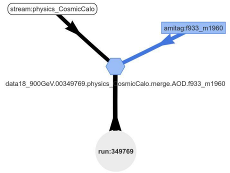
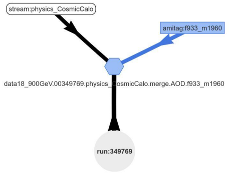
Phoenix+


stream:physics_CosmicCalo
amitag:f933_m1960
data18_900GeV.00349769.physics_CosmicCalo.merge.AOD.f933_m1960
run:349769

- Motivation
- Architecture
- Interface
- API
- Implementation
- Prototype
- GUI Status
- Notes

Julius Hrivnac, LAL
EI WS, 3-5 June 2019, LAL

Motivation


stream:physics_CosmicCalo
amitag:f933_m1960
data18_900GeV.00349769.physics_CosmicCalo.merge.AOD.f933_m1960
run:349769

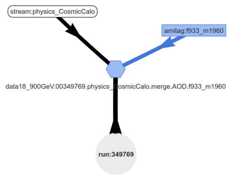
Adding flexibility to Phoenix storage.

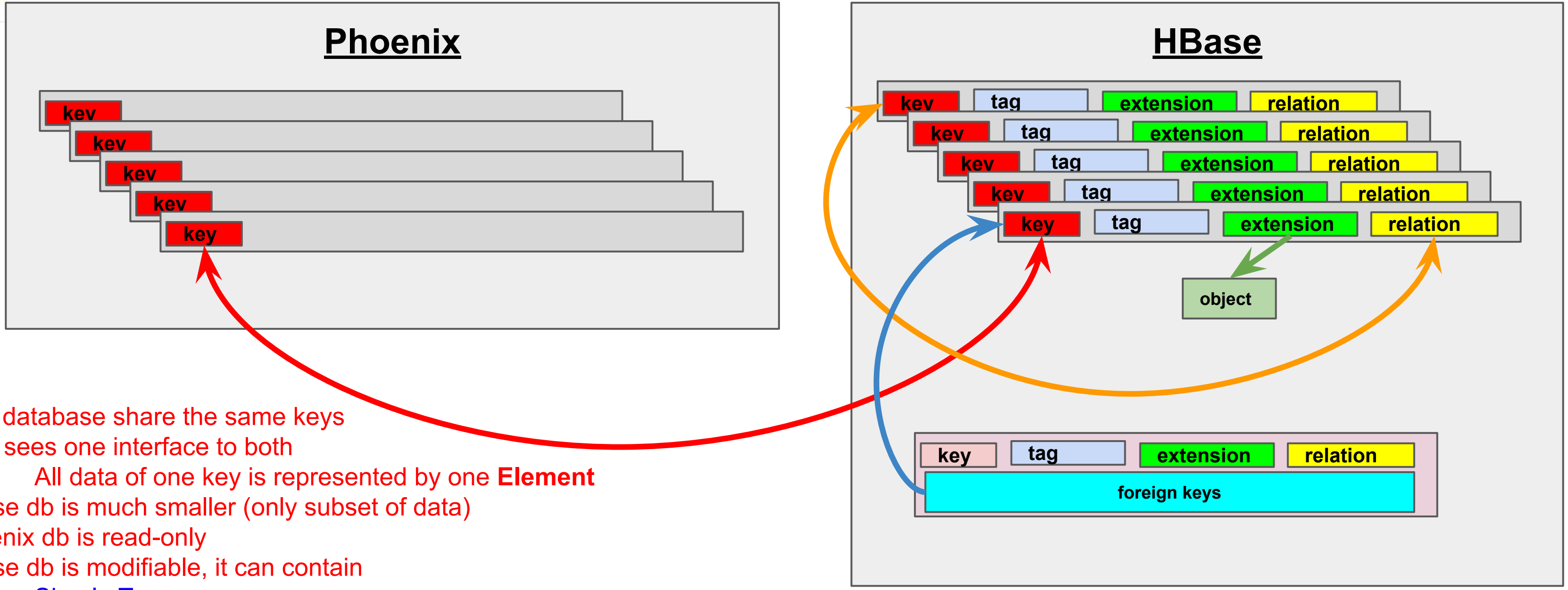
- Migrating to Phoenix, with **only SQL** interface, we lose the flexibility present in the current system
  - This makes implementation rigid & fragile wrt new or changed requirements
  - Yes, you can always create new SQL table with foreign keys, etc.
  - But designing complex SQL tablespace is tricky and easily goes haywired
  - Event more dangerous is adding new SQL features as needs come
  - Complex SQL queries are very difficult to optimise (remember TagDB)
    - Formulating a new SQL query is always risky
  - NoSQL storage, however, is made exactly for the iterative, unprevisible evolution

