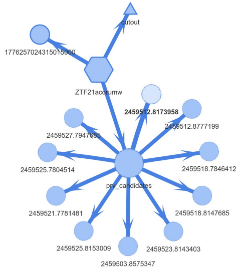
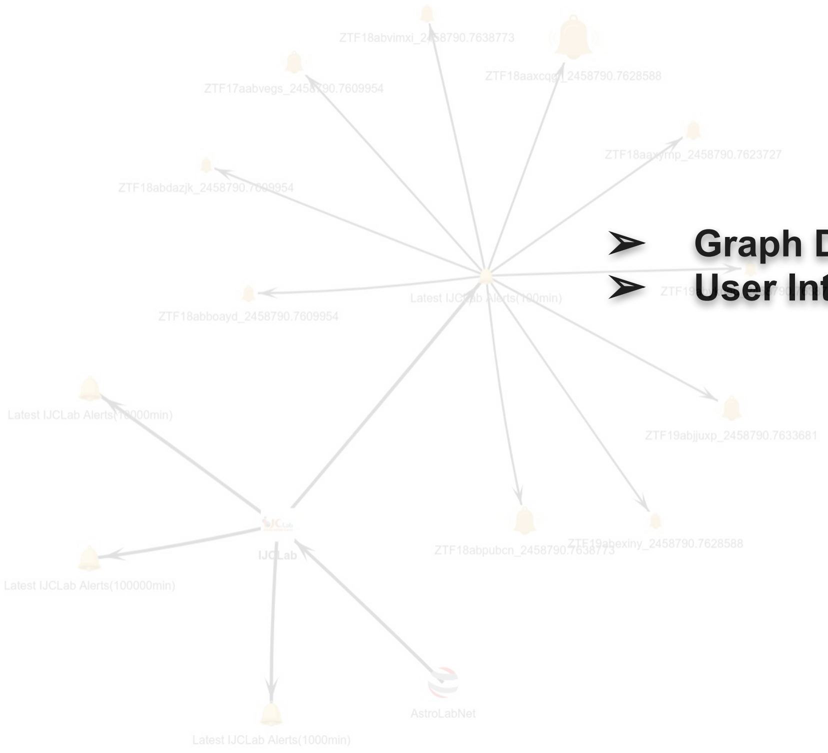
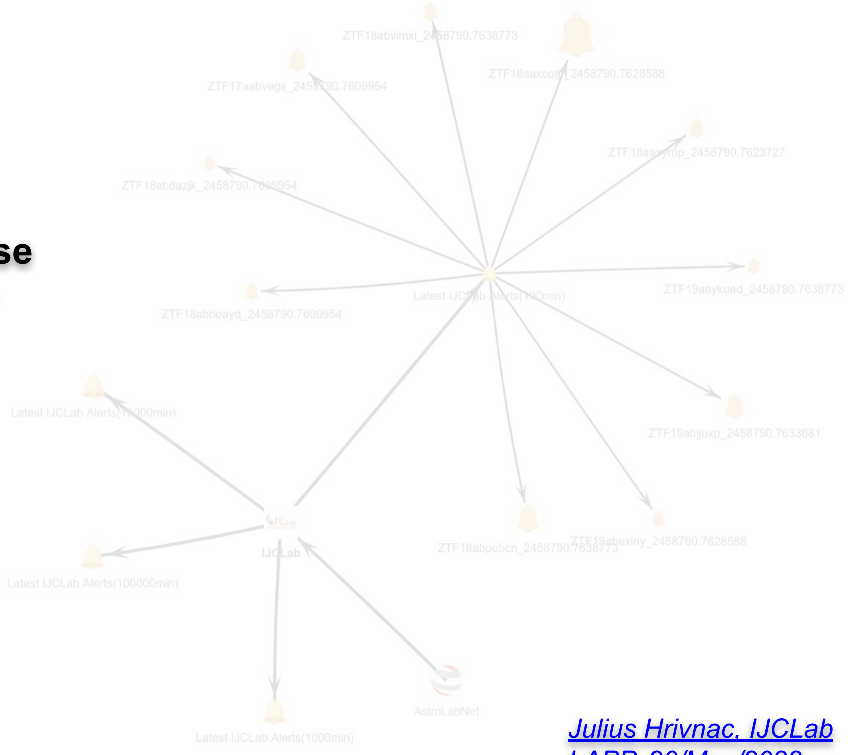
Fink in Graph Database



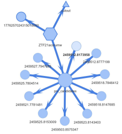

ZTF18abvimxi_2458790.7638773
ZTF18aaxcqqd_2458790.7628588
ZTF17aabvegs_2458790.7609954
ZTF18aaxymp_2458790.7623727
ZTF18abdazjk_2458790.7609954
Latest IJCLab Alerts(100min)
ZTF18abboayd_2458790.7609954
ZTF19abjjuxp_2458790.7633681
Latest IJCLab Alerts(10000min)
IJCLab
Latest IJCLab Alerts(100000min)
ZTF18abpubcn_2458790.7638773
ZTF19abexiny_2458790.7628588
AstroLabNet
Latest IJCLab Alerts(1000min)

- ➢ **Graph Database**
- ➢ **User Interface**


ZTF18abvimxi_2458790.7638773
ZTF18aaxcqqd_2458790.7628588
ZTF17aabvegs_2458790.7609954
ZTF18aaxymp_2458790.7623727
ZTF18abdazjk_2458790.7609954
ZTF19abykued_2458790.7638773
Latest IJCLab Alerts(100min)
ZTF18abboayd_2458790.7609954
ZTF19abjjuxp_2458790.7633681
Latest IJCLab Alerts(10000min)
IJCLab
Latest IJCLab Alerts(100000min)
ZTF18abpubcn_2458790.7638773
ZTF19abexiny_2458790.7628588
AstroLabNet
Latest IJCLab Alerts(1000min)

