



Performance testing framework for the ATLAS EventIndex

Elizaveta Cherepanova^{1*}, Fedor Prokoshin¹

1: JINR Dubna

GRID'2021

July 5, 2021

[*elizaveta.cherepanova@cern.ch](mailto:elizaveta.cherepanova@cern.ch)

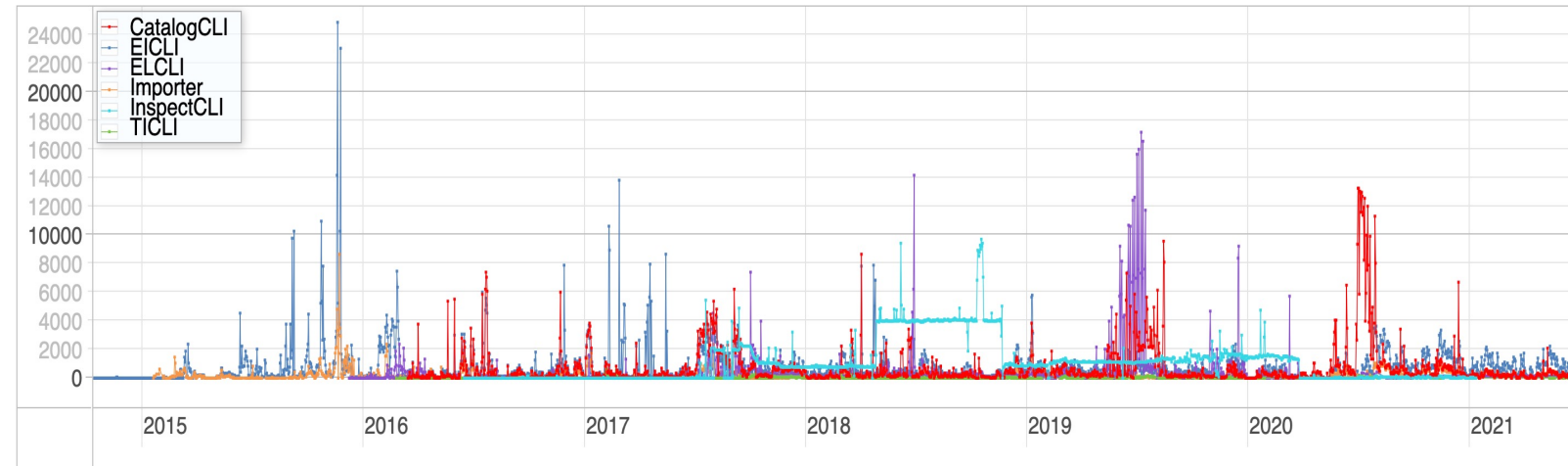
- EventIndex performance
- Event Lookup tests
- EventIndex tests
 - without selection
 - with trigger filters
- Local EI launch vs remote EI launch
- Summary + Future plans

The Hadoop system runs a variety of tasks:

