Performance and optimization

This document provides an overview of techniques and tools that can help get your Django code running more efficiently - faster, and using fewer system resources.

Introduction

Generally one's first concern is to write code that *works*, whose logic functions as required to produce the expected output. Sometimes, however, this will not be enough to make the code work as *efficiently* as one would like.

In this case, what's needed is something - and in practice, often a collection of things - to improve the code's performance without, or only minimally, affecting its behavior.

General approaches

What are you optimizing *for*?

It's important to have a clear idea what you mean by 'performance'. There is not just one metric of it.

Improved speed might be the most obvious aim for a program, but sometimes other performance improvements might be sought, such as lower memory consumption or fewer demands on the database or network.

Improvements in one area will often bring about improved performance in another, but not always; sometimes one can even be at the expense of another. For example, an improvement in a program's speed might cause it to use more memory. Even worse, it can be self-defeating - if the speed improvement is so memory-hungry that the system starts to run out of memory, you'll have done more harm than good.

There are other trade-offs to bear in mind. Your own time is a valuable resource, more precious than CPU time. Some improvements might be too difficult to be worth implementing, or might affect the portability or maintainability of the code. Not all performance improvements are worth the effort.

So, you need to know what performance improvements you are aiming for, and you also need to know that you have a good reason for aiming in that direction - and for that you need:

Performance benchmarking

It's no good just guessing or assuming where the inefficiencies lie in your code.

Django tools

django-debug-toolbar is a very handy tool that provides insights into what your code is doing and how much time it spends doing it. In particular it can show you all the SQL queries your page is generating, and how long each one has taken.

Third-party panels are also available for the toolbar, that can (for example) report on cache performance and template rendering times.

Third-party services

There are a number of free services that will analyze and report on the performance of your site's pages from the perspective of a remote HTTP client, in effect simulating the experience of an actual user.

These can't report on the internals of your code, but can provide a useful insight into your site's overall performance, including aspects that can't be adequately measured from within Django environment. Examples include:

- Yahoo's Yslow
- Google PageSpeed

There are also several paid-for services that perform a similar analysis, including some that are Django-aware and can integrate with your codebase to profile its performance far more comprehensively.

Get things right from the start

Some work in optimization involves tackling performance shortcomings, but some of the work can be built-in to what you'd do anyway, as part of the good practices you should adopt even before you start thinking about improving performance.

In this respect Python is an excellent language to work with, because solutions that look elegant and feel right usually are the best performing ones. As with most skills, learning what "looks right" takes practice, but one of the most useful guidelines is:

Work at the appropriate level

Django offers many different ways of approaching things, but just because it's possible to do something in a certain way doesn't mean that it's the most appropriate way to do it. For example, you might find that you could calculate the same thing - the number of items in a collection, perhaps - in a `QuerySet`, in Python, or in a template.

However, it will almost always be faster to do this work at lower rather than higher levels. At higher levels the system has to deal with objects through multiple levels of abstraction and layers of machinery.

That is, the database can typically do things faster than Python can, which can do them faster than the template language can:

```
# QuerySet operation on the database
# fast, because that's what databases are good at
my_bicycles.count()

# counting Python objects
# slower, because it requires a database query anyway, and processing
# of the Python objects
len(my_bicycles)
```

```
<!--
Django template filter
slower still, because it will have to count them in Python anyway,
and because of template language overheads
-->
{{ my_bicycles|length }}
```

Generally speaking, the most appropriate level for the job is the lowest-level one that it is comfortable to code for.

Note: The example above is merely illustrative.

Firstly, in a real-life case you need to consider what is happening before and after your count to work out what's an optimal way of doing it *in that particular context*. The database optimization documents describes a case where counting in the template would be better.

Secondly, there are other options to consider: in a real-life case, `{{ my_bicycles.count }}`, which invokes the `QuerySet` `count()` method directly from the template, might be the most appropriate choice.

Caching

Often it is expensive (that is, resource-hungry and slow) to compute a value, so there can be huge benefit in saving the value to a quickly accessible cache, ready for the next time it's required.

It's a sufficiently significant and powerful technique that Django includes a comprehensive caching framework, as well as other smaller pieces of caching functionality.

The caching framework

Django's caching framework offers very significant opportunities for performance gains, by saving dynamic content so that it doesn't need to be calculated for each request.

For convenience, Django offers different levels of cache granularity: you can cache the output of specific views, or only the pieces that are difficult to produce, or even an entire site.

Implementing caching should not be regarded as an alternative to improving code that's performing poorly because it has been written badly. It's one of the final steps toward producing well-performing code, not a shortcut.

`cached_property`

It's common to have to call a class instance's method more than once. If that function is expensive, then doing so can be wasteful.

Using the `cached_property` decorator saves the value returned by a property; the next time the function is called on that instance, it will return the saved value rather than re-computing it. Note that this only works on methods that take `self` as their only argument and that it changes the method to a property.

Certain Django components also have their own caching functionality; these are discussed below in the sections related to those components.

Understanding laziness

Laziness is a strategy complementary to caching. Caching avoids recomputation by saving results; laziness delays computation until it's actually required.

Laziness allows us to refer to things before they are instantiated, or even before it's possible to instantiate them. This has numerous uses.

For example, lazy translation can be used before the target language is even known, because it doesn't take place until the translated string is actually required, such as in a rendered template.

Laziness is also a way to save effort by trying to avoid work in the first place. That is, one aspect of laziness is not doing anything until it has to be done, because it may not turn out to be necessary after all. Laziness can therefore have performance implications, and the more expensive the work concerned, the more there is to gain through laziness.

