# Visualization of dCache accounting information with state-of-the-art Data Analysis Tools

Tigran Mkrtchyan
for DESY dCache operating Team

# Outline

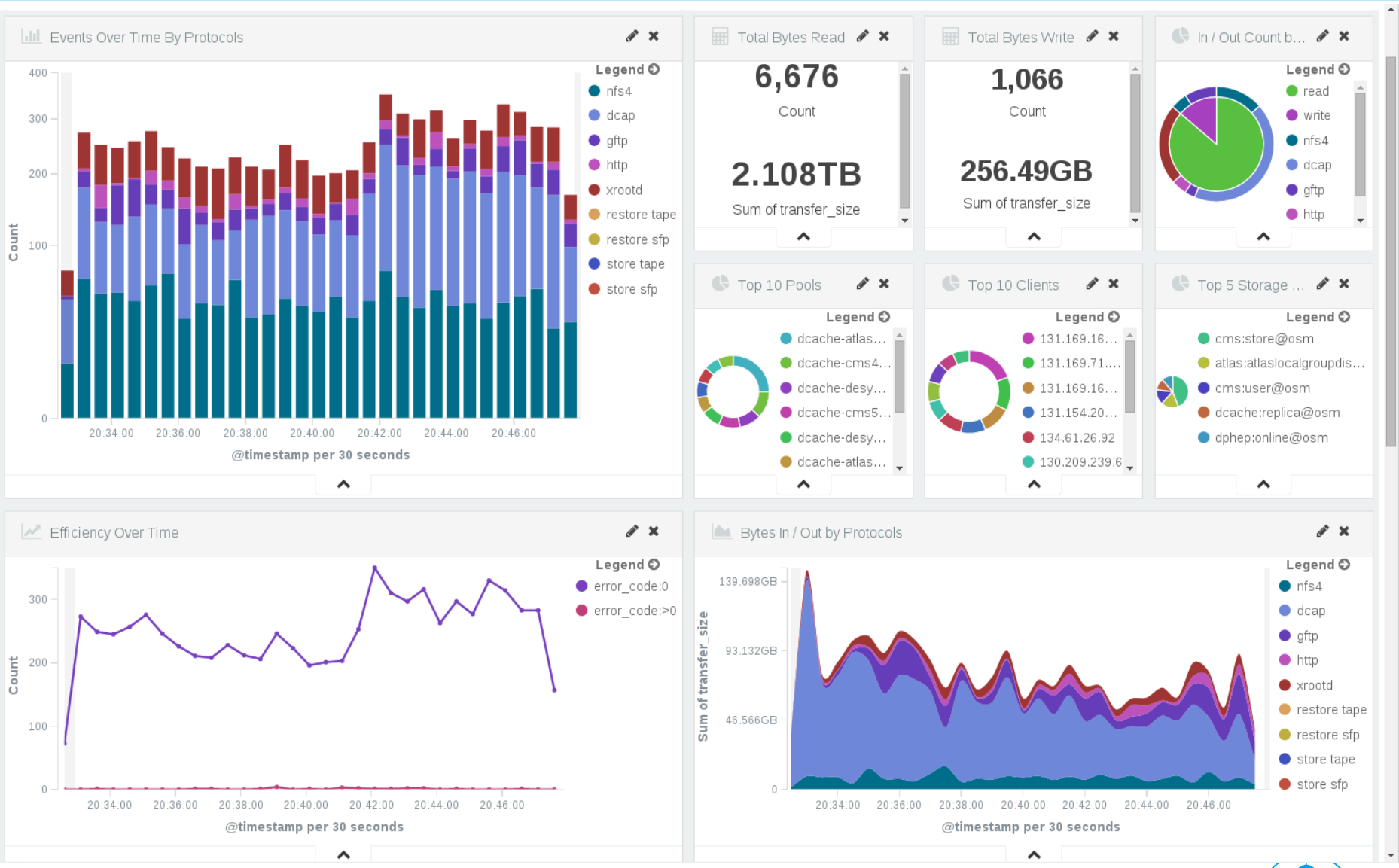
root@dcache-se-desy03:~

```
03.02 00:00:07 [pool:dcache-desy18-05:transfer] [0000F96E04EFC0A74501B0B8265FE1C454B2,616248] [/pnfs/desy.de/desy/generated/SET_testfile] desy:generated@osm 616248 7 false {GFtp-2.0 131.169.223.91 50712} [door:GFTP-dcache-door-desy09-AAUtBLoJ0XA:1456873207925000] {0:""}
03.02 00:00:07 [door:GFTP-dcache-door-desy09-AAUtBLoJ0XA@gridftp-dcache-door-desy09Domain:request] ["/O=GermanGrid/OU=DESY/CN=Andreas Gellrich":43700:4370:131.169.223.91] [0000F96E04EFC0A74501B0B8265FE1C454B2,616248] [/pnfs/desy.de/desy/generated/SET_testfile] desy:generated@osm 29 0 {0:""}
03.02 00:00:16 [pool:dcache-desy14-03:transfer] [00007675386F2FB84F8AABC4D9E71B8DC8EE,16777216] [/pnfs/desy.de/belle/belle2/generated/2016-03-02/file0bb0feeb-a627-4a3a-a728-e12e5c155395] belle:generated@osm 16777216 188 true {GFtp-2.0 131.169.223.91 37423} [door:GFTP-dcache-door-desy12-AAUtBLqIEcA:1456873216217000] {0:""}
03.02 00:00:16 [door:GFTP-dcache-door-desy12-AAUtBLqIEcA@gridftp-dcache-door-desy12Domain:request] ["/O=GermanGrid/OU=DESY/CN=Andreas Gellrich":21065:5296:131.169.223.91] [00007675386F2FB84F8AABC4D9E71B8DC8EE,16777216] [/pnfs/desy.de/belle/belle2/generated/2016-03-02/file0bb0feeb-a627-4a3a-a728-e12e5c155395] belle:generated@osm 393 0 {0:""}
03.02 00:00:17 [door:SRM-dcache-se-desy03@srm-dcache-se-desy03Domain:request] ["/O=GermanGrid/OU=DESY/CN=Andreas Gellrich":21065:5296:131.169.223.91] [00007675386F2FB84F8AABC4D9E71B8DC8EE,16777216] [/pnfs/desy.de/belle/belle2/generated/2016-03-02/file0bb0feeb-a627-4a3a-a728-e12e5c155395] belle:generated@osm 0 0 {0:""}
03.02 00:00:18 [pool:dcache-desy05-07:transfer] [0000729D4C30AA9741EEAC439BF6DD8D20FF,273874663] [/pnfs/desy.de/belle/belle2/TMP/belle/MC/fab/merge/release-00-05-03/DBxxxxxxxx/MC5/prod00000141/s00/e0001/4S/r00000/1290000017/sub00/mdst_001244_prod00000141_task00001244.root] belle:generated@osm 273874663 47028 false {GFtp-2.0 202.13.207.3 60451} [door:GFTP-dcache-door-desy13-AAUtBLemOwg:1456873171376000] {0:""}
03.02 00:00:18 [door:GFTP-dcache-door-desy13-AAUtBLemOwg@gridftp-dcache-door-desy13Domain:request] ["/C=JP/O=KEK/OU=CRC/OU=IPNS/CN=Takanori Hara":23202:5296:202.13.207.3] [0000729D4C30AA9741EEAC439BF6DD8D20FF,273874663] [/pnfs/desy.de/belle/belle2/TMP/belle/MC/fab/merge/release-00-05-03/DBxxxxxxxx/MC5/prod00000141/s00/e0001/4S/r00000/1290000017/sub00/mdst_001244_prod00000141_task00001244.root] belle:generated@osm 47050 0 {0:""}
03.01 22:38:53 [pool:dcache-desy38-02@dcache-desy38-02Domain:store] [0000F8EE066A90E94D058E64884ECF23133B,4295032832] [Unknown] afs:afs-backup@osm 489760112 {0:""}
03.02 00:00:38 [pool:dcache-desy19-05:transfer] [00007816B030EF5E4C5CB7801E617FF797BA,1737396342] [/pnfs/desy.de/olympus/generated/tracked/secondrun/081915/6528.root] olympus:generated@osm 1737396342 21258 false {Http-1.1:olymp-wgs01.desy.de:0:WebDAV-dcache-door-desy14:webdav-dcache-door-desy14Domain:/pnfs/desy.de/olympus/generated/tracked/secondrun/081915/6528.root} [door:WebDAV-dcache-door-desy14:AAUtBLqRovg:1456873216787000] {0:""}
03.02 00:00:38 [door:WebDAV-dcache-door-desy14@webdav-dcache-door-desy14Domain:request] ["":-1:-1:131.169.71.142] [00007816B030EF5E4C5CB7801E617FF797BA,1737396342] [/pnfs/desy.de/olympus/generated/tracked/secondrun/081915/6528.root] olympus:generated@osm 21275 0 {0:""}
