Django-MySQL Documentation

Release 3.2.0

Adam Johnson

Aug 19, 2019

Contents

Django-MySQL is a non-inventively named package that helps you use some MySQL/MariaDB-specific features in the world of Django.

If you're new, check out the *Exposition* to see all the features in action, or get started with *Requirements and Installation*. Otherwise, take your pick:

CHAPTER 1

Exposition

Every feature in whistle-stop detail.

1.1 Checks

Extra checks added to Django's check framework to ensure your Django and MySQL configurations are optimal.

```
$ ./manage.py check
?: (django_mysql.W001) MySQL strict mode is not set for database connection 'default'
...
```

Read more

1.2 QuerySet Extensions

Django-MySQL comes with a number of extensions to `QuerySet` that can be installed in a number of ways - e.g. adding the `QuerySetMixin` to your existing `QuerySet` subclass.

1.2.1 Approximate Counting

`SELECT COUNT(*) ...` can become a slow query, since it requires a scan of all rows; the `approx_count` functions solves this by returning the estimated count that MySQL keeps in metadata. You can call it directly:

```
Author.objects.approx_count()
```

Or if you have pre-existing code that calls `count()` on a `QuerySet` you pass it, such as the Django Admin, you can set the `QuerySet` to do try `approx_count` first automatically:

```
qs = Author.objects.all().count_tries_approx()
# Now calling qs.count() will try approx_count() first
```

Read more

1.2.2 Query Hints

Use MySQL's query hints to optimize the SQL your `QuerySets` generate:

```
Author.objects.straight_join().filter(book_set__title__startswith="The ")
# Does SELECT STRAIGHT_JOIN ...
```

Read more

1.2.3 'Smart' Iteration

Sometimes you need to modify every single instance of a model in a big table, without creating any long running queries that consume large amounts of resources. The 'smart' iterators traverse the table by slicing it into primary key ranges which span the table, performing each slice separately, and dynamically adjusting the slice size to keep them fast:

```
# Some authors to fix
bad_authors = Author.objects.filter(address="Nowhere")

# Before: bad, we can't fit all these in memory
for author in bad_authors.all():
    pass

# After: good, takes small dynamically adjusted slices, wraps in atomic()
for author in bad_authors.iter_smart():
    author.address = ""
    author.save()
    author.send_apology_email()
```

Read more

1.2.4 Integration with pt-visual-explain

For interactive debugging of queries, this captures the query that the `QuerySet` represents, and passes it through `EXPLAIN` and `pt-visual-explain` to get a visual representation of the query plan:

```
>>> Author.objects.all().pt_visual_explain()
Table scan
rows           1020
+- Table
   table          myapp_author
```

Read more

1.2.5 MySQL `HANDLER` API

MySQL's `HANDLER` commands give simple NoSQL-style read access to rows faster than normal SQL queries, with the ability to perform index lookups or page-by-page scans. This extension adds an ORM-based API for handlers:

```
with Author.objects.handler() as handler:
    for author in handler.iter(chunk_size=1000):
        author.send_apology_email()
```

Read more

1.3 Model Fields

Fields that use MySQL-specific features!

1.3.1 JSON Field

Implements MySQL 5.7+'s JSON data type for storing arbitrary JSON data:

```
class APIResponse(Model):
    url = models.CharField(max_length=200)
    data = JSONField()
```

```
>>> APIResponse.objects.create(url='/api/twoots/1/', data={
    'id': '123',
    'message': 'Loving #django and #mysql',
    'coords': [34.4, 56.2]
})
>>> APIResponse.objects.filter(data__coords__0=34.4)
[<APIResponse: /api/twoots/1/>]
```

Read more

1.3.2 Dynamic Columns Field

Use MariaDB's Dynamic Columns for storing arbitrary, nested dictionaries of values:

```
class ShopItem(Model):
    name = models.CharField(max_length=200)
    attrs = DynamicField()
```

```
>>> ShopItem.objects.create(name='Camembert', attrs={'smelliness': 15})
>>> ShopItem.objects.create(name='Brie', attrs={'smelliness': 5, 'squishiness': 10})
>>> ShopItem.objects.filter(attrs__smelliness_INTEGER__gte=10)
[<ShopItem: Camembert>]
```

Read more

1.3.3 List Fields

Two field classes that allow you to store lists of items in a comma-separated string:

```
class Person(Model):
    name = CharField(max_length=32)
    post_nominals = ListTextField(
```

(continues on next page)

(continued from previous page)

```
        base_field=CharField(max_length=32)
    )
```

```
>>> Person.objects.filter(post_nominals__contains='PhD')
[<Person: Horatio>, <Person: Severus>]
```

Read more

1.3.4 Set Fields

Two field classes that allow you to store sets of items in a comma-separated string:

```
class Post(Model):
    name = CharField(max_length=32)
    tags = SetTextField(
        base_field=CharField(max_length=10)
    )
```

```
>>> Post.objects.create(name='First post', tags={'thoughts', 'django'})
>>> Post.objects.filter(tags__contains='django')
[<Post: First post>]
```

Read more

1.3.5 EnumField

A field class for using MySQL's `ENUM` type, which allows strings that are restricted to a set of choices to be stored in a space efficient manner:

```
class BookCover(Model):
    color = EnumField(choices=['red', 'green', 'blue'])
```

Read more

1.3.6 Resizable Text/Binary Fields

Django's `TextField` and `BinaryField` fields are fixed at the MySQL level to use the maximum size class for the `BLOB` and `TEXT` data types - these fields allow you to use the other sizes, and migrate between them:

```
class BookBlurb(Model):
    blurb = SizedTextField(size_class=3)
    # Has a maximum length of 16MiB, compared to plain TextField which has
    # a limit of 4GB (!)
```

Read more

1.3.7 BIT(1) Boolean Fields

Some database systems, such as the Java Hibernate ORM, don't use MySQL's `bool` data type for storing boolean flags and instead use `BIT(1)`. This field class allows you to interact with those fields:

```
class HibernateModel(Model):
    some_bool = Bit1BooleanField()
    some_nullable_bool = NullBit1BooleanField()
```

Read more

1.4 Field Lookups

ORM extensions to built-in fields:

```
>>> Author.objects.filter(name__sounds_like='Robert')
[<Author: Robert>, <Author: Rupert>]
```

Read more

1.5 Aggregates

MySQL's powerful `GROUP_CONCAT` statement is added as an aggregate, allowing you to bring back the concatenation of values from a group in one query:

```
>>> author = Author.objects.annotate(
...     book_ids=GroupConcat('books__id')
... ).get(name="William Shakespeare")
>>> author.book_ids
"1,2,5,17,29"
```

Read more

1.6 Database Functions

MySQL-specific database functions for the ORM:

```
>>> Author.objects.annotate(
...     full_name=ConcatWS('first_name', 'last_name', separator=' ')
... ).first().full_name
"Charles Dickens"
```

Read more

1.7 Migration Operations

MySQL-specific operations for django migrations:

```
from django.db import migrations
from django_mysql.operations import InstallPlugin


class Migration(migrations.Migration):
```

(continues on next page)

(continued from previous page)

```
    dependencies = []

    operations = [
        InstallPlugin("metadata_lock_info", "metadata_lock_info.so")
    ]
```

Read more

1.8 Cache

An efficient backend for Django's cache framework using MySQL features:

```
cache.set("my_key", "my_value")  # Uses only one query
cache.get_many(["key1", "key2"])  # Only one query to do this too!
cache.set("another_key", some_big_value)  # Compressed above 5kb by default
```

Read more

1.9 Locks

Use MySQL as a locking server for arbitrarily named locks:

```
with Lock("ExternalAPI", timeout=10.0):
    do_some_external_api_stuff()
```

Read more

1.10 Status

Easy access to global or session status variables:

```
if global_status.get('Threads_running') > 100:
    raise BorkError("Server too busy right now, come back later")
```

Read more

1.11 Management Commands

`dbparams` helps you include your database parameters from settings in commandline tools with `dbparams`:

```
$ mysqldump $(python manage.py dbparams) > dump.sql
```

`fix_datetime_columns` helps you fix your `DateTimeFields` that don't have microseconds after an upgrade to MySQL 5.6+/MariaDB 5.3+:

```
$ python manage.py fix_datetime_columns
ALTER TABLE `app1_table1`
    MODIFY COLUMN `created_time` datetime(6) DEFAULT NULL;
```

Read more

1.12 Utilities

Fingerprint queries quickly with the `pt-fingerprint` wrapper:

```
>>> pt_fingerprint("SELECT * FROM myapp_author WHERE id = 5")
"select * from myapp_author where id = 5"
```

Read more

1.13 Test Utilities

Set some MySQL server variables on a test case for every method or just a specific one:

```
class MyTests(TestCase):

    @override_mysql_variables(SQL_MODE="ANSI")
    def test_it_works_in_ansi_mode(self):
        self.run_it()
```

Read more

CHAPTER 2

Requirements and Installation

2.1 Requirements

These are the supported, tested versions of Django-MySQL's requirements:

- Python: 3.5, 3.6, 3.7
- Django: 1.11, 2.0, 2.1, 2.2
- MySQL: 5.6, 5.7 / MariaDB: 10.0, 10.1, 10.2, 10.3
- mysqlclient: 1.3

2.2 Installation

Install it with **pip**:

```
$ pip install django-mysql
```

Or add it to your project's `requirements.txt`.

Add `'django_mysql'` to your `INSTALLED_APPS` setting:

```
INSTALLED_APPS = (
    ...
    'django_mysql',
)
```

Django-MySQL comes with some extra checks to ensure your configuration for Django + MySQL is optimal. It's best to run these now you've installed to see if there is anything to fix:

```
$ ./manage.py check
```

For help fixing any warnings, see *Checks*.

2.3 Extending your QuerySets

Half the fun features are extensions to `QuerySet`. You can add these to your project in a number of ways, depending on what is easiest for your code - all imported from `django_mysql.models`.

class Model

The simplest way to add the `QuerySet` extensions - this is a subclass of Django's `Model` that sets `objects` to use the Django-MySQL extended `QuerySet` (below) via `QuerySet.as_manager()`. Simply change your model base to get the goodness:

```
# from django.db.models import Model - no more!
from django_mysql.models import Model

class MySuperModel(Model):
    pass  # TODO: come up with startup idea.
```

class QuerySet

The second way to add the extensions - use this to replace your model's default manager:

```
from mythings import MyBaseModel
from django_mysql.models import QuerySet

class MySuperDuperModel(MyBaseModel):
    objects = QuerySet.as_manager()
    # TODO: what fields should this model have??
```

class QuerySetMixin

The third way to add the extensions, and the container class for the extensions. Add this mixin to your custom `QuerySet` class to add in all the fun:

```
from django.db.models import Model
from django_mysql.models import QuerySetMixin
from stackoverflow import CopyPasteQuerySet

class MySplendidQuerySet(QuerySetMixin, CopyPasteQuerySet):
    pass

class MySplendidModel(Model):
    objects = MySplendidQuerySet.as_manager()
    # TODO: profit
```

add_QuerySetMixin(*queryset*)

A final way to add the extensions, useful when you don't control the model class - for example with built in Django models. This function creates a subclass of a `QuerySet`'s class that has the `QuerySetMixin` added in and applies it to the `QuerySet`:

```
from django.contrib.auth.models import User
from django_mysql.models import add_QuerySetMixin

qs = User.objects.all()
qs = add_QuerySetMixin(qs)
# Now qs has all the extensions!
```

The extensions are described in *QuerySet Extensions*.

CHAPTER 3

Checks

Django-MySQL runs some extra checks as part of Django's check framework to ensure your configuration for Django + MySQL is optimal. If triggered, the checks give a brief message, and a link here for documentation on how to fix it.

Note: A reminder: as per the Django docs, you can silence individual checks in your settings. For example, if you determine `django_mysql.W002` doesn't require your attention, add the following to `settings.py`:

```
SILENCED_SYSTEM_CHECKS = [
    'django_mysql.W002',
]
```

3.1 django_mysql.W001: Strict Mode

MySQL's Strict Mode fixes many data integrity problems in MySQL, such as data truncation upon insertion, by escalating warnings into errors. It is strongly recommended you activate it.

Docs: MySQL / MariaDB.

It is configured as part of `sql_mode`, a system variable contains a list of comma-separated modes to activate. Please check the value of your install and update it as necessary to add `STRICT_TRANS_TABLES` (the default in MySQL 5.7 onwards) - the following instructions assume it is set to the empty string initially, which is the MySQL 5.5 default.

The easiest way to change `sql_mode` for your app is to set it from the `init_command` that is run on each new connection by `MySQLdb`:

```
DATABASES = {
    'default': {
        'ENGINE': 'django.db.backends.mysql',
        'NAME': 'my_database',
        'OPTIONS': {
            'init_command': "SET sql_mode='STRICT_TRANS_TABLES'",
```

(continues on next page)

(continued from previous page)

```
        },
    }
}
```

This sets it for your app, but it does not set it for other connections to your database server. There is also a roundtrip added for sending a command to the server on each new connection.

You can change `sql_mode` permanently by using `SET GLOBAL` from an admin user on the server, plus changing your configuration files so the setting survives a server restart. For more information, see Using System Variables system variables in the MySQL documentation.

3.2 django_mysql.W002: InnoDB Strict Mode

InnoDB Strict Mode is similar to the general Strict Mode, but for InnoDB. It escalates several warnings around InnoDB-specific statements into errors. Normally this just affects per-table settings for compression. It's recommended you activate this, but it's not very likely to affect you if you don't.

Docs: MySQL / MariaDB.

As above, the easiest way to set this is to add `SET` to `init_command` in your `DATABASES` setting:

```
DATABASES = {
    'default': {
        'ENGINE': 'django.db.backends.mysql',
        'NAME': 'my_database',
        'OPTIONS': {
            'init_command': 'SET innodb_strict_mode=1',
        },
    }
}
```

Note: If you use this along with the `init_command` for W001, combine them as `SET sql_mode='STRICT_TRANS_TABLES', innodb_strict_mode=1`.

Also, as above for `django_mysql.W001`, it's better that you set it permanently for the server with `SET GLOBAL` and a configuration file change.

3.3 django_mysql.W003: utf8mb4

MySQL's `utf8` character set does not include support for the largest, 4 byte characters in UTF-8; this basically means it cannot support emoji and custom Unicode characters. The `utf8mb4` character set was added to support all these characters, and there's really little point in not using it. Django currently suggests using the `utf8` character set for backwards compatibility, but it's likely to move in time.

It's strongly recommended you change to the `utf8mb4` character set and convert your existing `utf8` data as well, unless you're absolutely sure you'll never see any of these 'supplementary' Unicode characters (note: it's very easy for users to type emoji on phone keyboards these days!).

Docs: MySQL / MariaDB.

Also see this classic blogpost: How to support full Unicode in MySQL databases.

The easiest way to set this up is to make a couple of changes to your `DATABASES` settings. First, add `OPTIONS` with `charset` to your MySQL connection, so `MySQLdb` connects using the `utf8mb4` character set. Second, add `TEST` with `COLLATION` and `CHARSET` as below, so Django creates the test database, and thus all tables, with the right character set:

```
DATABASES = {
    'default': {
        'ENGINE': 'django.db.backends.mysql',
        'NAME': 'my_database',
        'OPTIONS': {
            # Tell MySQLdb to connect with 'utf8mb4' character set
            'charset': 'utf8mb4',
        },
        # Tell Django to build the test database with the 'utf8mb4' character set
        'TEST': {
            'CHARSET': 'utf8mb4',
            'COLLATION': 'utf8mb4_unicode_ci',
        }
    }
}
```

Note this does not transform the database, tables, and columns that already exist. Follow the examples in the 'How to' blog post link above to fix your database, tables, and character set. It's planned to add a command to Django-MySQL to help you do this, see Issue 216.

CHAPTER 4

QuerySet Extensions

MySQL-specific Model and QuerySet extensions. To add these to your `Model`/`Manager`/`QuerySet` trifecta, see *Requirements and Installation*. Methods below are all `QuerySet` methods; where standalone forms are referred to, they can be imported from `django_mysql.models`.

4.1 Approximate Counting

`approx_count` (*fall_back=True*, *return_approx_int=True*, *min_size=1000*)

By default a QuerySet's *count()* method runs *SELECT COUNT(*)* on a table. Whilst this is fast for `MyISAM` tables, for `InnoDB` it involves a full table scan to produce a consistent number, due to MVCC keeping several copies of rows when under transaction. If you have lots of rows, you will notice this as a slow query - Percona have some more details.

This method returns the approximate count found by running `EXPLAIN SELECT COUNT(*) ...`. It can be out by 30-50% in the worst case, but in many applications it is closer, and is good enough, such as when presenting many pages of results but users will only practically scroll through the first few. For example:

```
>>> Author.objects.count()  # slow
509741
>>> Author.objects.approx_count()  # fast, with some error
531140
```

Three arguments are accepted:

`fall_back=True`

If `True` and the approximate count cannot be calculated, `count()` will be called and returned instead, otherwise `ValueError` will be raised.

The approximation can only be found for `objects.all()`, with no filters, `distinct()` calls, etc., so it's reasonable to fall back.

`return_approx_int=True`

When `True`, an `int` is not returned (excpet when falling back), but instead a subclass called `ApproximateInt`. This is for all intents and purposes an `int`, apart from when cast to `str`, it renders

as e.g. **'Approximately 12345'** (internationalization ready). Useful for templates you can't edit (e.g. the admin) and you want to communicate that the number is not 100% accurate. For example:

```
>>> print(Author.objects.approx_count())  # ApproximateInt
Approximately 531140
>>> print(Author.objects.approx_count() + 0)  # plain int
531140
>>> print(Author.objects.approx_count(return_approx_int=False))  # plain int
531140
```

min_size=1000
The threshold at which to use the approximate algorithm; if the approximate count comes back as less that this number, `count()` will be called and returned instead, since it should be so small as to not bother your database. Set to `0` to disable this behaviour and always return the approximation.

