mysqlsmom Documentation

Release 0.1.6

MCTW

Oct 31, 2018

Contents:

MysqlsMom

CHAPTER 1

Mysql Elasticsearch binlog es

Python mysqlsmom :

- python2.7
- redis
- *Mysql binlog-format=row*

2.1

```
pip install mysqlsmom
```

elasticsearch 2.45.4elasticsearch

```
pip install --upgrade elasticsearch==5.4
```

2.2

```
mom new test_mom/init_config.py -t init --force
```

```
└── test_mom
    └── init_config.py
```

```
vim ./test_mom/init_config.py  #
```

2.3

```
mom run -c ./test_mom/init_config.py
```

2.4

pip install mysqlsmom

git clone python mysqlsmom.py ./config/xxx.py

4.1 *binlog*

1. MySqlbinlogredis(binlog)

2.
```
mom new test_mom/binlog_config.py -t binlog --force
```

3. test_mom/binlog_config.py

```python
# coding=utf-8

STREAM = "BINLOG"
SERVER_ID = 99  # binlogSERVER_ID
SLAVE_UUID = __name__

# binlogMySql
BINLOG_CONNECTION = {
    'host': '127.0.0.1',
    'port': 3306,
    'user': 'root',
    'passwd': ''
}

# es
NODES = [{"host": "127.0.0.1", "port": 9200}]

TASKS = [
    {
        "stream": {
            "database": "test_db",  # [table]
            "table": "person"  # binlog
        },
        "jobs": [
            {
                "actions": ["insert", "update"],
                "pipeline": [
```

```
                    {"only_fields": {"fields": ["id", "name", "age"]}},  # eses
                    {"set_id": {"field": "id"}}  # es_id id()
                ],
                "dest": {
                    "es": {
                        "action": "upsert",
                        "index": "test_index",  #  index
                        "type": "test",          #  type
                        "nodes": NODES
                    }
                }
            }
        ]
    }
]
```