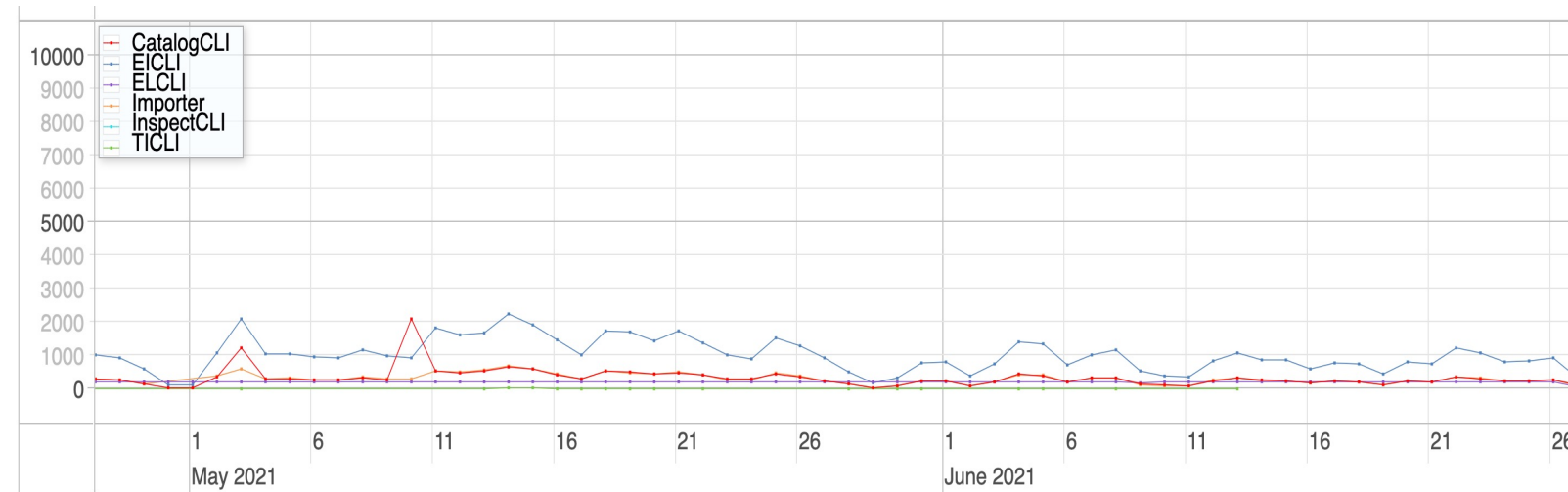
- importing and cataloguing data
- running consistency checks
- establishing links between related datasets
- responding to users' queries

Performance tests are required to track the current system state

Daily access statistics of the Hadoop system between May 2015 and June 2021



Daily access statistics of the Hadoop system between April 2021 and June 2021



Data for tests

- Data for yy → WW analysis: <https://its.cern.ch/jira/browse/DATREP-169> (2015-2018). ~50000 keys
 - Key (runNumber-eventNumber) lists for several datasets in each year (10 000 keys)
- Key lists for 1 chosen dataset in a year. ~1 million events
- Dataset id lists with different number of events (~10K, 100K, 1M, 10M events)

Types of queries

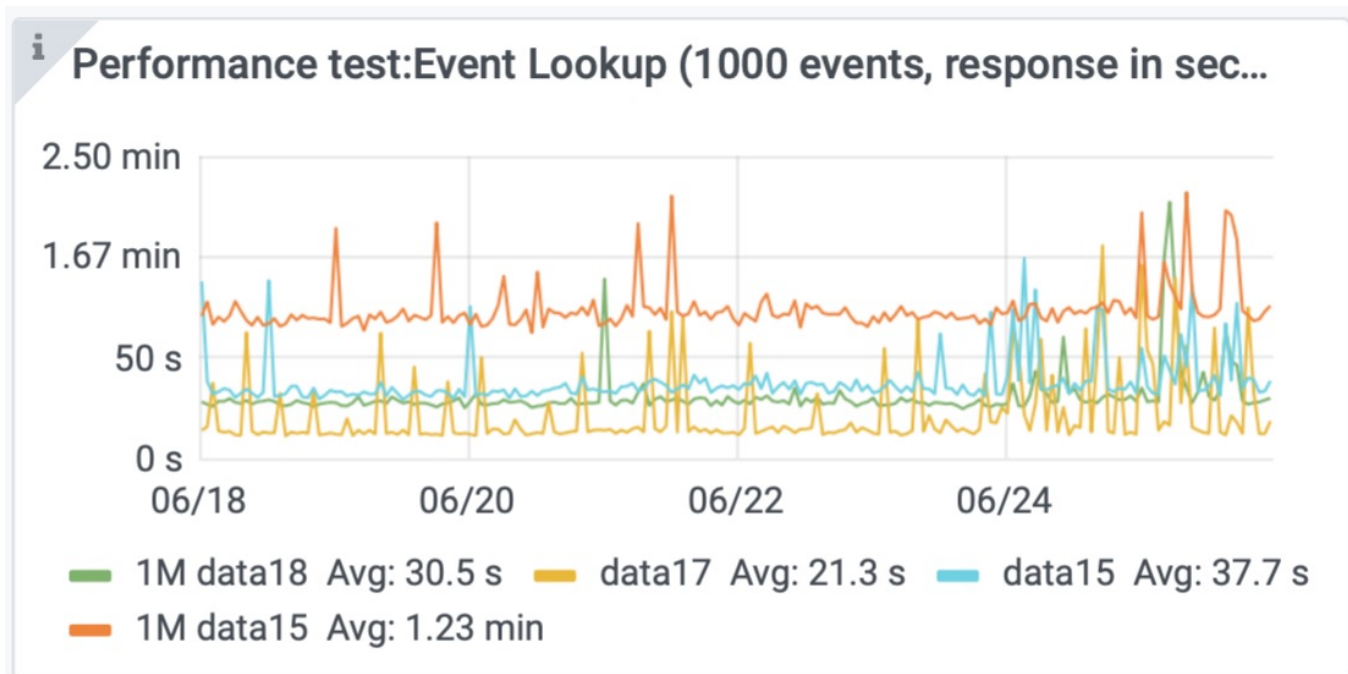
- Event Picking (Event Lookup, EventIndex)
 - Fast search of events using key pairs
 - Return events GUIDs, full dataset names, data types (RAW, AOD, DAOD)
- Search for events in a specified dataset (EventIndex)
 - Return full information about an event (GUID, data type, production step, LumiBlock, time, trigger chains, etc.)
 - Filters can be applied

