



MyDumper

PERCONA

Databases run better with Percona



Simplest way to masquerade logical backups

using MyDumper

David Ducos
Maintainer

How am I?

How did I end up here?

Why do I love it?

Agenda

Introduction

Masquerade backups

MyDumper configuration

Add functions to the code



Introduction

What's MyDumper?

Why do we need to masquerade backups?

Introduction – MyDumper

What is MyDumper?

MyDumper is a MySQL Multithreaded Logical Backup Tool.

MyDumper? The project name

mydumper? takes backups

myloader? restore backups

MyDumper is Open Source and maintained by the community, it is not a Percona, MariaDB or MySQL product.

Introduction – Why Data Masking?

Ensure Quality for testings

Data privacy and Security

Safeguarding sensitive information (GDPR & HIPAA)

How to obfuscate personally identifiable information?



Masquerade backups

What do we need?

Masquerade backups – Definition

What do we need to perform a data masking?

- Database
- Table
- Column

Masquerade backups – Example

Example from `sakila`.`customer`:

```
`customer_id` int unsigned,  
`store_id` tinyint unsigned,  
`first_name` varchar(45),  
`last_name` varchar(45),  
`email` varchar(50) DEFAULT NULL,  
`address_id` smallint unsigned,  
`active` tinyint(1) DEFAULT '1',  
`create_date` datetime,  
`last_update` timestamp,
```

first_name = MARY => PATRICIA

last_name = SMITH => JOHNSON

email = MARY.SMITH@sakilacustomer.org

=> PATRICIA.JOHNSON@sakilacustomer.org

Masquerade backups

What are we missing?

Masquerade backups

The transformation function

Masquerade backups – How mydumper works?

1. mydumper reads the data
2. Save it on disk

Masquerade backups – How mydumper works?

1. mydumper reads the data
2. Applies a function
3. Save it on disk



Configuration

How we configure mydumper?

Configuration - Syntax

```
[`database_A`.`table_A`]
```

```
`column1`=<function> <parameters>
```

```
`column2`=<function> <parameters>
```

```
`column3`=<function> <parameters>
```

```
[`database_B`.`table_B`]
```

```
`column1`=<function> <parameters>
```

```
`column2`=<function> <parameters>
```

```
`column3`=<function> <parameters>
```


Configuration - Example

```
[`sbtest`.`sbtest1`]
```

```
`k`=random_int
```

```
`c`=random_string
```

```
`pad`=random_uuid
```

```
[`sakila`.`customer`]
```

```
`first_name` = random_format <file first_name.txt>
```

```
`last_name` = random_format <file last_name.txt>
```

```
`email`=constant concat(first_name,concat(".",concat(last_name,"@sakilacustomer.or
```

Recommendation: Use `--defaults-extra-file`

Configuration - Available Functions

Basic

- random_string / random_string_with_mem
- random_int / random_int_with_mem
- random_uuid / random_uuid_with_mem

Advance

- constant
- apply
- random_format

Configuration - Basic functions

- `random_int / random_int_with_mem`
- `random_string / random_string_with_mem`
- `random_uuid / random_uuid_with_mem`

Configuration – Basic functions – Example

```
[`sbtest`.`sbtest1`]
```

```
`k`=random_int
```

```
`c`=random_string
```

```
`pad`=random_uuid
```

```
INSERT INTO `sbtest1` (`id`,`k`,`c`,`pad`) VALUES(1,  
18,  
"83868641912-28773972837-60736120486-...  
"67847967377-48000963322-62604785301...  
)
```

```
INSERT INTO `sbtest1` (`id`,`k`,`c`,`pad`) VALUES(1,  
20,  
"vockffpyzgxlmxoqgyzifpjljyabstunhyzodqm...  
"4a6aaad401d-2125f34b97c-cf426e28272...  
)
```

Configuration - Advance functions

- constant
- apply
- random_format

Configuration - constant

```
[`sakila`.`customer`]
```

```
`email`=constant
```

```
concat(first_name,concat(".",concat(last_name,"@sakilacustomer.org"))))
```

```
INSERT INTO `customer` VALUES(1,1,  
"MARY","SMITH",  
"MARY.SMITH@sakilacustomer.org",  
5,1,"2006-02-14 22:04:36",  
"2006-02-15 04:57:20")
```

```
INSERT INTO `customer` VALUES (1,1,  
"MARY","SMITH",  
concat(first_name,concat(".",concat(last_name,"  
@sakilacustomer.org")))),  
5,1,"2006-02-14 22:04:36",  
"2006-02-15 04:57:20")
```

Configuration - apply

```
[`sakila`.`customer`]
```

```
`last_update`=apply 'date_add(' ', interval 1 year)'
```

```
INSERT INTO `customer` VALUES(1,1,  
"MARY","SMITH",  
"MARY.SMITH@sakilacustomer.org",  
5,1,"2006-02-14 22:04:36",  
"2006-02-15 04:57:20")
```

```
INSERT INTO `customer` VALUES (1,1,  
"MARY","SMITH",  
"MARY.SMITH@sakilacustomer.org",  
5,1,"2006-02-14 22:04:36",  
date_add("2006-02-15 04:57:20", interval 1 year))
```

Configuration - random_format

- <FILE filename>

```
`email`=random_format <file first_name.txt>'<file last_name.txt>'@'<file domains.txt>  
jo.coy@google.com
```

- <STRING length>

```
`pad`=random_format <string 9>'-'<string 9>'-'<string 9>'-'<string 9>'-'<string 9>  
voclgakfl-ebdhakfbr-kyoupwqh-avdzkgnhl-wosjfyckz
```

- <NUMBER length>

```
`phone`=random_format '+1 ('<number 3>)' <number 3>'-'<number 4>  
+1(827)734-0185
```




Add functions

What happen if you have a special use case?

Add new functions

- Function that is going to be executed
- Function that will parse the parameters

Add new functions – Example

in src/mydumper/mydumper_masquerade.c

get_function_pointer_for::

```
if (g_str_has_prefix(function_char, "random_format")) {  
    return &random_format_function;
```

init_function_pointer::

```
if (g_str_has_prefix(value, "random_format")) {  
    parse_value(fp, g_strdup(&(fp->value[14])));
```

Conclusions

Perform data masking is a must

MyDumper makes it simple!



Thank You!

Simplest way to masquerade logical backups with MyDumper

Questions?