

ゆるふわ非構造 #1 (LT 5min)
fluentdで~~非構造化データ~~とゆるふわっと戯れる会
前処理後のデータ

2014/09/28

高橋達

@nora96o

自己紹介

名前:

高橋 達 (たかはし とおる)

Twitter:

@nora96o

普段:

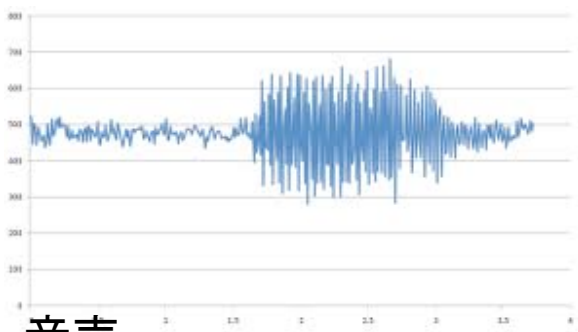
ビッグデータ分析基盤のシステム構築・支援

Like:

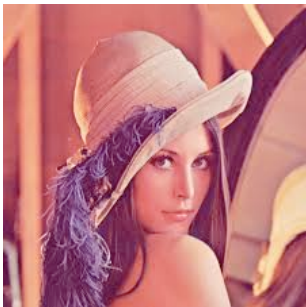
TreasureData / AWS / Ruby / fluentd / Tableau



前処理後のデータはどうするの？



音声



画像

HTML



Webページ

「彼女は大学で勉強します」を形態素解析すると？

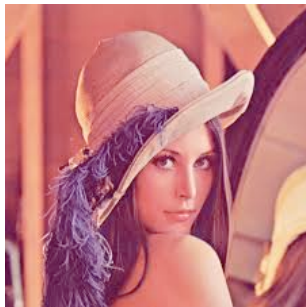
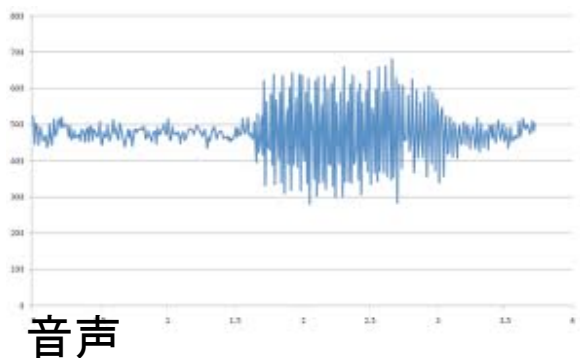
文字列	読み	原形	品詞の種類	活用の種類	活用形
彼女	カノジョ	彼女	代名詞 人称		
は	ワ	は	副助詞 は		
大学	ダイガク	大学	普通名詞		
で	デ	で	格助詞 で		
勉強	ベンキョウ	勉強	さ変名詞		
し	シ	する	動詞	さ変	連用
ます	マス	ます	助動詞	ます	終止

テキスト

```
- - [09/Sep/2014:00:50:21 +0900] "GET /kibana/app/dashboards/default.json?1410191421455 HTTP/1.1" 200 4140
- - [09/Sep/2014:00:50:25 +0900] "GET /kibana/app/dashboards/noted.json?1410191425722 HTTP/1.1" 200 4101
- - [09/Sep/2014:00:50:29 +0900] "GET /kibana/app/dashboards/default.json?1410191429288 HTTP/1.1" 200 4140
- - [09/Sep/2014:00:50:30 +0900] "GET /kibana/app/dashboards/logstash.json?1410191430626 HTTP/1.1" 200 4674
```

システムログ

前処理後のデータはどうするの？



画像

このあたりの前処理後の話

HTML



Webページ

「ます」を形態素解析すると？

文字列	読み		活用の種類	活用形
彼女	カノジョ	彼女	代名詞	
は	ワ	は	副助詞	は
大学	ダイガク	大学	普通名詞	
で	デ	で	格助詞	で
勉強	ベンキョウ	勉強	さ変名詞	
し	シ	する	動詞	さ変
ます	マス	ます	助動詞	ます