03.02 00:00:43 [pool:dcache-desy17-05:transfer] [0000DC13DB508EAE4641BB54D7169837DF9A,1345869749] [/pnfs/desy.de/olympus/generated/tracked/secondrun/081915/6550.root] olympus:generated@osm 1345869749 20783 false {Http-1.1:olymp-wgs01.desy.de:0:WebDAV-dcache-door-desy14:webdav-dcache-door-desy14Domain:/pnfs/desy.de/olympus/generated/tracked/secondrun/081915/6550.root} [door:WebDAV-dcache-door-desy14:AAUtBLrmYkg:1456873222341000] {0:""}
03.02 00:00:43 [door:WebDAV-dcache-door-desy14@webdav-dcache-door-desy14Domain:request] ["":-1:-1:131.169.71.142] [0000DC13DB508EAE4641BB54D7169837DF9A,1345869749] [/pnfs/desy.de/olympus/generated/tracked/secondrun/081915/6550.root] olympus:generated@osm 20799 0 {0:""}
03.02 00:00:48 [door:SRM-dcache-se-desy03@srm-dcache-se-desy03Domain:request] ["/C=JP/O=KEK/OU=CRC/CN=HAYASAKA Kiyoshi":21065:5296:202.13.193.136] [,0] [/pnfs/desy.de/belle/belle2/TMP/belle/monitor/urandom100MB.1456873218.dat] <Unknown> 0 0 {666:"Request aborted by client."}
03.02 00:00:50 [pool:dcache-desy14-03:transfer] [00007675386F2FB84F8AABC4D9E71B8DC8EE,16777216] [/pnfs/desy.de/belle/belle2/generated/2016-03-02/file0bb0feeb-a627-4a3a-a728-e12e5c155395] belle:generated@osm 16777216 117 false {GFtp-2.0 131.169.223.91 51079} [door:GFTP-dcache-door-desy12-AAUtBLyLNMA:1456873249959000] {0:""}
03.02 00:00:50 [door:GFTP-dcache-door-desy12-AAUtBLyLNMA@gridftp-dcache-door-desy12Domain:request] ["/O=GermanGrid/OU=DESY/CN=Andreas Gellrich":21065:5296:131.169.223.91] [00007675386F2FB84F8AABC4D9E71B8DC8EE,16777216] [/pnfs/desy.de/belle/belle2/generated/2016-03-02/file0bb0feeb-a627-4a3a-a728-e12e5c155395] belle:generated@osm 138 0 {0:""}
03.02 00:00:59 [door:SRM-dcache-se-desy03@srm-dcache-se-desy03Domain:remove] ["/O=GermanGrid/OU=DESY/CN=Andreas Gellrich":21065:5296:131.169.223.91] [00007675386F2FB84F8AABC4D9E71B8DC8EE,0] [/pnfs/desy.de/belle/belle2/generated/2016-03-02/file0bb0feeb-a627-4a3a-a728-e12e5c155395] <Unknown> 0 0 {0:""}
03.02 00:01:02 [pool:dcache-desy38-02@dcache-desy38-02Domain:remove] [0000AA9238B3E38F456397798B361766D37D,4295032832] [Unknown] afs:afs-backup@osm {0:""}
03.02 00:01:06 [pool:dcache-desy11-01:transfer] [00003D2E6DCC689F463A9A0F2820B2AEAE34,275870325] [/pnfs/desy.de/belle/belle2/TMP/belle/MC/fab/merge/release-00-05-03/DBxxxxxxxx/MC5/prod00000141/s00/e0001/4S/r00000/1290000017/sub02/mdst_020590_prod00000141_task00020590.root] belle:generated@osm 275870325 159532 false {GFtp-2.0 202.13.207.3 46053} [door:GFTP-dcache-door-desy10-AAUtBLPLuCg:1456873106735000] {0:""}
03.02 00:01:06 [door:GFTP-dcache-door-desy10-AAUtBLPLuCg@gridftp-dcache-door-desy10Domain:request] ["/C=JP/O=KEK/OU=CRC/OU=IPNS/CN=Takanori Hara":23202:5296:202.13.207.3] [00003D2E6DCC689F463A9A0F2820B2AEAE34,275870325] [/pnfs/desy.de/belle/belle2/TMP/belle/MC/fab/merge/release-00-05-03/DBxxxxxxxx/MC5/prod00000141/s00/e0001/4S/r00000/1290000017/sub02/mdst_020590_prod00000141_task00020590.root] belle:generated@osm 159556 0 {0:""}
03.01 22:38:53 [pool:dcache-desy38-02@dcache-desy38-02Domain:store] [0000E6536EF87BB046CC993F37C604C36533,4295032832] [Unknown] afs:afs-backup@osm 493734712 {0:""}
03.02 00:01:12 [pool:dcache-desy37-01@dcache-desy37-01Domain:remove] [00008A7F4A2157E343CA9349955DB5E7A1EA,194615] [Unknown] desy:zimbra@dcache {0:""}
/var/lib/dcache/billing/2016/03/billing-2016.03.02
```