Looking events using key pairs
(running hourly in acron)

2021/06/18 – 2021/06/26

dataset	nkeys	Avg time
data15 (1M)	1000	1.23 min
data15 (mix)	1000	37.7 s
data17 (mix)	1000	21.3 s
data18 (1M)	10	4.18 s
data18 (1M)	100	7.49 s
data18 (1M)	1 000	30.5 s
data18 (1M)	10 000	3.65 min
data18 (1M)	50 000	16.07 min

*Green ones are displayed



Response times of the EventIndex Hadoop server to event lookup queries [\[Grafana link\]](#)

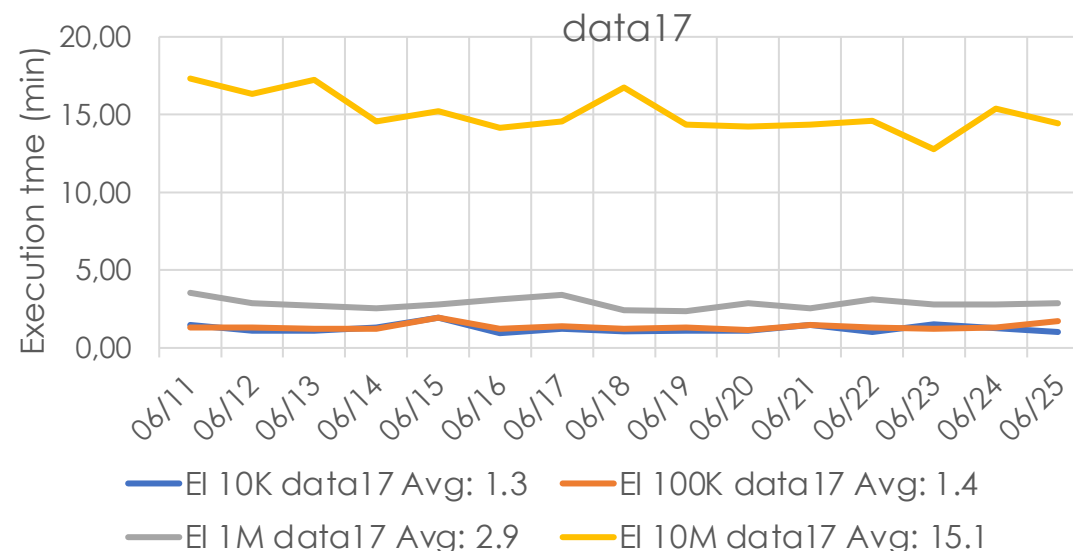
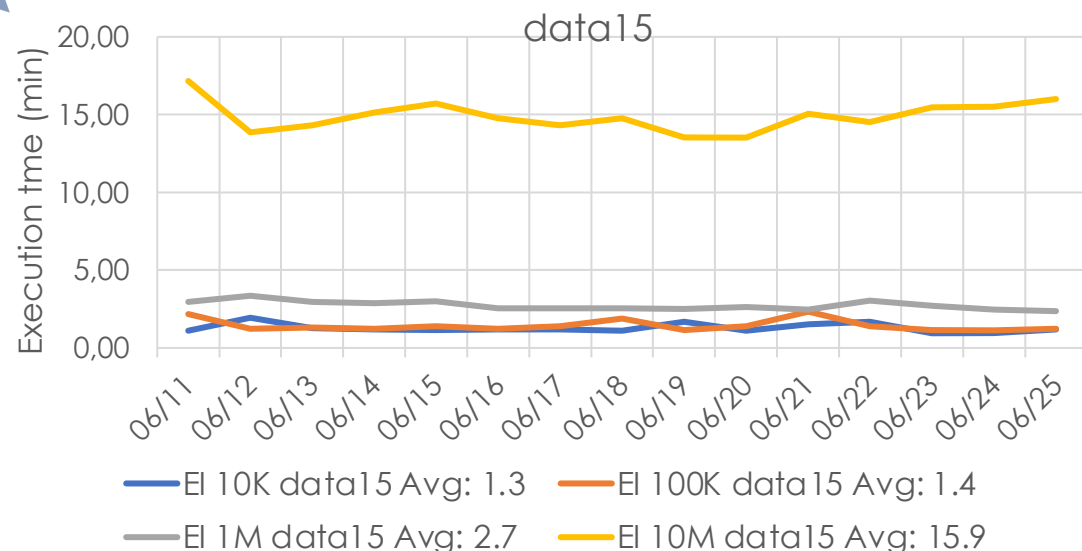
Query example:

```
el -f /afs/cern.ch/user/a/atlevind/work/DB/data/performance_tests/key_lists/EventIndexList_data15_RegionC_1000.txt
```