4.
```
mom run -c ./test_mom/binlog_config.py
```

elasticsearch

MySql

MySqles

4.2

Mysql `update_time` *binlog*

1.
```
mom new test_mom/cron_config.py -t cron --force
```

2. test_mom/cron_config.py

```python
# coding=utf-8

STREAM = "CRON"

#
CONNECTION = {
    'host': '127.0.0.1',
    'port': 3306,
    'user': 'root',
    'passwd': ''
}

# redis
REDIS = {
    "host": "127.0.0.1",
    "port": 6379,
    "db": 0,
    "password": "password",  #
}

#  BULK_SIZE elasticsearch1
BULK_SIZE = 1
```
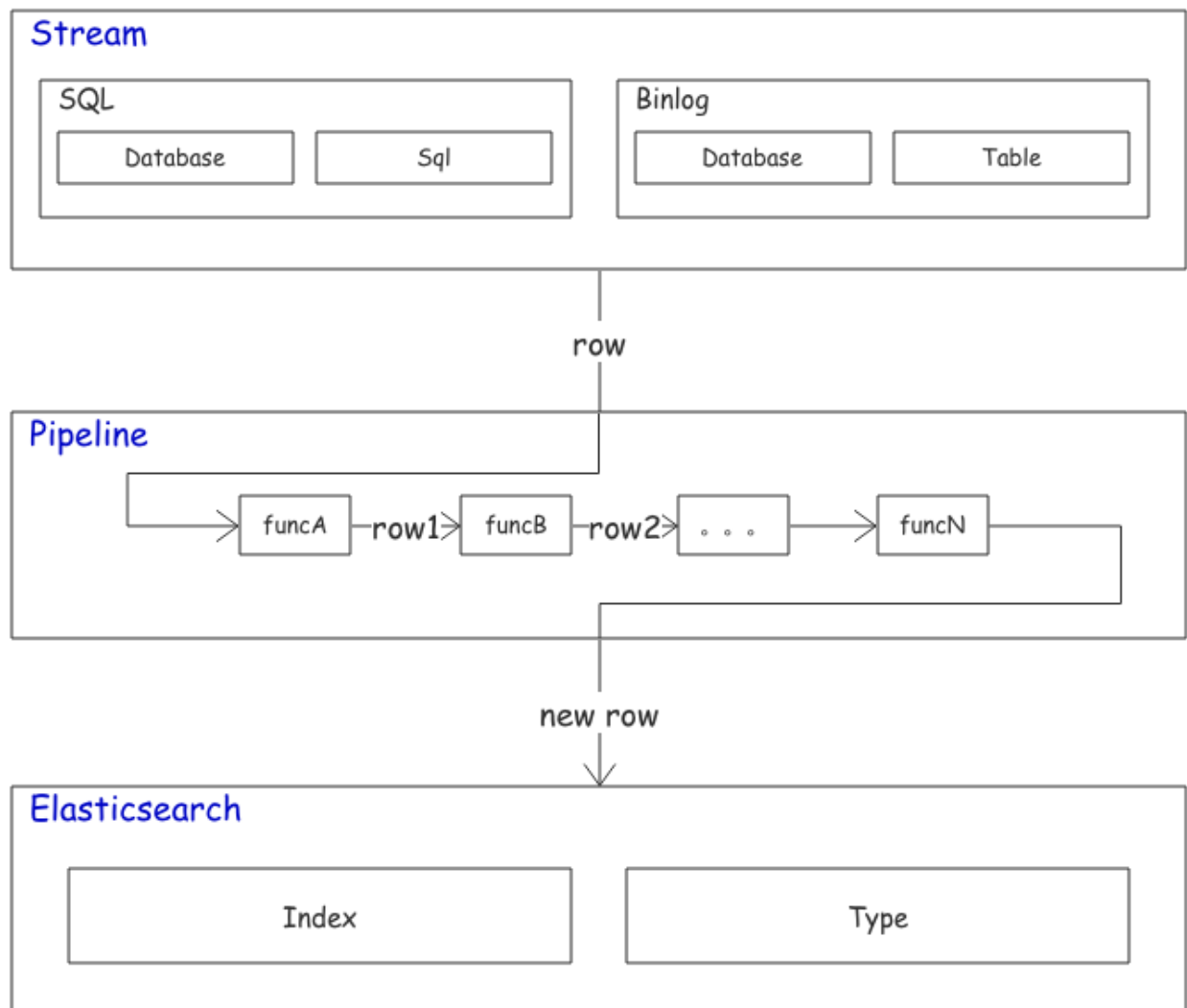
```python
# elasticsearch
NODES = [{"host": "127.0.0.1", "port": 9200}]

TASKS = [
    {
        "stream": {
            "database": "test_db",  # sql
            "sql": "select id, name from person where update_time >= ?",  # sql
→elasticsearch
            "seconds": 10,  #  seconds ,
            "init_time": "2018-08-15 18:05:47"  #
        },
        "jobs": [
            {
                "pipeline": [
                    {"set_id": {"field": "id"}}  #  id  es id
                ],
                "dest": {
                    "es": {
                        "action": "upsert",
                        "index": "test_index",   #  index
                        "type": "test"           #  type
                    }
                }
            }
        ]
    }
]
```

3.
```
mom run -c ./test_mom/cron_config.py
```

CHAPTER 5

CHAPTER 6

Mysqlsmom

Mysqlsmom

- *row_handlers.py* Mysql
- *row_filters.py*
- *config/*

Python Github issue QQ

6.1

```
...
TASKS = [
    # 1
    {
        "stream": {
            "database": "1",
            "table": "1"
        },
        "jobs": [...]
    }
    # 2
    {
        "stream": {
            "database": "2",
            "table": "2"
        },
        "jobs": [...]
    }
]
```

Mysql Connection

6.2

```
...
TASKS = [
    {
        "stream": {...},
        "jobs": [
            {
                "actions": [...],
                "pipeline": [...],
                "dest": [
                    # 1
                    {
                        "es": {"action": "upsert", "index": "1", "type": "1", "nodes
↪": NODES},
                    },
                    # 2
                    {
                        "es": {"action": "upsert", "index": "2", "type": "2", "nodes
↪": NODES},
                    }
                ]
            }
        ]
    },
    ...
]
```

pipeline

```
...
TASKS = [
    {
        "stream": {...},
        "jobs": [
            {
                "actions": {...},
                "pipeline": [...],  #
                "dest": {"es": {"index": "1", ...}}  # 1
            },
            {
                "actions": {...},
                "pipeline": [...],  # pipeline
                "dest": {"es": {"index": "2", ...}}  # 2
            }
        ]
    }
]
```

- *TASKS*

- *jobs*

- *dest* es

6.3

Mysql **pipeline**

```
"pipeline": [
  {"only_fields": {"fields": ["id", "name"]}},  #  id  name
    {"set_id": {"field": "id"}}  #  id es_id
]
```

