

RRDtoolからInfluxDBにデータ移行してみた

v1 Jan.26.2018

Atsushi Fujiwara
Global Operations Department.
Rakuten, Inc.

自己紹介



藤原 敦史

所属： 楽天株式会社

社内向けプラットフォームサービスの運用チーム

担当業務：

以前はネットワークエンジニア

現在は社内システムのパフォーマンス監視など

興味があるもの：

データ収集、可視化、自動化

InfluxDB、Grafana、Fluentd

RRD ファイルに格納されている情報

```
$ rrdtool info (filename)
```

```
filename = "(filename)"
```

```
rrd_version = "0003"
```

```
step = 300
```

```
last_update = 1515986183
```

```
header_size = 2912
```

```
ds[ds0].index = 0
```

```
ds[ds0].type = "COUNTER"
```

```
ds[ds0].minimal_heartbeat = 600
```

```
...
```

```
ds[ds1].last_ds = "8826395"
```

```
ds[ds1].value = 0.000000000000e+00
```

```
ds[ds1].unknown_sec = 83
```

```
rra[0].cf = "AVERAGE"
```

```
rra[0].rows = 800
```

```
rra[0].cur_row = 368
```

```
rra[0].pdp_per_row = 1
```

```
rra[0].xff = 5.000000000000e-01
```

```
...
```

データの時間間隔(秒)

データの最終更新時刻(UNIX時間)

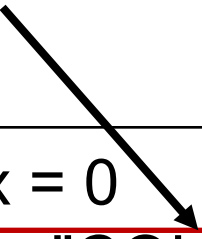
DS定義情報

RRA定義情報

DS(Data Source)

対象のデータ取得元の特徴を定義しておく

- 取りうる最大、最小値
- 観測された最新値
- データのタイプ(COUNTER, GAUGE など)



```
ds[ds0].index = 0
ds[ds0].type = "COUNTER"
ds[ds0].minimal_heartbeat = 600
ds[ds0].min = 0.000000000000e+00
ds[ds0].max = 1.250000000000e+09
ds[ds0].last_ds = "46732397779"
ds[ds0].value = 0.000000000000e+00
ds[ds0].unknown_sec = 83
```

データ移行のためには個数が分かっていたら十分

RRA(Round Robin Archive)