Open-Close Architecture

- Existing data-centric Web Service
  - Which has quite awkward connection with current EI Core
  - But UI WS part can be easily re-engineered for different backend
- Proposed solution:
  - Extend Phoenix storage with pure HBase storage
    - Sharing the same keys
  - So keeping Phoenix advantages (speed, SQL interface) for RO data
  - While opening for new possibilities, adaptability to changing environment
  - Phoenix for static data, HBase for dynamic data

Architecture

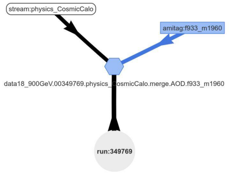

stream:physics_CosmicCalo
amitag:f933_m1960
data18_900GeV.00349769.physics_CosmicCalo.merge.AOD.f933_m1960
run:349769



- ➢ Both database share the same keys
- ➢ User sees one interface to both
    - ○ All data of one key is represented by one **Element**
- ➢ HBase db is much smaller (only subset of data)
- ➢ Phoenix db is read-only
- ➢ HBase db is modifiable, it can contain
    - ○ Simple **Tags**
        - ■ They can be also used in search filter
    - ○ **Extensions** with any object
        - ■ E.g. Trigger statistics and overlap, duplicated events list,...
    - ○ **Relations** to other elements (like Graph DB)
        - ■ E.g. overlaps between datasets

- ➢ HBase can also contain Elements without Phoenix partner: **Hubs**
    - ○ They represent pre-defined **virtual collections** of Elements
        - ■ E.g. Amitag, Stream, Run, Project,...
    - ○ They can be extended and searched in the same way as other Elements
- ➢ Ad-hoc virtual collections can be build also using Tags

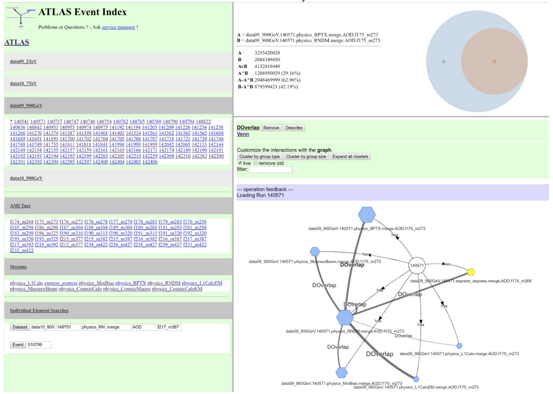
Interface



stream:physics_CosmicCalo
amitag:f933_m1960
data18_900GeV.00349769.physics_CosmicCalo.merge.AOD.f933_m1960
run:349769

- ➢ API - done
- ➢ REST Web Service (also serving CLI) - done
  - ○ Serving text; xml or json can be added
- ➢ Graphical Web Service - under development
  - ○ Adapting existing GraphDB WS

```
ElementFactory ef = …;
Dataset dprototype1 = new Dataset();
dprototype1.set("runnumber", 140571).
           .set(“project”, “data09_900GeV”).
           …;
Dataset dataset1 = (Dataset)ef.search(dprototype1).get(0);
…
Dataset dataset2 = (Dataset)ef.search(dprototype2).get(0);
dataset2.add(new Overlap(10, 30, 50, 40, dataset1));
dataset2.add(new Tag(“mytag”, “myvalue”));
ef.update(dataset2);
```



ATLAS Event Index

Problems or Questions ? - Ask service manager !

