Large Scale OpenAFS Performance Monitoring with Graphite

Sine Nomine Associates

August 20, 2015

Graphite

- Store and graph time-series metrics
- Clustering features for scaling
- Metric namespace
- Render API
- Python stack

Graphite

- Apache License
- Widely deployed
- Many tools that work with graphite
- I/O intensive (one file per metric!)

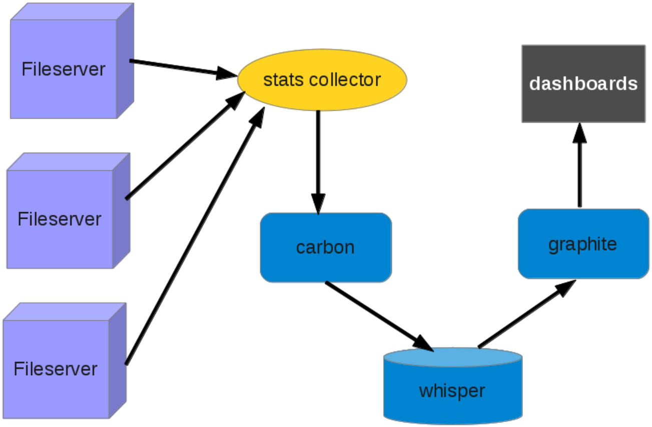
Graphite

- Carbon - daemons to accept and store metrics
- Whisper - library/file format to store time-series metrics
- Graphite - web app provides render API and basic dashboard

Deployment Experience

- Deployed to a separate server
- Partition for whisper
- Graphite RPM install almost worked, ymmv
    - some django setup headaches
    - several small patches needed
- Configure carbon data retention policies

Feeding AFS metrics to Graphite



Feeding metrics to graphite

- It is very easy to feed metrics to graphite
- Lots of tools to feed system metrics
- Metrics are sent to carbon via tcp (or udp)
- One metric per line: (path, value, timestamp)

```
afs.example.afs01.rx.callswaited  31324 1420520400
```

Important
Data points should be feed to graphite at the same rate as the highest data retention frequency!

Gathering AFS fileserver metrics

- rxdebug: rx_GetServerDebug() rx_GetServerStats()
- vice stats GetStatistics64
- xstats RXAFS_GetXStats()
- xstat collections
  - 0: not implemented on the fileserver
  - 1: partial performance stats (subset of 2)
  - 2: performance stats
  - 3: callback stats
- vos partition information
- audit log fids and hosts
  - sys-v msg queue

Carbon Configuration

- Retention buckets
  - regex of paths
  - frequency:duration
- Aggregation methods
  - regex of metric
  - how to rollup data

Important
Use validate-storage-schemas to check retention policies.

Carbon Aggregation

- sum for counters (rx calls waited, packets sent)
- average for gauges (size, used percent)

Important

If you have more than one retention bucket, you must provide the method to aggregate data! The carbon default is average (gauge). Most afs metrics are counts.

Metric Namespace

- Decide on metric namespace conventions early
- Changes later will break render API calls
- Dot is reserved as a separator!

Example:

```
afs.<cellname>.<server>.part.<part>.<metric>
```

Whisper files

- Metrics are stored in regular files in a binary format.
- Makes it easy to archive the whole set
- Nice for "offline" analysis
- Whisper tools are handy to dump header info and raw data!

```
whisper-dump /var/lib/carbon/whisper/afs/example/foo/bar.w
```

Render API

- The render API is the heart of graphite!
- Rich set of functions
    - combine, transform, calculate, filter
- Rest API
- Json for data dumps
- Create and share a set of standard render calls
- Ad hoc exploration