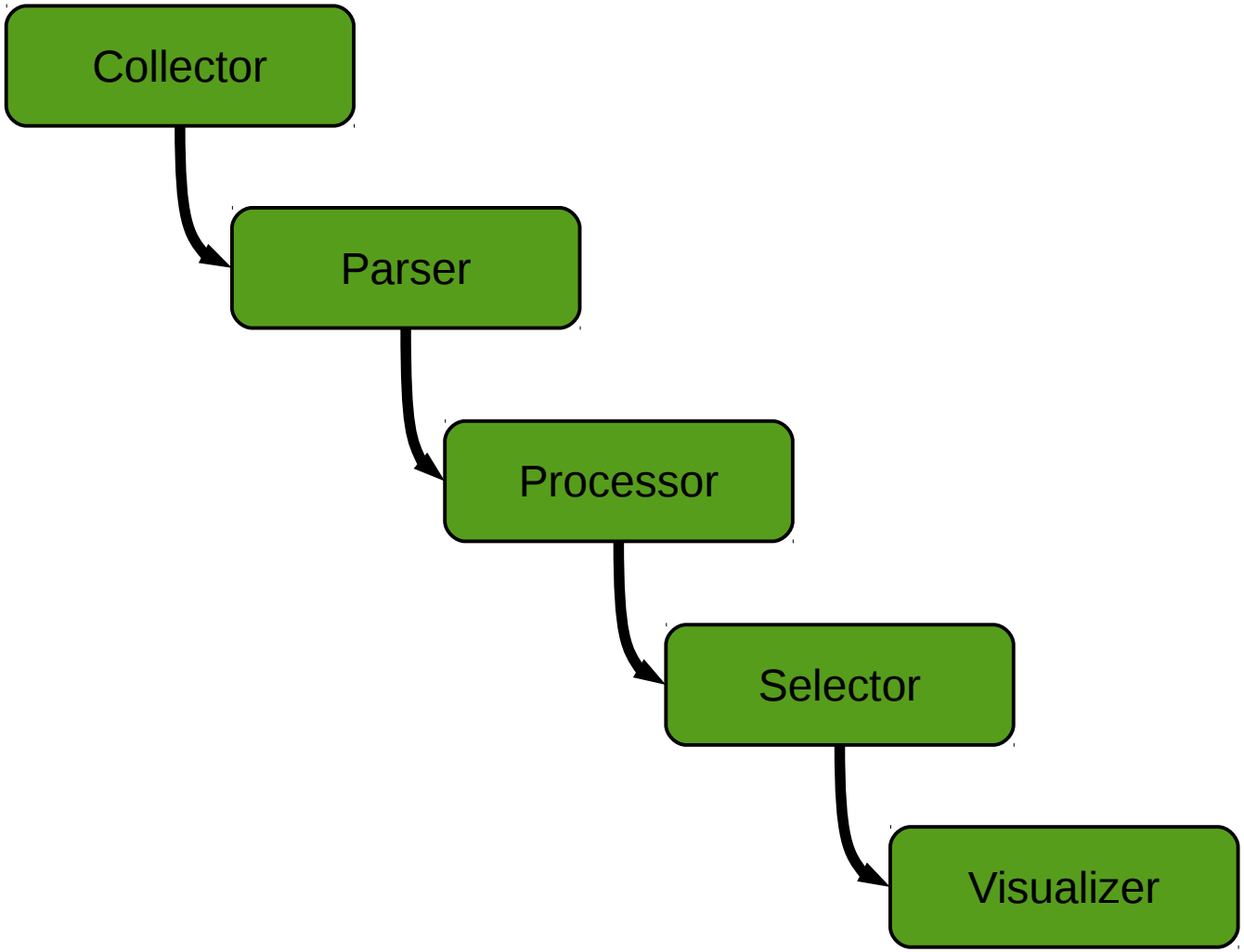


DESY

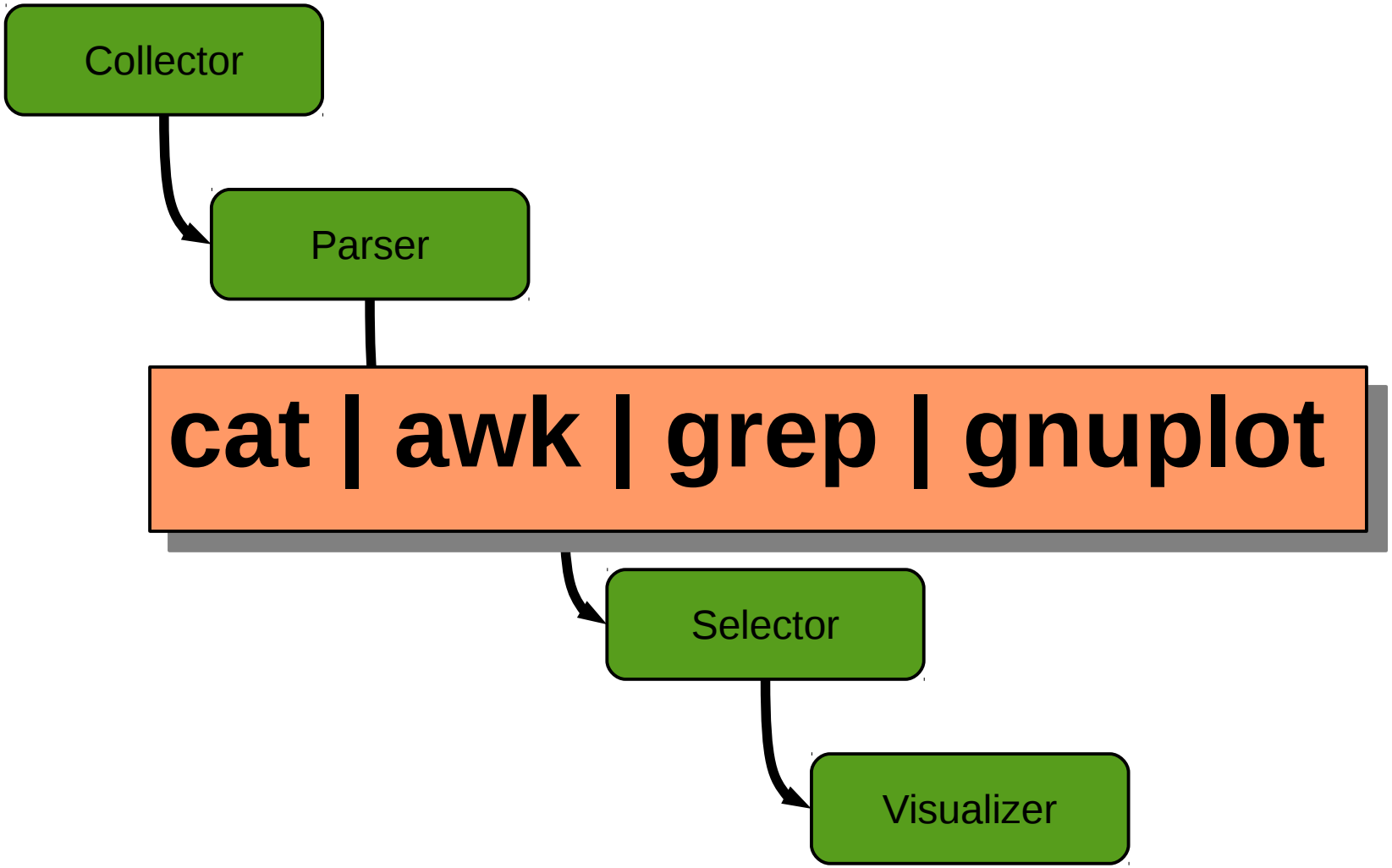
# Outline


Events Over Time By Protocols
Count
@timestamp per 30 seconds
Legend
nfs4
dcap
gftp
http
xrootd
restore tape
restore sfp
store tape
store sfp
Total Bytes Read
6,676
Count
2.108TB
Sum of transfer_size
Total Bytes Write
1,066
Count
256.49GB
Sum of transfer_size
In / Out Count b…
read
write
Top 10 Pools
Top 10 Clients
Top 5 Storage …
cms:store@osm
atlas:atlaslocalgroupdis…
cms:user@osm
dcache:replica@osm
dphep:online@osm
Efficiency Over Time
error_code:0
error_code:>0
Bytes In / Out by Protocols
Sum of transfer_size
139.698GB
93.132GB
46.566GB