テキスト

```
- - [09/Sep/2014:00:50:21 +0900] "GET /kibana/app/dashboards/default.json?1410191421455 HTTP/1.1" 200 414
- - [09/Sep/2014:00:50:25 +0900] "GET /kibana/app/dashboards/noted.json?1410191425722 HTTP/1.1" 200 410
- - [09/Sep/2014:00:50:29 +0900] "GET /kibana/app/dashboards/default.json?1410191429288 HTTP/1.1" 200
- - [09/Sep/2014:00:50:30 +0900] "GET /kibana/app/dashboards/logstash.json?1410191430626 HTTP/1.1" 200 414
```

システムログ

前処理後のデータ連携も一手間が必要

「彼女は大学で勉強します」を形態素解析すると？

文字列	読み	原形	品詞の種類	活用の種類	活用形
彼女	カノジョ	彼女	代名詞	人称	
は	ワ	は	助動詞	は	
大学	ダイガク	大学	普通名詞		
で	デ	で	格助詞	で	
勉強	ベンキョウ	勉強	さ変名詞		
し	シ	する	動詞	さ変	
ます	マス	ます	助動詞	ます	

Python Apps

HTML



Ruby Apps

Bash script

Alerting

- Nagios

Analytics

- MongoDB
- MySQL
- Jubatus

Archive

- Amazon S3

前処理以外はゆるふわっとシンプルにしたい

「彼女は大学で勉強します」を形態素解析すると？

文字列	読み	原形	品詞の種類	活用の種類	活用形
彼女	カノジョ	彼女	代名詞	人称	
は	ワ	は	副助詞	は	
大学	ダイガク	大学	普通名詞		
で	デ	で	格助詞	で	
勉強	ベンキョウ	勉強	さ変名詞		
し	シ	する	動詞	さ変	
ます	マス	ます	助動詞	ます	

Python Apps

HTML



Ruby Apps

Bash script

Alerting

- Nagios

Analytics

- MongoDB
- MySQL
- Jubatus

Archive

- Amazon S3

fluentdでデータフローをシンプルに

「彼女は大学で勉強します」を形態素解析すると？

文字列	読み	原形	品詞の種類	活用の種類	活用形
彼女	カノジョ	彼女	代名詞 人称		
は	ワ	は	副助詞 は		
大学	ダイガク	大学	普通名詞		
で	デ	で	格助詞 で		
勉強	ベンキョウ	勉強	さ変名詞		
し	シ	する	動詞	さ変	
ます	マス	ます	助動詞	ます	

