PRO
  - should set up a second one for DEV/PRO, fallback a possibly load-balancing
- Backup
  - HBase tables are backuped nightly
  - HDFS files not backuped
- Query Spaces:
  - system designed in such way, that results of all searches are registered as new TagFiles and can be used for further re-fined searches
  - the mechanism is in place, but has not been activated due to more urgent tasks

Summary

- Chosen technology:
  - Two-level storage: HBase catalog + HDFS Map TagFiles
  - Web Service for remote access
- Strong points:
  - Flexibility: new requirements can be easily added, technology changes can be incorporated
  - Performance scaling: for most cases constant or at worst linear
  - Mainstream technology
- Weak points:
  - New technology (to us), still learning how to use it efficiently (and not to break it)
  - Changing technology: new tools arrive very often
- Development strategy:
  - first functionality, then speed
  - but perturbed by changing requirements
- External dependencies:
  - Hadoop
  - Apache & Tomcat
  - everything else is easily replaceable
- Constrains & obstacles:
  - Security requirements (Web Service authentication)
  - Configuration of some CERN services