The default of `1000` is a bit pessimistic - most tables won't take long when calling `COUNT(*)` on tens of thousands of rows, but it *could* be slow for very wide tables.

count_tries_approx (*activate=True*, *fall_back=True*, *return_approx_int=True*, *min_size=1000*)
This is the 'magic' method to make pre-existing code, such as Django's admin, work with `approx_count`. Calling `count_tries_approx` sets the QuerySet up such that then calling `count` will call `approx_count` instead, with the given arguments.

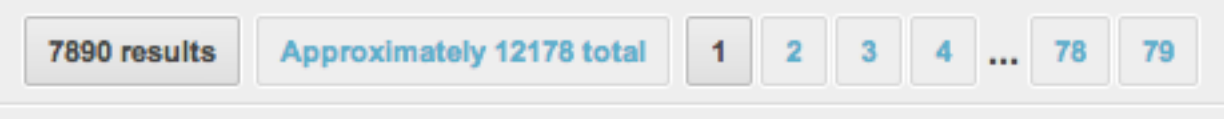
To unset this, call `count_tries_approx` with `activate=False`.

To 'fix' an Admin class with this, simply do the following (assuming `Author` inherits from `django_mysql`'s `Model`):

```
class AuthorAdmin(ModelAdmin):

    def get_queryset(self, request):
        qs = super(AuthorAdmin, self).get_queryset(request)
        return qs.count_tries_approx()
```

You'll be able to see this is working on the pagination due to the word **'Approximately'** appearing:



You can do this at a base class for all your `ModelAdmin` subclasses to apply the magical speed increase across your admin interface.

4.2 Query Hints

The following methods add extra features to the ORM which allow you to access some MySQL-specific syntax. They do this by inserting special comments which pass through Django's ORM layer and get re-written by a function that wraps the lower-level `cursor.execute()`.

Because not every user wants these features and there is a (small) overhead to every query, you must activate this feature by adding to your settings:

```
DJANGO_MYSQL_REWRITE_QUERIES = True
```

Once you've done this, the following methods will work.

label(*comment*)

Allows you to add an arbitrary comment to the start of the query, as the second thing after the keyword. This can be used to 'tag' queries so that when they show in the *slow_log* or another monitoring tool, you can easily back track to the python code generating the query. For example, imagine constructing a QuerySet like this:

```
qs = Author.objects.label("AuthorListView").all()
```

When executed, this will have SQL starting:

```
SELECT /*AuthorListView*/ ...
```

You can add arbitrary labels, and as many of them as you wish - they will appear in the order added. They will work in `SELECT` and `UPDATE` statements, but not in `DELETE` statements due to limitations in the way Django performs deletes.

You should not pass user-supplied data in for the comment. As a basic protection against accidental SQL injection, passing a comment featuring `*/` will raise a `ValueError`, since that would prematurely end the comment. However due to executable comments, the comment is still prone to some forms of injection.

However this is a feature - by not including spaces around your string, you may use this injection to use executable comments to add hints that are otherwise not supported, or to use the new MySQL 5.7 optimizer hints.

straight_join()

Adds the `STRAIGHT_JOIN` hint, which forces the join order during a `SELECT`. Note that you can't force Django's join order, but it tends to be in the order that the tables get mentioned in the query.

Example usage:

```
# Note from Adam: sometimes the optimizer joined books -> author, which
# is slow. Force it to do author -> books.
Author.objects.distinct().straight_join().filter(books__age=12)[:10]
```

Docs: MySQL / MariaDB.

The MariaDB docs also have a good page Index Hints: How to Force Query Plans" which covers some cases when you might want to use `STRAIGHT_JOIN`.

sql_small_result()

Adds the `SQL_SMALL_RESULT` hint, which avoids using a temporary table in the case of a `GROUP BY` or `DISTINCT`.

Example usage:

```
# Note from Adam: we have very few distinct birthdays, so using a
# temporary table is slower
Author.objects.values('birthday').distinct().sql_small_result()
```

Docs: MySQL / MariaDB.

sql_big_result()

Adds the `SQL_BIG_RESULT` hint, which forces using a temporary table in the case of a `GROUP BY` or `DISTINCT`.

Example usage:

```
# Note from Adam: for some reason the optimizer didn't use a temporary
# table for this, so we force it
Author.objects.distinct().sql_big_result()
```

Docs: MySQL / MariaDB.

sql_buffer_result()
Adds the SQL_BUFFER_RESULT hint, which forces the optimizer to use a temporary table to process the result. This is useful to free locks as soon as possible.

Example usage:

```
# Note from Adam: seeing a lot of throughput on this table. Buffering
# the results makes the queries less contentious.
HighThroughputModel.objects.filter(x=y).sql_buffer_result()
```

Docs: MySQL / MariaDB.

sql_cache()
Adds the SQL_CACHE hint, which means the result set will be stored in the Query Cache. This only has an effect when the MySQL system variable query_cache_type is set to 2 or DEMAND.

Example usage:

```
# Fetch recent posts, cached in MySQL for speed
recent_posts = BlogPost.objects.sql_cache().order_by('-created')[:5]
```

Docs: MySQL / MariaDB.

sql_no_cache()
Adds the SQL_NO_CACHE hint, which means the result set will not be fetched from or stored in the Query Cache. This only has an effect when the MySQL system variable query_cache_type is set to 1 or ON.

Example usage:

```
# Avoid caching all the expired sessions, since we're about to delete
# them
deletable_session_ids = (
    Session.objects.sql_no_cache()
                   .filter(expiry__lt=now())
                   .values_list('id', flat=True)
)
```

Docs: MySQL / MariaDB.

sql_calc_found_rows()
Adds the SQL_CALC_FOUND_ROWS hint, which means the total count of matching rows will be calculated when you only take a slice. You can access this count with the found_rows attribute of the QuerySet after filling its result cache, by e.g. iterating it.

This can be faster than taking the slice and then again calling .count() to get the total count.

Example usage:

```
>>> can_drive = Customer.objects.filter(age=21).sql_calc_found_rows()[:10]
>>> len(can_drive)  # Fetches the first 10 from the database
10
>>> can_drive.found_rows  # The total number of 21 year old customers
1942
```

Docs: MySQL / MariaDB.

use_index(**index_names*, *for_=None*, *table_name=None*)
Adds a USE INDEX hint, which affects the index choice made by MySQL's query optimizer for resolving the query.

Note that index names on your tables will normally have been generated by Django and contain a hash fragment. You will have to check your database schema to determine the index name.

If you pass any non-existent index names, MySQL will raise an error. This means index hints are especially important to test in the face of future schema changes.

`for_` restricts the scope that the index hint applies to. By default it applies to all potential index uses during the query; you may supply one of `'JOIN'`, `'ORDER BY'`, or `'GROUP BY'` to restrict the index hint to only be used by MySQL for index selection in their respective stages of query execution. For more information see the MySQL/MariaDB docs (link below).

`table_name` is the name of the table that the hints are for. By default, this will be the name of the table of the model that the `QuerySet` is for, however you can supply any other table that may be joined into the query (from e.g. `select_related()`). Be careful - there is no validation on the table name, and if it does not exist in the final query it will be ignored. Also it is injected raw into the resultant SQL, so you should not use user data otherwise it may open the potential for SQL injection.

Note that `USE INDEX` accepts no index names to mean 'use no indexes', i.e. table scans only.

Example usage:

```
# SELECT ... FROM `author` USE INDEX (`name_12345`) WHERE ...
>>> Author.objects.use_index('name_12345').filter(name='John')
# SELECT ... FROM `author` USE INDEX (`name_12345`, `name_age_678`) WHERE ...
>>> Author.objects.use_index('name_12345', 'name_age_678').filter(name='John')
# SELECT ... FROM `author` USE INDEX FOR ORDER BY (`name_12345`) ... ORDER BY␣
↪`name`
>>> Author.objects.use_index('name_12345', for_='ORDER BY').order_by('name')
# SELECT ... FROM `book` INNER JOIN `author` USE INDEX (`authbook`) ...
>>> Book.objects.select_related('author').use_index('authbook', table_name='author
↪')
```

Docs: MySQL / MariaDB.

force_index(**index_names*, *for_=None*)
Similar to the above `use_index()`, but adds a `FORCE INDEX` hint. Note that unlike `use_index()` you must supply at least one index name. For more information, see the MySQL/MariaDB docs.

ignore_index(**index_names*, *for_=None*)
Similar to the above `use_index()`, but adds an `IGNORE INDEX` hint. Note that unlike `use_index()` you must supply at least one index name. For more information, see the MySQL/MariaDB docs.

4.3 'Smart' Iteration

Here's a situation we've all been in - we screwed up, and now we need to fix the data. Let's say we accidentally set the address of all authors without an address to "Nowhere", rather than the blank string. How can we fix them??

The simplest way would be to run the following:

```
Author.objects.filter(address="Nowhere").update(address="")
```

Unfortunately with a lot of rows ('a lot' being dependent on your database server and level of traffic) this will stall other access to the table, since it will require MySQL to read all the rows and to hold write locks on them in a single query.

To solve this, we could try updating a chunk of authors at a time; such code tends to get ugly/complicated pretty quickly:

```
min_id = 0
max_id = 1000
biggest_author_id = Author.objects.order_by('-id')[0].id
while True:
    Author.objects.filter(id__gte=min_id, id__lte=BLA BLA BLA

# I'm not even going to type this all out, it's so much code
```

Here's the solution to this boilerplate with added safety features - 'smart' iteration! There are two classes; one yields chunks of the given `QuerySet`, and the other yields the objects inside those chunks. Nearly every data update can be thought of in one of these two methods.

class SmartChunkedIterator(*queryset*, *atomically=True*, *status_thresholds=None*, *pk_range=None*, *chunk_time=0.5*, *chunk_size=2*, *chunk_min=1*, *chunk_max=10000*, *report_progress=False*, *total=None*)

Implements a smart iteration strategy over the given `queryset`. There is a method `iter_smart_chunks` that takes the same arguments on the `QuerySetMixin` so you can just:

```
bad_authors = Author.objects.filter(address="Nowhere")
for author_chunk in bad_authors.iter_smart_chunks():
    author_chunk.update(address="")
```

Iteration proceeds by yielding primary-key based slices of the queryset, and dynamically adjusting the size of the chunk to try and take `chunk_time` seconds. In between chunks, the *`wait_until_load_low()`* method of *`GlobalStatus`* is called to ensure the database is not under high load.

> **Warning:** Because of the slicing by primary key, there are restrictions on what `QuerySets` you can use, and a `ValueError` will be raised if the queryset doesn't meet that. Specifically, only `QuerySets` on models with integer-based primary keys, which are unsliced, and have no `order_by` will work.

There are a lot of arguments and the defaults have been picked hopefully sensibly, but please check for your case though!

queryset
The queryset to iterate over; if you're calling via `.iter_smart_chunks` then you don't need to set this since it's the queryset you called it on.

atomically=True
If true, wraps each chunk in a transaction via django's `transaction.atomic()`. Recommended for any write processing.

status_thresholds=None
A dict of status variables and their maximum tolerated values to be checked against after each chunk with *`wait_until_load_low()`*.

When set to `None`, it lets *`GlobalStatus`* use its default of `'Threads_running': 5}`. Set to an empty dict to disable status checking (not really recommended, it doesn't add much overhead and can will probably save your butt one day).

pk_range=None
Controls the primary key range to iterate over with slices. By default, with `pk_range=None`, the QuerySet will be searched for its minimum and maximum `pk` values before starting. On QuerySets that match few rows, or whose rows aren't evenly distributed, this can still execute a long blocking table scan to find these two rows. You can remedy this by giving a value for `pk_range`:

- If set to `'all'`, the range will be the minimum and maximum PK values of the entire table, excluding any filters you have set up - that is, for `Model.objects.all()` for the given `QuerySet`'s model.

- If set to a 2-tuple, it will be unpacked and used as the minimum and maximum values respectively.

Note: The iterator determines the minimum and maximum at the start of iteration and does not update them whilst iterating, which is normally a safe assumption, since if you're "fixing things" you probably aren't creating any more bad data. If you do need to process *every* row then set `pk_range` to have a maximum far greater than what you expect would be reached by inserts that occur during iteration.

`chunk_time=0.5`
The time in seconds to aim for each chunk to take. The chunk size is dynamically adjusted to try and match this time, via a weighted average of the past and current speed of processing. The default and algorithm is taken from the analogous `pt-online-schema-change` flag –chunk-time.

`chunk_size=2`
The initial size of the chunk that will be used. As this will be dynamically scaled and can grow fairly quickly, the initial size of 2 should be appropriate for most use cases.

`chunk_min=1`
The minimum number of objects in a chunk. You do not normally need to tweak this since the dynamic scaling works very well, however it might be useful if your data has a lot of "holes" or if there are other constraints on your application.

`chunk_max=10000`
The maximum number of objects in a chunk, a kind of sanity bound. Acts to prevent harm in the case of iterating over a model with a large 'hole' in its primary key values, e.g. if only ids 1-10k and 100k-110k exist, then the chunk 'slices' could grow very large in between 10k and 100k since you'd be "processing" the non-existent objects 10k-100k very quickly.

`report_progress=False`
If set to true, display out a running counter and summary on `sys.stdout`. Useful for interactive use. The message looks like this:

```
AuthorSmartChunkedIterator processed 0/100000 objects (0.00%) in 0 chunks
```

And uses `\r` to erase itself when re-printing to avoid spamming your screen. At the end `Finished!` is printed on a new line.

`total=None`
By default the total number of objects to process will be calculated with `approx_count()`, with `fall_back` set to `True`. This `count()` query could potentially be big and slow.

`total` allows you to pass in the total number of objects for processing, if you can calculate in a cheaper way, for example if you have a read-replica to use.

`class SmartIterator`
A convenience subclass of `SmartChunkedIterator` that simply unpacks the chunks for you. Can be accessed via the `iter_smart` method of `QuerySetMixin`.

For example, rather than doing this:

```
bad_authors = Author.objects.filter(address="Nowhere")
for authors_chunk in bad_authors.iter_smart_chunks():
    for author in authors_chunk:
        author.send_apology_email()
```

You can do this:

```
bad_authors = Author.objects.filter(address="Nowhere")
for author in bad_authors.iter_smart():
    author.send_apology_email()
```

All the same arguments as `SmartChunkedIterator` are accepted.

class SmartPKRangeIterator

A subclass of `SmartChunkedIterator` that doesn't return the chunk's `QuerySet` but instead returns the start and end primary keys for the chunk. This may be useful when you want to iterate but the slices need to be used in a raw SQL query. Can be accessed via the `iter_smart_pk_ranges` method of `QuerySetMixin`.

For example, rather than doing this:

```
for authors_chunk in Author.objects.iter_smart_chunks():
    limits = author_chunk.aggregate(min_pk=Min('pk'), max_pk=Max('pk'))
    authors = Author.objects.raw("""
        SELECT name from app_author
        WHERE id >= %s AND id <= %s
    """, (limits['min_pk'], limits['max_pk']))
    # etc...
```

...you can do this:

```
for start_pk, end_pk in Author.objects.iter_smart_pk_ranges():
    authors = Author.objects.raw("""
        SELECT name from app_author
        WHERE id >= %s AND id < %s
    """, (start_pk, end_pk))
    # etc...
```

In the first format we were forced to perform a dumb query to determine the primary key limits set by `SmartChunkedIterator`, due to the `QuerySet` not otherwise exposing this information.

Note: There is a **subtle** difference between the two versions. In the first the end boundary, `max_pk`, is a closed bound, whereas in the second, the `end_pk` from `iter_smart_pk_ranges` is an open bound. Thus the <= changes to a <.

All the same arguments as `SmartChunkedIterator` are accepted.

4.4 Integration with pt-visual-explain

How does MySQL *really* execute a query? The `EXPLAIN` statement (docs: MySQL / MariaDB), gives a description of the execution plan, and the `pt-visual-explain` tool can format this in an understandable tree.

This function is a shortcut to turn a `QuerySet` into its visual explanation, making it easy to gain a better understanding of what your queries really end up doing.

pt_visual_explain(*display=True*)

Call on a `QuerySet` to print its visual explanation, or with `display=False` to return it as a string. It prepends the SQL of the query with 'EXPLAIN' and passes it through the `mysql` and `pt-visual-explain` commands to get the output. You therefore need the MySQL client and Percona Toolkit installed where you run this.

Example:

```
>>> Author.objects.all().pt_visual_explain()
Table scan
rows           1020
+- Table
   table          myapp_author
```

Can also be imported as a standalone function if you want to use it on a `QuerySet` that does not have the `QuerySetMixin` added, e.g. for built-in Django models:

```
>>> from django_mysql.models import pt_visual_explain
>>> pt_visual_explain(User.objects.all())
Table scan
rows           1
+- Table
   table          auth_user
```

4.5 Handler

MySQL's `HANDLER` commands give you simple NoSQL-style read access, faster than full SQL queries, with the ability to perform index lookups or paginated scans (docs: MySQL / MariaDB).

This extension adds an ORM-based API for handlers. You can instantiate them from a `QuerySet` (and thus from *.objects*), and open/close them as context managers:

```
with Author.objects.handler() as handler:

    first_author_by_pk = handler.read()

    first_ten_authors_by_pk = handler.read(limit=10)

    for author in handler.iter(chunk_size=1000):
        author.send_apology_email()
```

The `.handler()` method simply returns a `Handler` instance; the class can be imported and applied to `QuerySets` from models without the extensions easily as well:

```
from django_mysql.models.handler import Handler

with Handler(User.objects.all()) as handler:

    for user in handler.iter(chunk_size=1000):
        user.send_notification_email()
```

Warning: `HANDLER` is lower level than `SELECT`, and has some optimizations that mean it permits 'for example' **dirty reads**. Check the database documentation and understand the consequences of this before you replace any SQL queries!

class Handler(*queryset*)

Implements a handler for the given queryset's model. The `WHERE` clause and query parameters, if they exist, will be extracted from the queryset's SQL. Since `HANDLER` statements can only operate on one table at a time, only relatively simple querysets can be used - others will result in a `ValueError` being raised.

A `Handler` is only opened and available for reads when used as a context manager. You may have multiple handlers open at once, even on the same table, but you cannot open the same one twice.

read(*index='PRIMARY'*, *value__LOOKUP=None*, *mode=None*, *where=None*, *limit=None*)

Returns the result of a `HANDLER .. READ` statement as a `RawQuerySet` for the given `queryset`'s model (which, like all `QuerySet`s, is lazy).