Python Apps

HTML



Ruby Apps



fluentd



filter / buffer / routing

Bash script

Alerting

- Nagios

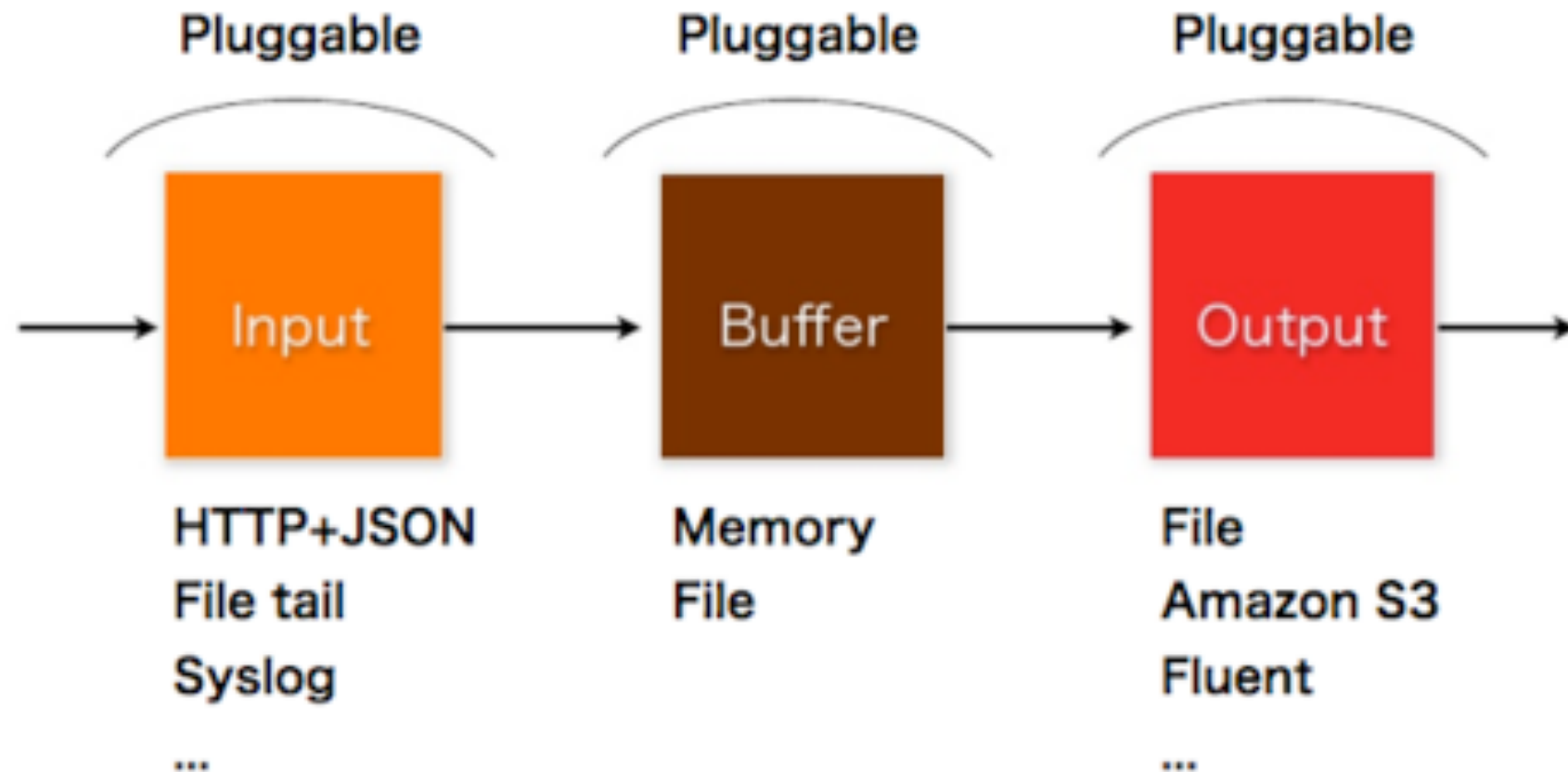
Analytics

- MongoDB
- MySQL
- Jubatus

Archive

- Amazon S3

Pluginで入力元/出力先を容易に対応可能



引用: Internal Architecture: Input -> Buffer -> Output

<http://blog.treasure-data.com/post/13047440992/fluentd-the-missing-log-collector-software>

fluentdをどうやって使うの？

Ruby

```
require 'fluent-logger'  
Fluent::Logger::FluentLogger.open(  
  nil, :host=>'localhost', :port=>24224  
)  
Fluent::Logger.post("myapp.access", {"key"=>"value"})
```

fluentd

<pre><source> type forward port 24224 </source></pre>	<pre><match myapp.access> type mongo host mongo.example.com capped capped_size 200m </match></pre>
---	--

まとめ

- 前処理自体は、頑張りましょう
- 前処理以外は、ゆるふわっといきましょう
- fluentdを使ってデータフローをシンプルに
 - Pluginで多種多様な入力元/出力先に対応可能
- 複数のアプリケーションでのデータ連携も容易に

そう、 fluentd ならね！