Asynchronous Programming

When a program needs to interact with the outside world, for example communicating with another machine over the internet, operations in the program may need to happen in an unpredictable order. Say your program needs to download a file. We would like to initiate the download operation, perform other operations while we wait for it to complete, and then resume the code that needs the downloaded file when it is available. This sort of scenario falls in the domain of asynchronous programming, sometimes also referred to as concurrent programming (since, conceptually, multiple things are happening at once).

To address these scenarios, Julia provides `Task`s (also known by several other names, such as symmetric coroutines, lightweight threads, cooperative multitasking, or one-shot continuations). When a piece of computing work (in practice, executing a particular function) is designated as a `Task`, it becomes possible to interrupt it by switching to another `Task`. The original `Task` can later be resumed, at which point it will pick up right where it left off. At first, this may seem similar to a function call. However there are two key differences. First, switching tasks does not use any space, so any number of task switches can occur without consuming the call stack. Second, switching among tasks can occur in any order, unlike function calls, where the called function must finish executing before control returns to the calling function.

Basic `Task` operations

You can think of a `Task` as a handle to a unit of computational work to be performed. It has a create-start-run-finish lifecycle. Tasks are created by calling the `Task` constructor on a 0-argument function to run, or using the `@task` macro:

```
julia> t = @task begin; sleep(5); println("done"); end
Task (runnable) @0x00007f13a40c0eb0
```

`@task x` is equivalent to `Task(()->x)`.

This task will wait for five seconds, and then print `done`. However, it has not started running yet. We can run it whenever we're ready by calling `schedule`:

```
julia> schedule(t);
```

If you try this in the REPL, you will see that `schedule` returns immediately. That is because it simply

adds `t` to an internal queue of tasks to run. Then, the REPL will print the next prompt and wait for more input. Waiting for keyboard input provides an opportunity for other tasks to run, so at that point `t` will start. `t` calls `sleep`, which sets a timer and stops execution. If other tasks have been scheduled, they could run then. After five seconds, the timer fires and restarts `t`, and you will see `done` printed. `t` is then finished.

The `wait` function blocks the calling task until some other task finishes. So for example if you type

```
julia> schedule(t); wait(t)
```

instead of only calling `schedule`, you will see a five second pause before the next input prompt appears. That is because the REPL is waiting for `t` to finish before proceeding.

It is common to want to create a task and schedule it right away, so the macro `@async` is provided for that purpose -- `@async x` is equivalent to `schedule(@task x)`.

Communicating with Channels

In some problems, the various pieces of required work are not naturally related by function calls; there is no obvious "caller" or "callee" among the jobs that need to be done. An example is the producer-consumer problem, where one complex procedure is generating values and another complex procedure is consuming them. The consumer cannot simply call a producer function to get a value, because the producer may have more values to generate and so might not yet be ready to return. With tasks, the producer and consumer can both run as long as they need to, passing values back and forth as necessary.

Julia provides a `Channel` mechanism for solving this problem. A `Channel` is a waitable first-in first-out queue which can have multiple tasks reading from and writing to it.

Let's define a producer task, which produces values via the `put!` call. To consume values, we need to schedule the producer to run in a new task. A special `Channel` constructor which accepts a 1-arg function as an argument can be used to run a task bound to a channel. We can then `take!` values repeatedly from the channel object:

```
julia> function producer(c::Channel)
           put!(c, "start")
           for n=1:4
               put!(c, 2n)
           end
           put!(c, "stop")
       end;
```

```
julia> chnl = Channel(producer);

julia> take!(chnl)
"start"

julia> take!(chnl)
2

julia> take!(chnl)
4

julia> take!(chnl)
6

julia> take!(chnl)
8

julia> take!(chnl)
"stop"
```

One way to think of this behavior is that `producer` was able to return multiple times. Between calls to `put!`, the producer's execution is suspended and the consumer has control.

The returned `Channel` can be used as an iterable object in a `for` loop, in which case the loop variable takes on all the produced values. The loop is terminated when the channel is closed.

```
julia> for x in Channel(producer)
           println(x)
       end
start
2
4
6
8
stop
```

Note that we did not have to explicitly close the channel in the producer. This is because the act of binding a `Channel` to a `Task` associates the open lifetime of a channel with that of the bound task. The channel object is closed automatically when the task terminates. Multiple channels can be bound to a task, and vice-versa.

While the `Task` constructor expects a 0-argument function, the `Channel` method that creates a task-bound channel expects a function that accepts a single argument of type `Channel`. A common pattern is

for the producer to be parameterized, in which case a partial function application is needed to create a 0 or 1 argument anonymous function.

For `Task` objects this can be done either directly or by use of a convenience macro:

```
function mytask(myarg)
    ...
end

taskHdl = Task(() -> mytask(7))
# or, equivalently
taskHdl = @task mytask(7)
```

To orchestrate more advanced work distribution patterns, `bind` and `schedule` can be used in conjunction with `Task` and `Channel` constructors to explicitly link a set of channels with a set of producer/consumer tasks.