Note: The `HANDLER` statements must select whole rows, therefore there is no way of optimizing by returning only certain columns (like `QuerySet`'s `only()`).

MySQL has three forms of `HANDLER .. READ` statements, but only the **first two forms** of `HANDLER .. READ` statements are supported - you can specify index lookups, or pagination. The third form, 'natural row order', only makes sense for MyISAM tables.

index='PRIMARY'

The name of the index of the table to read, defaulting to the primary key. You must provide the index name as known by MySQL, not the names of the indexed column[s] as Django's `db_index` and `index_together` let you specify. This will only be checked by MySQL so an `OperationalError` will be raised if you specify a wrong name.

Both single and multi-column indexes are supported.

value__LOOKUP=None

The 'first form' of `HANDLER .. READ` supports index lookups. `value__LOOKUP` allows you to specify a lookup on `index` using the same style as Django's ORM, and is mutually exclusive with `mode`. You may only have one index lookup on a `read` - other conditions must be filtered with `where`. For example:

```
# Read objects with primary key <= 100
handler.read(value__lte=100, limit=100)
```

The valid lookups are:

- `value__lt=x` - index value < x
- `value__lte=x` - index value <= x
- `value=x`, `value__exact=x` - index value = x
- `value__gte=x` - index value >= x
- `value__gt=x` - index value > x

For single-column indexes, specify the value; for multi-column indexes, specify an iterable of values, one for each column, in index order. For example:

```
grisham = handler.read(index='full_name_idx',
                       value=('John', 'Grisham'))
```

mode=None

The 'second form' of `HANDLER .. READ` supports paging over a table, fetching one batch of results at a time whilst the handler object on MySQL's end retains state, somewhat like a 'cursor'. This is mutually exclusive with `value__LOOKUP`, and if neither is specified, this is the default.

There are four modes:

- `first` - commence iteration at the start
- `next` - continue ascending/go forward one page
- `last` - commence iteration at the end (in reverse)
- `prev` - continue descending/go backward one page

To iterate forwards, use `'first'` and then repeatedly `'next'`. To iterate backwards, use `'last'` and then repeatedly `'prev'`. The page size is set with `limit`.

N.B. the below `iter` method below is recommended for most iteration.

where=None

HANDLER .. READ statements support WHERE clauses for columns on the same table, which apply after the index filtering. By default the WHERE clause from the queryset used to construct the Handler will be applied. Passing a different QuerySet as where allows you to read with different filters. For example:

```
with Author.objects.handler() as handler:

    old = Author.objects.filter(age__gte=50)
    first_old_author = handler.read(where=old)[0]

    young = Author.objects.filter(age__lte=50)
    first_young_author = handler.read(where=young)[0]
```

limit=None

By default every HANDLER .. READ statement returns *only* the first row. Specify limit to retrieve a different number of rows.

iter(*index='PRIMARY'*, *where=None*, *chunk_size=100*, *reverse=False*)

Iterate over a table via the named index, one chunk at a time, yielding the individual objects. Acts as a wrapper around repeated calls to read.

index='PRIMARY'

The name of the index to iterate over. As detailed above, this must be the index name on MySQL.

where=None

A QuerySet for filter conditions, the same as read's where argument, as detailed above.

chunk_size=100

The size of the chunks to read during iteration.

reverse=False

The direction of iteration over the index. By default set to True, the index will be iterated in ascending order; set to False, the index will be iterated in descending order.

Sets mode on read to either FIRST then repeatedly NEXT or LAST then repeatedly PREV respectively.

> **Warning:** You can only have one iteration happening at a time per Handler, otherwise on the MySQL side it loses its position. There is no checking for this in Handler class.

CHAPTER 5

Model Fields

More MySQL and MariaDB specific ways to store data!

Field classes should always be imported from `django_mysql.models`, similar to the home of Django's native fields in `django.db.models`.

5.1 JSONField

MySQL 5.7 comes with a JSON data type that stores JSON in a way that is queryable and updatable in place. This is ideal for data that varies widely, or very sparse columns, or just for storing API responses that you don't have time to turn into the relational format.

Docs: MySQL.

Django-MySQL supports the JSON data type and related functions through `JSONField` plus some *JSON database functions*.

class JSONField(***kwargs*)

A field for storing JSON. The Python data type may be either `str`, `int`, `float`, `dict`, or `list` - basically anything that is supported by `json.dumps`. There is no restriction between these types - this may be surprising if you expect it to just store JSON objects/`dicts`.

So for example, the following all work:

```
mymodel.myfield = "a string"
mymodel.myfield = 1
mymodel.myfield = 0.3
mymodel.myfield = ["a", "list"]
mymodel.myfield = {"a": "dict"}
```

This field requires Django 1.8+ and MySQL 5.7+. Both requirements are checked by the field and you'll get sensible errors for them when Django's checks run if you're not up to date on either.

Warning: If you give the field a `default`, ensure it's a callable, such as a function, or the `dict` or `list` classes themselves. Incorrectly using a mutable object, such as `default={}`, creates a single object that is shared between all instances of the field. There's a field check that errors if a plain `list` or `dict` instance is used for `default`, so there is some protection against this.

The correct way to provide a rich default like `{'foo': 'bar'}` is to define a module level function that returns it, so it can be serialized in migrations. For example:

```
def my_default():
    return {'foo': 'bar'}

class MyModel(Model):
    attrs = JSONField(default=my_default)
```

5.1.1 JSONFields in Forms

By default this uses the simple Django-MySQL form field *JSONField*, which simply displays the JSON in an HTML `<textarea>`.

5.1.2 Querying JSONField

You can query by object keys as well as array positions. In cases where names collide with existing lookups, you might want to use the *JSONExtract* database function.

Warning: Most of the standard lookups don't make sense for `JSONField` and so have been made to fail with `NotImplementedError`. There is scope for making some of them work in the future, but it's non-trivial. Only the lookups documented below work.

Also be careful with the key lookups. Since any string could be a key in a JSON object, any lookup name other than the standard ones or those listed below will be interpreted as a key lookup. No errors are raised. Be extra careful for typing mistakes, and always check your queries, e.g. `myfield__eaxct` as a typo of `myfield__exact` will not do what the author intended!

We'll use the following example model:

```
from django_mysql.models import JSONField, Model

class ShopItem(Model):
    name = models.CharField(max_length=200)
    attrs = JSONField()

    def __str__(self):
        return self.name
```

Exact Lookups

To query based on an exact match, just use an object of any JSON type.

For example:

```
>>> ShopItem.objects.create(name='Gruyère', attrs={'smelliness': 5})
>>> ShopItem.objects.create(name='Feta', attrs={'smelliness': 3, 'crumbliness': 10})
>>> ShopItem.objects.create(name='Hack', attrs=[1, 'arbitrary', 'data'])

>>> ShopItem.objects.filter(attrs={'smelliness': 5})
[<ShopItem: Gruyère>]
>>> ShopItem.objects.filter(attrs__exact={'smelliness': 3, 'crumbliness': 10})
[<ShopItem: Feta>]
>>> ShopItem.objects.filter(attrs=[1, 'arbitrary', 'data'])
[<ShopItem: Hack>]
```

Ordering Lookups

MySQL defines an ordering on JSON objects - see the docs for more details. The ordering rules can make sense for some types (e.g. strings, arrays), however they can also be confusing if your data is of mixed types, so be careful. You can use the ordering by querying with Django's built-in `gt`, `gte`, `lt`, and `lte` lookups.

For example:

```
>>> ShopItem.objects.create(name='Cheshire', attrs=['Dense', 'Crumbly'])
>>> ShopItem.objects.create(name='Double Gloucester', attrs=['Semi-hard'])

>>> ShopItem.objects.filter(attrs__gt=['Dense', 'Crumbly'])
[<ShopItem: Double Gloucester>]
>>> ShopItem.objects.filter(attrs__lte=['ZZZ'])
[<ShopItem: Cheshire>, <ShopItem: Double Gloucester>]
```

Key, Index, and Path Lookups

To query based on a given dictionary key, use that key as the lookup name:

```
>>> ShopItem.objects.create(name='Gruyère', attrs={
    'smelliness': 5,
    'origin': {
        'country': 'Switzerland',
    }
    'certifications': ['Swiss AOC', 'Swiss AOP'],
})
>>> ShopItem.objects.create(name='Feta', attrs={'smelliness': 3, 'crumbliness': 10})

>>> ShopItem.objects.filter(attrs__smelliness=3)
[<ShopItem: Feta>]
```

Multiple keys can be chained together to form a path lookup:

```
>>> ShopItem.objects.filter(attrs__origin__country='Switzerland')
[<ShopItem: Gruyère>]
```

If the key is an integer, it will be interpreted as an index lookup in an array:

```
>>> ShopItem.objects.filter(attrs__certifications__0='Swiss AOC')
[<ShopItem: Gruyère>]
```

If the key you wish to query is not valid for a Python keyword argument (e.g. it contains unicode characters), or it clashes with the name of another field lookup, use the *`JSONExtract`* database function to fetch it.

Key Presence Lookups

To query to check if an object has a given key, use the `has_key` lookup:

```
# Find all ShopItems with a hardness rating
>>> ShopItem.objects.filter(attrs__has_key='hardness')
[]
# Find all ShopItems missing certification information
>>> ShopItem.objects.exclude(attrs__has_key='certifications')
[<ShopItem: Feta>]
```

To check if an object has several keys, use the `has_keys` lookup with a list of keys:

```
# Find all ShopItems with both origin and certification information
>>> ShopItem.objects.filter(attrs__has_keys=['origin', 'certifications'])
[<ShopItem: Gruyère>]
```

To find objects with one of several keys, use the `has_any_keys` lookup with a list of keys:

```
# Find all ShopItems with either a smelliness or a hardness rating
>>> ShopItem.objects.filter(attrs__has_any_keys=['smelliness', 'hardness'])
[<ShopItem: Gruyère>, <ShopItem: Feta>]
```

Length Lookup

This is very similar to the `functions:JSONLength` database function. You can use it to filter based upon the length of the JSON documents in the field, using the MySQL `JSON_LENGTH` function.

As per the MySQL documentation, the length of a document is determined as follows:

- The length of a scalar is 1.
- The length of an array is the number of array elements.
- The length of an object is the number of object members.
- The length does not count the length of nested arrays or objects.

Docs: MySQL.

For example:

```
# Find all the ShopItems with nothing in 'attrs'
>>> ShopItems.objects.filter(attrs__length=0)
[]
# Find all the ShopItems with >50 keys in 'attrs'
>>> ShopItems.objects.filter(attrs__length__gt=50)
[<ShopItem: Incredible Cheese>]
```

Containment Lookups

The `contains` lookup is overriden on `JSONField` to support the MySQL `JSON_CONTAINS` function. This allows you to search, for example, JSON objects that contain at least a given set of key-value pairs. Additionally you can do the inverse with `contained_by`, i.e. find values where the objects are contained by a given value.

The definition of containment is, as per the MySQL docs:

- A candidate scalar is contained in a target scalar if and only if they are comparable and are equal. Two scalar values are comparable if they have the same `JSON_TYPE()` types, with the exception that values of types `INTEGER` and `DECIMAL` are also comparable to each other.
- A candidate array is contained in a target array if and only if every element in the candidate is contained in some element of the target.
- A candidate nonarray is contained in a target array if and only if the candidate is contained in some element of the target.
- A candidate object is contained in a target object if and only if for each key in the candidate there is a key with the same name in the target and the value associated with the candidate key is contained in the value associated with the target key.

Docs: MySQL.

For example:

```
# Find all ShopItems with a crumbliness of 10 and a smelliness of 5
>>> ShopItems.objects.filter(attrs__contains={
    'crumbliness': 10,
    'smelliness': 5,
})
[<ShopItem: Feta>]

# Find all ShopItems that have either 0 properties, or 1 or more of the given␣
↪properties
>>> ShopItems.objects.filter(attrs__contained_by={
    'crumbliness': 10,
    'hardness': 1,
    'smelliness': 5,
})
[<ShopItem: Feta>]
```

5.2 DynamicField

MariaDB has a feature called **Dynamic Columns** that allows you to store different sets of columns for each row in a table. It works by storing the data in a blob and having a small set of functions to manipulate this blob. (Docs).

Django-MySQL supports the *named* Dynamic Columns of MariaDB 10.0+, as opposed to the *numbered* format of 5.5+. It uses the mariadb-dyncol python package to pack and unpack Dynamic Columns blobs in Python rather than in MariaDB (mostly due to limitations in the Django ORM).

class DynamicField(*spec=None*, ***kwargs*)

A field for storing Dynamic Columns. The Python data type is `dict`. Keys must be `strs` and values must be one of the supported value types in `mariadb-dyncol`:

- `str`
- `int`
- `float`
- `datetime.date`
- `datetime.datetime`
- `datetime.datetime`
- A nested dict conforming to this spec too

Note that there are restrictions on the range of values supported for some of these types, and that `decimal.Decimal` objects are not yet supported though they are valid in MariaDB. For more information consult the `mariadb-dyncol` documentation.

Values may also be `None`, though they will then not be stored, since dynamic columns do not store `NULL`, so you should use `.get()` to retrieve values that may be `None`.

To use this field, you'll need to:

1. Use MariaDB 10.0.2+
2. Install `mariadb-dyncol` (`pip install mariadb-dyncol`)
3. Use either the `utf8mb4` or `utf8` character set for your database connection.

These are all checked by the field and you will see sensible errors for them when Django's checks run if you have a `DynamicField` on a model.

spec

This is an optional type specification that checks that the named columns, if present, have the given types. It is validated against on `save()` to ensure type safety (unlike normal Django validation which is only used in forms). It is also used for type information for lookups (below).

`spec` should be a `dict` with string keys and values that are the type classes you expect. You can also nest another such dictionary as a value for validating nested dynamic columns.

For example:

```
import datetime

class SpecModel(Model):
    attrs = DynamicField(spec={
        'an_integer_key': int,
        'created_at': datetime.datetime,
        'nested_columns': {
            'lat': int,
            'lon': int,
        }
    })
```

This will enforce the following rules:

- `instance.attrs['an_integer_key']`, if present, is an `int`
- `instance.attrs['created_at']`, if present, is an `datetime.datetime`
- `instance.attrs['nested_columns']`, if present, is a `dict`
- `instance.attrs['nested_columns']['lat']`, if present, is an `int`
- `instance.attrs['nested_columns']['lon']`, if present, is an `int`

Trying to save a `DynamicField` with data that does not match the rules of its `spec` will raise `TypeError`. There is no automatic casting, e.g. between `int` and `float`. Note that columns not in `spec` will still be allowed and have no type enforced.

For example:

```
>>> SpecModel.objects.create(attrs={'an_integer_key': 1})  # Fine
>>> SpecModel.objects.create(attrs={'an_integer_key': 2.0})
Traceback (most recent call last):
...
```

(continues on next page)

(continued from previous page)

```
TypeError: Key 'an_integer_key' should be of type 'int'
>>> SpecModel.objects.create(attrs={'non_spec_key': 'anytype'})  # Fine
```

5.2.1 DynamicFields in Forms

By default a `DynamicField` has no form field, because there isn't really a practical way to edit its contents. If required, is possible to add extra form fields to a `ModelForm` that then update specific dynamic column names on the instance in the form's `save()`.

5.2.2 Querying DynamicField

You can query by names, including nested names. In cases where names collide with existing lookups (e.g. you have a column named `'exact'`), you might want to use the *`ColumnGet`* database function. You can also use the *`ColumnAdd`* and *`ColumnDelete`* functions for atomically modifying the contents of dynamic columns at the database layer.

We'll use the following example model:

```
from django_mysql.models import DynamicField, Model

class ShopItem(Model):
    name = models.CharField(max_length=200)
    attrs = DynamicField(spec={
        'size': str,
    })

    def __str__(self):
        return self.name
```

Exact Lookups

To query based on an exact match, just use a dictionary.

For example:

```
>>> ShopItem.objects.create(name='Camembert', attrs={'smelliness': 15})
>>> ShopItem.objects.create(name='Cheddar', attrs={'smelliness': 15, 'hardness': 5})

>>> ShopItem.objects.filter(attrs={'smelliness': 15})
[<ShopItem: Camembert>]
>>> ShopItem.objects.filter(attrs={'smelliness': 15, 'hardness': 5})
[<ShopItem: Cheddar>]
```

Name Lookups

To query based on a column name, use that name as a lookup with one of the below SQL types added after an underscore. If the column name is in your field's `spec`, you can omit the SQL type and it will be extracted automatically - this includes keys in nested `dicts`.

The list of SQL types is:

- `BINARY` - `dict` (a nested `DynamicField`)

- CHAR - str
- DATE - datetime.date
- DATETIME - datetime.datetime
- DOUBLE - float
- INTEGER - int
- TIME - datetime.time

These will also use the correct Django ORM field so chained lookups based on that type are possible, e.g. dynamicfield__age_INTEGER__gte=20.

Beware that getting a named column can always return NULL if the column is not defined for a row.