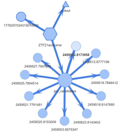
Julius Hrivnac, IJCLab
LAPP, 20/May/2022

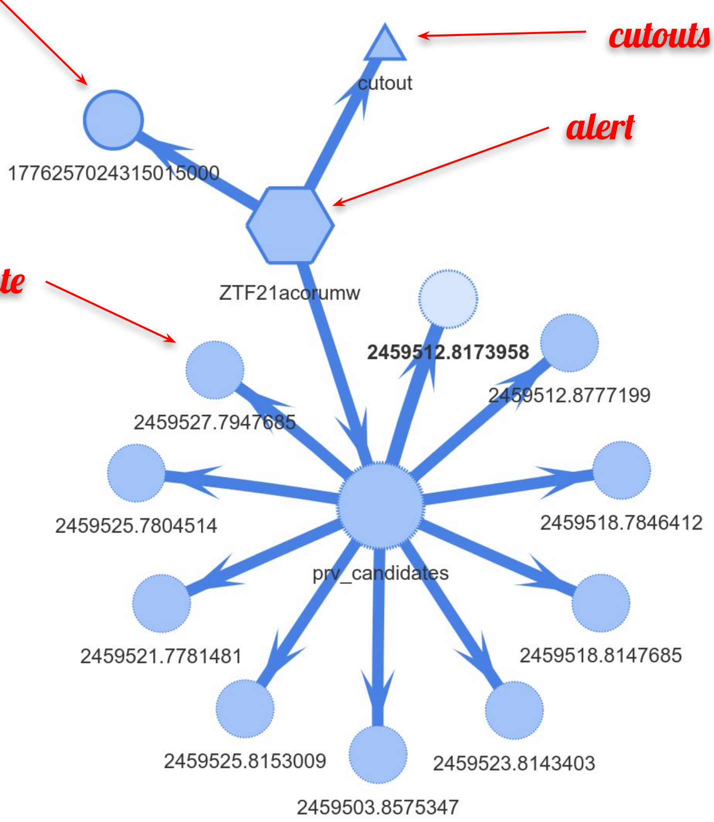
Fink Data Storage



- The current implementation = one big HBase table + a set of small HBase index tables for fast search
- The new implementation = this table is converted into Graph
  - Rows become Vertices
  - Relations become Edges between Vertices
    - Which are now explicite, directly stored in the database
- Structure and relations are moved from the code to the storage
- Both Vertices and Edges have properties
  - Some are defined in a Schema, others can be freely added
  - Also new Vertices and Edges can be added and modified
- Indexes may be attached to any property for faster search
- Graph DB is slower on injection, similar on search, very fast on navigation, very slow on deletion