Python provides a number of tools for lazy evaluation, particularly through the generator and generator expression constructs. It's worth reading up on laziness in Python to discover opportunities for making use of lazy patterns in your code.

Laziness in Django

Django is itself quite lazy. A good example of this can be found in the evaluation of `QuerySets`. QuerySets are lazy. Thus a `QuerySet` can be created, passed around and combined with other `QuerySets`, without actually incurring any trips to the database to fetch the items it describes. What gets passed around is the `QuerySet` object, not the collection of items that - eventually - will be required from the database.

On the other hand, certain operations will force the evaluation of a QuerySet. Avoiding the premature evaluation of a `QuerySet` can save making an expensive and unnecessary trip to the database.

Django also offers a `keep_lazy()` decorator. This allows a function that has been called with a lazy argument to behave lazily itself, only being evaluated when it needs to be. Thus the lazy argument - which could be an expensive one - will not be called upon for evaluation until it's strictly required.

Databases

Database optimization

Django's database layer provides various ways to help developers get the best performance from their databases. The database optimization documentation gathers together links to the relevant documentation and adds various tips that outline the steps to take when attempting to optimize your database usage.

Other database-related tips

Enabling Persistent connections can speed up connections to the database accounts for a significant part of the request processing time.

This helps a lot on virtualized hosts with limited network performance, for example.

HTTP performance

Middleware

Django comes with a few helpful pieces of middleware that can help optimize your site's performance. They include:

`ConditionalGetMiddleware`

Adds support for modern browsers to conditionally GET responses based on the `ETag` and `Last-Modified` headers. It also calculates and sets an ETag if needed.

`GZipMiddleware`

Compresses responses for all modern browsers, saving bandwidth and transfer time. Note that GZipMiddleware is currently considered a security risk, and is vulnerable to attacks that nullify the protection provided by TLS/SSL. See the warning in `GZipMiddleware` for more information.

Sessions

Using cached sessions

Using cached sessions may be a way to increase performance by eliminating the need to load session data from a slower storage source like the database and instead storing frequently used session data in memory.

Static files

Static files, which by definition are not dynamic, make an excellent target for optimization gains.

`ManifestStaticFilesStorage`

By taking advantage of web browsers' caching abilities, you can eliminate network hits entirely for a given file after the initial download.

`ManifestStaticFilesStorage` appends a content-dependent tag to the filenames of static files to make it safe for browsers to cache them long-term without missing future changes - when a file changes, so will the tag, so browsers will reload the asset automatically.

"Minification"

Several third-party Django tools and packages provide the ability to "minify" HTML, CSS, and JavaScript. They remove unnecessary whitespace, newlines, and comments, and shorten variable names, and thus reduce the size of the documents that your site publishes.

Template performance

Note that:

- using `{% block %}` is faster than using `{% include %}`
- heavily-fragmented templates, assembled from many small pieces, can affect performance

The cached template loader

Enabling the `cached template loader` often improves performance drastically, as it avoids compiling each template every time it needs to be rendered.

Using different versions of available software

It can sometimes be worth checking whether different and better-performing versions of the software that you're using are available.

These techniques are targeted at more advanced users who want to push the boundaries of performance of an already well-optimized Django site.

However, they are not magic solutions to performance problems, and they're unlikely to bring better than marginal gains to sites that don't already do the more basic things the right way.

Note: It's worth repeating: **reaching for alternatives to software you're already using is never the first answer to performance problems**. When you reach this level of optimization, you need a formal benchmarking solution.

Newer is often - but not always - better

It's fairly rare for a new release of well-maintained software to be less efficient, but the maintainers can't anticipate every possible use-case - so while being aware that newer versions are likely to perform better, don't assume that they always will.

This is true of Django itself. Successive releases have offered a number of improvements across the system, but you should still check the real-world performance of your application, because in some cases you may find that changes mean it performs worse rather than better.

Newer versions of Python, and also of Python packages, will often perform better too - but measure, rather than assume.

Note: Unless you've encountered an unusual performance problem in a particular version, you'll generally find better features, reliability, and security in a new release and that these benefits are far more significant than any performance you might win or lose.

Alternatives to Django's template language

For nearly all cases, Django's built-in template language is perfectly adequate. However, if the bottlenecks in your Django project seem to lie in the template system and you have exhausted other opportunities to remedy this, a third-party alternative may be the answer.

Jinja2 can offer performance improvements, particularly when it comes to speed.

Alternative template systems vary in the extent to which they share Django's templating language.

Note: *If* you experience performance issues in templates, the first thing to do is to understand exactly why. Using an alternative template system may prove faster, but the same gains may also be available without going to that trouble - for example, expensive processing and logic in your templates could be done more efficiently in your views.

Alternative software implementations

It may be worth checking whether Python software you're using has been provided in a different implementation that can execute the same code faster.

However: most performance problems in well-written Django sites aren't at the Python execution level, but rather in inefficient database querying, caching, and templates. If you're relying on poorly-written Python code, your performance problems are unlikely to be solved by having it execute faster.

Using an alternative implementation may introduce compatibility, deployment, portability, or maintenance issues. It goes without saying that before adopting a non-standard implementation you should ensure it provides sufficient performance gains for your application to outweigh the potential risks.

With these caveats in mind, you should be aware of:

PyPy

PyPy is an implementation of Python in Python itself (the 'standard' Python implementation is in C). PyPy can offer substantial performance gains, typically for heavyweight applications.

A key aim of the PyPy project is compatibility with existing Python APIs and libraries. Django is compatible, but you will need to check the compatibility of other libraries you rely on.

C implementations of Python libraries

Some Python libraries are also implemented in C, and can be much faster. They aim to offer the same APIs. Note that compatibility issues and behavior differences are not unknown (and not always immediately evident).