ATLAS

data09_2TeV

data10_7TeV

data09_900GeV

* 140541 140571 140737 140747 140748 140754 140762 140765 140769 140790 140794 140822 140836 140842 140953 140955 140974 140975 141192 141194 141203 141209 141226 141234 141238 141266 141270 141374 141387 141398 141401 141403 141534 141561 141562 141563 141565 141688 141689 141691 141695 141700 141702 141704 141705 141706 141707 141718 141721 141730 141746 141748 141749 141755 141811 141818 141841 141994 141998 141999 142042 142065 142133 142144 142149 142154 142155 142157 142159 142161 142165 142166 142171 142174 142189 142190 142191 142192 142193 142194 142195 142199 142203 142205 142210 142259 142309 142310 142383 142390 142391 142392 142394 142395 142397 142400 142404 142405 142406

data10_900GeV

AMI Tags

f174_m268 f175_m273 f176_m273 f176_m278 f177_m278 f178_m283 f179_m283 f170_m258 f185_m298 f186_m298 f187_m304 f188_m304 f189_m304 f180_m288 f181_m293 f181_m298 f183_m298 f196_m325 f190_m310 f190_m315 f190_m320 f191_m315 f191_m320 f192_m320 f193_m320 f193_m325 f215_m377 f215_m382 f215_m387 f216_m382 f216_m387 f217_m387 f217_m392 f218_m392 f212_m377 f234_m422 f236_m427 f238_m427 f239_m427 f231_m422 f232_m422

Streams

physics_L1Calo express_express physics_MinBias physics_BPTX physics_RNDM physics_L1CaloEM physics_MuonswBeam physics_CosmicCalo physics_CosmicMuons physics_CosmicCaloEM

Individual Element Searches

Dataset: data10_900 | 149751 | physics_RN | merge | AOD | f217_m387

Event: 510799

A = data09_900GeV.140571.physics_BPTX.merge.AOD.f175_m273
B = data09_900GeV.140571.physics_RNDM.merge.AOD.f175_m273

A	3253420028
B	2084349450
AvB	4132819449
A^B	1204950029 (29.16%)
A-A^B	2048469999 (62.96%)
B-A^B	879399421 (42.19%)

DOverlap Remove Describe
Venn

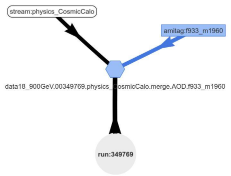
Customize the interactions with the **graph**.
Cluster by group type | Cluster by group size | Expand all clusters
☑ live ☐ remove old
filter:

--- operation feedback ---
Loading Run 140571

A
B
data09_900GeV.140571.physics_BPTX.merge.AOD.f175_m273
data09_900GeV.140571.physics_MuonswBeam.merge.AOD.f175_m273
140571
data09_900GeV.140571.express_express.merge.AOD.f174_m268
data09_900GeV.140571.physics_RNDM.merge.AOD.f175_m273
data09_900GeV.140571.physics_L1Calo.merge.AOD.f175_m273
data09_900GeV.140571.physics_MinBias.merge.AOD.f175_m273
data09_900GeV.140571.physics_L1CaloEM.merge.AOD.f175_m273
has
DOverlap

http://localhost:8888/EventIndex/REST.jsp?element=Event&search=eventnumber:57555&tag=golden:mychoice

API

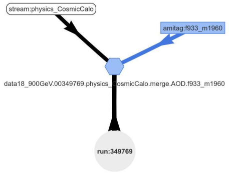

stream:physics_CosmicCalo
amitag:f933_m1960
data18_900GeV.00349769.physics_CosmicCalo.merge.AOD.f933_m1960
run:349769