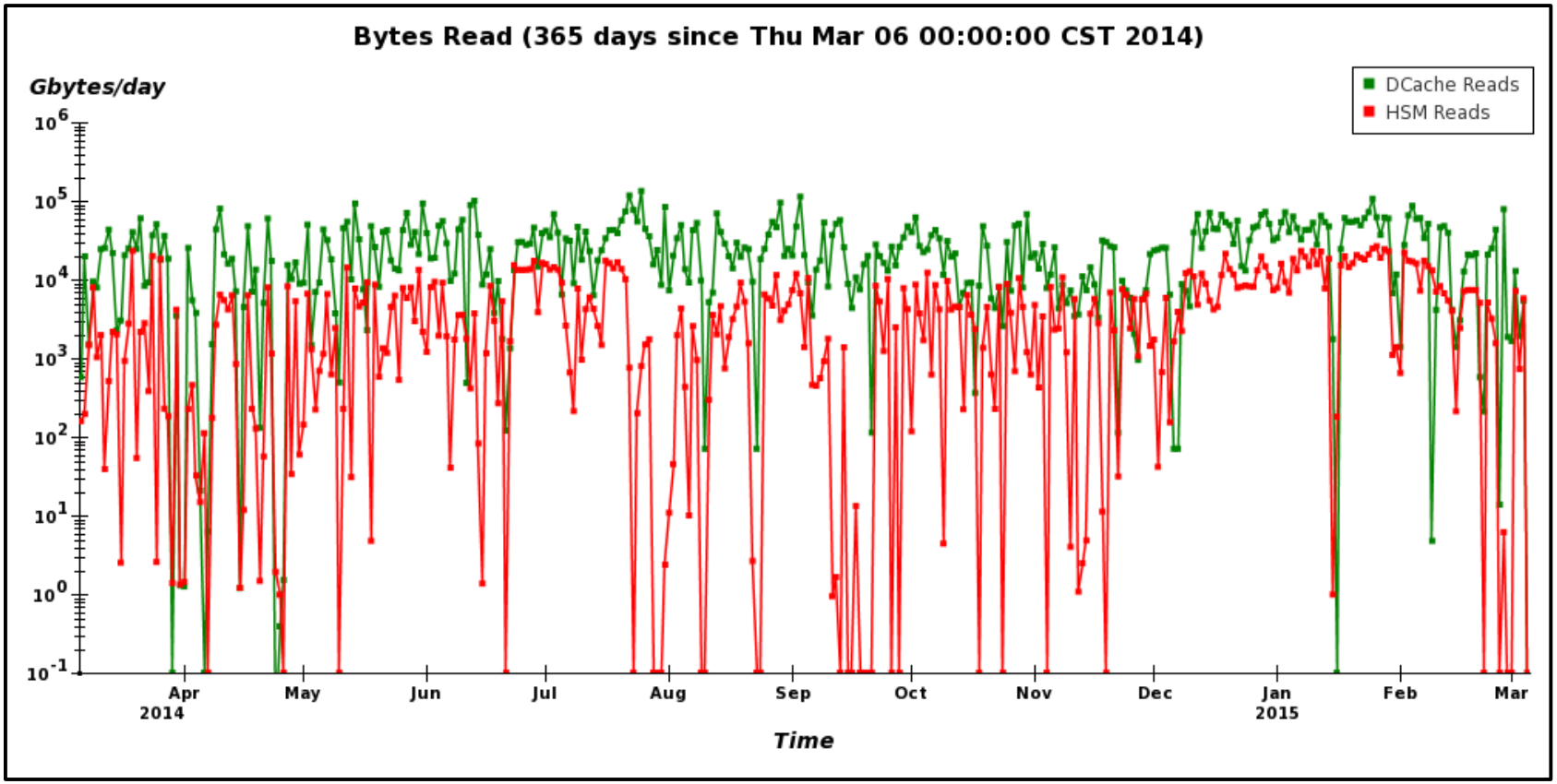
# The Flow



# The Flow (typical)



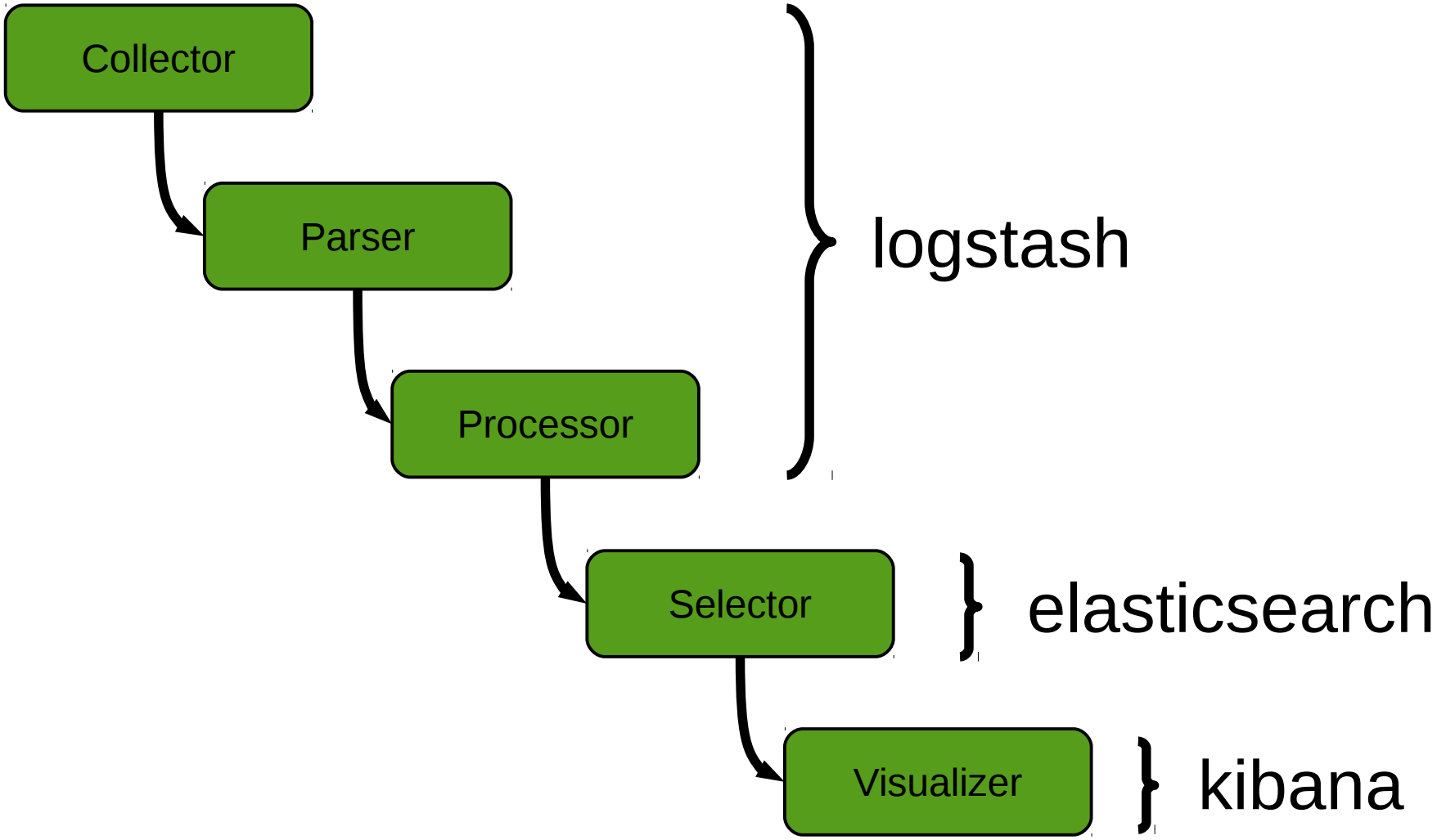
# Result


Bytes Read (365 days since Thu Mar 06 00:00:00 CST 2014)
Gbytes/day
DCache Reads
HSM Reads
$10^6$
$10^5$
$10^4$
$10^3$
$10^2$
$10^1$
$10^0$
$10^{-1}$
Apr
2014
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2015
Feb
Mar
Time

# Scaling problems

- ~20GB billing files/day
- ~50.000.000 records/day
  - ~500 records/sec
- 7 dCache instances
- need to adopt scripts for different needs
- need for a ***'State at Glance'***

# The Flow


Collector
Parser
Processor
Selector
Visualizer
logstash
elasticsearch
kibana

# Logstash

- Collect logs from any source
- parse them
- gets the right timestamp
- index them
- and move it into a central place

# Logstash anatomy

```
input {
  # read log events
}

filter {
  # parse, fix formats, mutate
}

output {
  # store processed events
}
```

# Logstash, single liner

```
$  echo "hello logstash" | logstash -e 'input { stdin{} } output { stdout {codec => rubydebug} }'
{
   "message" => "hello logstash",
   "@version" => "1",
   "@timestamp" => "2016-03-06T22:49:37.797Z",
    "host" => "dcache-lab"
}
```

# Real life example

03.02 08:35:49 [pool:dcache-desy23-05:transfer] [00009A23BB6D280F46A7A6C12AC67F5EA897,59419220] [Unknown] desy:generated@osm 90112 1195 false {Http-1.1:dcache-infra03.desy.de:0:WebDAV-dcache-door-desy13:webdav-dcache-door-desy13Domain:/pnfs/desy.de/desy/dcache.org/2.1/dcache-server_2.1.1-1_all.deb} [door:WebDAV-dcache-door-desy13@webdav-dcache-door-desy13Domain:1399012548236-1399012548243] {0:""}

# Parse

```
filter {
  grok {
    match => [ "message", "%{TRANSFER_CLASSIC}" ]
    remove_field => [ "message" ]
  }

  date {
    match => [ "billing_time", "MM.dd HH:mm:ss" ]
    timezone => "CET"
    remove_field => [ "billing_time" ]
  }
}
```

# Parse

- Regexp like syntax
- Lot of ready patterns for common cases
- supports labels and types

# Parser, example

[00009A23BB6D280F46A7A6C12AC67F5EA897,59419220]
[00380000000000000559888,46305280]

PNFSID_NEW (?:[A-F0-9]{36})
PNFSID_OLD (?:[A-F0-9]{24})
PNFSID %{PNFSID_OLD}|%{PNFSID_NEW}

PNFSID_SIZE \[%{PNFSID:**pnfsid**},%{**NONNEGINT**:size:**int**}\]

# Parser, example

{DCap-3.0,131.169.74.175:34232}

PROTO (?:%{**DATA**}-[0-9]\.[0-9])

PROTOCOL \{%{PROTO:**proto**}(:)(%{**IPORHOST**:**remote_host**})(:)(%{**NONNEGINT**:**remote_port**:int})

# Full parser

TRANSFER_CLASSIC %{BILLING_TIME:billing_time} %{CELL_AND_TYPE} %{PNFSID_SIZE} %{PATH} %{SUNIT} %{TRANSFER_SIZE} %{TRANSFER_TIME} %{IS_WRITE} %{PROTOCOL} %{DOOR} %{ERROR}

# Real Life example


```
{
  "@version" => "1",
  "@timestamp" => "2016-03-02T06:35:49.000Z",
  "type" => "dcache-billing",
  "host" => "ani",
  "path" => "/var/lib/dcache/billing/2016/03/billing-2016-03-02.log",
  "pool_name" => "dcache-desy23-05",
  "bill_type" => "transfer",
  "pnfsid" => "00009A23BB6D280F46A7A6C12AC67F5EA897",
  "size" => 59419220,
  "file_path" => "/pnfs/desy.de/desy/dcache.org/2.1/dcache-server_2.1.1-1_all.deb",
  "sunit" => "desy:generated@osm",
  "transfer_size" => 90112,
  "transfer_time" => 1195,
  "is_write" => "false",
  "proto" => "Http-1.1",
  "remote_host" => "dcache-infra03.desy.de",
  "remote_port" => 0,
  "payload" => ":WebDAV-dcache-door-desy13:webdav-dcache-door-desy13Domain:",
  "initiator_type" => "door",
  "initiator" => "WebDAV-dcache-door-desy13@webdav-dcache-door-desy13Domain:1399012548236-1399012548243",
  "error_code" => 0
}
```

# And store it in....

```
output {
  elasticsearch {
    host => "elastic-search-master-node"
    index => "logstash-%{+YYYY.MM.dd}"
  }
}
```

# Elasticsearch

- Open-source full-text search engine
- Schema-free JSON documents
- Powerful JSON based REST-API
- Distributed
  - data can be divided into shards
  - each shard can have zero or more replicas
- Node can be Master-node, Data-node or both
- Can be used as a NoSQL database

# Document, Index and type

- Document is a basic unit of information
- Documents are expressed in JSON
- Each log entry corresponds to a document
- Index is a collection of documents
- An index is identified by a name (or alias)
- Name is used to refer to the index when performing actions
- Type is a logical category/partition of an index
- Type is defined for documents that have a set of common fields

(something like DATABASE (index), ROW(document) and TABLE(type) in RDBMS)