One Alert





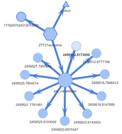
candidate

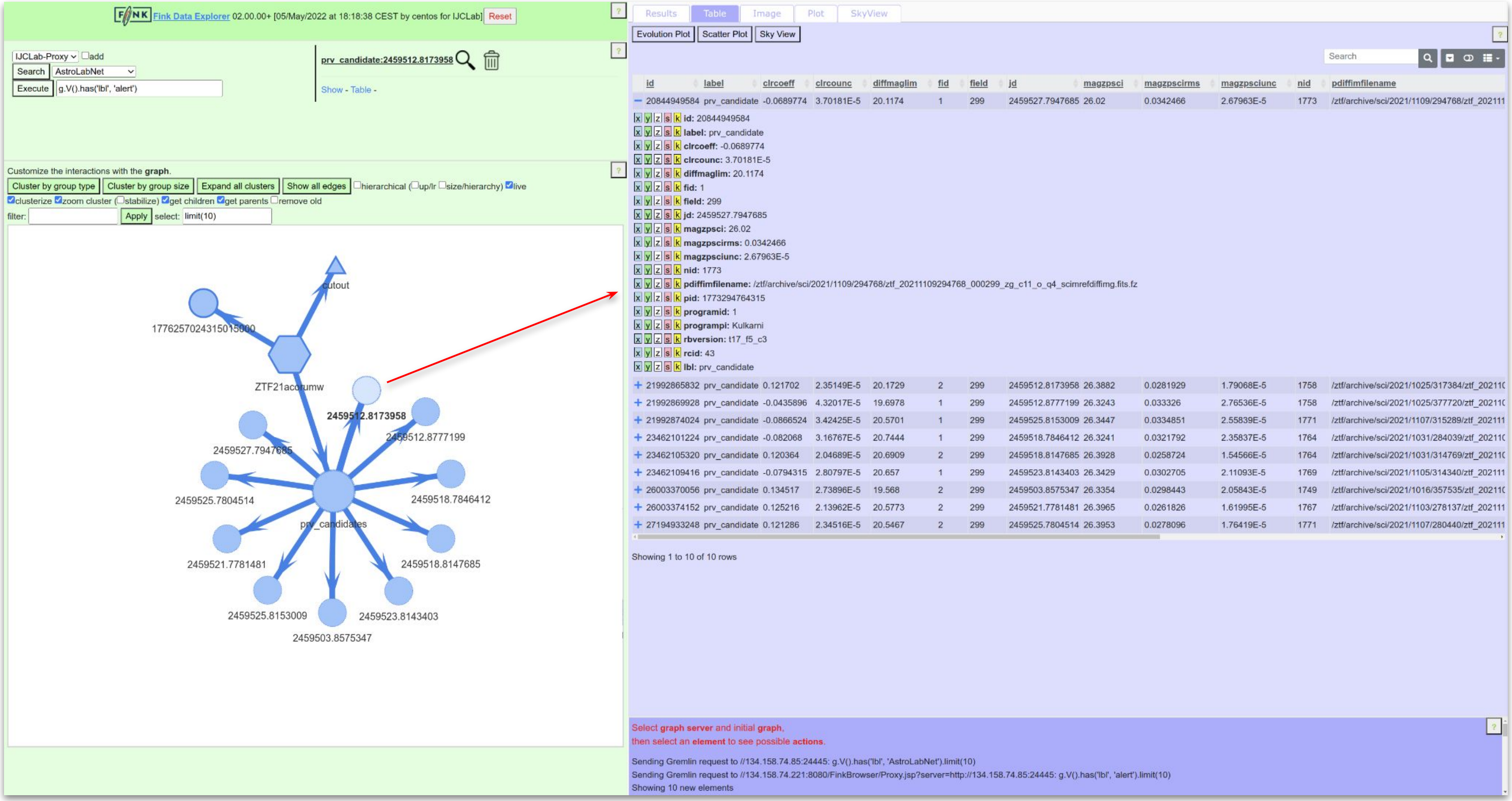
cutouts

alert

prv_candidate

Inside Alert





	id	label	clrcoeff	clrcounc	diffmaglim	fid	field	jd	magzpsci	magzpscirms	magzpsciunc	nid	pdiffimfilename
−	20844949584	prv_candidate	-0.0689774	3.70181E-5	20.1174	1	299	2459527.7947685	26.02	0.0342466	2.67963E-5	1773	/ztf/archive/sci/2021/1109/294768/ztf_202111

x y z s k **id:** 20844949584
x y z s k **label:** prv_candidate
x y z s k **clrcoeff:** -0.0689774
x y z s k **clrcounc:** 3.70181E-5
x y z s k **diffmaglim:** 20.1174
x y z s k **fid:** 1
x y z s k **field:** 299
x y z s k **jd:** 2459527.7947685
x y z s k **magzpsci:** 26.02
x y z s k **magzpscirms:** 0.0342466
x y z s k **magzpsciunc:** 2.67963E-5
x y z s k **nid:** 1773
x y z s k **pdiffimfilename:** /ztf/archive/sci/2021/1109/294768/ztf_20211109294768_000299_zg_c11_o_q4_scimrefdiffimg.fits.fz
x y z s k **pid:** 1773294764315
x y z s k **programid:** 1
x y z s k **programpi:** Kulkarni
x y z s k **rbversion:** t17_f5_c3
x y z s k **rcid:** 43
x y z s k **lbl:** prv_candidate

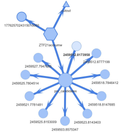
	id	label	clrcoeff	clrcounc	diffmaglim	fid	field	jd	magzpsci	magzpscirms	magzpsciunc	nid	pdiffimfilename
+	21992865832	prv_candidate	0.121702	2.35149E-5	20.1729	2	299	2459512.8173958	26.3882	0.0281929	1.79068E-5	1758	/ztf/archive/sci/2021/1025/317384/ztf_20211(
+	21992869928	prv_candidate	-0.0435896	4.32017E-5	19.6978	1	299	2459512.8777199	26.3243	0.033326	2.76536E-5	1758	/ztf/archive/sci/2021/1025/377720/ztf_20211(
+	21992874024	prv_candidate	-0.0866524	3.42425E-5	20.5701	1	299	2459525.8153009	26.3447	0.0334851	2.55839E-5	1771	/ztf/archive/sci/2021/1107/315289/ztf_202111
+	23462101224	prv_candidate	-0.082068	3.16767E-5	20.7444	1	299	2459518.7846412	26.3241	0.0321792	2.35837E-5	1764	/ztf/archive/sci/2021/1031/284039/ztf_20211(
+	23462105320	prv_candidate	0.120364	2.04689E-5	20.6909	2	299	2459518.8147685	26.3928	0.0258724	1.54566E-5	1764	/ztf/archive/sci/2021/1031/314769/ztf_20211(
+	23462109416	prv_candidate	-0.0794315	2.80797E-5	20.657	1	299	2459523.8143403	26.3429	0.0302705	2.11093E-5	1769	/ztf/archive/sci/2021/1105/314340/ztf_202111
+	26003370056	prv_candidate	0.134517	2.73896E-5	19.568	2	299	2459503.8575347	26.3354	0.0298443	2.05843E-5	1749	/ztf/archive/sci/2021/1016/357535/ztf_20211(
+	26003374152	prv_candidate	0.125216	2.13962E-5	20.5773	2	299	2459521.7781481	26.3965	0.0261826	1.61995E-5	1767	/ztf/archive/sci/2021/1103/278137/ztf_202111
+	27194933248	prv_candidate	0.121286	2.34516E-5	20.5467	2	299	2459525.7804514	26.3953	0.0278096	1.76419E-5	1771	/ztf/archive/sci/2021/1107/280440/ztf_202111

Showing 1 to 10 of 10 rows

Select **graph server** and initial **graph**,
then select an **element** to see possible **actions**.

