

概述：

Ranger maskk 主要是用来对敏感信息进行保护，不让显示敏感信息。

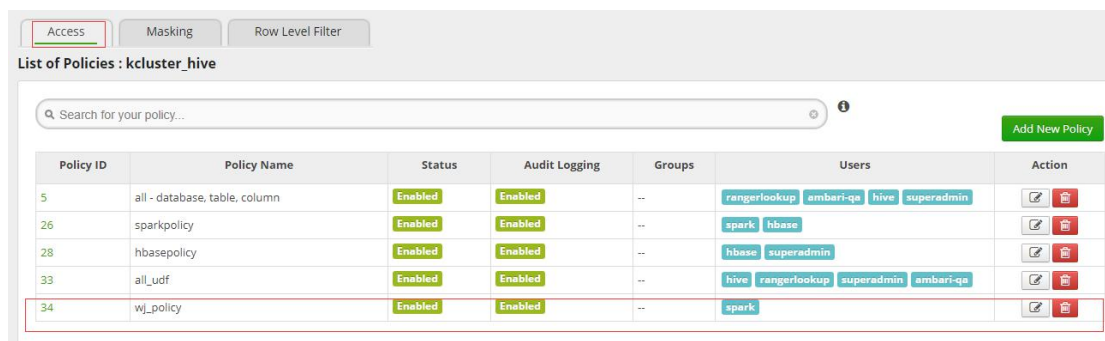
操作 mask 特性

使用 spark 用户，让 spark 用户对 hive 的 wj 库，t1 表的 name 字段产生遮罩，没法展现：

（1）通过 Ambari 页面进入 Ranger UI 管理页面



（2）让 Spark 用户对 wj 库具有所有权限吧，应该读写权限就行，这里贪省事了



Policy Name * enabled

databases * include

table * include

Hive Column * include

Audit Logging YES

Select Group	Select User	Permissions	Delegate Admin	
<input type="text" value="Select Group"/>	<input type="text" value="spark"/>	<div> select update Create Drop Alter </div> <div> Index Lock All Read Write </div>	<input type="checkbox"/>	<input type="checkbox"/>
+				

这里缓存 spark 票据，访问

jdbc:hive2://10.110.18.48:10000/default;principal=hive/idap-agent-48.idap.com@IDAP.COM 是
可以看到 wj 库下面 t1,t2,t3 的

```
2,230 rows selected (8.511 seconds)
0: jdbc:hive2://10.110.18.48:10000/default> select * from t2 limit 1;

1 row selected (0.199 seconds)
```

(3) 然后配置遮罩层

Access Masking Row Level Filter

List of Policies : kcluster_hive

Policy Details :

Policy Type **Masking**

Policy ID **35**

Policy Name * wj_mask_policy

enabled

Hive Database *

x wj

Hive Table *

x t1

Hive Column *

x name

Audit Logging

YES

Select Group	Select User	Access Types	Select Masking Option	
Select Group	x spark	select	Redact	

注意：这里有很多遮罩操作，这里我会一一测试：

测试 Redact

Select Masking Option

- ☒ Redact
- ☐ Partial mask: show last 4
- ☐ Partial mask: show first 4
- ☐ Hash
- ☐ Nullify
- ☐ Unmasked (retain original value)
- ☐ Date: show only year
- ☐ Custom

选择 Redact，进入 10.110.18.48 测试环境

由于我们是对 wj 库的 t1 表的名字字段遮罩的，我们就查看 t1 表：

```
0: jdbc:hive2://10.110.18.48:10000/default> select * from t1;
+-----+-----+-----+-----+
| t1.id | t1.name | t1.month | t1.day |
+-----+-----+-----+-----+
| 1     | xxx     | 6        | 3      |
| 2     | xx      | 6        | 3      |
| 3     | xx      | 6        | 3      |
| 4     | xx      | 6        | 3      |
| 5     | xx      | 6        | 3      |
| 5     | xx      | 6        | 3      |
+-----+-----+-----+-----+
```

可以看到 Redact 操作对 name 字段打码了，我们无法知道 name 信息。Redact 是编辑的意思。

测试 Partial mask: show last 4

Select Masking Option

☐ Redact

☒ Partial mask: show last 4

☐ Partial mask: show first 4

☐ Hash

☐ Nullify

☐ Unmasked (retain original value)