# Shards and Replicas

- Index can be subdivide into multiple pieces
- Each piece called shard
- Each shard is an independent "index" and can be hosted on any node in the cluster.
  - allows horizontally split/scale data volume
  - allows distribute operations across shards
- You can make one or more copies of index’s shards called replicas
  - provides high availability in case a shard/node fails
  - allows to scale out search volume/throughput since searches can be executed on all replicas in parallel

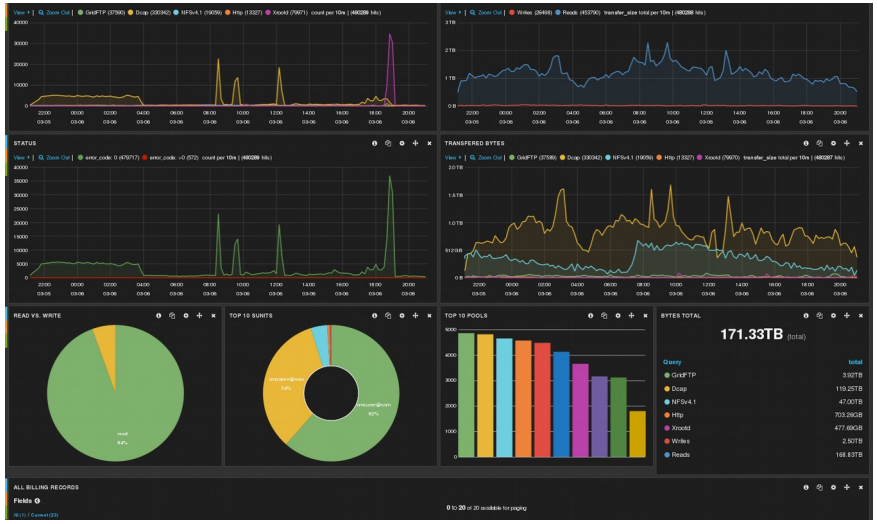
# CURD

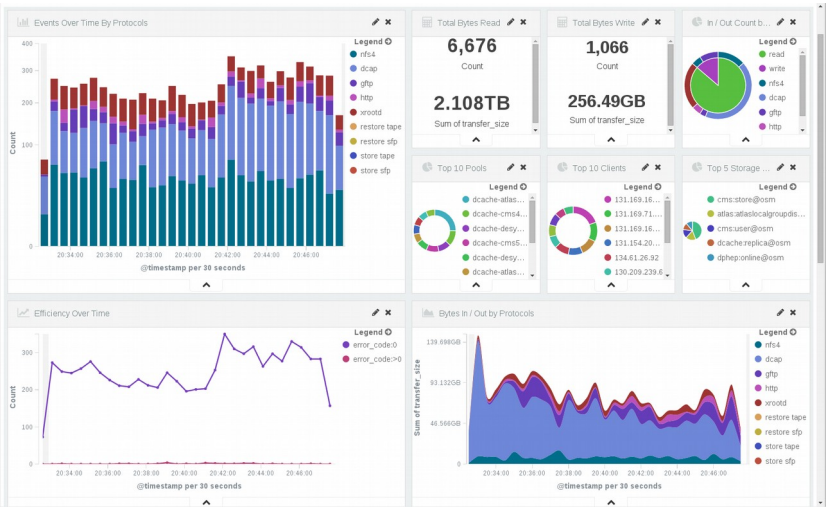
> REST API

- **POST** – create document, index
- **GET** – search/read document
- **PUT/PATCH** – update document
- **DELETE** – delete document, index

# Kibana

- Flexible analysis and visualization platform
- Real-time summary and charting of streaming data
- Intuitive interface for a variety of users
- Instant sharing and embedding of dashboards

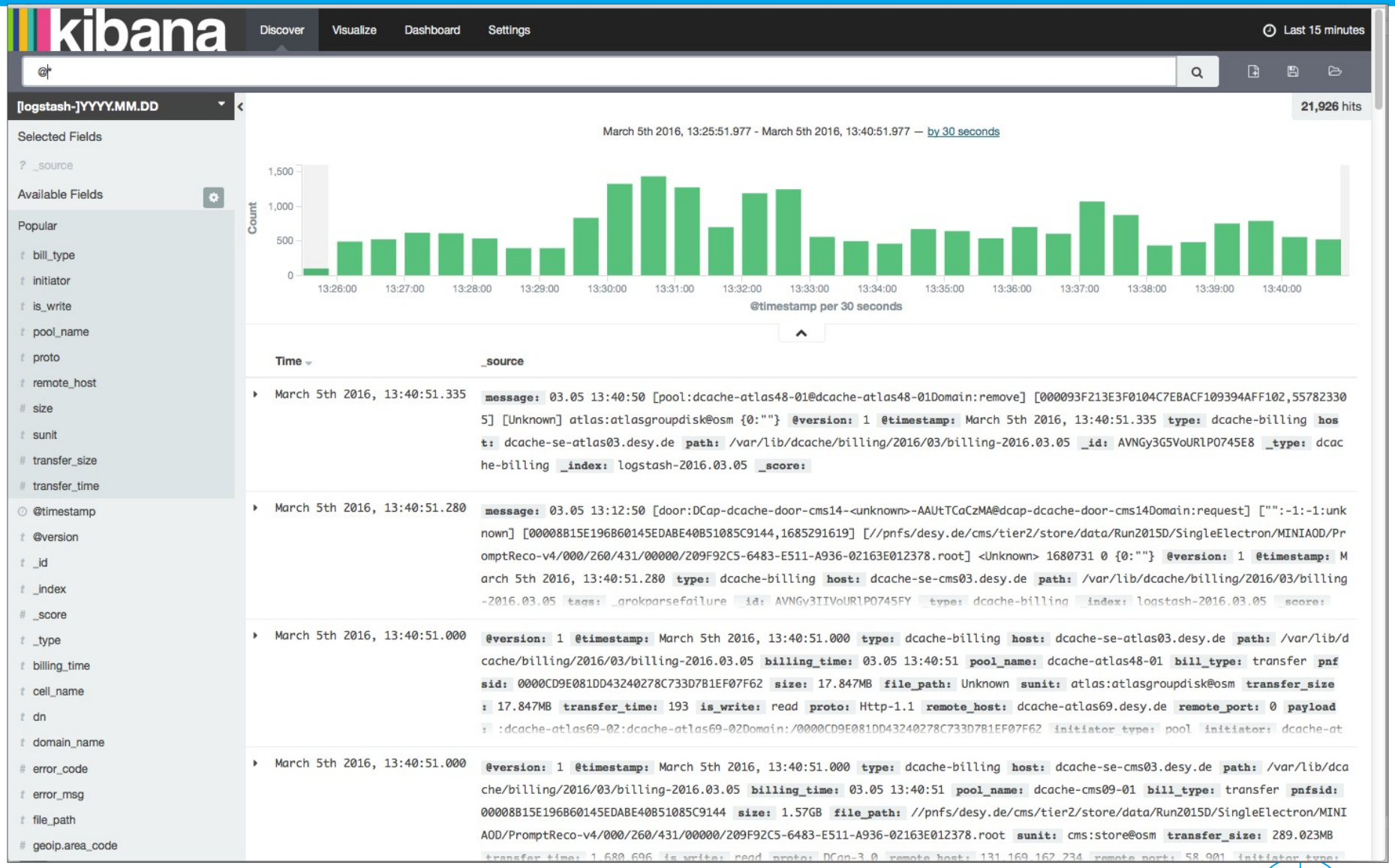


# Get started

- Dump data into elasticsearch
- Use discovery panel (or simple dashboard in Kibana3)
- Play with data
  - search and aggregate