データ格納領域の定義

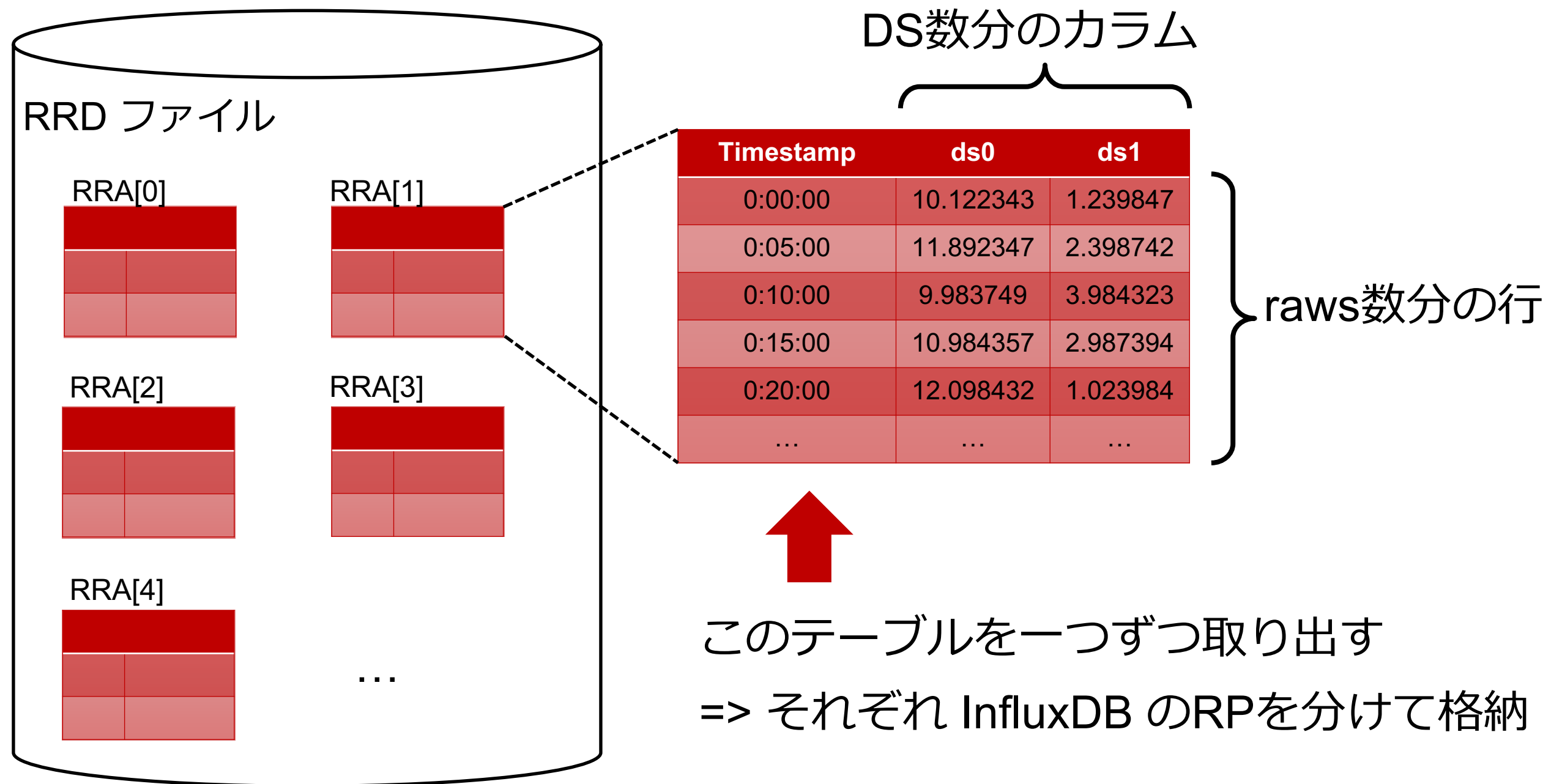
```
rra[0].cf = "AVERAGE"  
rra[0].rows = 800  
rra[0].cur_row = 368  
rra[0].pdp_per_row = 1  
rra[0].xff = 5.000000000000e-01  
rra[0].cdp_prep[0].value = NaN  
rra[0].cdp_prep[0].unknown_datapoints = 0  
rra[0].cdp_prep[1].value = NaN  
rra[0].cdp_prep[1].unknown_datapoints = 0
```

データ集約関数
(AVERAGE,MIN,MAX,LAST)

最大データ行数

データ集約する行数

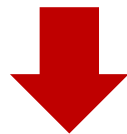
データエクスポートのイメージ



メタデータを補完

RRDファイルは1 インタフェースにつき1 ファイル

- ホスト名、インタフェース名などは通常ファイル名として現れる



複数ファイルをInfluxDBにまとめて入れる

- ホスト名、インタフェース名などをタグとして付与してやる必要がある
- その他、必要に応じてラック、建物などのロケーション情報など

やってみた

RRD ファイルのデータを InfluxDB に移行する

<https://qiita.com/atfujiwara/private/680a6d1127dd012c00b4>

Grafanaで可視化

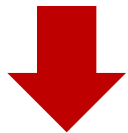


継続してデータを入れるには

Caution

DS type がGAUGEのもの以外は、そのままでは継続してデータを入れることができない

- RRD ファイルに格納されているのは前回の値との差分(COUNTER場合)などを計算した後の値
- 継続してデータを入れるには同等の処理が必要



どうするか

- 一旦 InfluxDB に生データを書き込んで、クエリーで集約処理する
- Kapacitor を経由して集約処理してから InfluxDB に書き込む

まとめ

RRD ファイルに含まれるデータ

- 各DSのメトリックデータ
=> fetch して整形してインポート
- データの集約方式や保持期間
=> InfluxDBのRPとして実装
=> 集約の処理については別途実装
- Grafana で可視化

これらの機能を網羅していた
RRDtool