1. *Create a prototype of the Element you want to search*
2. *Fill in known values*
   a. *You can use SQL for Phoenix part*
   b. *You may choose which backend (Phoenix, HBase or both) is used for searching and data filling*
3. *Send it to the ElementFactory*
4. *Get a set of satisfying Elements, with all values filled (from both Phoenix and HBase)*
5. *Add Tags, Relations of Extensions to Elements*
   a. *DOverlap is_a Relation*
   b. *TStat is_a Extention*
6. *Update via ElementFactory*
   a. *HBase will be updated*

```
ElementFactory ef = ...;
Dataset dprototype1 = new Dataset();
dprototype1.set("runnumber", 140571).
           .set(“project”, “data09_900GeV”).
           ...;
Dataset dataset1 = (Dataset)ef.search(dprototype1).get(0);
...
Dataset dataset2 = (Dataset)ef.search(dprototype2).get(0);
dataset2.add(new DOverlap(10, 30, 50, 40, dataset1));
dataset2.add(new Tag(“mytag”, “myvalue”));
dataset2.add(new TStat(......));
ef.update(dataset2);
```

- ➢ REST and GUI WS are build on top of this API
  - ○ Not all functionality is (yet) interfaced
- ➢ Similar design to the current Core system

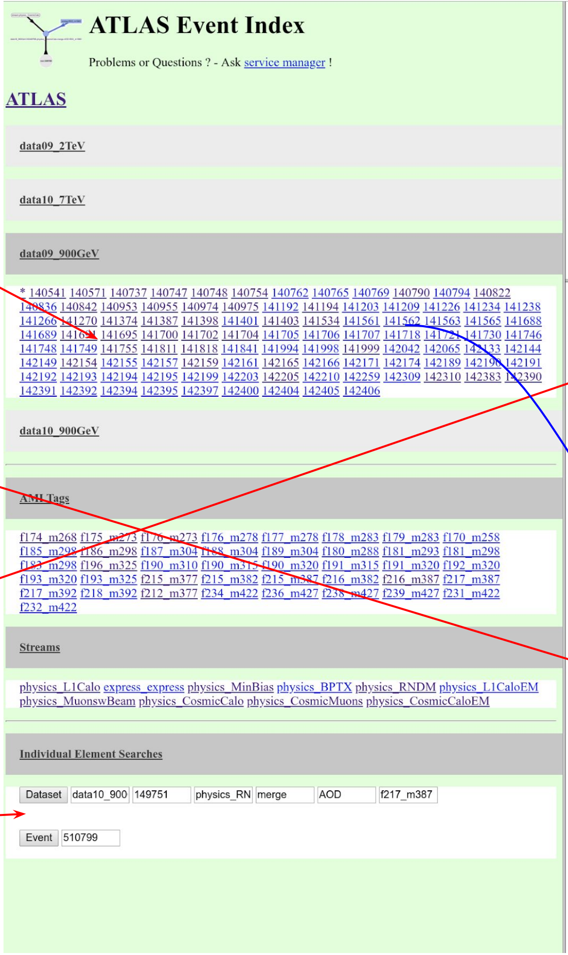
Implementation


stream:physics_CosmicCalo
amitag:f933_m1960
data18_900GeV.00349769.physics_CosmicCalo.merge.AOD.f933_m1960
run:349769

- ➢ Prototype running on the LAL (partial) replica of the CERN Phoenix storage
  - ○ The same schema
  - ○ Subset of data
  - ○ No authentication, simple configuration
- ➢ Can easily use any other Phoenix schema
- ➢ Can be moved to CERN
  - ○ Authentication, configuration ?...
- ➢ Included in the **Core EI** GIT repository

Prototype

1. Select period, run, ami tag, stream, dataset or event
   a. Or any other Hub (virtual collection)
2. Get it, together with all related entities and information stored in EI
3. Each entity has a set of possible actions to perform (EI or external)
4. Each entity can be annotated
5. Detailed search is possible too
   a. Could expose also SQL part



ATLAS Event Index

Problems or Questions ? - Ask service manager !