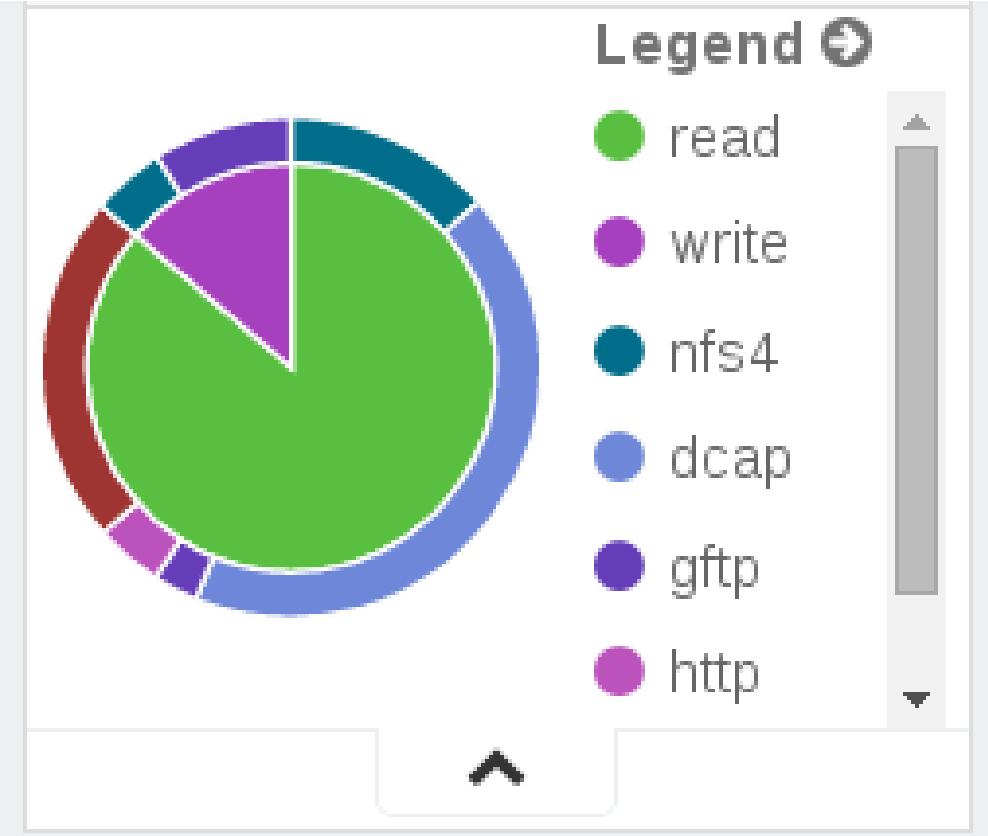
# Get started


kibana
Discover
Visualize
Dashboard
Settings
Last 15 minutes
@*
[logstash-]YYYY.MM.DD
Selected Fields
? _source
Available Fields
Popular
t bill_type
t initiator
t is_write
t pool_name
t proto
t remote_host
# size
t sunit
# transfer_size
# transfer_time
@timestamp
t @version
t _id
t _index
# _score
t _type
t billing_time
t cell_name
t dn
t domain_name
# error_code
t error_msg
t file_path
# geoip.area_code
21,926 hits
March 5th 2016, 13:25:51.977 - March 5th 2016, 13:40:51.977 — by 30 seconds
Count
1,500
1,000
500
0
13:26:00
13:27:00
13:28:00
13:29:00
13:30:00
13:31:00
13:32:00
13:33:00
13:34:00
13:35:00
13:36:00
13:37:00
13:38:00
13:39:00
13:40:00
@timestamp per 30 seconds
Time
_source
March 5th 2016, 13:40:51.335
message: 03.05 13:40:50 [pool:dcache-atlas48-01@dcache-atlas48-01Domain:remove] [000093F213E3F0104C7EBACF109394AFF102,557823305] [Unknown] atlas:atlasgroupdisk@osm {0:""} @version: 1 @timestamp: March 5th 2016, 13:40:51.335 type: dcache-billing host: dcache-se-atlas03.desy.de path: /var/lib/dcache/billing/2016/03/billing-2016.03.05 _id: AVNGy3G5VoURlPO745E8 _type: dcache-billing _index: logstash-2016.03.05 _score:
March 5th 2016, 13:40:51.280
message: 03.05 13:12:50 [door:DCap-dcache-door-cms14-<unknown>-AAUtTCaCzMA@dcap-dcache-door-cms14Domain:request] ["":-1:-1:unknown] [00008B15E196B60145EDABE40B51085C9144,1685291619] [//pnfs/desy.de/cms/tier2/store/data/Run2015D/SingleElectron/MINIAOD/PromptReco-v4/000/260/431/00000/209F92C5-6483-E511-A936-02163E012378.root] <Unknown> 1680731 0 {0:""} @version: 1 @timestamp: March 5th 2016, 13:40:51.280 type: dcache-billing host: dcache-se-cms03.desy.de path: /var/lib/dcache/billing/2016/03/billing-2016.03.05 tags: _grokparsefailure id: AVNGy3IIVoURlPO745FY type: dcache-billing index: logstash-2016.03.05 score:
March 5th 2016, 13:40:51.000
@version: 1 @timestamp: March 5th 2016, 13:40:51.000 type: dcache-billing host: dcache-se-atlas03.desy.de path: /var/lib/dcache/billing/2016/03/billing-2016.03.05 billing_time: 03.05 13:40:51 pool_name: dcache-atlas48-01 bill_type: transfer pnfsid: 0000CD9E081DD43240278C733D7B1EF07F62 size: 17.847MB file_path: Unknown sunit: atlas:atlasgroupdisk@osm transfer_size: 17.847MB transfer_time: 193 is_write: read proto: Http-1.1 remote_host: dcache-atlas69.desy.de remote_port: 0 payload: :dcache-atlas69-02:dcache-atlas69-02Domain:/0000CD9E081DD43240278C733D7B1EF07F62 initiator_type: pool initiator: dcache-at
March 5th 2016, 13:40:51.000
@version: 1 @timestamp: March 5th 2016, 13:40:51.000 type: dcache-billing host: dcache-se-cms03.desy.de path: /var/lib/dcache/billing/2016/03/billing-2016.03.05 billing_time: 03.05 13:40:51 pool_name: dcache-cms09-01 bill_type: transfer pnfsid: 00008B15E196B60145EDABE40B51085C9144 size: 1.57GB file_path: //pnfs/desy.de/cms/tier2/store/data/Run2015D/SingleElectron/MINIAOD/PromptReco-v4/000/260/431/00000/209F92C5-6483-E511-A936-02163E012378.root sunit: cms:store@osm transfer_size: 289.023MB transfer_time: 1,680,696 is_write: read proto: DCap-3.0 remote_host: 131.169.162.234 remote_port: 58,901 initiator_type:

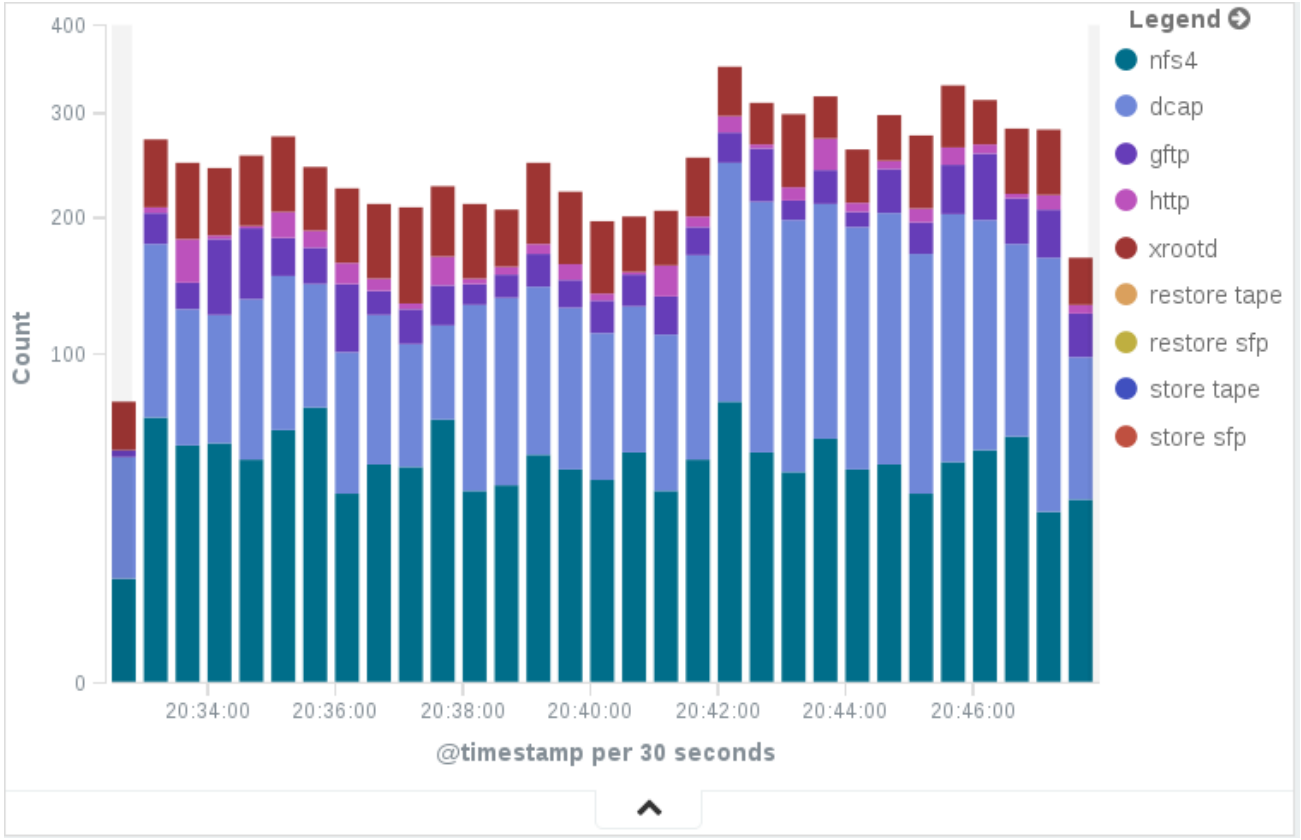
# The building blocks

- Search
- Aggregation
- Visualization

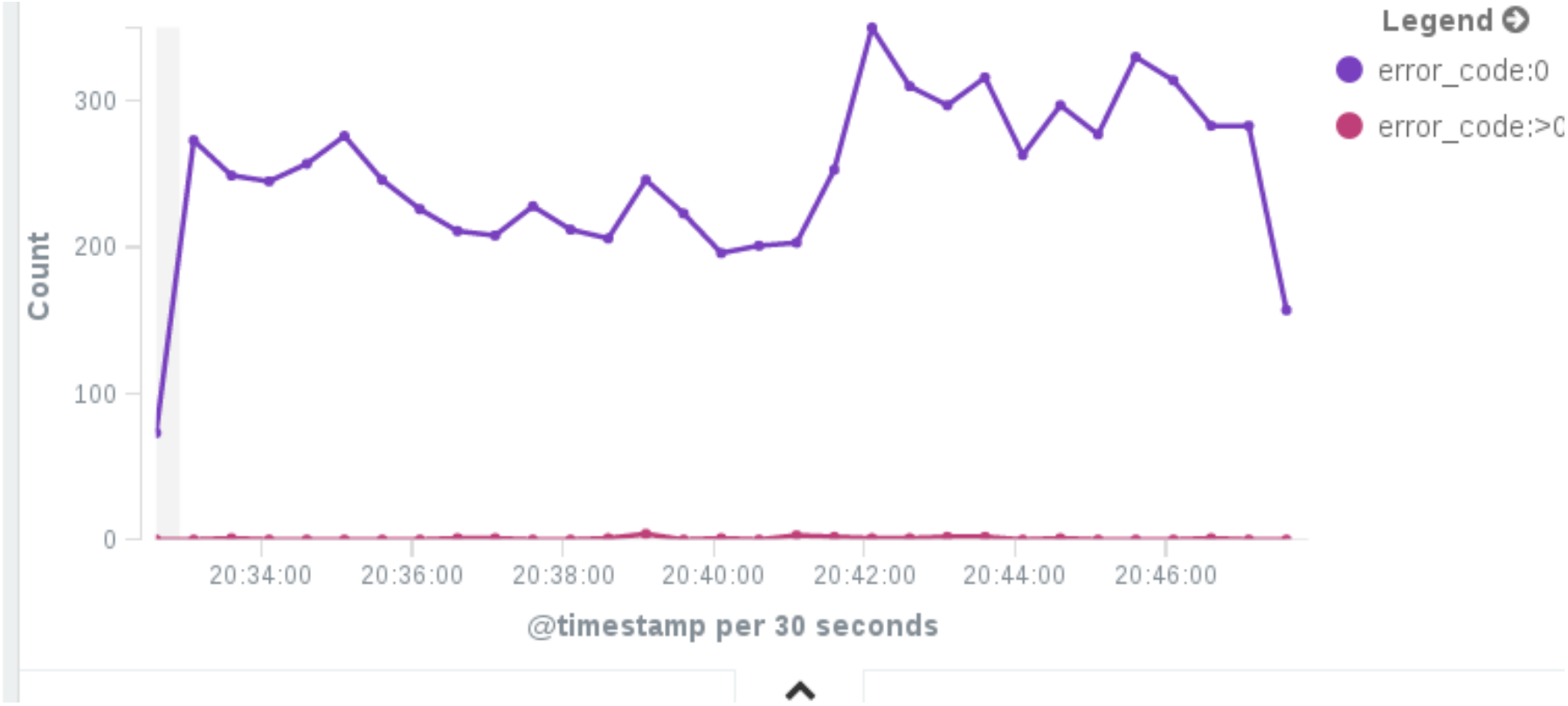
# Example



# Example


Legend
nfs4
dcap
gftp
http
xrootd
restore tape
restore sfp
store tape
store sfp
Count
400
300
200
100
0
20:34:00
20:36:00
20:38:00
20:40:00
20:42:00
20:44:00
20:46:00
@timestamp per 30 seconds

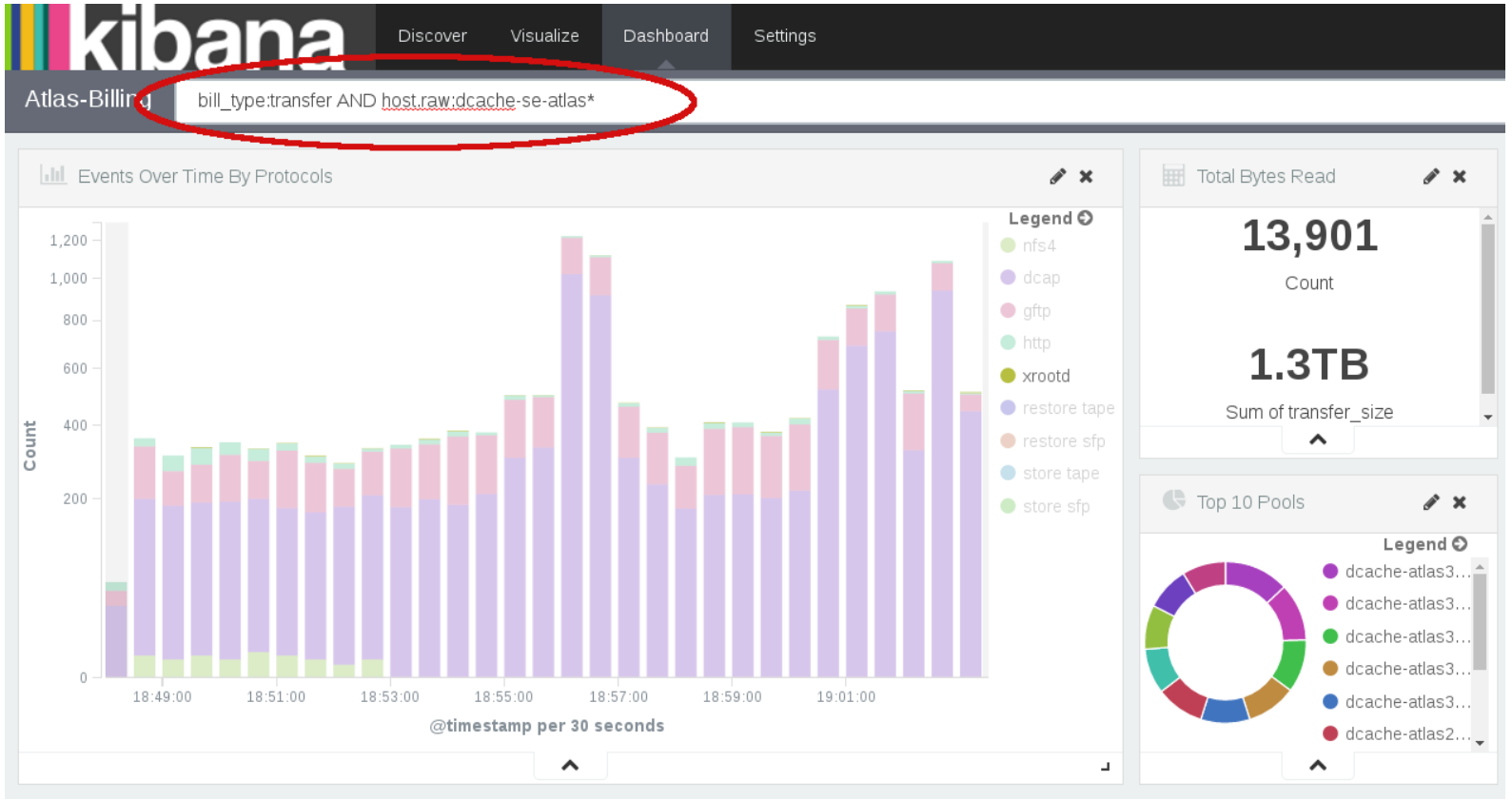
# Example


Legend
error_code:0
error_code:>0
Count
300
200
100
0
20:34:00
20:36:00
20:38:00
20:40:00
20:42:00
20:44:00
20:46:00
@timestamp per 30 seconds