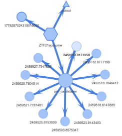
```
Sending Gremlin request to //134.158.74.85:24445: g.V().has('lbl', 'AstroLabNet').limit(10)
Sending Gremlin request to //134.158.74.221:8080/FinkBrowser/Proxy.jsp?server=http://134.158.74.85:24445: g.V().has('lbl', 'alert').limit(10)
Showing 10 new elements
```

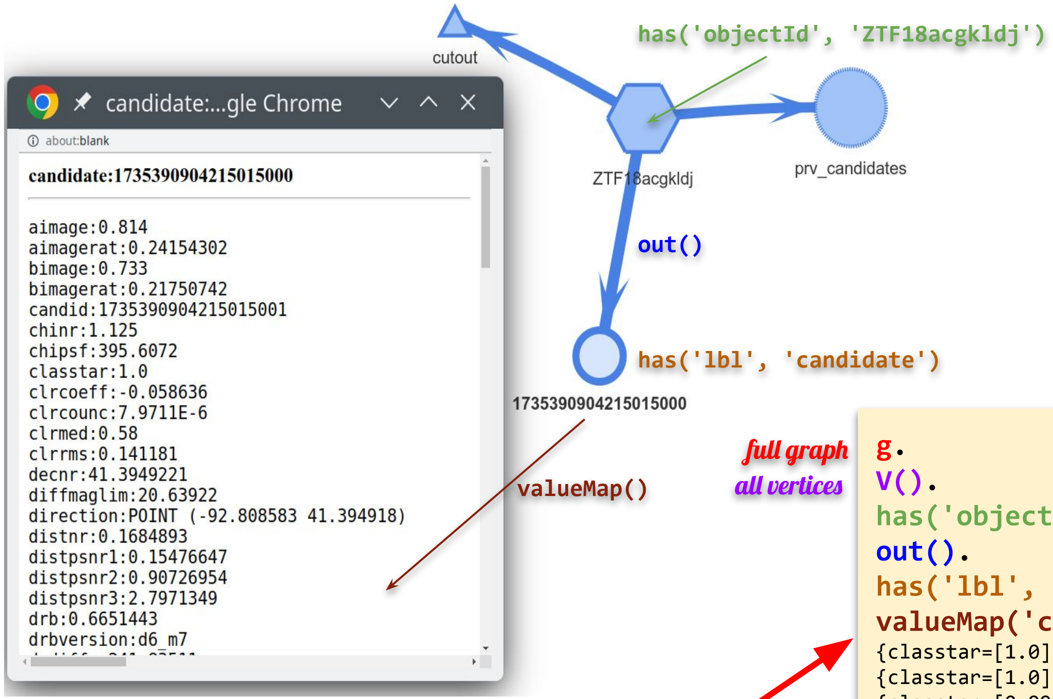
Gremlin Query Language



- Data can be accessed via **Gremlin query language**
  - Based on Groovy language
  - Functional-style syntax, where function = relation and function execution = relation navigation
  - Available for almost all major programming languages
    - But best suited for languages, which are already naturally functional
  - Simple searches are very intuitive, but sophisticated operation are possible
  - Allows complex graph navigation, mathematical and statistical operations and full functional processing of graphs

Gremlin Search





candidate:1735390904215015000

```
aimage:0.814
aimagerat:0.24154302
bimage:0.733
bimagerat:0.21750742
candid:1735390904215015001
chinr:1.125
chipsf:395.6072
classtar:1.0
clrcoeff:-0.058636
clrcounc:7.9711E-6
clrmed:0.58
clrrms:0.141181
decnr:41.3949221
diffmaglim:20.63922
direction:POINT (-92.808583 41.394918)
distnr:0.1684893
distpsnr1:0.15476647
distpsnr2:0.90726954
distpsnr3:2.7971349
drb:0.6651443
drbversion:d6_m7
```

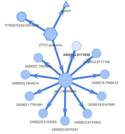
full graph

all vertices

```
g.
V().
has('objectId', 'ZTF18acgkldj').
out().
has('lbl', 'candidate').
valueMap('classtar', 'jd', 'direction')
{classtar=[1.0],   direction=[POINT (-92.808682 41.394857)], jd=[2459484.926331] }
{classtar=[1.0],   direction=[POINT (-92.808593 41.394903)], jd=[2459496.9461458]}
{classtar=[0.991], direction=[POINT (-92.808662 41.394983)], jd=[2459530.9093403]}
…
```

```
g.V().has('objectId', 'ZTF18acgkldj').out().has('lbl', 'candidate').valueMap('classtar', 'jd', 'direction')
```

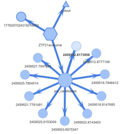
Gremlin for Fink



- ➢ Frequently used and typical queries will be implemented as server-side function to be available to all clients
- ➢ Typical user request:
  - ○ Server-side selection function
  - ○ + Further refining selection
  - ○ + Set of values to return
  - ○ + Further math or graphics
- ➢ Any Gremlin code is possible
  - ○ With some kind of user authentication and authorisation

```
g.V().has('objectId', 'ZTF18acgkldj').out().has('lbl', 'candidate').valueMap('classtar', 'jd', 'direction')
      candidates('ZTF18acgkldj').                                                valueMap('classtar', 'jd', 'direction')
```

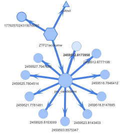
server-side selection function



Gremlin on Server

```
# gives 10 first vertices 0.1 degree around direction 57.5 x -1.97 between two jd times
# implemented as a server-side function
geosearch(57.5, -1.97, 0.1, 2359300.7629977, 2559317.7015982, 10).has('lbl', 'candidate').valueMap(...)
# internally contains (with protection against overuse and optimisation code):
g.V().has('direction', geoWithin(Geoshape.circle(dec, 180 - ra, dist))).has('jd', inside(jdmin, jdmax))
```