EventIndex queries (no selection)

- Key search
 - 10 keys for 2015-2018 data
 - Query example: `ei -query 'id:EI17.1 runNumber:00339435 status:~bad' -key '00339435-002273429414' -filter 'RunNumber_EventNumber db0 db1 db2'`
- Id search
 - Dataset sizes ~10K (50K), 100K, 1M, 10M events for 2015-2018 data
 - Query example: `ei -query 'id:EI15.1.data15_13TeV.physics_Main.merge.DESDM_MCP.f628_m1501.00279284' -mr 'true' -filter 'RunNumber_EventNumber db0 db1 db2'`
- Running daily in acron (no Grafana dashboard)

2021/06/11 – 2021/06/25



EventIndex queries (trigger selection)

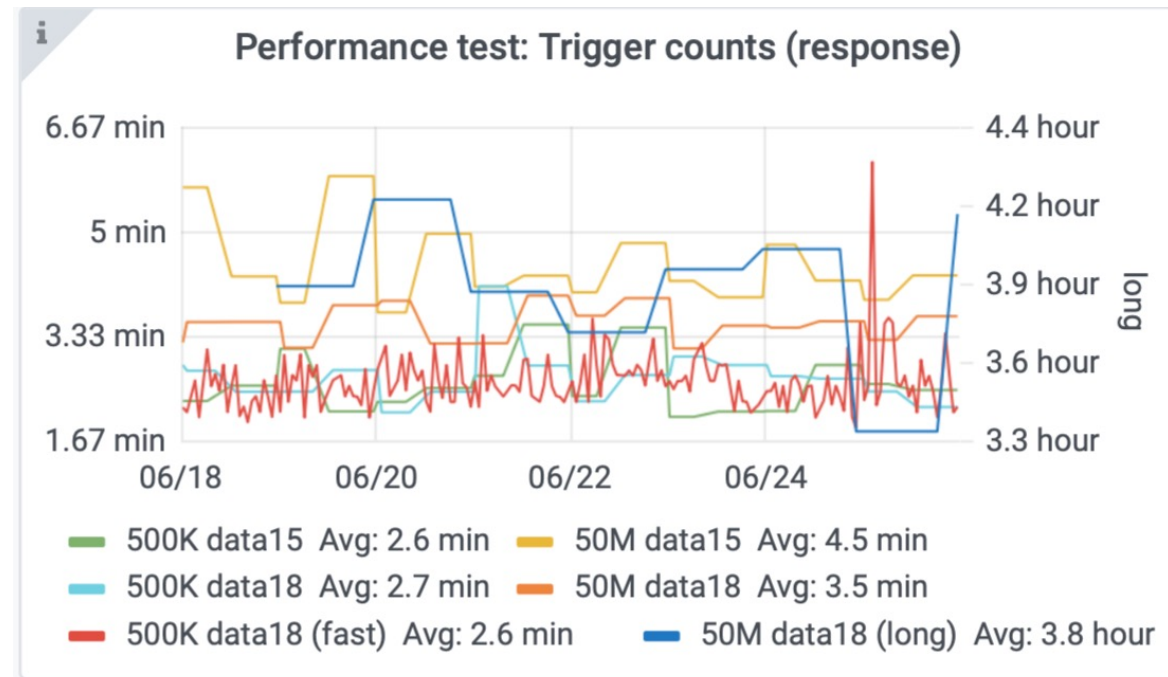
Looking events that passed specified trigger in selected dataset