For example:

```
>>> ShopItem.objects.create(name='T-Shirt', attrs={'size': 'Large'})
>>> ShopItem.objects.create(name='Rocketship', attrs={
...     'speed_mph': 300,
...     'dimensions': {'width_m': 10, 'height_m': 50}
... })

# Basic template: DynamicField + '__' + column name + '_' + SQL type
>>> ShopItem.objects.filter(attrs__size_CHAR='Large')
[<ShopItem: T-Shirt>]

# As 'size' is in the field's spec, there is no need to give the SQL type
>>> ShopItem.objects.filter(attrs__size='Large')
[<ShopItem: T-Shirt>]

# Chained lookups are possible based on the data type
>>> ShopItem.objects.filter(attrs__speed_mph_INTEGER__gte=100)
[<ShopItem: Rocketship>]

# Nested keys can be looked up
>>> ShopItem.objects.filter(attrs__dimensions_BINARY__width_m_INTEGER=10)
[<ShopItem: Rocketship>]

# Nested DynamicFields can be queried as ``dict``s, as per the ``exact`` lookup
>>> ShopItem.objects.filter(attrs__dimensions_BINARY={'width_m': 10, 'height_m': 50})
[<ShopItem: Rocketship>]

# Missing keys are always NULL
>>> ShopItem.objects.filter(attrs__blablabla_INTEGER__isnull=True)
[<ShopItem: T-Shirt>, <ShopItem: Rocketship>]
```

5.3 List Fields

Two fields that store lists of data, grown-up versions of Django's CommaSeparatedIntegerField, cousins of django.contrib.postgres's ArrayField. There are two versions: ListCharField, which is based on CharField and appropriate for storing lists with a small maximum size, and ListTextField, which is based on TextField and therefore suitable for lists of (near) unbounded size (the underlying LONGTEXT MySQL datatype has a maximum length of 2^{32} - 1 bytes).

class ListCharField(*base_field*, *size=None*, ***kwargs*)

A field for storing lists of data, all of which conform to the base_field.

base_field
The base type of the data that is stored in the list. Currently, must be `IntegerField`, `CharField`, or any subclass thereof - except from `ListCharField` itself.

size
Optionally set the maximum numbers of items in the list. This is only checked on form validation, not on model save!

As `ListCharField` is a subclass of `CharField`, any `CharField` options can be set too. Most importantly you'll need to set `max_length` to determine how many characters to reserve in the database.

Example instantiation:

```
from django.db.models import CharField, Model
from django_mysql.models import ListCharField

class Person(Model):
    name = CharField()
    post_nominals = ListCharField(
        base_field=CharField(max_length=10),
        size=6,
        max_length=(6 * 11)  # 6 * 10 character nominals, plus commas
    )
```

In Python simply set the field's value as a list:

```
>>> p = Person.objects.create(name='Horatio', post_nominals=['PhD', 'Esq.'])
>>> p.post_nominals
['PhD', 'Esq.']
>>> p.post_nominals.append('III')
>>> p.post_nominals
['PhD', 'Esq.', 'III']
>>> p.save()
```

Validation on save()

When performing the list-to-string conversion for the database, `ListCharField` performs some validation, and will raise `ValueError` if there is a problem, to avoid saving bad data. The following are invalid:

- Any member containing a comma in its string representation
- Any member whose string representation is the empty string

The default form field is *SimpleListField*.

class ListTextField(*base_field*, *size=None*, ***kwargs*)
The same as `ListCharField`, but backed by a `TextField` and therefore much less restricted in length. There is no `max_length` argument.

Example instantiation:

```
from django.db.models import IntegerField, Model
from django_mysql.models import ListTextField

class Widget(Model):
    widget_group_ids = ListTextField(
        base_field=IntegerField(),
        size=100,  # Maximum of 100 ids in list
    )
```

5.3.1 Querying List Fields

Warning: These fields are not built-in datatypes, and the filters use one or more SQL functions to parse the underlying string representation. They may slow down on large tables if your queries are not selective on other columns.

contains

The `contains` lookup is overridden on `ListCharField` and `ListTextField` to match where the set field contains the given element, using MySQL's `FIND_IN_SET` function (docs: MariaDB / MySQL docs).

For example:

```
>>> Person.objects.create(name='Horatio', post_nominals=['PhD', 'Esq.', 'III'])
>>> Person.objects.create(name='Severus', post_nominals=['PhD', 'DPhil'])
>>> Person.objects.create(name='Paulus', post_nominals=[])

>>> Person.objects.filter(post_nominals__contains='PhD')
[<Person: Horatio>, <Person: Severus>]

>>> Person.objects.filter(post_nominals__contains='Esq.')
[<Person: Horatio>]

>>> Person.objects.filter(post_nominals__contains='DPhil')
[<Person: Severus>]

>>> Person.objects.filter(Q(post_nominals__contains='PhD') & Q(post_nominals__
↪contains='III'))
[<Person: Horatio>]
```

Note: `ValueError` will be raised if you try `contains` with a list. It's not possible without using `AND` in the query, so you should add the filters for each item individually, as per the last example.

len

A transform that converts to the number of items in the list. For example:

```
>>> Person.objects.filter(post_nominals__len=0)
[<Person: Paulus>]

>>> Person.objects.filter(post_nominals__len=2)
[<Person: Severus>]

>>> Person.objects.filter(post_nominals__len__gt=2)
[<Person: Horatio>]
```

Index lookups

This class of lookups allows you to index into the list to check if the first occurrence of a given element is at a given position. There are no errors if it exceeds the `size` of the list. For example:

```
>>> Person.objects.filter(post_nominals__0='PhD')
[<Person: Horatio>, <Person: Severus>]

>>> Person.objects.filter(post_nominals__1='DPhil')
[<Person: Severus>]

>>> Person.objects.filter(post_nominals__100='VC')
[]
```

Warning: The underlying function, `FIND_IN_SET`, is designed for *sets*, i.e. comma-separated lists of unique elements. It therefore only allows you to query about the *first* occurrence of the given item. For example, this is a non-match:

```
>>> Person.objects.create(name='Cacistus', post_nominals=['MSc', 'MSc'])
>>> Person.objects.filter(post_nominals__1='MSc')
[]  # Cacistus does not appear because his first MSc is at position 0
```

This may be fine for your application, but be careful!

Note: `FIND_IN_SET` uses 1-based indexing for searches on comma-based strings when writing raw SQL. However these indexes use 0-based indexing to be consistent with Python.

Note: Unlike the similar feature on `django.contrib.postgres`'s `ArrayField`, 'Index transforms', these are lookups, and only allow direct value comparison rather than continued chaining with the base-field lookups. This is because the field is not a native list type in MySQL.

5.3.2 `ListF()` expressions

Similar to Django's `F` expression, this allows you to perform an atomic add and remove operations on list fields at the database level:

```
>>> from django_mysql.models import ListF
>>> Person.objects.filter(post_nominals__contains="PhD").update(
...     post_nominals=ListF('post_nominals').append('Sr.')
... )
2
>>> Person.objects.update(
...     post_nominals=ListF('post_nominals').pop()
... )
3
```

Or with attribute assignment to a model:

```
>>> horatio = Person.objects.get(name='Horatio')
>>> horatio.post_nominals = ListF('post_nominals').append('DSocSci')
>>> horatio.save()
```

class ListF(*field_name*)

You should instantiate this class with the name of the field to use, and then call one of its methods.

Note that unlike `F`, you cannot chain the methods - the SQL involved is a bit too complicated, and thus only single operations are supported.

append(*value*)
Adds the value of the given expression to the (right hand) end of the list, like `list.append`:

```
>>> Person.objects.create(name='Horatio', post_nominals=['PhD', 'Esq.', 'III
↪'])
>>> Person.objects.update(
...     post_nominals=ListF('post_nominals').append('DSocSci')
... )
>>> Person.objects.get().full_name
"Horatio Phd Esq. III DSocSci"
```

appendleft(*value*)
Adds the value of the given expression to the (left hand) end of the list, like `deque.appendleft`:

```
>>> Person.objects.update(
...     post_nominals=ListF('post_nominals').appendleft('BArch')
... )
>>> Person.objects.get().full_name
"Horatio BArch Phd Esq. III DSocSci"
```

pop()
Takes one value from the (right hand) end of the list, like `list.pop`:

```
>>> Person.objects.update(
...     post_nominals=ListF('post_nominals').pop()
... )
>>> Person.objects.get().full_name
"Horatio BArch Phd Esq. III"
```

popleft()
Takes one value off the (left hand) end of the list, like `deque.popleft`:

```
>>> Person.objects.update(
...     post_nominals=ListF('post_nominals').popleft()
... )
>>> Person.objects.get().full_name
"Horatio Phd Esq. III"
```

Warning: All the above methods use SQL expressions with user variables in their queries, all of which start with `@tmp_`. This shouldn't affect you much, but if you use user variables in your queries, beware for any conflicts.

5.4 Set Fields

Two fields that store sets of a base field in comma-separated strings - cousins of Django's `CommaSeparatedIntegerField`. There are two versions: `SetCharField`, which is based on `CharField` and appropriate for storing sets with a small maximum size, and `SetTextField`, which is based on `TextField` and therefore suitable for sets of (near) unbounded size (the underlying `LONGTEXT` MySQL datatype has a maximum length of 2^{32} - 1 bytes).

SetCharField(base_field, size=None, **kwargs):

A field for storing sets of data, which all conform to the `base_field`.

base_field

The base type of the data that is stored in the set. Currently, must be `IntegerField`, `CharField`, or any subclass thereof - except from `SetCharField` itself.

size

Optionally set the maximum number of elements in the set. This is only checked on form validation, not on model save!

As `SetCharField` is a subclass of `CharField`, any `CharField` options can be set too. Most importantly you'll need to set `max_length` to determine how many characters to reserve in the database.

Example instantiation:

```
from django.db.models import IntegerField, Model
from django_mysql.models import SetCharField

class LotteryTicket(Model):
    numbers = SetCharField(
        base_field=IntegerField(),
        size=6,
        max_length=(6 * 3)  # 6 two digit numbers plus commas
    )
```

In Python simply set the field's value as a set:

```
>>> lt = LotteryTicket.objects.create(numbers={1, 2, 4, 8, 16, 32})
>>> lt.numbers
{1, 2, 4, 8, 16, 32}
>>> lt.numbers.remove(1)
>>> lt.numbers.add(3)
>>> lt.numbers
{32, 3, 2, 4, 8, 16}
>>> lt.save()
```

Validation on save()

When performing the set-to-string conversion for the database, `SetCharField` performs some validation, and will raise `ValueError` if there is a problem, to avoid saving bad data. The following are invalid:

- If there is a comma in any member's string representation
- If the empty string is stored.

The default form field is *SimpleSetField*.

SetTextField(base_field, size=None, **kwargs):

The same as `SetCharField`, but backed by a `TextField` and therefore much less restricted in length. There is no `max_length` argument.

Example instantiation:

```
from django.db.models import IntegerField, Model
from django_mysql.models import SetTextField

class Post(Model):
    tags = SetTextField(
```

(continues on next page)

(continued from previous page)

```
        base_field=CharField(max_length=32),
    )
```

5.4.1 Querying Set Fields

Warning: These fields are not built-in datatypes, and the filters use one or more SQL functions to parse the underlying string representation. They may slow down on large tables if your queries are not selective on other columns.

contains

The `contains` lookup is overridden on `SetCharField` and `SetTextField` to match where the set field contains the given element, using MySQL's `FIND_IN_SET` (docs: MariaDB / MySQL).

For example:

```
>>> Post.objects.create(name='First post', tags={'thoughts', 'django'})
>>> Post.objects.create(name='Second post', tags={'thoughts'})
>>> Post.objects.create(name='Third post', tags={'tutorial', 'django'})

>>> Post.objects.filter(tags__contains='thoughts')
[<Post: First post>, <Post: Second post>]

>>> Post.objects.filter(tags__contains='django')
[<Post: First post>, <Post: Third post>]

>>> Post.objects.filter(Q(tags__contains='django') & Q(tags__contains='thoughts'))
[<Post: First post>]
```

Note: `ValueError` will be raised if you try `contains` with a set. It's not possible without using `AND` in the query, so you should add the filters for each item individually, as per the last example.

len

A transform that converts to the number of items in the set. For example:

```
>>> Post.objects.filter(tags__len=1)
[<Post: Second post>]

>>> Post.objects.filter(tags__len=2)
[<Post: First post>, <Post: Third post>]

>>> Post.objects.filter(tags__len__lt=2)
[<Post: Second post>]
```

5.4.2 `SetF()` expressions

Similar to Django's `F` expression, this allows you to perform an atomic add or remove on a set field at the database level:

```
>>> from django_mysql.models import SetF
>>> Post.objects.filter(tags__contains="django").update(tags=SetF('tags').add(
↪'programming'))
2
>>> Post.objects.update(tags=SetF('tags').remove('thoughts'))
2
```

Or with attribute assignment to a model:

```
>>> post = Post.objects.earliest('id')
>>> post.tags = SetF('tags').add('python')
>>> post.save()
```

class SetF(*field_name*)

You should instantiate this class with the name of the field to use, and then call one of its two methods with a value to be added/removed.

Note that unlike `F`, you cannot chain the methods - the SQL involved is a bit too complicated, and thus you can only perform a single addition or removal.

add(*value*)

Takes an expression and returns a new expression that will take the value of the original field and add the value to the set if it is not contained:

```
post.tags = SetF('tags').add('python')
post.save()
```

remove(*value*)

Takes an expression and returns a new expression that will remove the given item from the set field if it is present:

```
post.tags = SetF('tags').remove('python')
post.save()
```

Warning: Both of the above methods use SQL expressions with user variables in their queries, all of which start with `@tmp_`. This shouldn't affect you much, but if you use user variables in your queries, beware for any conflicts.

5.5 EnumField

Using a `CharField` with a limited set of strings leads to inefficient data storage since the string value is stored over and over on disk. MySQL's `ENUM` type allows a more compact representation of such columns by storing the list of strings just once and using an integer in each row to refer to which string is there. `EnumField` allows you to use the `ENUM` type with Django.

Docs: MySQL / MariaDB.

class EnumField(*choices*, ***kwargs*)

A subclass of Django's `Charfield` that uses a MySQL `ENUM` for storage.

`choices` is a standard Django argument for any field class, however it is required for `EnumField`. It can either be a list of strings, or a list of two-tuples of strings, where the first element in each tuple is the value used, and the second the human readable name used in forms. For example:

```
from django_mysql.models import EnumField

class BookCover(Model):
    color = EnumField(choices=['red', 'green', 'blue'])

class Book(Model):
    color = EnumField(choices=[
      ('red', 'Bright Red'),
      ('green', 'Vibrant Green'),
      'blue',  # human readable name will be set to "blue"
    ])
```

Warning: It is possible to append new values to `choices` in migrations, as well as edit the *human readable* names of existing choices.

However, editing or removing existing choice values will error if MySQL Strict Mode is on, and replace the values with the empty string if it is not.

Also the empty string has strange behaviour with `ENUM`, acting somewhat like `NULL`, but not entirely; therefore it's recommended you have Strict Mode on.

5.6 Resizable Text/Binary Fields

Django's `TextField` and `BinaryField` fields are fixed at the MySQL level to use the maximum size class for the `BLOB` and `TEXT` data types. This is fine for most applications, however if you are working with a legacy database, or you want to be stricter about the maximum size of data that can be stored, you might want one of the other sizes.

The following field classes are simple subclasses that allow you to provide an extra parameter to determine which size class to use. They work with migrations, allowing you to swap them for the existing Django class and then use a migration to change their size class. This might help when taking over a legacy database for example.

Warning: One caveat on migrations - you won't be able to use a *default* properly at sizes other than `LONGTEXT`/`LONGBLOB` until Django 1.9 which includes a fix from Django Ticket 24846. This is anyway mostly due to a MySQL limitation - `DEFAULT` cannot be specified, other than the empty string, for `TEXT` and `BLOB` columns.

Docs: MySQL / MariaDB.

class SizedTextField(*size_class*, ***kwargs*)

A subclass of Django's `TextField` that allows you to use the other sizes of `TEXT` data type. Set `size_class` to:

- `1` for a `TINYTEXT` field, which has a maximum length of 255 bytes
- `2` for a `TEXT` field, which has a maximum length of 65,535 bytes
- `3` for a `MEDIUMTEXT` field, which has a maximum length of 16,777,215 bytes (16MiB)
- `4` for a `LONGTEXT` field, which has a maximum length of 4,294,967,295 bytes (4GiB)

class SizedBinaryField(*size_class*, ***kwargs*)

A subclass of Django's `BinaryField` that allows you to use the other sizes of `BLOB` data type. Set `size_class` to:

- `1` for a `TINYBLOB` field, which has a maximum length of 255 bytes
- `2` for a `BLOB` field, which has a maximum length of 65,535 bytes
- `3` for a `MEDIUMBLOB` field, which has a maximum length of 16,777,215 bytes (16MiB)
- `4` for a `LONGBLOB` field, which has a maximum length of 4,294,967,295 bytes (4GiB)

5.7 BIT(1) Boolean Fields

Some database systems, such as the Java Hibernate ORM, don't use MySQL's `bool` data type for storing boolean flags and instead use `BIT(1)`. Django's default `BooleanField` and `NullBooleanField` classes can't work with this.

The following subclasses are boolean fields that work with `BIT(1)` columns that will help when connecting to a legacy database. If you are using `inspectdb` to generate models from the database, use these to replace the `TextField` output for your `BIT(1)` columns.

class Bit1BooleanField
A subclass of Django's `BooleanField` that uses the `BIT(1)` column type instead of `bool`.

class NullBit1BooleanField
A subclass of Django's `NullBooleanField` that uses the `BIT(1)` column type instead of `bool`.

CHAPTER 6

Field Lookups

ORM extensions for filtering. These are all automatically added for the appropriate field types when `django_mysql` is in your `INSTALLED_APPS`. Note that lookups specific to included *model fields* are documented with the field, rather than here.

6.1 Case-sensitive String Comparison

MySQL string comparison has a case-sensitivity dependent on the collation of your tables/columns, as the Django manual describes. However, it is possible to query in a case-sensitive manner even when your data is not stored with a case-sensitive collation, using the `BINARY` keyword. The following lookup adds that capability to the ORM for `CharField`, `TextField`, and subclasses thereof.

6.1.1 case_exact

Exact, case-sensitive match for character columns, no matter the underlying collation:

```
>>> Author.objects.filter(name__case_exact="dickens")
[]
>>> Author.objects.filter(name__case_exact="Dickens")
[<Author: Dickens>]
```

6.2 Soundex

MySQL implements the Soundex algorithm with its `SOUNDEX` function, allowing you to find words sounding similar to each other (in English only, regrettably). These lookups allow you to use that function in the ORM and are added for `CharField` and `TextField`.