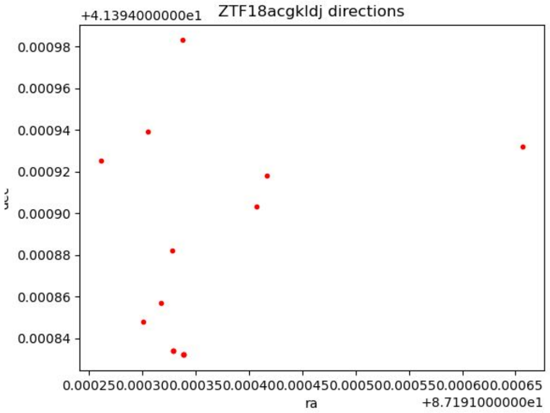
server-side selection function



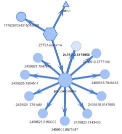
API

Simple Python Example

```
import sys

import jpype
import jpype.imports
from jpype import JImplements, JOverride, JImplementationFor

import matplotlib.pyplot as plt

# ../dist/FinkBrowser.exe.jar
jpype.startJVM(jpype.getDefaultJVMPath(), "-ea", "-Djava.class.path=" + sys.argv[1], convertStrings=False)

from com.Lomikel.Januser                    import StringGremlinClient
from com.astrolabsoftware.FinkBrowser.Utils import Init

Init.init()

client = StringGremlinClient("graph-server", 24444);

results = client.interpret("candidates('ZTF18acgkldj').elementMap('direction')");

ra  = []
dec = []

for element in results:
  dec += [element.getObject().get("direction").getPoint().getLatitude()]
  ra  += [element.getObject().get("direction").getPoint().getLongitude() + 180]

plt.plot(ra, dec, 'r.')
plt.title('ZTF18acgkldj directions')
plt.xlabel('ra')
plt.ylabel('dec')
plt.show()
client.close()

jpype.shutdownJVM()
```

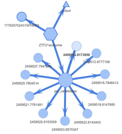


Direct|String API



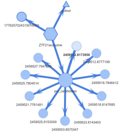
```
# instead of
client = StringGremlinClient("graph-server", 24444);
results = client.interpret("candidates('ZTF18acgkldj').elementMap('direction')");

# we can do
client = DirectGremlinClient("graph-server", 24444);
g = client.g();
query = g.V().has('lbl', 'alert').limit(4).values(objectId);
results = client.submit(query);
# advantage: results is an actual object,
#            while above it was just a string with JSON content
# problem: cannot use server-side functions and objects,
#          which are unknown to client
```

End User Access



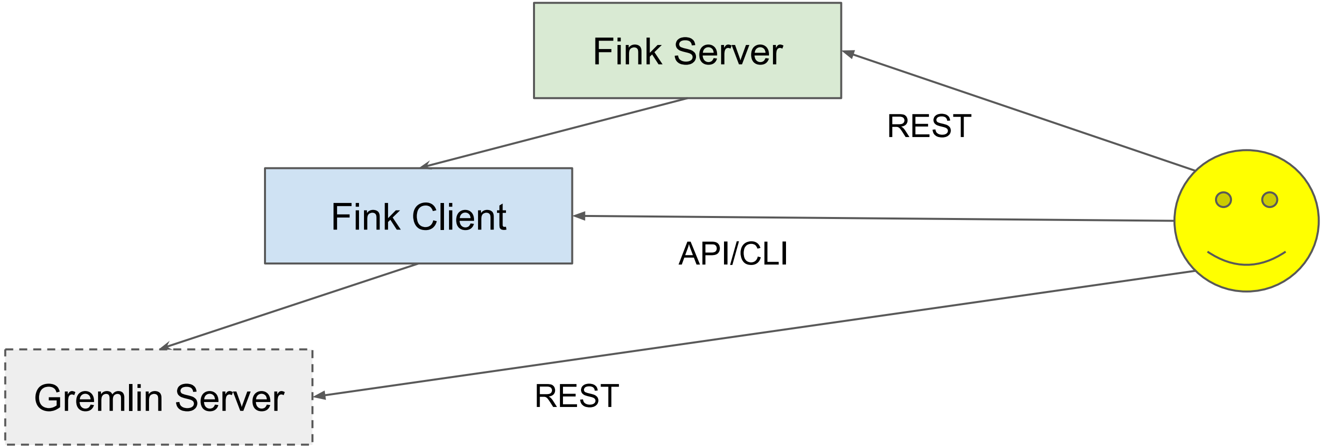
- ➢ All queries can be issued using standard Gremlin clients (in all popular languages)
- ➢ Requests can be send directly to Gremlin server
- ➢ **A special client** also available in several incarnation, providing some pre/post-processing, overuse protection and connection handling
  - ○ Java executable
  - ○ Linux native executable
  - ○ GUI (platform independent)
  - ○ REST Web Service
  - ○ Python, Java, Scala, Groovy,... API
  - ○ Jupyter API
- ➢ The same answer in CLI and from REST WS

```
# Direct connection to Graph server (gives very verbose JSON answer, not all queries supported)
curl 'http://graph-server:24444/gremlin'
  -XPOST -d '{"gremlin":"candidates(\"ZTF18acgkldj\").valueMap(\"classtar\", \"jd\", \"direction\")"}'
# Connection to Fink server
curl 'http://fink-server:8080/FinkBrowser/Fremlin.jsp'
  -get --data-urlencode 'gremlin=candidates("ZTF18acgkldj").valueMap("classtar", "jd", "direction")'
# Java client
java -jar FinkBrowser.exe.jar --gremlin 'candidates("ZTF18acgkldj").valueMap("classtar", "jd", "direction")'
# Native Linux client
FinkBrowser.exe --gremlin 'candidates("ZTF18acgkldj").valueMap("classtar", "jd", "direction")'
```