2021/06/18 – 2021/06/25

dataset	(% of events passed trigger)	Running
data15 (500K, 1M, 10M, 50M)	<1	twice a day
data16 (500K, 1M, 10M, 50M)	<1	twice a day
data17 (500K, 1M, 10M, 50M)	<1	twice a day
data18 (500K, 1M, 10M, 50M)	<1	twice a day
data18 (500K) fast	<1	hourly
data18 (50M) long	~40	once a day

*Green ones are displayed

Query example: `ei -query id:EI18.1.data18_13TeV.physics_Main.merge.AOD.f1002_m2037.00364214 -mr 'chainFired(\"L1_EM22VHI\")' -filter RunNumber_EventNumber`



Response times queries retrieving information on events that passed trigger selection [\[Grafana link\]](#)

- Test results have been written to JSON files
- Example output file content:

```
...
{
  "command": "ei",
  "dataType": "AOD",
  "nevents": 566259,
  "query": "hadoop jar /afs/cern.ch/user/a/atlevind/work/DB/Atlas-Event-Index-
Core4import/lib/EIHadoop.jar net.hep.atlas.Database.EIHadoop.Apps.EICLI -query
id:EI18.1.data18_13TeV.express_express.merge.AOD.f937_m1972.00351455 -mr
'chainFired(\"L1_LAR-J\")' -filter RunNumber_EventNumber",
  "result": 103,
  "sizeTitle": "500K",
  "status": "ok",
  "time": 172.1403670310974,
  "timestamp": "2021-04-14_23:46:37 (CEST)"
},
...
```