☐ Date: show only year

☐ Custom

```
0: jdbc:hive2://10.110.18.48:10000/default> select * from t1;
+-----+-----+-----+-----+
| t1.id | t1.name | t1.month | t1.day |
+-----+-----+-----+-----+
| 1 | xxxxxxxxxxxxxxxxxxxxxxxaosi | 6 | 3 |
| 2 | xxxxxxxxxxxxxxxxxxxxxxxaosi | 6 | 3 |
| 3 | xxxxxxxxxxxxxxxxxxxxxxxaosi | 6 | 3 |
| 4 | xxxxxxxxxxxxxxxxxxxxxxxaosi | 6 | 3 |
| 5 | xxxxxxxxxxxxxxxxxxxxxxxaosi | 6 | 3 |
+-----+-----+-----+-----+
5 rows selected (1.514 seconds)
```

可以看到，它只显示后面四个字节，前面字节打码了。

测试 hash

Select Masking Option

☒ Redact

☐ Partial mask: show last 4

☐ Partial mask: show first 4

☒ Hash

☐ Nullify

☐ Unmasked (retain original value)

☐ Date: show only year

☐ Custom

```
0: jdbc:hive2://10.110.18.48:10000/default> select * from wj.tl;
+-----+-----+-----+-----+
| tl.id | tl.name | tl.month | tl.day |
+-----+-----+-----+-----+
| 1 | 16a300c9379b618919b0679869bb6f50 | 6 | 3 |
| 2 | 16a300c9379b618919b0679869bb6f50 | 6 | 3 |
| 3 | 16a300c9379b618919b0679869bb6f50 | 6 | 3 |
| 4 | 16a300c9379b618919b0679869bb6f50 | 6 | 3 |
| 5 | d41d8cd98f00b204e9800998ecf8427e | 6 | 3 |
+-----+-----+-----+-----+
5 rows selected (0.519 seconds)
```

测试 Partial mask:show first 4

Select Masking Option

☐ Redact

☐ Partial mask: show last 4

☒ Partial mask: show first 4

☐ Hash

☐ Nullify

☐ Unmasked (retain original value)

☐ Date: show only year

☐ Custom

```
0: jdbc:hive2://10.110.18.48:10000/default> select * from t1;
```

t1.id	t1.name	t1.month	t1.day
1	yanzxxxxxxxxxxxxxxxxxxxxxx	6	3
2	yanzxxxxxxxxxxxxxxxxxxxxxx	6	3
3	yanzxxxxxxxxxxxxxxxxxxxxxx	6	3
4	yanzxxxxxxxxxxxxxxxxxxxxxx	6	3
5	yanzxxxxxxxxxxxxxxxxxxxxxx	6	3

```
5 rows selected (1.09 seconds)
```

测试 Nulify

Select Masking Option

☒ Redact
 ☐ Partial mask: show last 4
 ☐ Partial mask: show first 4
 ☐ Hash
 ☒ Nullify
 ☐ Unmasked (retain original value)
 ☐ Date: show only year
 ☐ Custom

```
0: jdbc:hive2://10.110.18.48:10000/default> select * from wj.t1;
```

t1.id	t1.name	t1.month	t1.day
1	NULL	6	3
2	NULL	6	3
3	NULL	6	3
4	NULL	6	3
5	NULL	6	3

```
5 rows selected (0.373 seconds)
```

测试 Unmasked(retain original value)

Select Masking Option

☒ Redact

☐ Partial mask: show last 4

☐ Partial mask: show first 4

☐ Hash

☐ Nullify

☒ Unmasked (retain original value)

☐ Date: show only year

☐ Custom

```

Transaction Isolation: TRANSACTION_REPEATABLE_READ
0: jdbc:hive2://10.110.18.48:10000/default> select * from wj.t1;
+-----+-----+-----+-----+
| t1.id | t1.name          | t1.month | t1.day |
+-----+-----+-----+-----+
| 1     | yanzhouwuwangnigulasizhaosi | 6        | 3      |
| 2     | yanzhouwuwangnigulasizhaosi | 6        | 3      |
| 3     | yanzhouwuwangnigulasizhaosi | 6        | 3      |
| 4     | yanzhouwuwangnigulasizhaosi | 6        | 3      |
| 5     | yanzhouwuwangnigulasizhaosi | 6        | 3      |
+-----+-----+-----+-----+
5 rows selected (0.317 seconds)

```

发现一个意外，虽然是不打码了，但是如果 name 字段包含中文，这里就显示空了。

测试 date:show only year

Select Masking Option

☒ Redact

☐ Partial mask: show last 4

☐ Partial mask: show first 4

☐ Hash

☐ Nullify

☐ Unmasked (retain original value)