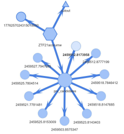


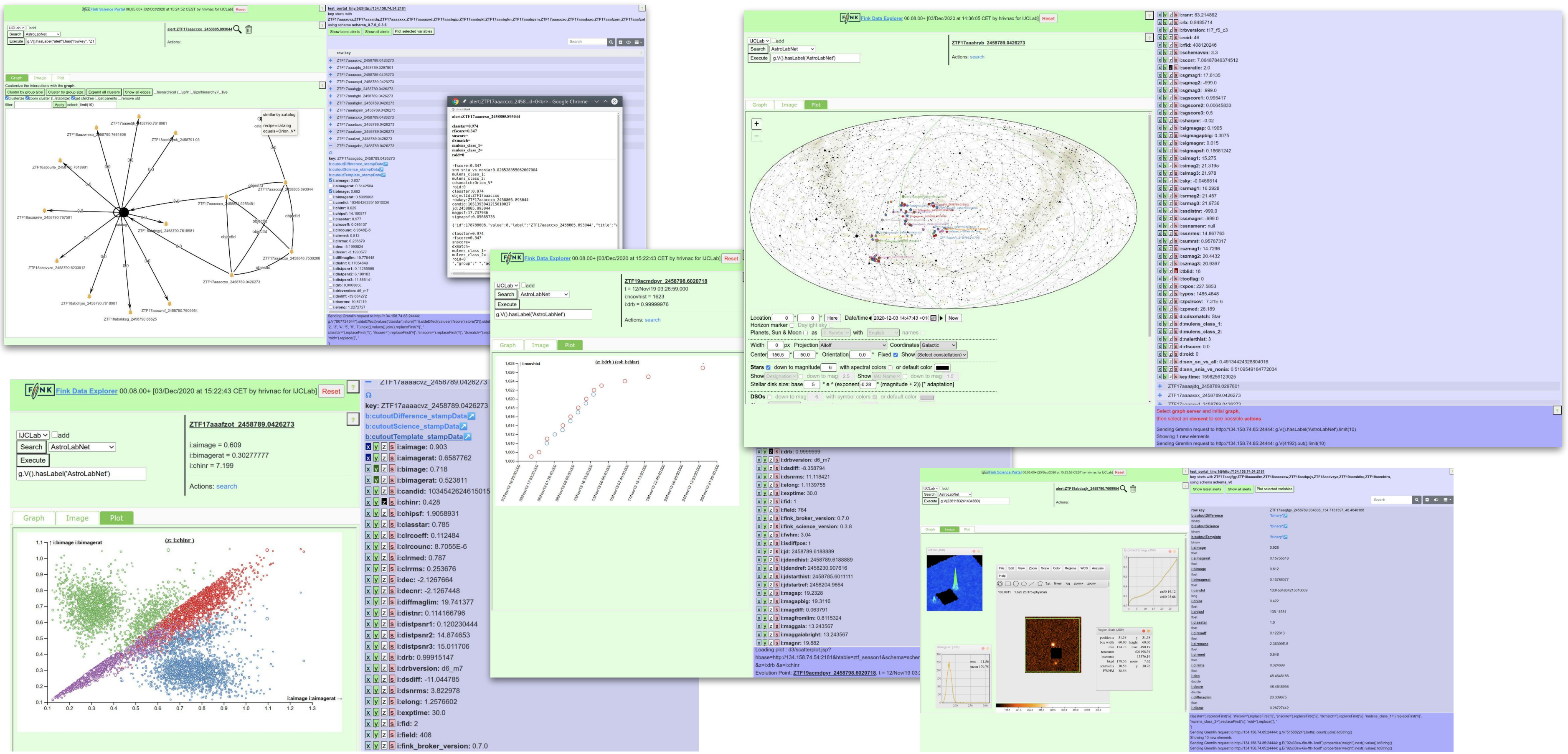
End User Access

- ➢ Differences wrt current system:
  - ○ The current system: many REST entry points with many options
  - ○ The graph system: one REST entry point with commands and options expressed as Gremlin fragments
- ➢ Requests for the whole world (g.V(), g.E()) forbidden for ordinary users
  - ○ All requests should start with a server-side selection function
  - ○ And be further customised by simple Gremlin code



Graph Web Service Prototypes





Fink Data Explorer 00.08.00+ [03/Dec/2020 at 14:36:05 CET by hrivnac for IJCLab] Reset

IJCLab ☐add
Search AstroLabNet
Execute g.V().hasLabel('AstroLabNet')

ZTF17aaahryb_2458789.0426273

Actions: search

Graph Image Plot

Location 0 ° 0 ° Here Date/time 2020-12-03 14:47:43 +01(Now
Horizon marker ☐ Daylight sky ☐
Planets, Sun & Moon ☐ as Symbol with English names ☐
Width 0 px Projection Aitoff Coordinates Galactic
Center 156.5 ° 50.0 ° Orientation 0.0 ° Fixed ☑ Show (Select constellation)
Stars ☑ down to magnitude 6 with spectral colors ☐ or default color
Show Designation ☐ down to mag 2.5 Show IAU Name ☐ down to mag 1.5
Stellar disk size: base 5 * e ^ (exponent -0.28 * (magnitude + 2)) [* adaptation]
DSOs ☐ down to mag 6 with symbol colors ☑ or default color