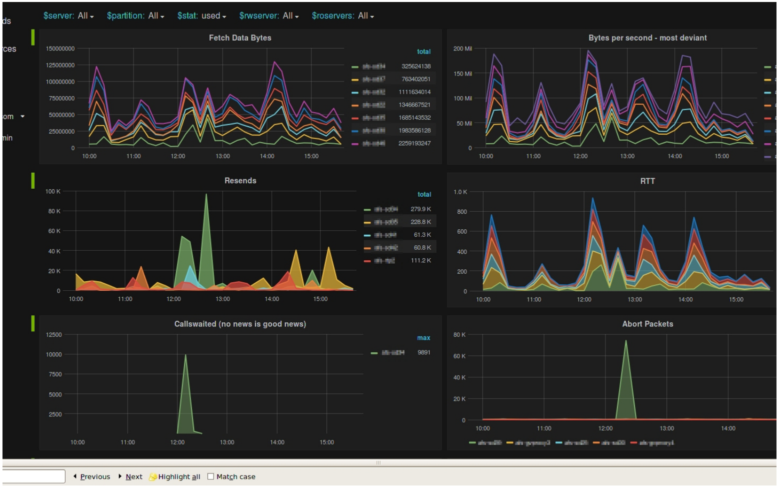
Render Example

```
http://graphite/render?
  title=Fetched bytes per second
  from=-14day
  until=now
  target=
    aliasByNode(
    highestAverage(
    scaleToSeconds(
    nonNegativeDerivative(
        afs.example.*.stats.xfer.FetchData.bytes.sum),1))
```

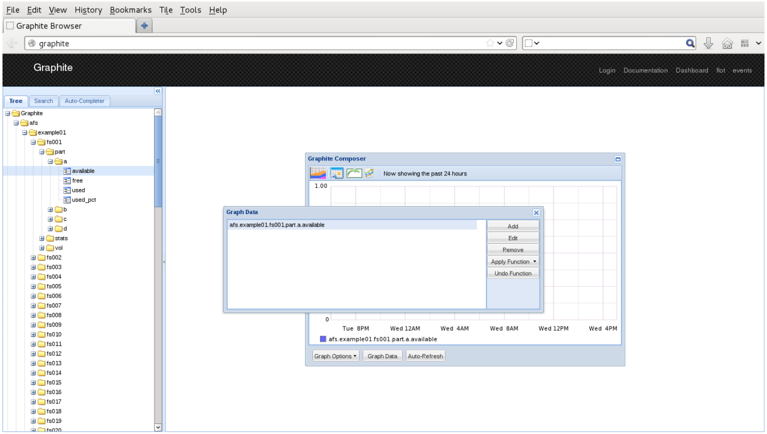
Grafana

- Popular optional client dashboard
- Easy to deploy
- Generally nicer than the built-in graphite dashboard
- Great for ad hoc graphs and exploration

Example

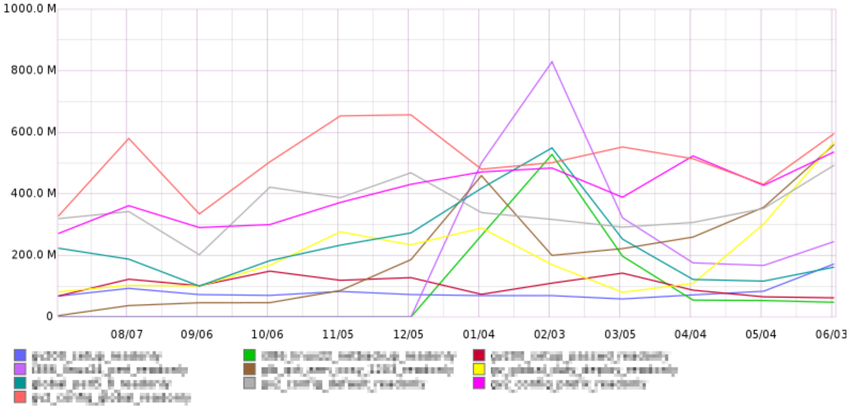

ds
ces
om
min
$server: All
$partition: All
$stat: used
$rwserver: All
$roservers: All
Fetch Data Bytes
total
[redacted] 325624138
[redacted] 763402051
[redacted] 1111634014
[redacted] 1346667521
[redacted] 1685143532
[redacted] 1983586128
[redacted] 2259193247
Bytes per second - most deviant
Resends
total
[redacted] 279.9 K
[redacted] 228.8 K
[redacted] 61.3 K
[redacted] 60.8 K
[redacted] 111.2 K
RTT
Callswaited (no news is good news)
max
[redacted] 9891
Abort Packets
Previous
Next
Highlight all
Match case