☒ Date: show only year

☐ Custom

Policy Details :

Policy Type	Masking
Policy Name *	testyear enabled
Hive Database *	x wj
Hive Table *	x t1
Hive Column *	x birthday
Audit Logging	YES

alter table t1 add birthday date;

update t1 set birthday=cast('1989-11-14' as date) where id=1;

然后切换回 spark 票据

```
Transaction Isolation: TRANSACTION_READ_COMMITTED
0: jdbc:hive2://10.110.18.48:10000/default> select * from wj.t1;
+-----+-----+-----+-----+-----+
| t1.id | t1.name          | t1.birthday | t1.month | t1.day |
+-----+-----+-----+-----+-----+
| 1     | xxxxxxxxxxxxxxxx | NULL       | 6        | 3      |
| 2     | xxxxxxxxxxxxxxxx | NULL       | 6        | 3      |
| 3     | xxxxxxxxxxxxxxxx | NULL       | 6        | 3      |
| 4     | xxxxxxxxxxxxxxxx | NULL       | 6        | 3      |
| 5     | 往年下雪xxxxxx大 | NULL       | 6        | 3      |
+-----+-----+-----+-----+-----+
5 rows selected (1.112 seconds)
```

这里特效没有显示出来，birthday 应该显示年才对，但所有的都是空

在这个测试里面，name 字段采用了 Mask.custom 特性，name 打码了，但还是不支持中文。

操作 Row level Filter

(1)添加一个行级别过滤

Policy Details :

Policy Type	Row Level Filter
Policy ID	37
Policy Name *	<input type="text" value="wj_row_test"/> enabled
Hive Database *	<input type="text" value="x wj"/>
Hive Table *	<input type="text" value="x t1"/>
Audit Logging	YES
Description	<input type="text"/>

Select Group	Select User	Access Types	Row Level Filter	
<input type="text" value="Select Group"/>	<input type="text" value="x spark"/>	<input type="button" value="select"/> <input type="button" value="edit"/>	<input type="text" value="id in(2,3,4)"/> <input type="button" value="edit"/>	<input type="button" value="x"/>
<input type="button" value="+"/>				

(2)切换 spark 用户票据

kinit -kt spark.headless.keytab spark-kcluster@IDAP.COM

(3)查看 t1 数据

```
Transaction Isolation: TRANSACTION_READ_COMMITTED
0: jdbc:hive2://10.110.18.48:10000/default> select * from wj.t1;
+-----+-----+-----+-----+-----+-----+
| t1.id | t1.name | t1.birthday | t1.month | t1.day | |
+-----+-----+-----+-----+-----+-----+
| 2 | yanzhouwangnigulasizhaosi | NULL | 6 | 3 | |
| 3 | yanzhouwangnigulasizhaosi | NULL | 6 | 3 | |
| 4 | yanzhouwangnigulasizhaosi | NULL | 6 | 3 | |
+-----+-----+-----+-----+-----+-----+
3 rows selected (0.989 seconds)
```

发现只有 id=2,3,4 的记录才可以查看

疑问

1、mask 对什么格式的 hive 表起作用

开始验证 text 格式表格

Policy Details :

Policy TypeMasking

Policy ID46

Policy Name *mask_text

enabled

Hive Database *wj

Hive Table *tb5text

Hive Column *name

Select Group	Select User	Access Types	Select Masking Option	
<div>Select Group</div>	<div>spark</div>	<div>select</div>	<div>Redact</div>	<div></div>
<div></div>				

现在对 name 字段打码了，然后再加行过滤：

Policy Details :

Policy Type	Row Level Filter
Policy ID	43
Policy Name *	<input type="text" value="row_text"/> enabled
Hive Database *	<input type="text" value="wj"/>
Hive Table *	<input type="text" value="tb5text"/>

Select Group	Select User	Access Types	Row Level Filter	
<input type="text" value="Select Group"/>	<input type="text" value="spark"/>	<input type="button" value="select"/> <input type="button" value="edit"/>	<input type="text" value="id=1"/> <input type="button" value="edit"/>	<input type="button" value="x"/>
<input type="button" value="+"/>				

总共 tb5text 有 4 条记录，然后用 text 查看：

```
0: jdbc:hive2://10.110.18.48:10000/default> select * from tb5text;
+-----+-----+-----+
| tb5text.id | tb5text.name | tb5text.age |
+-----+-----+-----+
| 1          | xx          | 28          |
+-----+-----+-----+
```