i:ranr: 83.214862
i:rb: 0.8485714
i:rbversion: t17_f5_c3
i:rcid: 46
i:rfid: 408120246
i:schemavsn: 3.3
i:scorr: 7.06487846374512
i:seeratio: 2.0
i:sgmag1: 17.6135
i:sgmag2: -999.0
i:sgmag3: -999.0
i:sgscore1: 0.995417
i:sgscore2: 0.00645833
i:sgscore3: 0.5
i:sharpnr: -0.02
i:sigmagap: 0.1905
i:sigmagapbig: 0.3075
i:sigmagnr: 0.015
i:sigmapsf: 0.18681242
i:simag1: 15.275
i:simag2: 21.3195
i:simag3: 21.978
i:sky: -0.0466814
i:srmag1: 16.2928
i:srmag2: 21.457
i:srmag3: 21.9736
i:ssdistnr: -999.0
i:ssmagnr: -999.0
i:ssnamenr: null
i:ssnrms: 14.867763
i:sumrat: 0.95787317
i:szmag1: 14.7296
i:szmag2: 20.4432
i:szmag3: 20.9367
i:tblid: 16
i:tooflag: 0
i:xpos: 227.5853
i:ypos: 1485.4648
i:zpclrcov: -7.31E-6
i:zpmed: 26.189
d:cdsxmatch: Star
d:mulens_class_1:
d:mulens_class_2:
d:nalerthist: 3
d:rfscore: 0.0
d:roid: 0
d:snn_sn_vs_all: 0.49134424328804016
d:snn_snia_vs_nonia: 0.5109549164772034
key:time: 1598256123025

+ ZTF17aaaajdq_2458789.0297801
+ ZTF17aaaaxxx_2458789.0426273

Select graph server and initial graph,
then select an element to see possible actions.

Sending Gremlin request to http://134.158.74.85:24444: g.V().hasLabel('AstroLabNet').limit(10)
Showing 1 new elements
Sending Gremlin request to http://134.158.74.85:24444: g.V(4192).out().limit(10)

Fink Data Explorer 00.08.00+ [03/Dec/2020 at 15:22:43 CET by hrivnac for IJCLab] Reset

IJCLab ☐add
Search AstroLabNet
Execute
g.V().hasLabel('AstroLabNet')

ZTF19acmdpyr_2458798.6020718
t = 12/Nov/19 03:26:59.000
i:ncovhist = 1623
i:drb = 0.99999976

Actions: search

Graph Image Plot

i:drb: 0.9999999
i:drbversion: d6_m7
i:dsdiff: -8.358794
i:dsnrms: 11.118421
i:elong: 1.1139755
i:exptime: 30.0
i:fid: 1
i:field: 764
i:fink_broker_version: 0.7.0
i:fink_science_version: 0.3.8
i:fwhm: 3.04
i:isdiffpos: t
i:jd: 2458789.6188889
i:jdendhist: 2458789.6188889
i:jdendref: 2458230.907616
i:jdstarthist: 2458785.6011111
i:jdstartref: 2458204.9664
i:magap: 19.2328
i:magapbig: 19.3116
i:magdiff: 0.063791
i:magfromlim: 0.8115324
i:maggaia: 13.243567
i:maggaiabright: 13.243567
i:magnr: 19.882

Loading plot : d3/scatterplot.jsp?
hbase=http://134.158.74.54:2181&htable=ztf_season1&schema=schem
&z=i:drb &s=i:chinr
Evolution Point: ZTF19acmdpyr_2458798.6020718, t = 12/Nov/19 03:2

Fink Data Explorer 00.08.00+ [03/Dec/2020 at 15:22:43 CET by hrivnac for IJCLab] Reset

IJCLab ☐add
Search AstroLabNet
Execute
g.V().hasLabel('AstroLabNet')

ZTF17aaafzot_2458789.0426273

i:aimage = 0.609
i:bimagerat = 0.30277777
i:chinr = 7.199

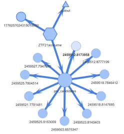
Actions: search

Graph Image Plot

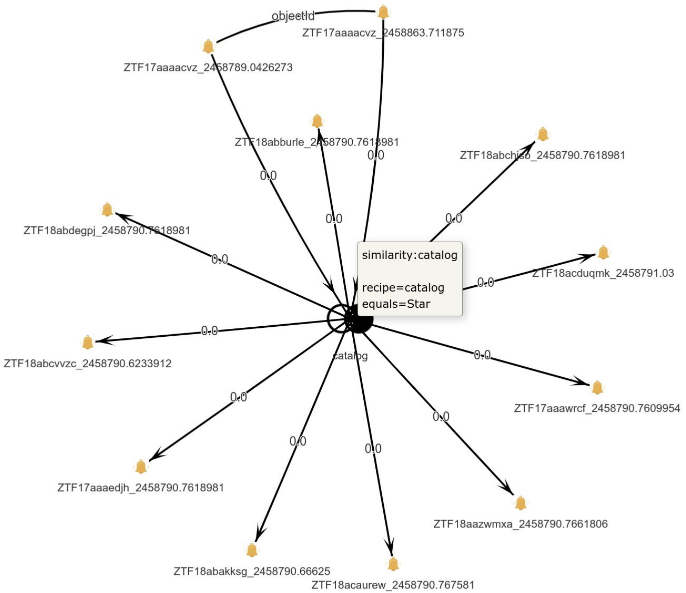
ZTF17aaaacvz_2458789.0426273
key: ZTF17aaaacvz_2458789.0426273
b:cutoutDifference_stampData
b:cutoutScience_stampData
b:cutoutTemplate_stampData
i:aimage: 0.903
i:aimagerat: 0.6587762
i:bimage: 0.718
i:bimagerat: 0.523811
i:candid: 1034542624615015
i:chinr: 0.428
i:chipsf: 1.9058931
i:classtar: 0.785
i:clrcoeff: 0.112484
i:clrcounc: 8.7055E-6
i:clrmed: 0.787
i:clrrms: 0.253676
i:dec: -2.1267664
i:decnr: -2.1267448
i:diffmaglim: 19.741377
i:distnr: 0.114166796
i:distpsnr1: 0.120230444
i:distpsnr2: 14.874653
i:distpsnr3: 15.011706
i:drb: 0.99915147
i:drbversion: d6_m7
i:dsdiff: -11.044785
i:dsnrms: 3.822978
i:elong: 1.2576602
i:exptime: 30.0
i:fid: 2
i:field: 408
i:fink_broker_version: 0.7.0