6.2.1 soundex

Match a given soundex string:

```
>>> Author.objects.filter(name__soundex='R163')
[<Author: Robert>, <Author: Rupert>]
```

SQL equivalent:

```
SELECT ... WHERE SOUNDEX(`name`) = 'R163'
```

6.2.2 sounds_like

Match the `SOUNDEX` of the given string:

```
>>> Author.objects.filter(name__sounds_like='Robert')
[<Author: Robert>, <Author: Rupert>]
```

SQL equivalent:

```
SELECT ... WHERE `name` SOUNDS LIKE 'Robert'
```

CHAPTER 7

Aggregates

MySQL-specific database aggregates for the ORM.

The following can be imported from `django_mysql.models`.

class BitAnd(*column*)
Returns an `int` of the bitwise `AND` of all input values, or 18446744073709551615 (a `BIGINT UNSIGNED` with all bits set to 1) if no rows match.

Docs: MySQL / MariaDB.

Example usage:

```
>>> Book.objects.create(bitfield=29)
>>> Book.objects.create(bitfield=15)
>>> Book.objects.all().aggregate(BitAnd('bitfield'))
{'bitfield__bitand': 13}
```

class BitOr(*column*)
Returns an `int` of the bitwise `OR` of all input values, or 0 if no rows match.

Docs: MySQL / MariaDB.

Example usage:

```
>>> Book.objects.create(bitfield=29)
>>> Book.objects.create(bitfield=15)
>>> Book.objects.all().aggregate(BitOr('bitfield'))
{'bitfield__bitor': 31}
```

class BitXor(*column*)
Returns an `int` of the bitwise `XOR` of all input values, or 0 if no rows match.

Docs: MySQL / MariaDB.

Example usage:

```
>>> Book.objects.create(bitfield=11)
>>> Book.objects.create(bitfield=3)
>>> Book.objects.all().aggregate(BitXor('bitfield'))
{'bitfield__bitxor': 8}
```

class GroupConcat(*column*, *distinct=False*, *separator=','*, *ordering=None*)

An aggregate that concatenates values from a column of the grouped rows. Useful mostly for bringing back lists of ids in a single query.

Docs: MySQL / MariaDB.

Example usage:

```
>>> from django_mysql.models import GroupConcat
>>> author = Author.objects.annotate(
...     book_ids=GroupConcat('books__id')
... ).get(name="William Shakespeare")
>>> author.book_ids
"1,2,5,17,29"
```

> **Warning:** MySQL will truncate the value at the value of `group_concat_max_len`, which by default is quite low at 1024 characters. You should probably increase it if you're using this for any sizeable groups.
>
> `group_concat_max_len` docs: MySQL / MariaDB.

Optional arguments:

distinct=False

If set to `True`, removes duplicates from the group.

separator=','

By default the separator is a comma. You can use any other string as a separator, including the empty string.

> **Warning:** Due to limitations in the Django aggregate API, this is not protected against SQL injection. Don't pass in user input for the separator.

ordering=None

By default no guarantee is made on the order the values will be in pre-concatenation. Set ordering to `'asc'` to sort them in ascending order, and `'desc'` for descending order. For example:

```
>>> Author.objects.annotate(
...     book_ids=GroupConcat('books__id', ordering='asc')
... )
```

CHAPTER 8

Database Functions

MySQL-specific database functions for the ORM.

The following can be imported from `django_mysql.models.functions`.

Note: Functions were only added to Django in version 1.8.

8.1 Comparison Functions

class Greatest(**expressions*)

Note: Django core has included this since Django 1.9 as `django.db.models.functions.Greatest`; it's preferable to use that instead.

With two or more arguments, returns the largest (maximum-valued) argument.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.filter(sales_eu=Greatest('sales_eu', 'sales_us'))
```

class Least(**expressions*)

Note: Django core has included this since Django 1.9 as `django.db.models.functions.Least`; it's preferable to use that instead.

With two or more arguments, returns the smallest (minimum-valued) argument.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.filter(sales_eu=Least('sales_eu', 'sales_us'))
```

8.2 Control Flow Functions

class If(*condition*, *true*, *false=None*)

Evaluates the expression `condition` and returns the value of the expression `true` if true, and the result of expression `false` if false. If `false` is not given, it will be `Value(None)`, i.e. `NULL`.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.annotate(
...     is_william=If(Q(name__startswith='William '), True, False)
... ).values_list('name', 'is_william')
[('William Shakespeare', True),
 ('Ian Fleming', False),
 ('William Wordsworth', True)]
```

8.3 Numeric Functions

class Abs(*expression*)

Returns the absolute (non-negative) value of `expression`. If `expression` is not a number, it is converted to a numeric type.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.annotate(abs_wealth=Abs('dollars'))
```

class Ceiling(*expression*)

Returns the smallest integer value not less than *expression*.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.annotate(years_ceiling=Ceiling('age'))
```

class CRC32(*expression*)

Computes a cyclic redundancy check value and returns a 32-bit unsigned value. The result is `NULL` if the argument is `NULL`. The argument is expected to be a string and (if possible) is treated as one if it is not.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.annotate(description_crc=CRC32('description'))
```

class Floor(*expression*)
Returns the largest integer value not greater than `expression`.

Docs: MySQL / MariaDB.

Usage example:

```
>> Author.objects.annotate(age_years=Floor('age'))
```

class Round(*expression*, *places=0*)
Rounds the argument `expression` to `places` decimal places. The rounding algorithm depends on the data type of `expression`. `places` defaults to 0 if not specified. `places` can be negative to cause `places` digits left of the decimal point of the value `expression` to become zero.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.annotate(kilo_sales=Round('sales', -3))
```

class Sign(*expression*)
Returns the sign of the argument as -1, 0, or 1, depending on whether `expression` is negative, zero, or positive.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.annotate(wealth_sign=Sign('wealth'))
```

8.4 String Functions

class ConcatWS(**expressions*, *separator=', '*)
`ConcatWS` stands for Concatenate With Separator and is a special form of `Concat` (included in Django). It concatenates all of its argument expressions as strings with the given `separator`. Since `NULL` values are skipped, unlike in `Concat`, you can use the empty string as a separator and it acts as a `NULL`-safe version of `Concat`.

If `separator` is a string, it will be turned into a `Value`. If you wish to join with the value of a field, you can pass in an `F` object for that field.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.annotate(sales_list=ConcatWS('sales_eu', 'sales_us'))
```

class ELT(*number*, *values*)
Given a numerical expression `number`, it returns the `number`th element from `values`, 1-indexed. If `number` is less than 1 or greater than the number of expressions, it will return `None`. It is the complement of the `Field` function.

Note that if `number` is a string, it will refer to a field, whereas members of `values` that are strings will be wrapped with `Value` automatically and thus interpreted as the given string. This is for convenience with the most common usage pattern where you have the list pre- loaded in python, e.g. a `choices` field. If you want to refer to a column, use Django's `F()` class.

Docs: MySQL / MariaDB.

Usage example:

```
>>> # Say Person.life_state is either 1 (alive), 2 (dead), or 3 (M.I.A.)
>>> Person.objects.annotate(
...     state_name=ELT('life_state', ['Alive', 'Dead', 'M.I.A.'])
... )
```

class Field(*expression*, *values*)

Given an `expression` and a list of strings `values`, returns the 1-indexed location of the `expression`'s value in `values`, or 0 if not found. This is commonly used with `order_by` to keep groups of elements together. It is the complement of the `ELT` function.

Note that if `expression` is a string, it will refer to a field, whereas if any member of `values` is a string, it will automatically be wrapped with `Value` and refer to the given string. This is for convenience with the most common usage pattern where you have the list of things pre-loaded in Python, e.g. in a field's `choices`. If you want to refer to a column, use Django's `F()` class.

Docs: MySQL / MariaDB.

Usage example:

```
>>> # Females, then males - but other values of gender (e.g. empty string) first
>>> Person.objects.all().order_by(
...     Field('gender', ['Female', 'Male'])
... )
```

8.5 XML Functions

class UpdateXML(*xml_target*, *xpath_expr*, *new_xml*)

Returns the XML fragment `xml_target` with the single match for `xpath_expr` replaced with the xml fragment `new_xml`. If nothing matches `xpath_expr`, or if multiple matches are found, the original `xml_target` is returned unchanged.

This can be used for single-query updates of text fields containing XML.

Note that if `xml_target` is given as a string, it will refer to a column, whilst if either `xpath_expr` or `new_xml` are strings, they will be used as strings directly. If you want `xpath_expr` or `new_xml` to refer to columns, use Django's `F()` class.

Docs: MySQL / MariaDB.

Usage example:

```
# Remove 'sagacity' from all authors' xml_attrs
>>> Author.objects.update(
...     xml_attrs=UpdateXML('xml_attrs', '/sagacity', '')
... )
```

class XMLExtractValue(*xml_frag*, *xpath_expr*)

Returns the text (`CDATA`) of the first text node which is a child of the element(s) in the XML fragment `xml_frag` matched by the XPath expression `xpath_expr`. In SQL this function is called `ExtractValue`; the class has the `XML` prefix to make it clearer what kind of values are it extracts.

Note that if `xml_frag` is given as a string, it will refer to a column, whilst if `xpath_expr` is a string, it will be used as a string. If you want `xpath_expr` to refer to a column, use Django's `F()` class.

Docs: MySQL / MariaDB.

Usage example:

```
# Count the number of authors with 'sagacity' in their xml_attrs
>>> num_authors_with_sagacity = Author.objects.annotate(
...     has_sagacity=XMLExtractValue('xml_attrs', 'count(/sagacity)')
... ).filter(has_sagacity='1').count()
```

8.6 Regexp Functions

Note: These work with MariaDB 10.0.5+ only, which includes PCRE regular expressions and these extra functions to use them. More information can be found in its documentation.

class RegexpInstr(*expression*, *regex*)

Returns the 1-indexed position of the first occurrence of the regular expression `regex` in the string value of `expression`, or 0 if it was not found.

Note that if `expression` is given as a string, it will refer to a column, whilst if `regex` is a string, it will be used as a string. If you want `regex` to refer to a column, use Django's `F()` class.

Docs: MariaDB.

Usage example:

```
>>> Author.objects.annotate(name_pos=RegexpInstr('name', r'ens')) \
...               .filter(name_pos__gt=0)
[<Author: Charles Dickens>, <Author: Robert Louis Stevenson>]
```

class RegexpReplace(*expression*, *regex*, *replace*)

Returns the string value of `expression` with all occurrences of the regular expression `regex` replaced by the string `replace`. If no occurrences are found, then subject is returned as is.

Note that if `expression` is given as a string, it will refer to a column, whilst if either `regex` or `replace` are strings, they will be used as strings. If you want `regex` or `replace` to refer to columns, use Django's `F()` class.

Docs: MariaDB.

Usage example:

```
>>> Author.objects.create(name="Charles Dickens")
>>> Author.objects.create(name="Roald Dahl")
>>> qs = Author.objects.annotate(
...     surname_first=RegexpReplace('name', r'^(.*) (.*)$', r'\2, \1')
... ).order_by('surname_first')
>>> qs
[<Author: Roald Dahl>, <Author: Charles Dickens>]
>>> qs[0].surname_first
"Dahl, Roald"
```

class RegexpSubstr(*expression*, *regex*)

Returns the part of the string value of `expression` that matches the regular expression `regex`, or an empty string if `regex` was not found.

Note that if `expression` is given as a string, it will refer to a column, whilst if `regex` is a string, it will be used as a string. If you want `regex` to refer to a column, use Django's `F()` class.

Docs: MariaDB.

Usage example:

```
>>> Author.objects.create(name="Euripides")
>>> Author.objects.create(name="Frank Miller")
>>> Author.objects.create(name="Sophocles")
>>> Author.objects.annotate(
...     name_has_space=CharLength(RegexpSubstr('name', r'\s'))
... ).filter(name_has_space=0)
[<Author: Euripides>, <Author: Sophocles>]
```

8.7 Encryption Functions

class MD5(*expression*)

Calculates an MD5 128-bit checksum for the string `expression`.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.annotate(description_md5=MD5('description'))
```

class SHA1(*expression*)

Calculates an SHA-1 160-bit checksum for the string `expression`, as described in RFC 3174 (Secure Hash Algorithm).

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.annotate(description_sha=SHA1('description'))
```

class SHA2(*expression*, *hash_len=512*)

Given a string `expression`, calculates a SHA-2 checksum, which is considered more cryptographically secure than its SHA-1 equivalent. The SHA-2 family includes SHA-224, SHA-256, SHA-384, and SHA-512, and the `hash_len` must correspond to one of these, i.e. 224, 256, 384 or 512. The default for `hash_len` is 512.

Docs: MySQL / MariaDB.

Usage example:

```
>>> Author.objects.annotate(description_sha256=SHA2('description', 256))
```

8.8 Information Functions

class LastInsertId(*expression=None*)

With no argument, returns the last value added to an auto-increment column, or set by another call to `LastInsertId` with an argument. With an argument, sets the 'last insert id' value to the value of the given expression, and returns that value. This can be used to implement simple `UPDATE ... RETURNING` style queries.

This function also has a class method:

get(*using=DEFAULT_DB_ALIAS*)

Returns the value set by a call to `LastInsertId()` with an argument, by performing a single query. It is stored per-connection, hence you may need to pass the alias of the connection that set the `LastInsertId` as `using`.

Note: Any queries on the database connection between setting `LastInsertId` and calling `LastInsertId.get()` can reset the value. These might come from Django, which can issue multiple queries for `update()` with multi-table inheritance, or for `delete()` with cascading.

Docs: MySQL / MariaDB.

Usage examples:

```
>>> Countable.objects.filter(id=1).update(counter=LastInsertId('counter') + 1)
1
>>> # Get the pre-increase value of 'counter' as stored on the server
>>> LastInsertId.get()
242

>>> Author.objects.filter(id=1, age=LastInsertId('age')).delete()
1
>>> # We can also use the stored value directly in a query
>>> Author.objects.filter(id=2).update(age=LastInsertId())
1
>>> Author.objects.get(id=2).age
35
```

8.9 JSON Database Functions

These are for MySQL 5.7+ only, for use with JSON data stored in a `CharField`, `TextField`, or most importantly, a *`JSONField`* (with which the functions are much faster due to the JSON being stored in a binary representation).

These functions use JSON paths to address content inside JSON documents - for more information on their syntax, refer to the MySQL documentation.

`class JSONExtract`(*expression*, **paths*, *output_field=None*)

Given `expression` that resolves to some JSON data, extract the given JSON paths. If there is a single path, the plain value is returned; if there is more than one path, the output is a JSON array with the list of values represented by the paths. If the expression does not match for a particular JSON object, returns `NULL`.

If only one path is given, `output_field` may also be given as a model field instance like `IntegerField()`, into which Django will load the value; the default is `JSONField()`, as it supports all return types including the array of values for multiple paths.

Note that if `expression` is a string, it will refer to a field, whereas members of `paths` that are strings will be wrapped with `Value` automatically and thus interpreted as the given string. If you want any of `paths` to refer to a field, use Django's `F()` class.

Docs: MySQL.

Usage examples:

```
>>> # Fetch a list of tuples (id, size_or_None) for all ShopItems
>>> ShopItem.objects.annotate(
...     size=JSONExtract('attrs', '$.size')
... ).values_list('id', 'size')
[(1, '3m'), (3, '5nm'), (8, None)]
>>> # Fetch the distinct values of attrs['colours'][0] for all items
>>> ShopItem.objects.annotate(
...     primary_colour=JSONExtract('attrs', '$.colours[0]')
```

(continues on next page)

(continued from previous page)

```
... ).distinct().values_list('primary_colour', flat=True)
['Red', 'Blue', None]
```

class JSONKeys(*expression*, *path=None*)

Given `expression` that resolves to some JSON data containing a JSON object, return the keys in that top-level object as a JSON array, or if `path` is given, return the keys at that path. If the path does not match, or if `expression` is not a JSON object (e.g. it contains a JSON array instead), returns `NULL`.

Note that if `expression` is a string, it will refer to a field, whereas if `path` is a string it will be wrapped with `Value` automatically and thus interpreted as the given string. If you want `path` to refer to a field, use Django's `F()` class.

Docs: MySQL.

```
>>> # Fetch the top-level keys for the first item
>>> ShopItem.objects.annotate(
...     keys=JSONKeys('attrs')
... ).values_list('keys', flat=True)[0]
['size', 'colours', 'age', 'price', 'origin']
>>> # Fetch the keys in 'origin' for the first item
>>> ShopItem.objects.annotate(
...     keys=JSONKeys('attrs', '$.origin')
... ).values_list('keys', flat=True)[0]
['continent', 'country', 'town']
```

class JSONLength(*expression*, *path=None*)

Given `expression` that resolves to some JSON data, return the length of that data, or if `path` is given, return the length of the data at that path. If the path does not match, or if `expression` is `NULL` it returns `NULL`.

As per the MySQL documentation, the length of a document is determined as follows:

- The length of a scalar is 1.
- The length of an array is the number of array elements.
- The length of an object is the number of object members.
- The length does not count the length of nested arrays or objects.

Note that if `expression` is a string, it will refer to a field, whereas if `path` is a string it will be wrapped with `Value` automatically and thus interpreted as the given string. If you want `path` to refer to a field, use Django's `F()` class.

Docs: MySQL.

```
>>> # Which ShopItems don't have more than three colours?
>>> ShopItem.objects.annotate(
...     num_colours=JSONLength('attrs', '$.colours')
... ).filter(num_colours__gt=3)
[<ShopItem: Rainbow Wheel>, <ShopItem: Hard Candies>]
```

class JSONInsert(*expression*, *data*)

Given `expression` that resolves to some JSON data, adds to it using the dictionary `data` of JSON paths to new values. If any JSON path in the `data` dictionary does not match, or if `expression` is `NULL`, it returns `NULL`. Paths that already exist in the original data are ignored.

Note that if `expression` is a string, it will refer to a field, whereas keys and values within the `pairs` dictionary will be wrapped with `Value` automatically and thus interpreted as the given string. If you want a key or value to refer to a field, use Django's `F()` class.

Docs: MySQL.

