Apache Access Logs Import Integration

You can tail Apache logs from Treasure Agent (td-agent), to continuously import the access logs into the cloud.

- Installing td-agent
- Modifying td-agent.conf
- Confirming Data Import
- Next Steps

Installing td-agent

Install td-agent on your application servers. td-agent is a daemon program dedicated to the streaming upload of any kind of the time-series data. td-agent is a Treasure Data maintained open source project available under the Fluentd project. td-agent extends Fluentd with custom plugins for Treasure Data.

- Install Fluentd
- Installing Treasure Agent

Modifying td-agent.conf

Specify your authentication key by setting the `apikey` option. You can view your API key from the TD Console.

To modify the file, go to `/etc/td-agent/td-agent.conf`. Set the `apikey` option in your td-agent.conf file.

Where *YOUR_API_KEY* is your actual apikey string.

```
# Tailing the Apache Log
<source>
  type tail
  path /var/log/httpd-access.log
  pos_file /var/log/td-agent/httpd-access.pos
  tag td.production.access
  format apache2
</source>

# Treasure Data Input and Output
<match td.*.*>
  type tdlog
  endpoint api.treasuredata.com
  apikey YOUR_API_KEY
  auto_create_table
  buffer_type file
  buffer_path /var/log/td-agent/buffer/td
  use_ssl true
</match>
```

Restart your agent when the following lines are in place.

```
$ sudo /etc/init.d/td-agent restart
```

td-agent keeps tailing the log, buffers the log (*var/log/td-agent/buffer/td*), and automatically uploads the log into the cloud.

Confirming Data Import

Sending a SIGUSR1 signal flushes td-agent's buffer and the upload starts immediately.

```
# generate access logs
$ curl http://host:port/

# flush the buffer
$ kill -USR1 `cat /var/run/td-agent/td-agent.pid`
```

To confirm that your data has been uploaded successfully, issue the `td tables` command as follows.

```
$ td tables
+-----------+-----------+------+-----------+
| Database  | Table     | Type | Count     |
+-----------+-----------+------+-----------+
| production | access   | log  | 1         |
+-----------+-----------+------+-----------+
```

Check `/var/log/td-agent.log` if it's not working correctly. `td-agent:td-agent` must have permission to read the logs.

td-agent handles log-rotation. td-agent keeps a record of the last position of the log, ensuring that each line is read exactly one time even if the td-agent process goes down. However, because the information is kept in a file, the "exactly once" guarantee breaks down if the file becomes corrupted.

Next Steps

We offer a schema mechanism that is more flexible than that of traditional RDBMSs. For queries, we leverage the Hive Query Language.

- Schema Management
- Writing Hive Queries
- Programatic Access with REST API and its Bindings