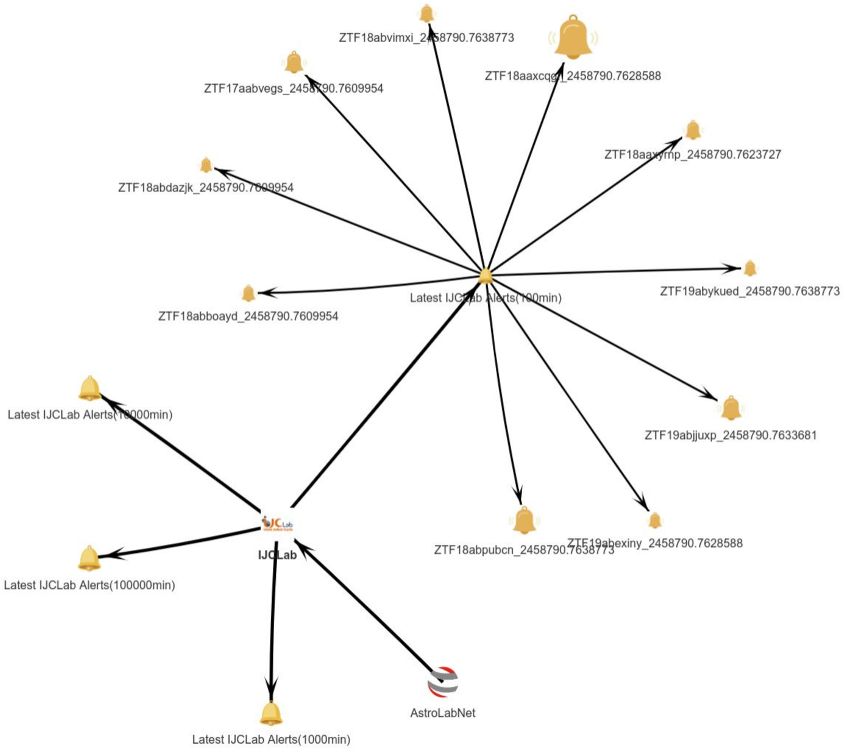
test_portal_tiny.1@http://134.158.74.54:2181
key starts with ZTF17aaajfgy,ZTF18aaacdlm,ZTF18aaacxew,ZTF18aadqujv,ZTF18acdvzye,ZTF19acmbtkq,ZTF19acmbtrn,
using schema schema_v0
Show latest alerts Show all alerts Plot selected variables

Other Elements

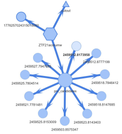


- ➢ An authorised user can add vertex/edge properties and new vertices and edges
- ➢ For example:
  - ○ To connect related alerts
  - ○ To create a virtual collection of alerts
  - ○ To annotate alerts with additional information
  - ○ To add elements with global information


objectId
ZTF17aaaacvz_2458863.711875
ZTF17aaaacvz_2458789.0426273
ZTF18abburle_2458790.7618981
ZTF18abchjsb_2458790.7618981
ZTF18abdegpj_2458790.7618981
similarity:catalog
recipe=catalog
equals=Star
ZTF18acduqmk_2458791.03
ZTF18abcvvzc_2458790.6233912
catalog
ZTF17aaawrcf_2458790.7609954
ZTF17aaaedjh_2458790.7618981
ZTF18aazwmxa_2458790.7661806
ZTF18abakksg_2458790.66625
ZTF18acaurew_2458790.767581
0.0


ZTF18abvimxi_2458790.7638773
ZTF18aaxcqq_2458790.7628588
ZTF17aabvegs_2458790.7609954
ZTF18aaxymp_2458790.7623727
ZTF18abdazjk_2458790.7609954
Latest IJCLab Alerts(100min)
ZTF19abykued_2458790.7638773
ZTF18abboayd_2458790.7609954
Latest IJCLab Alerts(10000min)
ZTF19abjjuxp_2458790.7633681
IJCLab
ZTF18abpubcn_2458790.7638773
ZTF19abexiny_2458790.7628588
Latest IJCLab Alerts(100000min)
Latest IJCLab Alerts(1000min)
AstroLabNet

GQL (Cypher)



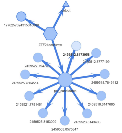
- ➢ SQL-like (declarative) graph query languages developed by Neo4J
- ➢ GQL can be run on top of Gremlin
  - ○ Not the other way around

Gremlin

```
g.V().has('objectId', 'ZTF18acgkldj').out().has('lbl', 'candidate').valueMap('classtar', 'jd', 'direction'
```

GQL

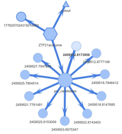
```
(a) - [:contains:] - (b:candidate)
WHERE a.objectId = 'ZTF18acgkldj'
RETURN b.classtar, b.jd, b.direction
```



Implementation

- Current implementation uses
  - JanusGraph database
  - HBase on Hadoop data storage
  - ElasticSearch indexing
- All those choices can change

Links



- Main ideas:
  - **Use Graph DB to provide flexibility**
  - **Expose directly Gremlin query language**
    - **In API, CLI, REST**
  - **Provide server-side functions with requested functionality**
  - **Use proxy-client to customise interface**
- Near future:
  - Re-implement existing functionality
  - Add additional features
  - Connect to the Fink Service
- More info about Graph databases:
  - Using Graph Databases
  - Gremlin Query language