- Data from output files go to Grafana monitors

Local EI launch vs remote EI launch

↓
On the Hadoop
cluster

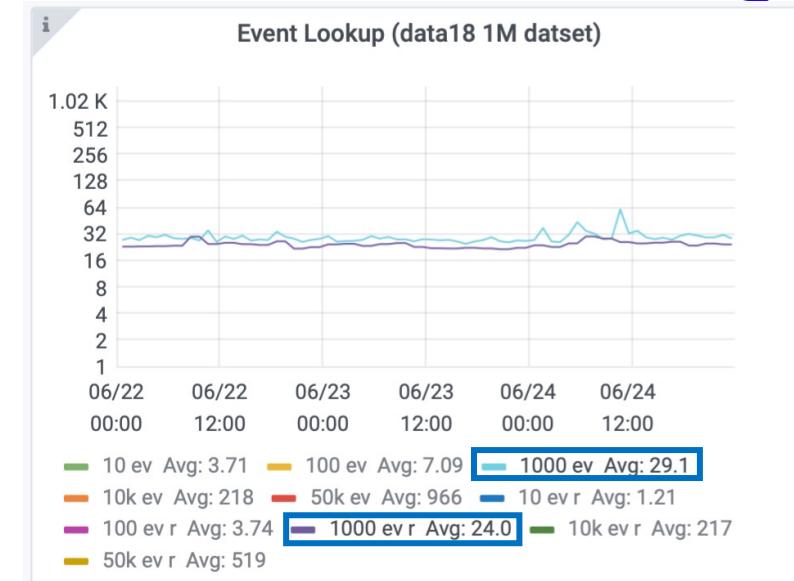
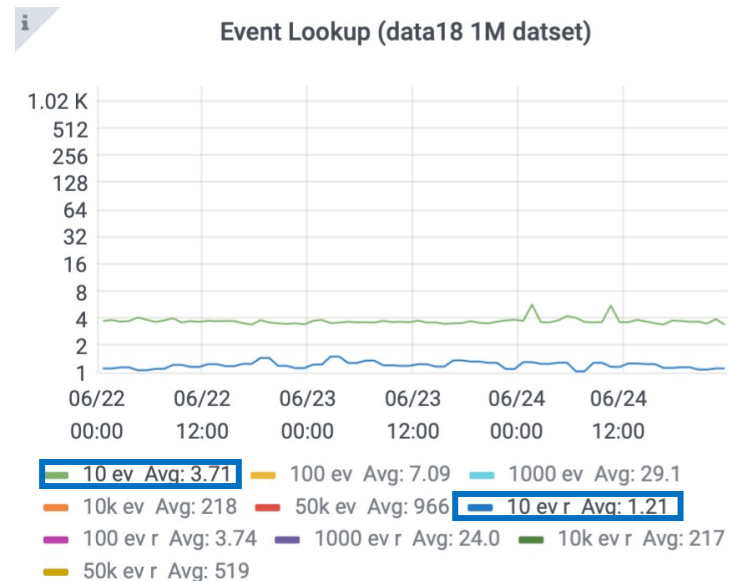
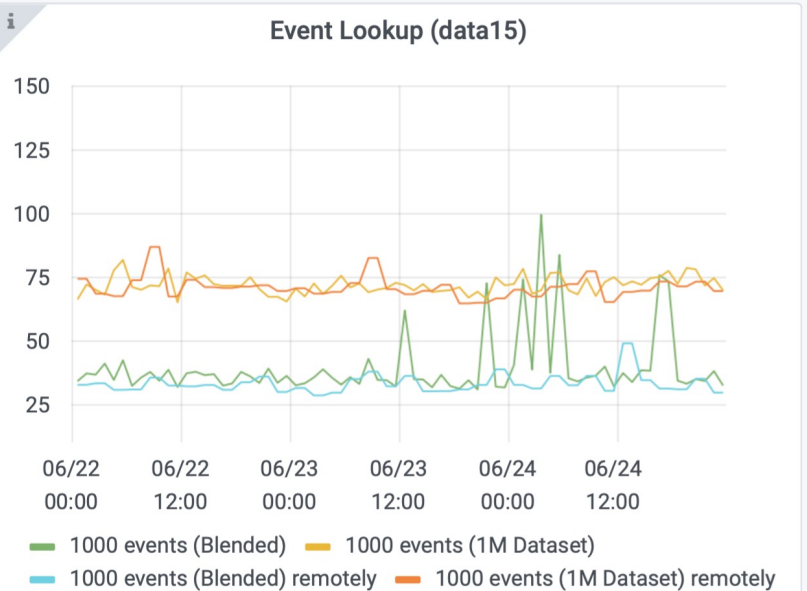
↓
On CERN Ixplus via
Tomcat server

← That is how users usually do

Remote:

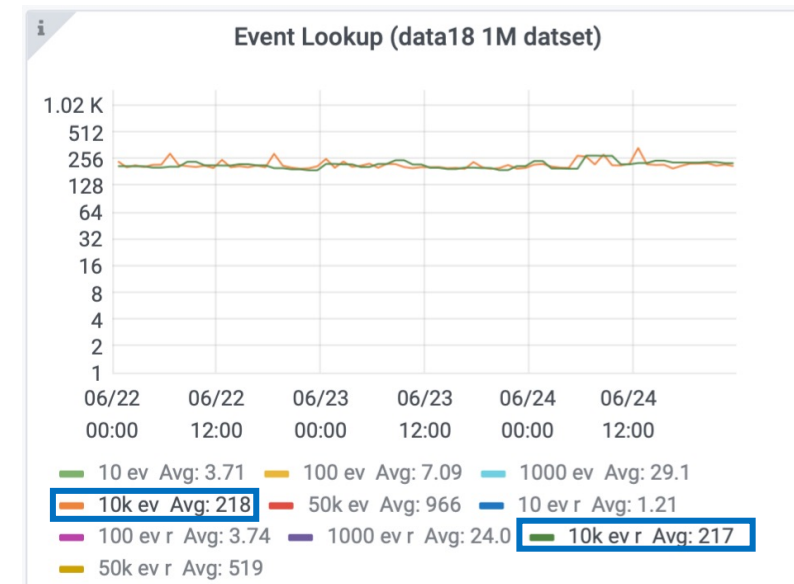
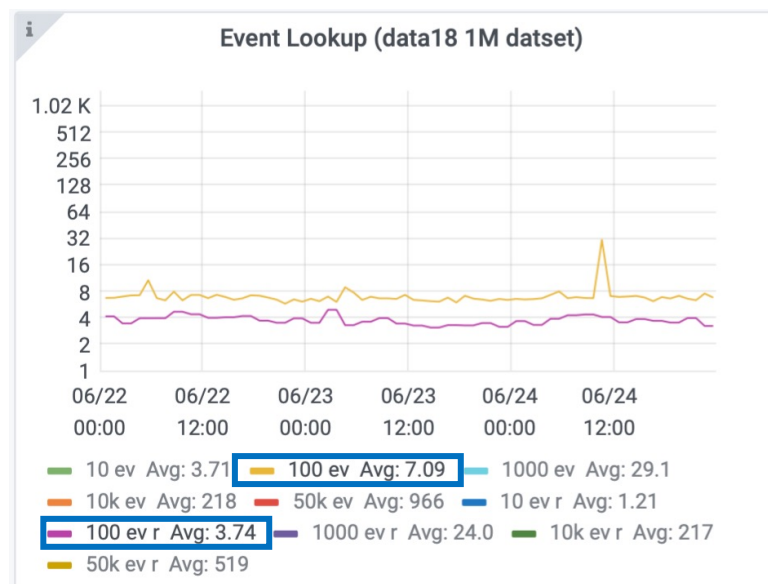
- Event Lookup queries
 - Twice an hour
 - Same original key list (data18) for 10, 100, 1000 and 10000 key searches, but other for 50000 key searches
- EventIndex search with trigger selections
 - Four datasets for every year → once a day
 - 'Fast' trigger search → once an hour

Local vs remote (Event Lookup)