ATLAS

data09_2TeV

data10_7TeV

data09_900GeV

* 140541 140571 140737 140747 140748 140754 140762 140765 140769 140790 140794 140822 140836 140842 140953 140955 140974 140975 141192 141194 141203 141209 141226 141234 141238 141266 141270 141374 141387 141398 141401 141403 141534 141561 141562 141563 141565 141688 141689 141691 141695 141700 141702 141704 141705 141706 141707 141718 141721 141730 141746 141748 141749 141755 141811 141818 141841 141994 141998 141999 142042 142065 142133 142144 142149 142154 142155 142157 142159 142161 142165 142166 142171 142174 142189 142190 142191 142192 142193 142194 142195 142199 142203 142205 142210 142259 142309 142310 142383 142390 142391 142392 142394 142395 142397 142400 142404 142405 142406

data10_900GeV

AMI Tags

f174_m268 f175_m273 f176_m273 f176_m278 f177_m278 f178_m283 f179_m283 f170_m258 f185_m298 f186_m298 f187_m304 f188_m304 f189_m304 f180_m288 f181_m293 f181_m298 f183_m298 f196_m325 f190_m310 f190_m315 f190_m320 f191_m315 f191_m320 f192_m320 f193_m320 f193_m325 f215_m377 f215_m382 f215_m387 f216_m382 f216_m387 f217_m387 f217_m392 f218_m392 f212_m377 f234_m422 f236_m427 f238_m427 f239_m427 f231_m422 f232_m422

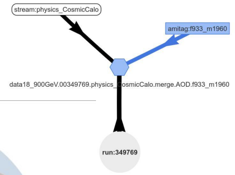
Streams

physics_L1Calo express_express physics_MinBias physics_BPTX physics_RNDM physics_L1CaloEM physics_MuonswBeam physics_CosmicCalo physics_CosmicMuons physics_CosmicCaloEM