6.4

Mysql elasticsearch

```
"pipeline": [
    # namename1age age1
  {"replace_fields": {"name": ["name1"], "age": ["age1"]}},
    {"set_id": {"field": "id"}}
]
```

pipeline

```
"pipeline": [
    #  name  name1
  {"replace_fields": {"name": ["name1"]}},
    #  age  age1
    {"replace_fields": {"age": ["age1"]}},
    {"set_id": {"field": "id"}}
]
```

es

name_default defaultname_raw name

```
"pipeline": [
  {"replace_fields": {"name": ["name_default", "name_raw"]}},
    {"set_id": {"field": "id"}}
]
```

es mappings name fields

6.5

Mysql "aaa|bbb|ccc": ["aaa", "bbb", "ccc"]

```
"pipeline": [
  # tags "aaa|bbb|ccc" tags  `|`
  {"split": {"field": "tags", "flag": "|"}},
    {"set_id": {"field": "id"}}
]
```

6.6

binlog elasticsearch

```
TASKS = [
    {
        "stream": {
            "database": "test_db",
            "table": "person"
        },
        "jobs": [
            #
            {
                "actions": ["insert", "update"],
                "pipeline": [
                    {"set_id": {"field": "id"}} # id es _id
                ],
                "dest": {
                    "es": {
                        "action": "upsert",
                        ...
                    }
                }
            },
            #
            {
                "actions": ["delete"], # binlog
                "pipeline": [{"set_id": {"field": "id"}}], # _id
                "dest": {
                    "es": {
                        "action": "delete", # es
                        ... # index type
                    }
                }
            }
        ]
    },
    ...
]
```

6.7

CHAPTER 7

row_handlers.py pipeline

pipeline Mysql pipeline es

binlog

1. `my_handlers.py`

2. `# CUSTOM_ROW_HANDLERS = "./my_handlers.py"`

3. `my_handlers.py` pipeline

7.1

`my_handlers.py`

```python
# coding=utf-8
import copy

#  my_field  my_value
def my_func(row, my_field, my_value):
    """
     row: {"id": 123, "name": ""}
     my_field: id, name
     my_value:  my_field  my_value
    """
    new_row = copy.deepcopy(row) #

    #
    new_row[my_field] = my_value

    return new_row # es

# 2 _id  _id  "123"
def my_func2(row):
    row = copy.deepcopy(row)
```

```
    if not row["_id"]:
        row["_id"] = "123"
    return row


# 3...
```

pipeline

```
...
"pipeline": [
            {
                "my_func": #
                #
                # row
                #  name   "Jack"
                {
                    "my_field": "name",
                    "my_value": "Jack"
                }
            },
            {"set_id": {"field": "id"}}  #  id  es  _id
        ],
...
```

8.1 es

```
pip install --upgrade elasticsearch==6.4.2
```

```
Could not find a version that satisfies the requirement elasticsearch==6.4.2 (from
→versions: 0.4.1, 0.4.2, 0.4.3, 0.4.4, 0.4.5, 1.0.0, 1.1.0, 1.1.1, 1.2.0, 1.3.0, 1.4.
→0, 1.5.0, 1.6.0, 1.7.0, 1.8.0, 1.9.0, 2.0.0, 2.1.0, 2.2.0, 2.3.0, 2.4.0, 2.4.1, 5.0.
→0, 5.0.1, 5.1.0, 5.2.0, 5.3.0, 5.4.0, 5.5.0, 5.5.1, 5.5.2, 5.5.3, 6.0.0, 6.1.1, 6.2.
→0, 6.3.0, 6.3.1)
No matching distribution found for elasticsearch==6.4.2
```

```
pip install --upgrade elasticsearch==6.3.1
```

8.2

1. ```
   pip install virtualenv
   ```

2. ```
   virtualenv ./venv
   ```

3. ```
   . ./venv/bin/activate
   ```

4. mysqlsmom

   ```
   pip install mysqlsmom
   pip install --upgrade elasticsearch==5.4   #
   ```

5. ```
   mom new ...
   ```

6. ```
   deactivate
   ```

mysqlsmom

8.3

1.

2. 2.1 IPMysqlIPIPbinlog

1.

issues24QQ: 569069847

CHAPTER 11

Indices and tables

- genindex
- modindex
- search