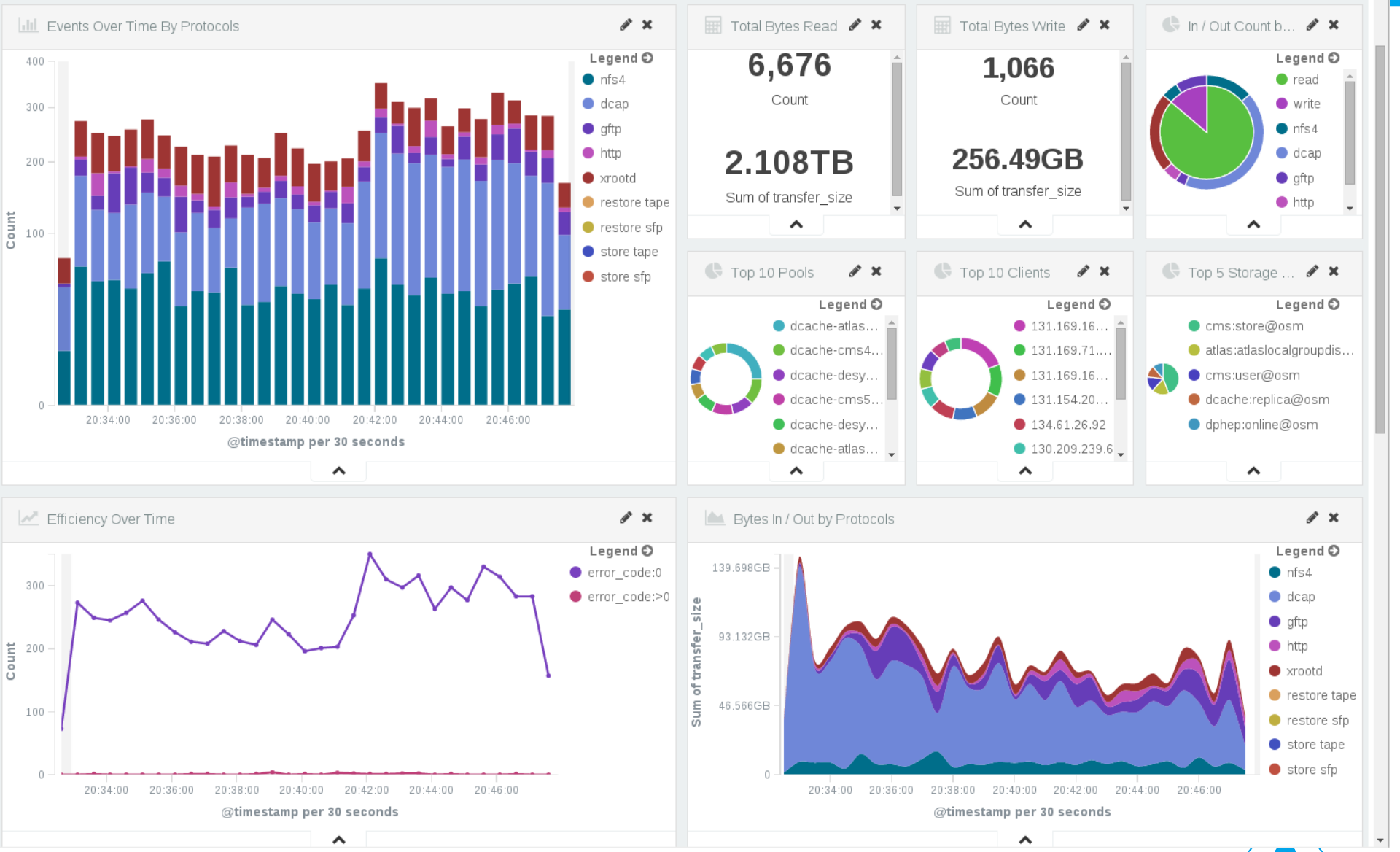
# Dashboard

- A collection of visualizations
- Visualizations may use different 'data sources'
- A search in a dashboard affects all visualizations

# Search in dashboard



# Transfers at glance



# Too good to be the truth

- Elasticsearch is very giddy component
- Number of active indexes is limited by file descriptors
  - #indexes * #shards * #replicas * #segments
  - In production we can have max ~180 days
- Can't be used to analyze historic data
- But good enough for live monitors
  - not a reporting too
- Updates brake backward compatibility
  - real pain with restrictions on field names for existing documents

# Development

- Iterative functional enhancements
  - Each Kibana adds great functionality
- Kibana3 → Kibana4
  - Different products
  - Different concepts
  - No migration path
- Kibana often requires new version of Elasticsearch
- Grafana – Visualization tool based on fork of Kibana

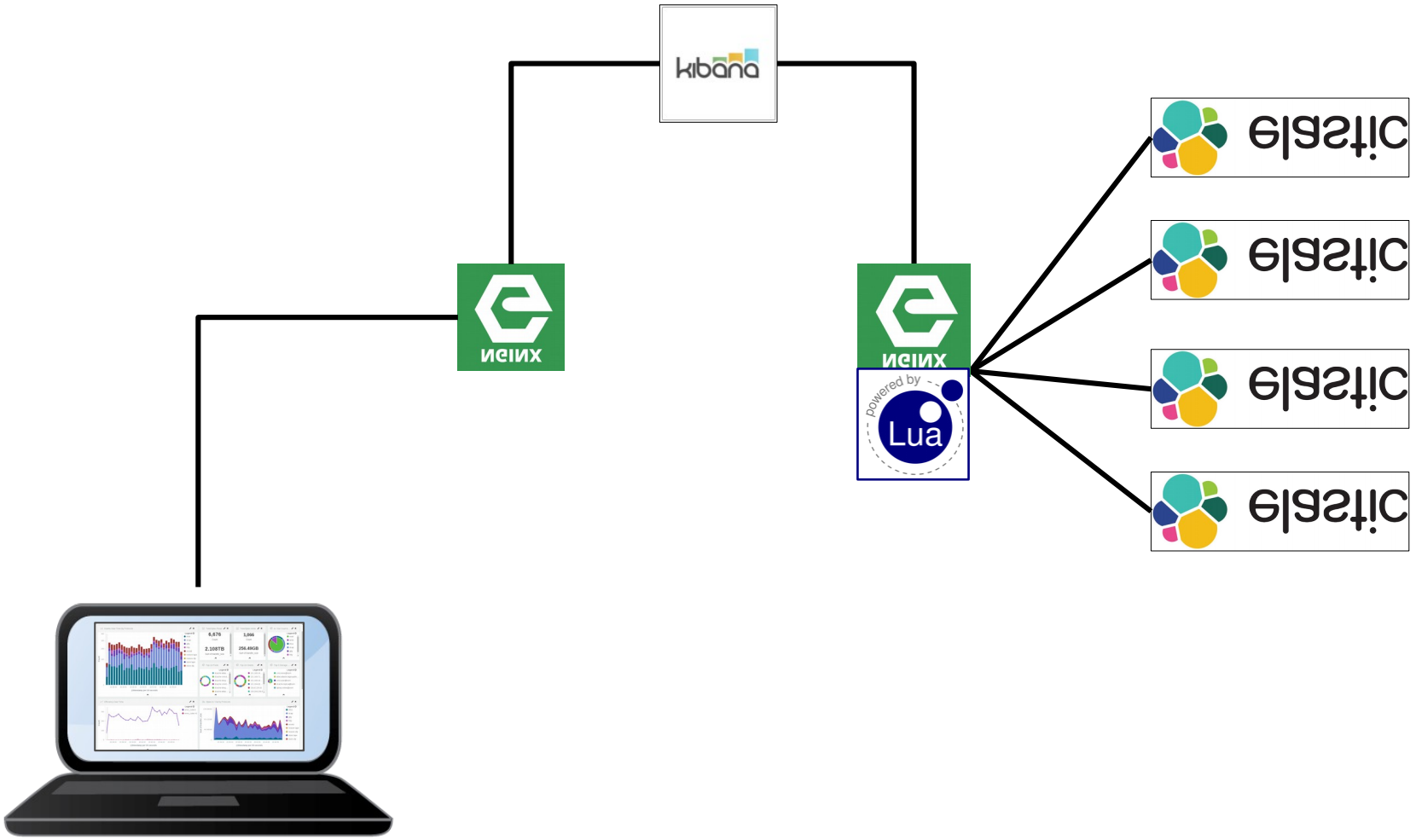
# Looking back (how to organize index)

- index => "logstash-%{+YYYY.MM.dd}"
  - typical search will use 1-7 indexes
  - typical search data overhead one day
  - limited number of indexes
  - discard granularity one day
- index => "logstash-%{+YYYY.MM}"
  - typical search will use 1 index
  - typical search data overhead one month
  - discard granularity one month
- your 'live view' defines which type of index you need

# Our infrastructure

- Elasticsearch
  - 2.0.0, 2.2.1 by end of month
  - 9 Nodes
  - Recycled hardware
  - All data replicated
- Kibana
  - 4.2.0,  4.4.2 by end of month
- Logstash
- nginx
  - OpenResty-1.9.7.3 with LuaJIT (Lua 5.1)

# Our Infrastructure



# Who-is-Who

- No Authentication/Authorization by default
- All data available as soon as you can get access to ES REST-API
- Shield – native commercial solution
  - Prices on request
- Different projects to solve this issue
  - Search Guard – similar to shield
  - Custom http request manipulations

# Summary

- Production services produce Gigabytes of log files per day
- Crunching the millions of numbers into a useful and handy information is not a simple task.
- Modern BigData tools looks promising approach to attack the problem
- Using widely used tools let as to adopt common practices used by other communities.

# Questions?