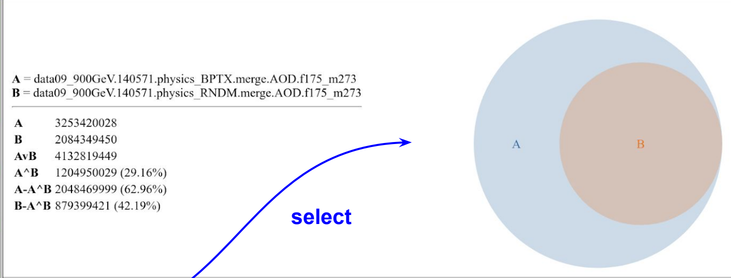
Individual Element Searches

Dataset: data10_900 | 149751 | physics_RN | merge | AOD | f217_m387

Event: 510799

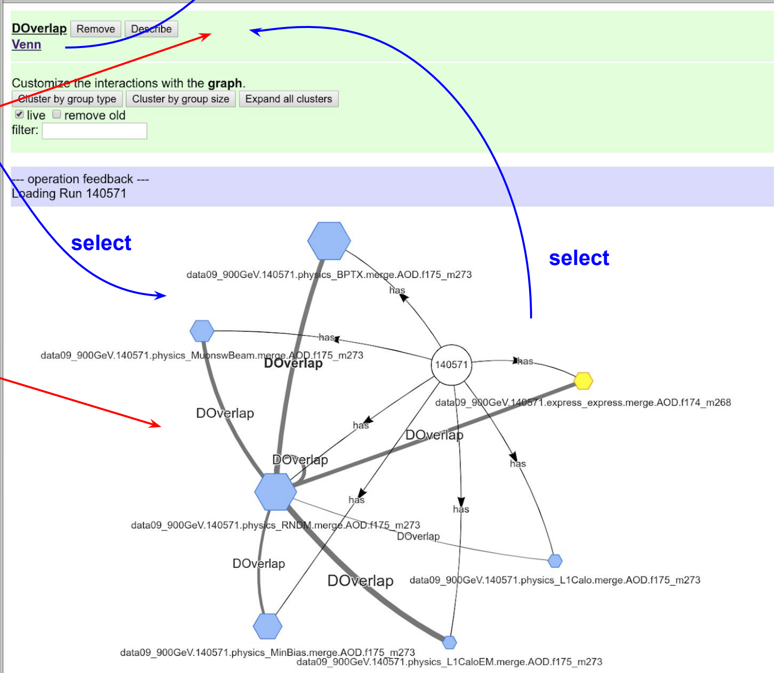


stream:physics_CosmicCalo
amitag:f933_m1960
data18_900GeV.00349769.physics_CosmicCalo.merge.AOD.f933_m1960
run:349769



A = data09_900GeV.140571.physics_BPTX.merge.AOD.f175_m273
B = data09_900GeV.140571.physics_RNDM.merge.AOD.f175_m273

A	3253420028
B	2084349450
AvB	4132819449
A^B	1204950029 (29.16%)
A-A^B	2048469999 (62.96%)
B-A^B	879399421 (42.19%)



DOverlap Remove Describe
Venn

Customize the interactions with the **graph**.

Cluster by group type | Cluster by group size | Expand all clusters

☑ live ☐ remove old

filter:

--- operation feedback ---
Loading Run 140571

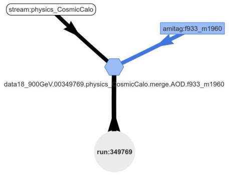
data09_900GeV.140571.physics_BPTX.merge.AOD.f175_m273
data09_900GeV.140571.physics_MuonswBeam.merge.AOD.f175_m273
140571
data09_900GeV.140571.express_express.merge.AOD.f174_m268
data09_900GeV.140571.physics_RNDM.merge.AOD.f175_m273
data09_900GeV.140571.physics_L1Calo.merge.AOD.f175_m273
data09_900GeV.140571.physics_MinBias.merge.AOD.f175_m273
data09_900GeV.140571.physics_L1CaloEM.merge.AOD.f175_m273
has
DOverlap

select

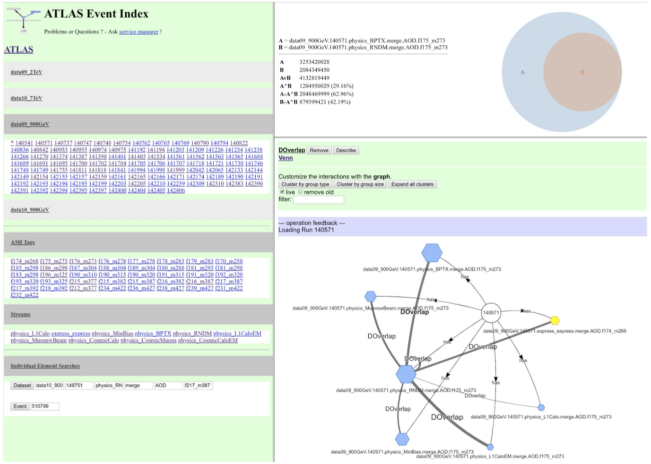
select

select

GUI Status


stream:physics_CosmicCalo
amitag:f933_m1960
data18_900GeV.00349769.physics_CosmicCalo.merge.AOD.f933_m1960
run:349769

- Similar as existing “data-centric browser”
- Much simpler implementation
- Very generic, mostly just reflects structure of data
  - Presentation using stylesheets
- Easily extensible with plugins or external apps
- Can work with any Phoenix schema
  - Or even completely without Phoenix, with all data in pure HBase
- All functionality available also via API and REST
- Interactive modifications (annotations) not yet available



ATLAS Event Index

Problems or Questions ? - Ask service manager !

