**Slide 2**

* For this example, let’s say the requests in the system depend on some external resource, like a database or Web API. When a request comes in, ASP.NET takes one of its thread pool threads and assigns it to that request. Because it’s written synchronously, the request handler will call that external resource synchronously. This blocks the request thread until the call to the external resource returns. Figure 1 illustrates a thread pool with two threads, one of which is blocked waiting for an external resource.
* Eventually, that external resource call returns, and the request thread resumes processing that request. When the request is complete and the response is ready to be sent, the request thread is returned to the thread pool.
* This is all well and good—until your ASP.NET server gets more requests than it has threads to handle. At this point, the extra requests have to wait for a thread to be available before they can run.
* In this situation, the first two requests are assigned threads from the thread pool. Each of these requests calls an external resource, blocking their threads. The third request has to wait for an available thread before it can even start processing, but the request is already in the system. Its timer is going, and it’s in danger of an **HTTP Error 503** (Service unavailable).
* But think about this for a second: That third request is waiting for a thread, when there are two other threads in the system effectively doing nothing. Those threads are just blocked waiting for an external call to return. They’re not doing any real work; they’re not in a running state and are not given any CPU time. Those threads are just being wasted while there’s a request in need. This is the situation addressed by asynchronous requests.

**Slide 3**

* Asynchronous code can scale further than blocking threads because it uses much less memory; every thread pool thread on a modern OS has a **1MB** stack, plus an **unpageable** kernel stack. That doesn’t sound like a lot until you start getting a whole lot of threads on your server. In contrast, the memory overhead for an asynchronous operation is much smaller. So, a request with an asynchronous operation has much less memory pressure than a request with a blocked thread. Asynchronous code allows you to use more of your memory for other things (caching, for example).
* Asynchronous code can scale faster than blocking threads because the thread pool has a limited **injection rate**. Currently, the rate is one thread every two seconds. This injection rate limit is a good thing; it avoids constant thread construction and destruction. However, consider what happens when a sudden flood of requests comes in. Synchronous code can easily get bogged down as the requests use up all available threads and the remaining requests have to wait for the thread pool to inject new threads. On the other hand, asynchronous code doesn’t need a limit like this; it’s “always on,” so to speak. Asynchronous code is more responsive to sudden swings in request volume.

**Slide 4**

Asynchronous programming is writing code that allows several things to happen at the same time without "blocking", or waiting for other things to complete. This is different from synchronous programming, in which everything happens in the order it is written (if you write code for a living, chances are it will be synchronous code).

\*Async methods are not related to Concurrency and parallelism, asynchrony is used to present the impression of concurrent or parallel tasking but effectively an asynchronous method call is normally used for a process that needs to do work away from the current application and we don't want to wait and block our application awaiting the response.

Benefits

Performance

Responsiveness

Key Definitions

1) async operator

- is applied to a method to indicate it is asynchronous

- it does not run this method on a thread pool thread

- it runs synchronously until it hits the await

2) await operator

- is applied to task in an asynchronous method to suspend execution of the method until the awaited task completes

- it takes an argument, an awaitable (an asynchronous operation)

- examines the awaitable to see if it is already completed.

- if the awaitable is not complete, then it acts asynchronously by telling the awaitable to run the remainder of the method when it completes, and then returns the async method.

- later on, when the awaitable completes, it will execute the remainder of the async method on a context that was captured before the await returned.

3) Task

- ongoing work

4) Thread

- execution of a single sequence of programming instructions

5) Thread Pool

- collection of worker threads that efficiently execute asynchronous callbacks on behalf of the application.

- primarily used to reduce the number of application threads and provide management of the worker threads.

Awaitables

- two types common to .NET framework

- Task

- Task<T> or Task<TResult>

- others

- Task.Yield: returns awaitables that are not tasks

- WinRT runtime (WIndows 8) has an unmanaged awaitable type.

\*- You can also create your own awaitable or use extension methods to make a non-awaitable type awaitable which is beyond the scope of this presentation

Return Types / Values

- Async methods can return Task<T>, Task, or void. In almost all cases, you want to return Task<T> or Task, and return void only when you have to.

- Task and Task<T> are awaitable while void is not, however Task acts as a void because it does not return a value

- You have to return void when you have async event handlers.

- To return a value, the method must be of Task<T> and must return a value of type T.

Context

- when you await a built-in awaitable, then the awaitable will capture the current "context" and later apply it to the remainder of the async method

- If SynchronizationContext.Current is not null, then it’s the current SynchronizationContext (UI or ASP.NET context)

- Otherwise its a TaskScheduler.Default (thread pool context)

Avoiding Context

- Most async methods will be designed with composition in mind: they await other operations, and each one represents an asynchronous operation itself.

- If you want to tell the awaiter to not capture the current context, you call the ConfigureAwait and pass false

- The important thing to note is that each level of async method calls has it's own context

- A good rule of thumb is to use ConfigureAwait(false) unless you know you do need the context

Async Composition

- So we considered serial composition, an async method waits for one operation at a time. it is also possible to start several operations and await for one (or all) of them to complete

Deadlock

Scenerio 1

- The Task<string>.Result property is strongly typed as a String.

- Thus, it can't return until it has the valid result string send back; so it blocks until the result is available.

- The UI thread blocks for LoadStringAsync to complete.

- However, LoadStringAsync’s implementation depends on being able to post code asynchronously back to the UI to be executed, and the task returned from LoadStringAsync won’t complete until it does.

- LoadStringAsync is waiting for MyButton\_Click to complete, and MyButton\_Click is waiting for LoadStringAsync to complete.

Conclusion

- As you can see asynchronous programming has many benefits but does add complexity to your code. Like all programming patterns it is best to weigh the benefits against the added complexity to see if it is the right fit for your application