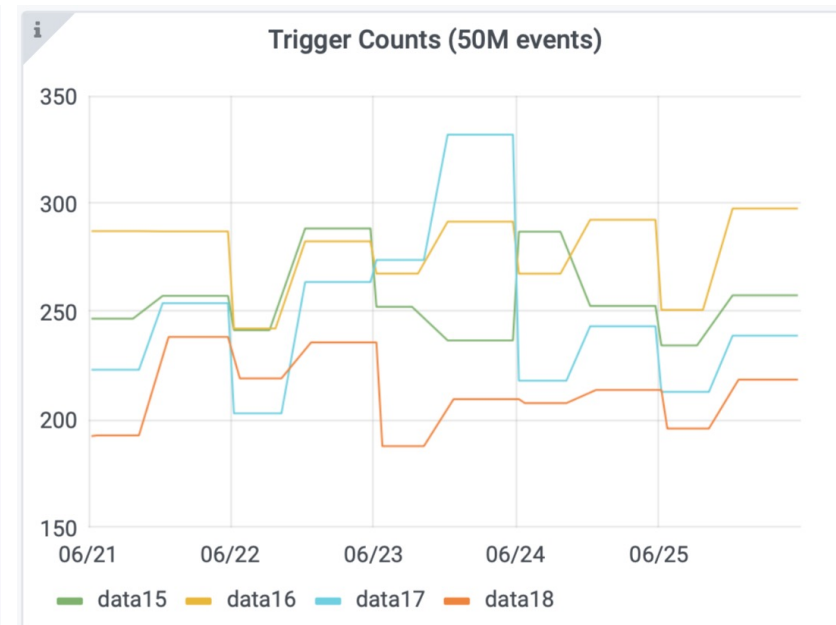
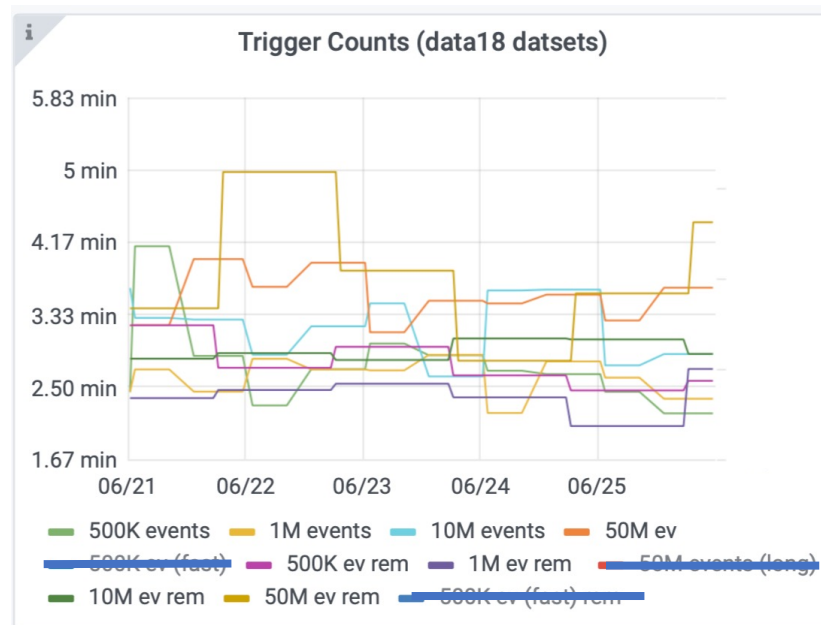
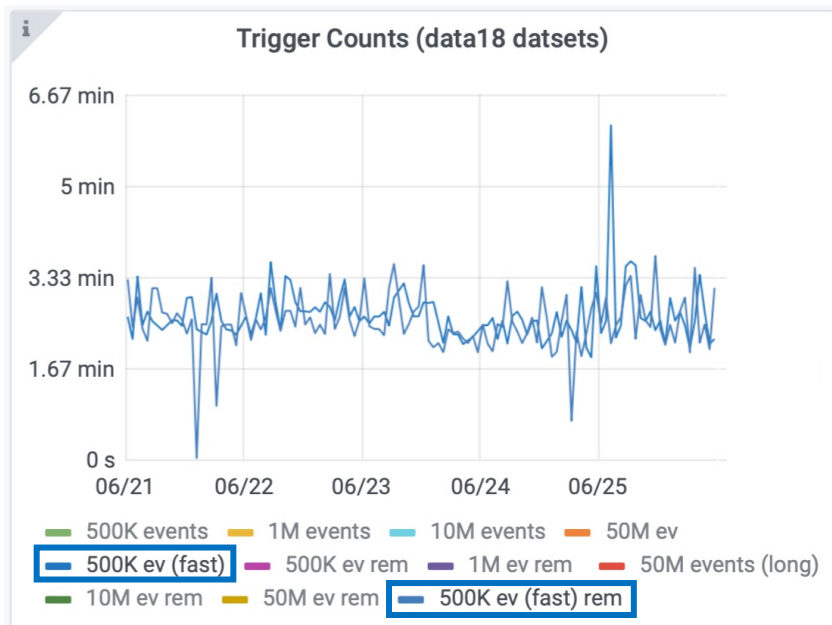
2021/06/22 – 2021/06/24

- Remotely queries are finishing faster in most cases for the Event Lookup search



Local vs remote (Trigger counts)

2021/06/21 – 2021/06/25



- No noticeable difference between two ways
- More plots on the Grafana tab: <https://monit-grafana.cern.ch/d/EceHNkcGk/atlas-ei-performance-tests?from=now-7d&orgId=6&to=now>

- Performance tests for the current system:
 - Three types of jobs are running in acron:
 - EL key search (once per hour) [1]
 - EI key search & EI full id search (once per day) [2]
 - EI search with trigger selection [3]
 - Only [1] and [3] are displayed on Grafana dashboards
 - Outputs are stored in the EI afs directory
 - Most part of locally running queries are also running remotely
- Plan to continue supporting and developing the tools
- Similar tests for the new system will be developed

Thank you for your attention!