ATLAS

data09_2TeV

data10_7TeV

data09_900GeV

* 140541 140571 140737 140747 140748 140754 140762 140765 140769 140790 140794 140822 140836 140842 140953 140955 140974 140975 141192 141194 141203 141209 141226 141234 141238 141266 141270 141374 141387 141398 141401 141403 141534 141561 141562 141563 141565 141688 141689 141691 141695 141700 141702 141704 141705 141706 141707 141718 141721 141730 141746 141748 141749 141755 141811 141818 141841 141994 141998 141999 142042 142065 142133 142144 142149 142154 142155 142157 142159 142161 142165 142166 142171 142174 142189 142190 142191 142192 142193 142194 142195 142199 142203 142205 142210 142259 142309 142310 142383 142390 142391 142392 142394 142395 142397 142400 142404 142405 142406

data10_900GeV

AMI Tags

f174_m268 f175_m273 f176_m273 f176_m278 f177_m278 f178_m283 f179_m283 f170_m258 f185_m298 f186_m298 f187_m304 f188_m304 f189_m304 f180_m288 f181_m293 f181_m298 f183_m298 f196_m325 f190_m310 f190_m315 f190_m320 f191_m315 f191_m320 f192_m320 f193_m320 f193_m325 f215_m377 f215_m382 f215_m387 f216_m382 f216_m387 f217_m387 f217_m392 f218_m392 f212_m377 f234_m422 f236_m427 f238_m427 f239_m427 f231_m422 f232_m422

Streams

physics_L1Calo express_express physics_MinBias physics_BPTX physics_RNDM physics_L1CaloEM physics_MuonswBeam physics_CosmicCalo physics_CosmicMuons physics_CosmicCaloEM

Individual Element Searches

Dataset data10_900 149751 physics_RN merge AOD f217_m387

Event 510799

A = data09_900GeV.140571.physics_BPTX.merge.AOD.f175_m273
B = data09_900GeV.140571.physics_RNDM.merge.AOD.f175_m273

A	3253420028
B	2084349450
AvB	4132819449
A^B	1204950029 (29.16%)
A-A^B	2048469999 (62.96%)
B-A^B	879399421 (42.19%)

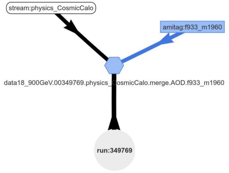
DOverlap Remove Describe
Venn

Customize the interactions with the **graph**.
Cluster by group type Cluster by group size Expand all clusters
☑ live ☐ remove old
filter:

--- operation feedback ---
Loading Run 140571

A
B
data09_900GeV.140571.physics_BPTX.merge.AOD.f175_m273
has
data09_900GeV.140571.physics_MuonswBeam.merge.AOD.f175_m273
DOverlap
140571
has
data09_900GeV.140571.express_express.merge.AOD.f174_m268
DOverlap
has
DOverlap
DOverlap
has
has
has
data09_900GeV.140571.physics_RNDM.merge.AOD.f175_m273
DOverlap
DOverlap
DOverlap
data09_900GeV.140571.physics_L1Calo.merge.AOD.f175_m273
data09_900GeV.140571.physics_MinBias.merge.AOD.f175_m273
data09_900GeV.140571.physics_L1CaloEM.merge.AOD.f175_m273

Notes


stream:physics_CosmicCalo
amitag:f933_m1960
data18_900GeV.00349769.physics_CosmicCalo.merge.AOD.f933_m1960
run:349769

- Semantics of SQL + NoSQL is not trivial
  - Search/Filter/Fill, search sequence, Scan/Get,...
- It would be more straightforward with just HBase backend
- On the other hand, the system works even without HBase part, with just Phoenix (but of course with less functionality)
  - Or just with any SQL database