Example

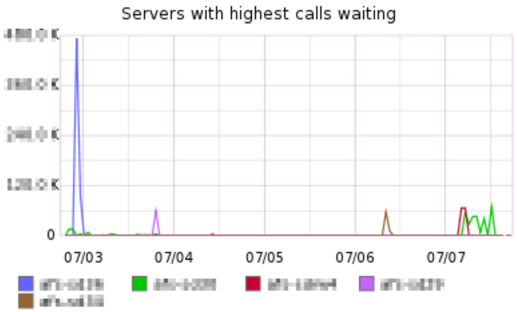


File Edit View History Bookmarks Tile Tools Help
Graphite Browser
graphite
Graphite
Login Documentation Dashboard flot events
Tree Search Auto-Completer
Graphite
afs
example01
fs001
part
a
available
free
used
used_pct
b
c
d
stats
vol
fs002
fs003
fs004
fs005
fs006
fs007
fs008
fs009
fs010
fs011
fs012
fs013
fs014
fs015
fs016
fs017
fs018
fs019
fs020
Graphite Composer
Now showing the past 24 hours
1.00
0
Tue 8PM Wed 12AM Wed 4AM Wed 8AM Wed 12PM Wed 4PM
afs.example01.fs001.part.a.available
Graph Data
afs.example01.fs001.part.a.available
Add
Edit
Remove
Apply Function
Undo Function
Graph Options
Graph Data
Auto-Refresh

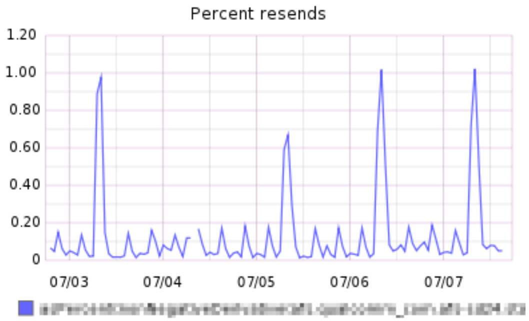
Example


1000.0 M
800.0 M
600.0 M
400.0 M
200.0 M
0
08/07
09/06
10/06
11/05
12/05
01/04
02/03
03/05
04/04
05/04
06/03

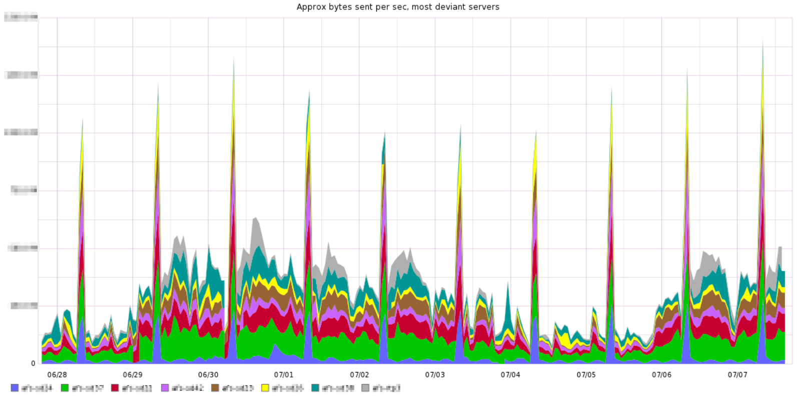
Example


Servers with highest calls waiting
0
07/03
07/04
07/05
07/06
07/07

Example


Percent resends
1.20
1.00
0.80
0.60
0.40
0.20
0
07/03
07/04
07/05
07/06
07/07

Example


Approx bytes sent per sec, most deviant servers
0
06/28
06/29
06/30
07/01
07/02
07/03
07/04
07/05
07/06
07/07

Questions

This slide intentionally left blank.