```
>>> # Add power_level = 0 for those items that don't have power_level
>>> ShopItem.objects.update(
...     attrs=JSONInsert('attrs', {'$.power_level': 0})
... )
```

class JSONReplace(*expression*, *data*)

Given `expression` that resolves to some JSON data, replaces existing paths in it using the dictionary `data` of JSON paths to new values. If any JSON path within the `data` dictionary does not match, or if `expression` is `NULL`, it returns `NULL`. Paths that do not exist in the original data are ignored.

Note that if `expression` is a string, it will refer to a field, whereas keys and values within the `pairs` dictionary will be wrapped with `Value` automatically and thus interpreted as the given string. If you want a key or value to refer to a field, use Django's `F()` class.

Docs: MySQL.

```
>>> # Reset all items' monthly_sales to 0 directly in MySQL
>>> ShopItem.objects.update(
...     attrs=JSONReplace('attrs', {'$.monthly_sales': 0})
... )
```

class JSONSet(*expression*, *data*)

Given `expression` that resolves to some JSON data, updates it using the dictionary `data` of JSON paths to new values. If any of the JSON paths within the data dictionary does not match, or if `expression` is `NULL`, it returns `NULL`. All paths can be modified - those that did not exist before and those that did.

Note that if `expression` is a string, it will refer to a field, whereas keys and values within the `data` dictionary will be wrapped with `Value` automatically and thus interpreted as the given string. If you want a key or value to refer to a field, use Django's `F()` class.

Docs: MySQL.

```
>>> # Modify 'size' value to '10m' directly in MySQL
>>> shop_item = ShopItem.objects.latest()
>>> shop_item.attrs = JSONSet('attrs', {'$.size': '10m'})
>>> shop_item.save()
```

class JSONArrayAppend(*expression*, *data*)

Given `expression` that resolves to some JSON data, adds to it using the dictionary `data` of JSON paths to new values. If a path selects an array, the new value will be appended to it. On the other hand, if a path selects a scalar or object value, that value is autowrapped within an array and the new value is added to that array. If any of the JSON paths within the data dictionary does not match, or if `expression` is `NULL`, it returns `NULL`.

Note that if `expression` is a string, it will refer to a field, whereas keys and values within the `data` dictionary will be wrapped with `Value` automatically and thus interpreted as the given string. If you want a key or value to refer to a field, use Django's `F()` class.

Docs: MySQL.

```
>>> # Append the string '10m' to the array 'sizes' directly in MySQL
>>> shop_item = ShopItem.objects.latest()
>>> shop_item.attrs = JSONArrayAppend('attrs', {'$.sizes': '10m'})
>>> shop_item.save()
```

8.10 Dynamic Columns Functions

These are MariaDB 10.0+ only, and for use with `DynamicField`.

class AsType(*expression*, *data_type*)

A partial function that should be used as part of a `ColumnAdd` expression when you want to ensure that `expression` will be stored as a given type `data_type`. The possible values for `data_type` are the same as documented for the `DynamicField` lookups.

Note that this is not a valid standalone function and must be used as part of `ColumnAdd` - see below.

class ColumnAdd(*expression*, *to_add*)

Given `expression` that resolves to a `DynamicField` (most often a field name), add/update with the dictionary `to_add` and return the new Dynamic Columns value. This can be used for atomic single-query updates on Dynamic Columns.

Note that you can add optional types (and you should!). These can not be drawn from the `spec` of the `DynamicField` due to ORM restrictions, so there are no guarantees about the types that will get used if you do not. To add a type cast, wrap the value with an `AsType` (above) - see examples below.

Docs: MariaDB.

Usage examples:

```
>>> # Add default 'for_sale' as INTEGER 1 to every item
>>> ShopItem.objects.update(
...     attrs=ColumnAdd('attrs', {'for_sale': AsType(1, 'INTEGER')})
... )
>>> # Fix some data
>>> ShopItem.objects.filter(attrs__size='L').update(
...     attrs=ColumnAdd('attrs', {'size': AsType('Large', 'CHAR')})
... )
```

class ColumnDelete(*expression*, **to_delete*)

Given `expression` that resolves to a `DynamicField` (most often a field name), delete the columns listed by the other expressions `to_delete`, and return the new Dynamic Columns value. This can be used for atomic single-query deletions on Dynamic Columns.

Note that strings in `to_delete` will be wrapped with `Value` automatically and thus interpreted as the given string - if they weren't, Django would interpret them as meaning "the value in this (non-dynamic) column". If you do mean that, use `F('fieldname')`.

Docs: MariaDB.

Usage examples:

```
>>> # Remove 'for_sail' and 'for_purchase' from every item
>>> ShopItem.objects.update(
...     attrs=ColumnDelete('attrs', 'for_sail', 'for_purchase')
... )
```

class ColumnGet(*expression*, *name*, *data_type*)

Given `expression` that resolves to a `DynamicField` (most often a field name), return the value of the column `name` when cast to the type `data_type`, or `NULL` / `None` if the column does not exist. This can be used to select a subset of column values when you don't want to fetch the whole blob. The possible values for `data_type` are the same as documented for the `DynamicField` lookups.

Docs: MariaDB.

Usage examples:

```
>>> # Fetch a list of tuples (id, size_or_None) for all items
>>> ShopItem.objects.annotate(
...     size=ColumnGet('attrs', 'size', 'CHAR')
... ).values_list('id', 'size')
>>> # Fetch the distinct values of attrs['seller']['url'] for all items
>>> ShopItem.objects.annotate(
...     seller_url=ColumnGet(ColumnGet('attrs', 'seller', 'BINARY'), 'url', 'CHAR
↪')
... ).distinct().values_list('seller_url', flat=True)
```

CHAPTER 9

Migration Operations

MySQL-specific migration operations that can all be imported from `django_mysql.operations`.

9.1 Install Plugin

class InstallPlugin(*name*, *soname*)

An `Operation` subclass that installs a MySQL plugin. Runs `INSTALL PLUGIN name SONAME soname`, but does a check to see if the plugin is already installed to make it more idempotent.

Docs: MySQL / MariaDB.

name

This is a required argument. The name of the plugin to install.

soname

This is a required argument. The name of the library to install the plugin from. Note that on MySQL you must include the extension (e.g. `.so`, `.dll`) whilst on MariaDB you may skip it to keep the operation platform-independent.

Example usage:

```
from django.db import migrations
from django_mysql.operations import InstallPlugin


class Migration(migrations.Migration):

    dependencies = []

    operations = [
        # Install https://mariadb.com/kb/en/mariadb/metadata_lock_info/
        InstallPlugin("metadata_lock_info", "metadata_lock_info.so")
    ]
```

9.2 Install SOName

class InstallSOName(*soname*)

MariaDB only.

An `Operation` subclass that installs a MariaDB plugin library. One library may contain multiple plugins that work together, this installs all of the plugins in the named library file. Runs `INSTALL SONAME soname`. Note that unlike `InstallPlugin`, there is no idempotency check to see if the library is already installed, since there is no way of knowing if all the plugins inside the library are installed.

Docs: MariaDB.

soname

This is a required argument. The name of the library to install the plugin from. You may skip the file extension (e.g. `.so`, `.dll`) to keep the operation platform-independent.

Example usage:

```
from django.db import migrations
from django_mysql.operations import InstallSOName


class Migration(migrations.Migration):

    dependencies = []

    operations = [
        # Install https://mariadb.com/kb/en/mariadb/metadata_lock_info/
        InstallSOName("metadata_lock_info")
    ]
```

9.3 Alter Storage Engine

class AlterStorageEngine(*name*, *to_engine*, *from_engine=None*)

An `Operation` subclass that alters the model's table's storage engine. Because Django has no knowledge of storage engines, you must provide the previous storage engine for the operation to be reversible.

name

This is a required argument. The name of the model to alter.

to_engine

This is a required argument. The storage engine to move the model to.

from_engine

This is an optional argument. The storage engine the model is moving from. If you do not provide this, the operation is not reversible.

Note: If you're using this to move from MyISAM to InnoDB, there's a page for you in the MariaDB knowledge base - Converting Tables from MyISAM to InnoDB.

Example usage:

```
from django.db import migrations
from django_mysql.operations import AlterStorageEngine

```

(continues on next page)

(continued from previous page)

```
class Migration(migrations.Migration):

    dependencies = []

    operations = [
        AlterStorageEngine("Pony", from_engine="MyISAM", to_engine="InnoDB")
    ]
```

CHAPTER 10

Form Fields

The following can be imported from `django_mysql.forms`.

10.1 JSONField

class JSONField

A field which accepts JSON encoded data for a *JSONField*. It is represented by an HTML `<textarea>`.

User friendly forms

`JSONField` is not particularly user friendly in most cases, however it is a useful way to format data from a client-side widget for submission to the server.

10.2 SimpleListField

class SimpleListField(*base_field*, *max_length=None*, *min_length=None*)

A simple field which maps to a list, with items separated by commas. It is represented by an HTML `<input>`. Empty items, resulting from leading, trailing, or double commas, are disallowed.

base_field

This is a required argument.

It specifies the underlying form field for the set. It is not used to render any HTML, but it does process and validate the submitted data. For example:

```
>>> from django import forms
>>> from django_mysql.forms import SimpleListField

>>> class NumberListForm(forms.Form):
...     numbers = SimpleListField(forms.IntegerField())
```

(continues on next page)

(continued from previous page)

```
>>> form = NumberListForm({'numbers': '1,2,3'})
>>> form.is_valid()
True
>>> form.cleaned_data
{'numbers': [1, 2, 3]}

>>> form = NumberListForm({'numbers': '1,2,a'})
>>> form.is_valid()
False
```

max_length
: This is an optional argument which validates that the list does not exceed the given length.

min_length
: This is an optional argument which validates that the list reaches at least the given length.

User friendly forms

`SimpleListField` is not particularly user friendly in most cases, however it's better than nothing.

10.3 SimpleSetField

class SimpleSetField(*base_field*, *max_length=None*, *min_length=None*)
: A simple field which maps to a set, with items separated by commas. It is represented by an HTML `<input>`. Empty items, resulting from leading, trailing, or double commas, are disallowed.

base_field
: This is a required argument.

  It specifies the underlying form field for the set. It is not used to render any HTML, but it does process and validate the submitted data. For example:

```
>>> from django import forms
>>> from django_mysql.forms import SimpleSetField

>>> class NumberSetForm(forms.Form):
...     numbers = SimpleSetField(forms.IntegerField())

>>> form = NumberSetForm({'numbers': '1,2,3'})
>>> form.is_valid()
True
>>> form.cleaned_data
{'numbers': set([1, 2, 3])}

>>> form = NumberSetForm({'numbers': '1,2,a'})
>>> form.is_valid()
False
```

max_length
: This is an optional argument which validates that the set does not exceed the given length.

min_length
: This is an optional argument which validates that the set reaches at least the given length.

User friendly forms

`SimpleSetField` is not particularly user friendly in most cases, however it's better than nothing.

CHAPTER 11

Validators

The following can be imported from `django_mysql.validators`.

class ListMaxLengthValidator
A subclass of django's `MaxLengthValidator` with list-specific wording.

class ListMinLengthValidator
A subclass of django's `MinLengthValidator` with list-specific wording.

class SetMaxLengthValidator
A subclass of django's `MaxLengthValidator` with set-specific wording.

class SetMinLengthValidator
A subclass of django's `MinLengthValidator` with set-specific wording.

CHAPTER 12

Cache

A MySQL-specific backend for Django's cache framework.

12.1 MySQLCache

An efficient cache backend using a MySQL table, an alternative to Django's database-agnostic `DatabaseCache`. It has the following advantages:

- Each operation uses only one query, including the `*_many` methods. This is unlike `DatabaseCache` which uses multiple queries for nearly every operation.
- Automatic client-side `zlib` compression for objects larger than a given threshold. It is also easy to subclass and add your own serialization or compression schemes.
- Faster probabilistic culling behaviour during write operations, which you can also turn off and execute in a background task. This can be a bottleneck with Django's `DatabaseCache` since it culls on every write operation, executing a `SELECT COUNT(*)` which requires a full table scan.
- Integer counters with atomic `incr()` and `decr()` operations, like the `MemcachedCache` backend.

12.1.1 Usage

To use, add an entry to your `CACHES` setting with:

- `BACKEND` set to `django_mysql.cache.MySQLCache`
- `LOCATION` set to `tablename`, the name of the table to use. This name can be whatever you want, as long as it's a valid table name that's not already being used in your database.

For example:

```
CACHES = {
    'default': {
        'BACKEND': 'django_mysql.cache.MySQLCache',
```

(continues on next page)

(continued from previous page)

```
        'LOCATION': 'my_super_cache'
    }
}
```

You then need to make the table. The schema is *not* compatible with that of `DatabaseCache`, so if you are switching, you will need to create a fresh table.

Use the management command `mysql_cache_migration` to output a migration that creates tables for all the `MySQLCache` instances you have configured. For example:

```
$ python manage.py mysql_cache_migration
from django.db import migrations


class Migration(migrations.Migration):

    dependencies = [
        # Add a dependency in here on an existing migration in the app you
        # put this migration in, for example:
        # ('myapp', '0001_initial'),
    ]

    operations = [
        migrations.RunSQL(
            """
            CREATE TABLE `my_super_cache` (
                cache_key varchar(255) CHARACTER SET utf8 COLLATE utf8_bin
                                       NOT NULL PRIMARY KEY,
                value longblob NOT NULL,
                value_type char(1) CHARACTER SET latin1 COLLATE latin1_bin
                                   NOT NULL DEFAULT 'p',
                expires BIGINT UNSIGNED NOT NULL
            );
            """,
            "DROP TABLE `my_super_cache`"
        ),
    ]
```

Save this to a file in the `migrations` directory of one of your project's apps, and add one of your existing migrations to the file's `dependencies`. You might want to customize the SQL at this time, for example switching the table to use the `MEMORY` storage engine.

Django requires you to install sqlparse to run the `RunSQL` operation in the migration, so make sure it is installed.

Once the migration has run, the cache is ready to work!

12.1.2 Multiple Databases

If you use this with multiple databases, you'll also need to set up routing instructions for the cache table. This can be done with the same method that is described for `DatabaseCache` in the Django manual, apart from the application name is `django_mysql`.

Note: Even if you aren't using multiple MySQL databases, it may be worth using routing anyway to put all your cache operations on a second connection - this way they won't be affected by transactions your main code runs.

12.1.3 Extra Details

`MySQLCache` is fully compatible with Django's cache API, but it also extends it and there are, of course, a few details to be aware of.

incr/decr

Like `MemcachedCache` (and unlike `DatabaseCache`), `incr` and `decr` are atomic operations, and can only be used with `int` values. They have the range of MySQL's `SIGNED BIGINT` (-9223372036854775808 to 9223372036854775807).

max_allowed_packet

MySQL has a setting called `max_allowed_packet`, which is the maximum size of a query, including data. This therefore constrains the size of a cached value, but you're more likely to run up against it first with the `get_many`/`set_many` operations.

The MySQL 5.5 default is 1 MB, and the MySQL 5.6 default is 4MB, with which most applications will be fine. You can tweak it as high as 1GB (if this isn't enough, you should probably be considering another solution!).

culling

MySQL is designed to store data forever, and thus doesn't have a direct way of setting expired rows to disappear. The expiration of old keys and the limiting of rows to `MAX_ENTRIES` is therefore performed in the cache backend by performing a cull operation when appropriate. This deletes expired keys first, then if more than `MAX_ENTRIES` keys remain, it deletes 1 / `CULL_FREQUENCY` of them. The options and strategy are described in in more detail in the Django manual.

Django's `DatabaseCache` performs a cull check on *every* write operation. This runs a `SELECT COUNT(*)` on the table, which means a full-table scan. Naturally, this takes a bit of time and becomes a bottleneck for medium or large cache table sizes of caching. `MySQLCache` helps you solve this in two ways:

1. The cull-on-write behaviour is probabilistic, by default running on 1% of writes. This is set with the `CULL_PROBABILITY` option, which should be a number between 0 and 1. For example, if you want to use the same cull-on-*every*-write behaviour as used by `DatabaseCache` (you probably don't), set `CULL_PROBABILITY` to 1.0:

```
CACHES = {
    'default': {
        'BACKEND': 'django_mysql.cache.MySQLCache',
        'LOCATION': 'some_table_name',
        'OPTIONS': {
            'CULL_PROBABILITY': 1.0
        }
    }
}
```

2. The `cull()` method is available as a public method so you can set up your own culling schedule in background processing, never affecting any user-facing web requests. Set `CULL_PROBABILITY` to 0, and then set up your task. For example, if you are using **celery** you could use a task like this:

```
@shared_task
def clear_caches():
```

(continues on next page)

(continued from previous page)

```
    caches['default'].cull()
    caches['other_cache'].cull()
```

This functionality is also available as the management command `cull_mysql_caches`, which you might run as a cron job. It performs `cull()` on all of your `MySQLCache` instances, or you can give it names to just cull those. For example, this:

```
$ python manage.py cull_mysql_caches default other_cache
```

...will call `caches['default'].cull()` and `caches['other_cache'].cull()`.

You can also disable the `MAX_ENTRIES` behaviour, which avoids the `SELECT COUNT(*)` entirely, and makes `cull()` only delete expired keys. To do this, set `MAX_ENTRIES` to -1:

```
CACHES = {
    'default': {
        'BACKEND': 'django_mysql.cache.MySQLCache',
        'LOCATION': 'some_table_name',
        'OPTIONS': {
            'MAX_ENTRIES': -1
        }
    }
}
```

Note that you should then of course monitor the size of your cache table well, since it has no bounds on its growth.

compression

Like the other Django cache backends, stored objects are serialized with `pickle` (except from integers, which are stored as integers so that the `incr()` and `decr()` operations will work). If pickled data has has a size in bytes equal to or greater than the threshold defined by the option `COMPRESS_MIN_LENGTH`, it will be compressed with `zlib` in Python before being stored, reducing the on-disk size in MySQL and network costs for storage and retrieval. The zlib level is set by the option `COMPRESS_LEVEL`.

`COMPRESS_MIN_LENGTH` defaults to 5000, and `COMPRESS_LEVEL` defaults to the `zlib` default of 6. You can tune these options - for example, to compress all objects >= 100 bytes at the maximum level of 9, pass the options like so:

```
CACHES = {
    'default': {
        'BACKEND': 'django_mysql.cache.MySQLCache',
        'LOCATION': 'some_table_name',
        'OPTIONS': {
            'COMPRESS_MIN_LENGTH': 100,
            'COMPRESS_LEVEL': 9
        }
    }
}
```

To turn compression off, set `COMPRESS_MIN_LENGTH` to 0. The options only affect new writes - any compressed values already in the table will remain readable.

custom serialization

You can implement your own serialization by subclassing `MySQLCache`. It uses two methods that you should override.

Values are stored in the table with two columns - `value`, which is the blob of binary data, and `value_type`, a single latin1 character that specifies the type of data in `value`. MySQLCache by default uses three codes for `value_type`:

- `i` - The blob is an integer. This is used so that counters can be deserialized by MySQL during the atomic `incr()` and `decr()` operations.
- `p` - The blob is a pickled Python object.
- `z` - The blob is a zlib-compressed pickled Python object.

For future compatibility, `MySQLCache` reserves all lower-case letters. For custom types you can use upper-case letters.

The methods you need to override (and probably call `super()` from) are:

encode(*obj*)
: Takes an object and returns a tuple `(value, value_type)`, ready to be inserted as parameters into the SQL query.

decode(*value*, *value_type*)
: Takes the pair of `(value, value_type)` as stored in the table and returns the deserialized object.

Studying the source of `MySQLCache` will probably give you the best way to extend these methods for your use case.

prefix methods

Three extension methods are available to work with sets of keys sharing a common prefix. Whilst these would not be efficient on other cache backends such as memcached, in an InnoDB table the keys are stored in order so range scans are easy.

To use these methods, it must be possible to reverse-map the "full" key stored in the databse to the key you would provide to `cache.get`, via a 'reverse key function'. If you have not set `KEY_FUNCTION`, `MySQLCache` will use Django's default key function, and can therefore default the reverse key function too, so you will not need to add anything.

However, if you have set `KEY_FUNCTION`, you will also need to supply `REVERSE_KEY_FUNCTION` before the prefix methods can work. For example, with a simple custom key function that ignores `key_prefix` and `version`, you might do this:

```
def my_key_func(key, key_prefix, version):
    return key  # Ignore key_prefix and version


def my_reverse_key_func(full_key):
    # key_prefix and version still need to be returned
    key_prefix = None
    version = None
    return key, key_prefix, version


CACHES = {
    'default': {
        'BACKEND': 'django_mysql.cache.MySQLCache',
        'LOCATION': 'some_table_name',
```

(continues on next page)

(continued from previous page)

```
        'KEY_FUNCTION': my_key_func,
        'REVERSE_KEY_FUNCTION': my_reverse_key_func
    }
}
```

Once you're set up, the following prefix methods can be used:

delete_with_prefix(*prefix*, *version=None*)
Deletes all keys that start with the string `prefix`. If `version` is not provided, it will default to the `VERSION` setting. Returns the number of keys that were deleted. For example:

```
>>> cache.set_many({'Car1': 'Blue', 'Car4': 'Red', 'Truck3': 'Yellow'})
>>> cache.delete_with_prefix('Truck')
1
>>> cache.get('Truck3')
None
```

Note: This method does not require you to set the reverse key function.

get_with_prefix(*prefix*, *version=None*)
Like `get_many`, returns a dict of key to value for all keys that start with the string `prefix`. If `version` is not provided, it will default to the `VERSION` setting. For example:

```
>>> cache.set_many({'Car1': 'Blue', 'Car4': 'Red', 'Truck3': 'Yellow'})
>>> cache.get_with_prefix('Truck')
{'Truck3': 'Yellow'}
>>> cache.get_with_prefix('Ca')
{'Car1': 'Blue', 'Car4': 'Red'}
>>> cache.get_with_prefix('')
{'Car1': 'Blue', 'Car4': 'Red', 'Truck3': 'Yellow'}
```

keys_with_prefix(*prefix*, *version=None*)
Returns a set of all the keys that start with the string `prefix`. If `version` is not provided, it will default to the `VERSION` setting. For example:

```
>>> cache.set_many({'Car1': 'Blue', 'Car4': 'Red', 'Truck3': 'Yellow'})
>>> cache.keys_with_prefix('Car')
set(['Car1', 'Car2'])
```

12.1.4 Changes

Versions 0.1.10 -> 0.2.0

Initially, in Django-MySQL version 0.1.10, `MySQLCache` did not force the columns to use case sensitive collations; in version 0.2.0 this was fixed. You can upgrade by adding a migration with the following SQL, if you replace `yourtablename`:

```
ALTER TABLE yourtablename
    MODIFY cache_key varchar(255) CHARACTER SET utf8 COLLATE utf8_bin
                                  NOT NULL,
    MODIFY value_type char(1) CHARACTER SET latin1 COLLATE latin1_bin
                              NOT NULL DEFAULT 'p';
```

Or as a reversible migration:

```
from django.db import migrations


class Migration(migrations.Migration):

    dependencies = []

    operations = [
        migrations.RunSQL(
            """
            ALTER TABLE yourtablename
                MODIFY cache_key varchar(255) CHARACTER SET utf8 COLLATE utf8_bin
                                              NOT NULL,
                MODIFY value_type char(1) CHARACTER SET latin1 COLLATE latin1_bin
                                          NOT NULL DEFAULT 'p'
            """,
            """
            ALTER TABLE yourtablename
                MODIFY cache_key varchar(255) CHARACTER SET utf8 NOT NULL,
                MODIFY value_type char(1) CHARACTER SET latin1 NOT NULL DEFAULT 'p'
            """
        )
    ]
```

CHAPTER 13

Locks

The following can be imported from `django_mysql.locks`.

class Lock (*name*, *acquire_timeout=10.0*, *using=None*)

MySQL can act as a locking server for arbitrary named locks (created on the fly) via its `GET_LOCK` function - sometimes called 'User Locks' since they are user-specific, and don't lock tables or rows. They can be useful for your code to limit its access to some shared resource.

This class implements a user lock and acts as either a context manager (recommended), or a plain object with `acquire` and `release` methods similar to `threading.Lock`. These call the MySQL functions `GET_LOCK`, `RELEASE_LOCK`, and `IS_USED_LOCK` to manage it. It is *not* re-entrant so don't write code that gains/releases the same lock more than once.

Basic usage:

```
from django_mysql.exceptions import TimeoutError
from django_mysql.locks import Lock

try:
    with Lock('my_unique_name', acquire_timeout=2.0):
        mutually_exclusive_process()
except TimeoutError:
    print "Could not get the lock"
```

For more information on user locks refer to the `GET_LOCK` documentation on MySQL or MariaDB.

Warning: As the documentation warns, user locks are unsafe to use if you have replication running and your replication format (`binlog_format`) is set to `STATEMENT`. Most environments have `binlog_format` set to `MIXED` because it can be more performant, but do check.

Warning: It's not very well documented, but you can only hold one lock per database connection at a time. Acquiring a lock releases any other lock you were holding.

Since there is no MySQL function to tell you if you are currently holding a lock, this class does not check that you only acquire one lock. It has been a more than 10 year feature request to hold more than one lock per connection, and has been finally announced in MySQL 5.7.5.

name
This is a required argument.

Specifies the name of the lock. Since user locks share a global namespace on the MySQL server, it will automatically be prefixed with the name of the database you use in your connection from DATABASES and a full stop, in case multiple apps are using different databases on the same server.

Whilst not documented, the length limit is somewhere between 1 and 10 million characters, so most sane uses should be fine.

acquire_timeout=10.0
The time in seconds to wait to acquire the lock, as will be passed to `GET_LOCK()`. Defaults to 10 seconds.

using=None
The connection alias from `DATABASES` to use. Defaults to Django's `DEFAULT_DB_ALIAS` to use your main database connection.

is_held()
Returns True iff a query to `IS_USED_LOCK()` reveals that this lock is currently held.

holding_connection_id()
Returns the MySQL `CONNECTION_ID()` of the holder of the lock, or `None` if it is not currently held.

acquire()
For using the lock as a plain object rather than a context manager, similar to `threading.Lock.acquire`. Note you should normally use `try` / `finally` to ensure unlocking occurs.

Example usage:

```
from django_mysql.locks import Lock

lock = Lock('my_unique_name')
lock.acquire()
try:
    mutually_exclusive_process()
finally:
    lock.release()
```

release()
Also for using the lock as a plain object rather than a context manager, similar to `threading.Lock.release`. For example, see above.

classmethod held_with_prefix(*prefix*, *using=DEFAULT_DB_ALIAS*)
Queries the held locks that match the given prefix, for the given database connection. Returns a dict of lock names to the `CONNECTION_ID()` that holds the given lock.

Example usage:

```
>>> Lock.held_with_prefix('Author')
{'Author.1': 451, 'Author.2': 457}
```

Note: Works with MariaDB 10.0.7+ only, when the `metadata_lock_info` plugin is loaded. You can install this in a migration using the *`InstallSOName`* operation, like so:

```
from django.db import migrations
from django_mysql.operations import InstallSOName


class Migration(migrations.Migration):
    dependencies = []

    operations = [
        # Install https://mariadb.com/kb/en/mariadb/metadata_lock_info/
        InstallSOName('metadata_lock_info')
    ]
```

---

class TableLock(*write=None*, *read=none*, *using=None*)

MySQL allows you to gain a table lock to prevent modifications to the data during reads or writes. Most applications don't need to do this since transaction isolation should provide enough separation between operations, but occasionally this can be useful, especially in data migrations or if you are using a non-transactional storage such as MyISAM.

This class implements table locking and acts as either a context manager (recommended), or a plain object with `acquire()` and `release()` methods similar to `threading.Lock`. It uses the transactional pattern from the MySQL manual to ensure all the necessary steps are taken to lock tables properly. Note that locking has no timeout and blocks until held.

Basic usage:

```
from django_mysql.locks import TableLock

with TableLock(read=[MyModel1], write=[MyModel2]):
    fix_bad_instances_of_my_model2_using_my_model1_data()
```

Docs: MySQL / MariaDB.

read

A list of models or raw table names to lock at the `READ` level. Any models using multi-table inheritance will also lock their parents.

write

A list of models or raw table names to lock at the `WRITE` level. Any models using multi-table inheritance will also lock their parents.

using=None

The connection alias from `DATABASES` to use. Defaults to Django's `DEFAULT_DB_ALIAS` to use your main database connection.

acquire()

For using the lock as a plain object rather than a context manager, similar to `threading.Lock.acquire`. Note you should normally use `try` / `finally` to ensure unlocking occurs.

Example usage:

```
from django_mysql.locks import TableLock

table_lock = TableLock(read=[MyModel1], write=[MyModel2])
table_lock.acquire()
try:
    fix_bad_instances_of_my_model2_using_my_model1_data()
finally:
    table_lock.release()
```

release()

Also for using the lock as a plain object rather than a context manager, similar to `threading.Lock.release`. For example, see above.

Note: Transactions are not allowed around table locks, and an error will be raised if you try and use one inside of a transaction. A transaction is created to hold the locks in order to cooperate with InnoDB. There are a number of things you can't do whilst holding a table lock, for example accessing tables other than those you have locked - see the MySQL/MariaDB documentation for more details.

Note: Table locking works on InnoDB tables only if the `innodb_table_locks` is set to 1. This is the default, but may have been changed for your environment.

CHAPTER 14

Status

MySQL gives you metadata on the server status through its `SHOW GLOBAL STATUS` and `SHOW SESSION STATUS` commands. These classes make it easy to get this data, as well as providing utility methods to react to it.

The following can all be imported from `django_mysql.status`.

class GlobalStatus(*name*, *using=None*)

Provides easy access to the output of `SHOW GLOBAL STATUS`. These statistics are useful for monitoring purposes, or ensuring queries your code creates aren't saturating the server.

Basic usage:

```
from django_mysql.status import global_status

# Wait until a quiet moment
while global_status.get('Threads_running') >= 5:
    time.sleep(1)

# Log all status variables
logger.log("DB status", extra=global_status.as_dict())
```

Note that `global_status` is a pre-existing instance for the default database connection from `DATABASES`. If you're using more than database connection, you should instantiate the class:

```
>>> from django_mysql.status import GlobalStatus
>>> GlobalStatus(using='replica1').get('Threads_running')
47
```

To see the names of all the available variables, refer to the documentation: MySQL / MariaDB. They vary based upon server version, plugins installed, etc.

using=None

The connection alias from `DATABASES` to use. Defaults to Django's `DEFAULT_DB_ALIAS` to use your main database connection.

get(*name*)
Returns the current value of the named status variable. The name may not include SQL wildcards (`%`). If it does not exist, `KeyError` will be raised.

The result set for `SHOW STATUS` returns values in strings, so numbers and booleans will be cast to their respective Python types - `int`, `float`, or `bool`. Strings are be left as-is.

get_many(*names*)
Returns a dictionary of names to current values, fetching them in a single query. The names may not include wildcards (`%`).

Uses the same type-casting strategy as `get()`.

as_dict(*prefix=None*)
Returns a dictionary of names to current values. If `prefix` is given, only those variables starting with the prefix will be returned. `prefix` should not end with a wildcard (`%`) since that will be automatically appended.

Uses the same type-casting strategy as `get()`.

wait_until_load_low(*thresholds={'Threads_running': 10}*, *timeout=60.0*, *sleep=0.1*)
A helper method similar to the logic in `pt-online-schema-change` for waiting with –max-load.

Polls global status every `sleep` seconds until every variable named in `thresholds` is at or below its specified threshold, or raises a *`django_mysql.exceptions.TimeoutError`* if this does not occur within `timeout` seconds. Set `timeout` to 0 to never time out.

`thresholds` defaults to `{'Threads_running': 10}`, which is the default variable used in `pt-online-schema-change`, but with a lower threshold of 10 that is more suitable for small servers. You will very probably need to tweak it to your server.

You can use this method during large background operations which you don't want to affect other connections (i.e. your website). By processing in small chunks and waiting for low load in between, you sharply reduce your risk of outage.

class SessionStatus(*name*, *connection_name=None*)
This class is the same as GlobalStatus apart from it runs `SHOW SESSION STATUS`, so *some* variables are restricted to the current connection only, rather than the whole server. For which, you should refer to the documentation: MySQL / MariaDB.

Also it doesn't have the `wait_until_load_low` method, which only makes sense in a global setting.

Example usage:

```
from django_mysql.status import session_status

read_operations = session_status.get("Handler_read")
```

And for a different connection:

```
from django_mysql.status import SessionStatus

replica1_reads = SessionStatus(using='replica1').get("Handler_read")
```

CHAPTER 15

Management Commands

MySQL-specific management commands. These are automatically available with your `manage.py` when you add `django_mysql` to your `INSTALLED_APPS`.

15.1 `dbparams` command

Outputs your database connection parameters in a form suitable for inclusion in other CLI commands, helping avoid copy/paste errors and accidental copying of passwords to shell history files. Knows how to output parameters in two formats - for `mysql` related tools, or the DSN format that some percona tools take. For example:

```
$ python manage.py dbparams && echo  # 'echo' adds a newline
--user=ausername --password=apassword --host=ahost.example.com mydatabase
$ mysql $(python manage.py dbparams)  # About the same as 'manage.py dbshell'
$ mysqldump $(python manage.py dbparams) | gzip -9 > backup.sql.gz  # Neat!
```

The format of parameters is:

```
python manage.py dbparams [--mysql | --dsn] <optional-connection-alias>
```

If the database alias is given, it should be alias of a connection from the `DATABASES` setting; defaults to 'default'. Only MySQL connections are supported - the command will fail for other connection vendors.

Mutually exclusive format flags:

15.1.1 `--mysql`

Default, so shouldn't need passing. Allows you to do, e.g.:

```
$ mysqldump $(python manage.py dbparams) | gzip -9 > backup.sql.gz
```

Which will translate to include all the relevant flags, including your database.

15.1.2 `--dsn`

Outputs the parameters in the DSN format, which is what many percona tools take, e.g.:

```
$ pt-duplicate-key-checker $(python manage.py dbparams --dsn)
```

Note: If you are using SSL to connect, the percona tools don't support SSL configuration being given in their DSN format; you must pass them via a MySQL configuration file instead. `dbparams` will output a warning on stderr if this is the case. For more info see the percona blog.

15.2 `fix_datetime_columns` command

This command scans your database and outputs the SQL necessary to fix any `datetime` columns into `datetime(6)` columns, as is necessary when upgrading to MySQL 5.6, or MariaDB 5.3+.

If you upgrade MySQL to 5.6, Django will have created all your `DateTimeField`s as `datetime`, although they can be turned into `datetime(6)` now with microsecond support. Even if microsecond support is not necessary for your database, it's best to convert them over so that you aren't surprised by a future migration that upgrades them; Django's migration system doesn't check the type in the database and will assume all `DateTimeField`s are already `datetime(6)`.

Example usage:

```
$ python manage.py fix_datetime_columns
ALTER TABLE `app1_table1`
    MODIFY COLUMN `created_time` datetime(6) DEFAULT NULL;
ALTER TABLE `app1_table2`
    MODIFY COLUMN `created_time` datetime(6) DEFAULT NULL,
    MODIFY COLUMN `updated_time` datetime(6) DEFAULT NULL;
```

You can run this SQL straight away with:

```
$ python manage.py fix_datetime_columns | python manage.py dbshell
```

However you might want to put the SQL into a file and run it as a Django migration, or use tools such as pt-online-schema-change.

The format of parameters is:

```
$ python manage.py fix_datetime_columns <optional-connection-alias>
```

If the database alias is given, it should be alias of a connection from the `DATABASES` setting; defaults to 'default'. Only MySQL connections are supported - the command will fail for other connection vendors.

CHAPTER 16

Utilities

The following can be imported from `django_mysql.utils`.

`connection_is_mariadb`(*connection*)
Given a Django database connection (from `django.db.connections`) return `True` if it is a connection to a MariaDB database else `False`. The result is cached to avoid unnecessary connections.

`pt_fingerprint`(*query*)
Given a string `query` containing a MySQL query, returns a 'fingerprint' of the query from the Percona `pt-fingerprint` tool (docs). You must therefore have `pt-fingerprint` installed.

Example usage:

```
>>> pt_fingerprint("SELECT a, b FROM myapp_author WHERE id = 55")
'select a, b from myapp_author where id = ?'
>>> pt_fingerprint("SELECT SLEEP(123)")
'select sleep(?)'
>>> pt_fingerprint("release savepoint `ax123`")
'release savepoint `a?`'
```

This is a complex subprocess wrapper that is suitable for processing many queries serially - it opens `pt-fingerprint` in a background thread, which accepts input line-by-line, and shuts it down after 60 seconds of not being used. It is therefore suitable for use one-query-at-a-time, even when batch processing hundreds of queries.

Note: Because this uses Python's `pty` library, it only works on Unix.

CHAPTER 17

Test Utilities

The following can be imported from `django_mysql.test.utils`.

override_mysql_variables(*using='default'*, ***options*)

Overrides MySQL system variables for a test method or for every test method in a class, similar to Django's `override_settings`. This can be useful when you're testing code that must run under multiple MySQL environments (like most of *django-mysql*). For example:

```
@override_mysql_variables(SQL_MODE="MSSQL")
class MyTests(TestCase):

    def test_it_works_in_mssql(self):
        run_it()

    @override_mysql_variables(SQL_MODE="ANSI")
    def test_it_works_in_ansi_mode(self):
        run_it()
```

During the first test, the `SQL_MODE` will be `MSSQL`, and during the second, it will be `ANSI`; each slightly changes the allowed SQL syntax, meaning they are useful to test.

Note: This only sets the system variables for the session, so if the tested code closes and re-opens the database connection the change will be reset.

---

using

The connection alias to set the system variables for, defaults to 'default'.

CHAPTER 18

Exceptions

Various exception classes that can be raised by `django_mysql` code. They can imported from the `django_mysql.exceptions` module.

exception TimeoutError
Indicates a database operation timed out in some way.

CHAPTER 19

Contributing

Contributions are welcome, and they are greatly appreciated! Every little bit helps, and credit will always be given.

You can contribute in many ways:

19.1 Types of Contributions

19.1.1 Report Bugs

Report bugs via Github Issues.

If you are reporting a bug, please include:

- Your versions of Django-MySQL, Django, and MySQL/MariaDB
- Any other details about your local setup that might be helpful in troubleshooting, e.g. operating system.
- Detailed steps to reproduce the bug.

19.1.2 Fix Bugs

Look through the GitHub issues for bugs. Anything tagged with "bug" is open to whoever wants to implement it.

19.1.3 Implement Features

Look through the GitHub issues for features. Anything tagged with "help wanted" and not assigned to anyone is open to whoever wants to implement it - please leave a comment to say you have started working on it, and open a pull request as soon as you have something working, so that Travis starts building it.

Issues without "help wanted" generally already have some code ready in the background (maybe it's not yet open source), but you can still contribute to them by saying how you'd find the fix useful, linking to known prior art, or other such help.

19.1.4 Write Documentation

Django-MySQL could always use more documentation, whether as part of the official Django-MySQL docs, in docstrings, or even on the web in blog posts, articles, and such. Write away!

19.1.5 Submit Feedback

The best way to send feedback is to file an issue via Github Issues.

If you are proposing a feature:

- Explain in detail how it would work.
- Keep the scope as narrow as possible, to make it easier to implement.
- Remember that this is a volunteer-driven project, and that contributions are welcome :)
- Link to any prior art, e.g. MySQL/MariaDB documentation that details the necessary database features

19.2 Get Started!

Ready to contribute? Here's how to set up Django-MySQL for local development.

1. Fork the Django-MySQL repo on GitHub.
2. Clone your fork locally:

```
$ git clone git@github.com:your_name_here/django-mysql.git
$ cd django-mysql/
```

3. Check you have a supported version of MySQL or MariaDB running and that the settings in `tests/settings.py` will work for connecting. This involves making sure you can connect from your terminal with the plain command `mysql` with no options, i.e. as your current user.

   On Ubuntu, this can be done with the commands below:

```
$ sudo apt-get install mysql-server-5.7
$ mysql -uroot -p -e "CREATE USER '$(whoami)'@localhost; GRANT ALL PRIVILEGES ON␣
↪*.* TO '$(whoami)'@localhost;"
# Enter the password for root you set in the apt dialog
```

   On Mac OS X, this can be done with something like:

```
$ brew install mariadb
$ mysql.server start
$ mysql -uroot -e "CREATE USER '$(whoami)'@localhost; GRANT ALL PRIVILEGES ON *.*␣
↪TO '$(whoami)'@localhost;"
```

   If you want to use a different user or add a password, you can patch the settings file in your local install.

5. Install `tox` and run the tests for Python 3.6 + Django 2.1:

```
$ pip install tox
$ tox -e py36-django21
```

   The `tox.ini` file defines a large number of test environments, for different Python and Django versions, plus for checking codestyle. During development of a feature/fix, you'll probably want to run just one plus the relevant codestyle:

```
$ tox -e py36-codestyle,py36-django21
```

You can run all the environments to check your code is okay for them with:

```
$ tox
```

6. To make changes, create a branch for local development:

```
$ git checkout -b name-of-your-bugfix-or-feature
```

...and hack away!

7. Commit your changes and push your branch to GitHub:

```
$ git add .
$ git commit -m "Your detailed description of your changes."
$ git push origin name-of-your-bugfix-or-feature
```

8. Submit a pull request through the GitHub website. This will trigger the Travis build which runs the tests against all supported versions of Python, Django, and MySQL/MariaDB.

19.3 Testing Tips

To only run a particular test file, you can run with the path to that file:

```
$ tox -- tests/testapp/test_some_feature.py
```

You can also pass other pytest arguments through `tox` after the `--` separator. There are lots of other useful features, most of which you can check out in the pytest docs!

CHAPTER 20

History

20.1 Pending

20.2 3.2.0 (2019-06-14)

- Update Python support to 3.5-3.7, as 3.4 has reached its end of life.
- Always cast SQL params to tuples in ORM code.

20.3 3.1.0 (2019-05-17)

- Remove authors file and documentation page. This was showing only 4 out of the 17 total contributors.
- Tested on Django 2.2. No changes were needed for compatibility.

20.4 3.0.0.post1 (2019-03-05)

- Remove universal wheel. Version 3.0.0 has been pulled from PyPI after being up for 3 hours to fix mistaken installs on Python 2.

20.5 3.0.0 (2019-03-05)

- Drop Python 2 support, only Python 3.4+ is supported now.

20.6 2.5.0 (2019-03-03)

- Drop Django 1.8, 1.9, and 1.10 support. Only Django 1.11+ is supported now.

20.7 2.4.1 (2018-08-18)

- Django 2.1 compatibility - no code changes were required, releasing for PyPI trove classifiers and documentation.

20.8 2.4.0 (2018-07-31)

- Added `JSONArrayAppend` database function that wraps the respective JSON-modifying function from MySQL 5.7.

20.9 2.3.1 (2018-07-22)

- Made `EnumField` escape its arguments in a `pymysql`-friendly fashion.

20.10 2.3.0 (2018-06-19)

- Started testing with MariaDB 10.3.
- Changed `GlobalStatus.wait_until_load_low()` to increase the default number of allowed running threads from 5 to 10, to account for the new default threads in MariaDB 10.3.
- Added `encoder` and `decoder` arguments to `JSONField` for customizing the way JSON is encoded and decoded from the database.
- Added a `touch` method to the `MySQLCache` to refresh cache keys, as added in Django 2.1.
- Use a temporary database connection in system checks to avoid application startup stalls.

20.11 2.2.2 (2018-04-24)

- Fixed some crashes from `DynamicField` instances without explicit `spec` definitions.
- Fixed a crash in system checks for `ListCharField` and `SetCharField` instances missing `max_length`.

20.12 2.2.1 (2018-04-14)

- Fixed `JSONField.deconstruct()` to not break the path for subclasses.

20.13 2.2.0 (2017-12-04)

- Add `output_field` argument to `JSONExtract` function.
- Improved DB version checks for `JSONField` and `DynamicField` so you can have just one connection that supports them.
- Django 2.0 compatibility.

20.14 2.1.1 (2017-10-10)

- Changed subprocess imports for compatibility with Google App Engine.
- (Insert new release notes below this line)
- Made `MySQLCache.set_many` return a list as per Django 2.0.

20.15 2.1.0 (2017-06-11)

- Django 1.11 compatibility
- Some fixes to work with new versions of `mysqlclient`

20.16 2.0.0 (2017-05-28)

- Fixed `JSONField` model field string serialization. This is a small backwards incompatible change.

  Storing strings mostly used to crash with MySQL error -1 "error totally whack", but in the case your string was valid JSON, it would store it as a JSON object at the MySQL layer and deserialize it when returned. For example you could do this:

```
>>> mymodel.attrs = '{"foo": "bar"}'
>>> mymodel.save()
>>> mymodel = MyModel.objects.get(id=mymodel.id)
>>> mymodel.attrs
{'foo': 'bar'}
```

  The new behaviour now correctly returns what you put in:

```
>>> mymodel.attrs
'{"foo": "bar"}'
```

- Removed the `connection.is_mariadb` monkey patch. This is a small backwards incompatible change. Instead of using it, use `django_mysql.utils.connection_is_mariadb`.

20.17 1.2.0 (2017-05-14)

- Only use Django's vendored six (`django.utils.six`). Fixes usage of `EnumField` and field lookups when `six` is not installed as a standalone package.

- Added `JSONInsert`, `JSONReplace` and `JSONSet` database functions that wraps the respective JSON-modifying functions from MySQL 5.7.
- Fixed `JSONField` to work with Django's serializer framework, as used in e.g. `dumpdata`.
- Fixed `JSONField` form field so that it doesn't overquote inputs when redisplaying the form due to invalid user input.

20.18 1.1.1 (2017-03-28)

- Don't allow NaN in `JSONField` because MySQL doesn't support it

20.19 1.1.0 (2016-07-22)

- Dropped Django 1.7 support
- Made the query hint functions raise `RuntimeError` if you haven't activated the query-rewriting layer in settings.

20.20 1.0.9 (2016-05-12)

- Fixed some features to work when there are non-MySQL databases configured
- Fixed `JSONField` to allow control characters, which MySQL does - but not in a top-level string, only inside a JSON object/array.

20.21 1.0.8 (2016-04-08)

- `SmartChunkedIterator` now fails properly for models whose primary key is a non-integer foreign key.
- `pty` is no longer imported at the top-level in `django_mysql.utils`, fixing Windows compatibility.

20.22 1.0.7 (2016-03-04)

- Added new `JSONField` class backed by the JSON type added in MySQL 5.7.
- Added database functions `JSONExtract`, `JSONKeys`, and `JSONLength` that wrap the JSON functions added in MySQL 5.7, which can be used with the JSON type columns as well as JSON data held in text/varchar columns.
- Added `If` database function for simple conditionals.

20.23 1.0.6 (2016-02-26)

- Now MySQL 5.7 compatible
- The final message from `SmartChunkedIterator` is now rounded to the nearest second.

- `Lock` and `TableLock` classes now have `acquire` and `release()` methods for using them as normal objects rather than context managers

20.24 1.0.5 (2016-02-10)

- Added `manage.py` command `fix_datetime_columns` that outputs the SQL necessary to fix any `datetime` columns into `datetime(6)`, as required when upgrading a database to MySQL 5.6+, or MariaDB 5.3+.
- `SmartChunkedIterator` output now includes the total time taken and number of objects iterated over in the final message.

20.25 1.0.4 (2016-02-02)

- Fixed the new system checks to actually work

20.26 1.0.3 (2016-02-02)

- Fixed `EnumField` so that it works properly with forms, and does not accept the `max_length` argument.
- `SmartChunkedIterator` output has been fixed for reversed iteration, and now includes a time estimate.
- Added three system checks that give warnings if the MySQL configuration can (probably) be improved.

20.27 1.0.2 (2016-01-24)

- New function `add_QuerySetMixin` allows addding the `QuerySetMixin` to arbitrary `QuerySets`, for when you can't edit a model class.
- Added field class `EnumField` that uses MySQL's `ENUM` data type.

20.28 1.0.1 (2015-11-18)

- Added `chunk_min` argument to `SmartChunkedIterator`

20.29 1.0.0 (2015-10-29)

- Changed version number to 1.0.0 to indicate maturity.
- Added `DynamicField` for using MariaDB's Named Dynamic Columns, and related database functions `ColumnAdd`, `ColumnDelete`, and `ColumnGet`.
- `SmartChunkedIterator` with `report_progress=True` correctly reports 'lowest pk so far' when iterating in reverse.
- Fix broken import paths during `deconstruct()` for subclasses of all fields: `ListCharField`, `ListTextField`, `SetCharField`, `SetTextField`, `SizedBinaryField` and `SizedTextField`

- Added XML database functions - `UpdateXML` and `XMLExtractValue`.

20.30 0.2.3 (2015-10-12)

- Allow `approx_count` on QuerySets for which only query hints have been used
- Added index query hints to `QuerySet` methods, via query-rewriting layer
- Added `ordering` parameter to `GroupConcat` to specify the `ORDER BY` clause
- Added index query hints to `QuerySet` methods, via query-rewriting layer
- Added `sql_calc_found_rows()` query hint that calculates the total rows that match when you only take a slice, which becomes available on the `found_rows` attribute
- Made `SmartChunkedIterator` work with `reverse()`'d `QuerySets`

20.31 0.2.2 (2015-09-03)

- `SmartChunkedIterator` now takes an argument `chunk_size` as the initial chunk size
- `SmartChunkedIterator` now allows models whose primary key is a `ForeignKey`
- Added `iter_smart_pk_ranges` which is similar to `iter_smart_chunks` but yields only the start and end primary keys for each chunks, in a tuple.
- Added prefix methods to `MySQLCache` - `delete_with_prefix`, `get_with_prefix`, `keys_with_prefix`
- Added `Bit1BooleanField` and `NullBit1BooleanField` model fields that work with boolean fields built by other databases that use the `BIT(1)` column type

20.32 0.2.1 (2015-06-22)

- Added Regexp database functions for MariaDB - `RegexpInstr`, `RegexpReplace`, and `RegexpSubstr`
- Added the option to not limit the size of a `MySQLCache` by setting `MAX_ENTRIES` = -1.
- `MySQLCache` performance improvements in *get*, *get_many*, and *has_key*
- Added query-rewriting layer added which allows the use of MySQL query hints such as `STRAIGHT_JOIN` via QuerySet methods, as well as adding label comments to track where queries are generated.
- Added `TableLock` context manager

20.33 0.2.0 (2015-05-14)

- More database functions added - `Field` and its complement `ELT`, and `LastInsertId`
- Case sensitive string lookup added as to the ORM for `CharField` and `TextField`
- Migration operations added - `InstallPlugin`, `InstallSOName`, and `AlterStorageEngine`
- Extra ORM aggregates added - `BitAnd`, `BitOr`, and `BitXor`

- `MySQLCache` is now case-sensitive. If you are already using it, an upgrade `ALTER TABLE` and migration is provided at the end of the cache docs.
- (MariaDB only) The `Lock` class gained a class method `held_with_prefix` to query held locks matching a given prefix
- `SmartIterator` bugfix for chunks with 0 objects slowing iteration; they such chunks most often occur on tables with primary key "holes"
- Now tested against Django master for cutting edge users and forwards compatibility

20.34 0.1.10 (2015-04-30)

- Added the `MySQLCache` backend for use with Django's caching framework, a more efficient version of `DatabaseCache`
- Fix a `ZeroDivision` error in `WeightedAverageRate`, which is used in smart iteration

20.35 0.1.9 (2015-04-20)

- `pt_visual_explain` no longer executes the given query before fetching its `EXPLAIN`
- New `pt_fingerprint` function that wraps the `pt-fingerprint` tool efficiently
- For `List` fields, the new `ListF` class allows you to do atomic append or pop operations from either end of the list in a single query
- For `Set` fields, the new `SetF` class allows you to do atomic add or remove operatiosn from the set in a single query
- The `@override_mysql_variables` decorator has been introduced which makes testing code with different MySQL configurations easy
- The `is_mariadb` property gets added onto Django's MySQL `connection` class automatically
- A race condition in determining the minimum and maximum primary key values for smart iteration was fixed.

20.36 0.1.8 (2015-03-31)

- Add `Set` and `List` fields which can store comma-separated sets and lists of a base field with MySQL-specific lookups
- Support MySQL's `GROUP_CONCAT` as an aggregate!
- Add a `functions` module with many MySQL-specific functions for the new Django 1.8 database functions feature
- Allow access of the global and session status for the default connection from a lazy singleton, similar to Django's `connection` object
- Fix a different recursion error on `count_tries_approx`

20.37 0.1.7 (2015-03-25)

- Renamed `connection_name` argument to `using` on `Lock`, `GlobalStatus`, and `SessionStatus` classes, for more consistency with Django.
- Fix recursion error on `QuerySetMixin` when using `count_tries_approx`

20.38 0.1.6 (2015-03-21)

- Added support for `HANDLER` statements as a `QuerySet` extension
- Now tested on Django 1.8
- Add `pk_range` argument for 'smart iteration' code

20.39 0.1.5 (2015-03-11)

- Added `manage.py` command `dbparams` for outputting database paramters in formats useful for shell scripts

20.40 0.1.4 (2015-03-10)

- Fix release process

20.41 0.1.3 (2015-03-08)

- Added `pt_visual_explain` integration on `QuerySet`
- Added soundex-based field lookups for the ORM

20.42 0.1.2 (2015-03-01)

- Added `get_many` to `GlobalStatus`
- Added `wait_until_load_low` to `GlobalStatus` which allows you to wait for any high load on your database server to dissipate.
- Added smart iteration classes and methods for `QuerySets` that allow efficient iteration over very large sets of objects slice-by-slice.

20.43 0.1.1 (2015-02-23)

- Added `Model` and `QuerySet` subclasses which add the `approx_count` method

20.44 0.1.0 (2015-02-12)

- First release on PyPI
- `Locks`
- `GlobalStatus` and `SessionStatus`

CHAPTER 21

Indices and tables

- genindex
- modindex
- search

Python Module Index

e

Index

T

U

W

X