More on Channels

A channel can be visualized as a pipe, i.e., it has a write end and a read end :

- Multiple writers in different tasks can write to the same channel concurrently via `put!` calls.
- Multiple readers in different tasks can read data concurrently via `take!` calls.
- As an example:

```
# Given Channels c1 and c2,
c1 = Channel(32)
c2 = Channel(32)

# and a function `foo` which reads items from c1, processes the item read
# and writes a result to c2,
function foo()
    while true
        data = take!(c1)
        [...]               # process data
        put!(c2, result)    # write out result
    end
end

# we can schedule `n` instances of `foo` to be active concurrently.
for _ in 1:n
    @async foo()
```

```
end
```

- Channels are created via the `Channel{T}(sz)` constructor. The channel will only hold objects of type `T`. If the type is not specified, the channel can hold objects of any type. `sz` refers to the maximum number of elements that can be held in the channel at any time. For example, `Channel(32)` creates a channel that can hold a maximum of 32 objects of any type. A `Channel{MyType}(64)` can hold up to 64 objects of `MyType` at any time.
- If a `Channel` is empty, readers (on a `take!` call) will block until data is available.
- If a `Channel` is full, writers (on a `put!` call) will block until space becomes available.
- `isready` tests for the presence of any object in the channel, while `wait` waits for an object to become available.
- A `Channel` is in an open state initially. This means that it can be read from and written to freely via `take!` and `put!` calls. `close` closes a `Channel`. On a closed `Channel`, `put!` will fail. For example:

```
julia> c = Channel(2);

julia> put!(c, 1) # `put!` on an open channel succeeds
1

julia> close(c);

julia> put!(c, 2) # `put!` on a closed channel throws an exception.
ERROR: InvalidStateException("Channel is closed.",:closed)
Stacktrace:
[...]
```

- `take!` and `fetch` (which retrieves but does not remove the value) on a closed channel successfully return any existing values until it is emptied. Continuing the above example:

```
julia> fetch(c) # Any number of `fetch` calls succeed.
1

julia> fetch(c)
1

julia> take!(c) # The first `take!` removes the value.
1

julia> take!(c) # No more data available on a closed channel.
ERROR: InvalidStateException("Channel is closed.",:closed)
Stacktrace:
[...]
```

Consider a simple example using channels for inter-task communication. We start 4 tasks to process data from a single `jobs` channel. Jobs, identified by an id (`job_id`), are written to the channel. Each task in this simulation reads a `job_id`, waits for a random amount of time and writes back a tuple of `job_id` and the simulated time to the results channel. Finally all the `results` are printed out.

```
julia> const jobs = Channel{Int}(32);

julia> const results = Channel{Tuple}(32);

julia> function do_work()
           for job_id in jobs
               exec_time = rand()
               sleep(exec_time)                # simulates elapsed time doing actual
                                               # typically performed externally.
               put!(results, (job_id, exec_time))
           end
       end;

julia> function make_jobs(n)
           for i in 1:n
               put!(jobs, i)
           end
       end;

julia> n = 12;

julia> @async make_jobs(n); # feed the jobs channel with "n" jobs

julia> for i in 1:4 # start 4 tasks to process requests in parallel
           @async do_work()
       end

julia> @elapsed while n > 0 # print out results
           job_id, exec_time = take!(results)
           println("$job_id finished in $(round(exec_time; digits=2)) seconds")
           global n = n - 1
       end
4 finished in 0.22 seconds
3 finished in 0.45 seconds
1 finished in 0.5 seconds
7 finished in 0.14 seconds
2 finished in 0.78 seconds
5 finished in 0.9 seconds
9 finished in 0.36 seconds
```

```
6 finished in 0.87 seconds
8 finished in 0.79 seconds
10 finished in 0.64 seconds
12 finished in 0.5 seconds
11 finished in 0.97 seconds
0.029772311
```

More task operations

Task operations are built on a low-level primitive called `yieldto`. `yieldto(task, value)` suspends the current task, switches to the specified `task`, and causes that task's last `yieldto` call to return the specified `value`. Notice that `yieldto` is the only operation required to use task-style control flow; instead of calling and returning we are always just switching to a different task. This is why this feature is also called "symmetric coroutines"; each task is switched to and from using the same mechanism.

`yieldto` is powerful, but most uses of tasks do not invoke it directly. Consider why this might be. If you switch away from the current task, you will probably want to switch back to it at some point, but knowing when to switch back, and knowing which task has the responsibility of switching back, can require considerable coordination. For example, `put!` and `take!` are blocking operations, which, when used in the context of channels maintain state to remember who the consumers are. Not needing to manually keep track of the consuming task is what makes `put!` easier to use than the low-level `yieldto`.

In addition to `yieldto`, a few other basic functions are needed to use tasks effectively.

- `current_task` gets a reference to the currently-running task.
- `istaskdone` queries whether a task has exited.
- `istaskstarted` queries whether a task has run yet.
- `task_local_storage` manipulates a key-value store specific to the current task.

Tasks and events

Most task switches occur as a result of waiting for events such as I/O requests, and are performed by a scheduler included in Julia Base. The scheduler maintains a queue of runnable tasks, and executes an event loop that restarts tasks based on external events such as message arrival.

The basic function for waiting for an event is `wait`. Several objects implement `wait`; for example, given a `Process` object, `wait` will wait for it to exit. `wait` is often implicit; for example, a `wait` can happen inside a call to `read` to wait for data to be available.

In all of these cases, `wait` ultimately operates on a `Condition` object, which is in charge of queueing and restarting tasks. When a task calls `wait` on a `Condition`, the task is marked as non-runnable, added to the condition's queue, and switches to the scheduler. The scheduler will then pick another task to run, or block waiting for external events. If all goes well, eventually an event handler will call `notify` on the condition, which causes tasks waiting for that condition to become runnable again.

A task created explicitly by calling `Task` is initially not known to the scheduler. This allows you to manage tasks manually using `yieldto` if you wish. However, when such a task waits for an event, it still gets restarted automatically when the event happens, as you would expect.

« Parallel Computing Multi-Threading »