验证 AVRO 格式表：

Policy Details :

Policy Type **Masking**

Policy ID **40**

Policy Name * **enabled**

Hive Database *

Hive Table *

Hive Column *

Select Group	Select User	Access Types	Select Masking Option	
<input type="text" value="Select Group"/>	<input type="text" value="x spark"/>	<input type="button" value="select"/> <input type="button" value="edit"/>	<input type="button" value="Redact"/> <input type="button" value="edit"/>	<input type="button" value="x"/>
<input type="button" value="+"/>				

然后看行过滤

Policy Details :

Policy Type **Row Level Filter**

Policy ID **41**

Policy Name * **enabled**

Hive Database *

Hive Table *

Select Group	Select User	Access Types	Row Level Filter	
Select Group	spark	select	city='zibo'	
+				

然后现在看看结果：

```
0: jdbc:hive2://10.110.18.48:10000/default> select * from tb6avro;
+-----+-----+-----+
| tb6avro.id | tb6avro.name | tb6avro.city |
+-----+-----+-----+
| 1          | 0x          | zibo          |
+-----+-----+-----+
1 row selected (0.326 seconds)
```

行过滤没有问题，但是遮罩这里 name 字段有了乱码

测试 parquet 格式 hive 表：

Policy Details :

Policy Type	Masking
Policy ID	39
Policy Name *	m_parquet enabled
Hive Database *	wj
Hive Table *	tb7parquet
Hive Column *	name

Select Group	Select User	Access Types	Select Masking Option	
Select Group	spark	select	Redact	
+				

然后看 row filter:

Policy Details :

Policy Type **Row Level Filter**

Policy ID **42**

Policy Name *

enabled ☒

Hive Database *

Hive Table *

Select Group	Select User	Access Types	Row Level Filter	
<input type="text" value="Select Group"/>	<input type="text" value="x spark"/>	<input type="button" value="select"/> <input type="button" value="edit"/>	<input type="text" value="city='zibo'"/> <input type="button" value="edit"/>	<input type="button" value="x"/>
<input type="button" value="+"/>				

然后看看结果：

```
1 row selected (0.020 seconds)
0: jdbc:hive2://10.110.18.48:10000/default> select * from tb7parquet;
+-----+-----+-----+
| tb7parquet.id | tb7parquet.name | tb7parquet.city |
+-----+-----+-----+
| 1             | 0x              | zibo             |
+-----+-----+-----+
1 row selected (0.583 seconds)
```

2、Row lever Filter 对什么格式的 hive 表起作用

```
0: jdbc:hive2://10.110.18.48:10000/default> select * from tb5text;
+-----+-----+-----+
| tb5text.id | tb5text.name | tb5text.age |
+-----+-----+-----+
| 1          | xx           | 28          |
+-----+-----+-----+
```

```
1 row selected (0.020 seconds)
0: jdbc:hive2://10.110.18.48:10000/default> select * from tb7parquet;
+-----+-----+-----+
| tb7parquet.id | tb7parquet.name | tb7parquet.city |
+-----+-----+-----+
| 1             | 0x              | zibo             |
+-----+-----+-----+
1 row selected (0.583 seconds)
```

```
0: jdbc:hive2://10.110.18.48:10000/default> select * from tb6avro;
+-----+-----+-----+
| tb6avro.id | tb6avro.name | tb6avro.city |
+-----+-----+-----+
| 1          | 0x           | zibo          |
+-----+-----+-----+
1 row selected (0.326 seconds)
```

可见 row filter 对所有的格式都支持。

3、Mask 对中文支持不够好？

```
0: jdbc:hive2://10.110.18.48:10000/default> select * from tb6avro;
+-----+-----+-----+
| tb6avro.id | tb6avro.name | tb6avro.city |
+-----+-----+-----+
| 1          | 0x           | zibo          |
+-----+-----+-----+
1 row selected (0.326 seconds)
```

```
1 row selected (0.326 seconds)
0: jdbc:hive2://10.110.18.48:10000/default> select * from tb7parquet;
+-----+-----+-----+
| tb7parquet.id | tb7parquet.name | tb7parquet.city |
+-----+-----+-----+
| 1             | 0x              | zibo             |
+-----+-----+-----+
1 row selected (0.583 seconds)
```

Name 里面包含有汉字，这里显示有乱码。

就算切换到 hive 票据，这里的 name 字段也看不到，说明 masking 对 name 字段有副作用，凡是包含有汉字的，都会破坏数据结构。