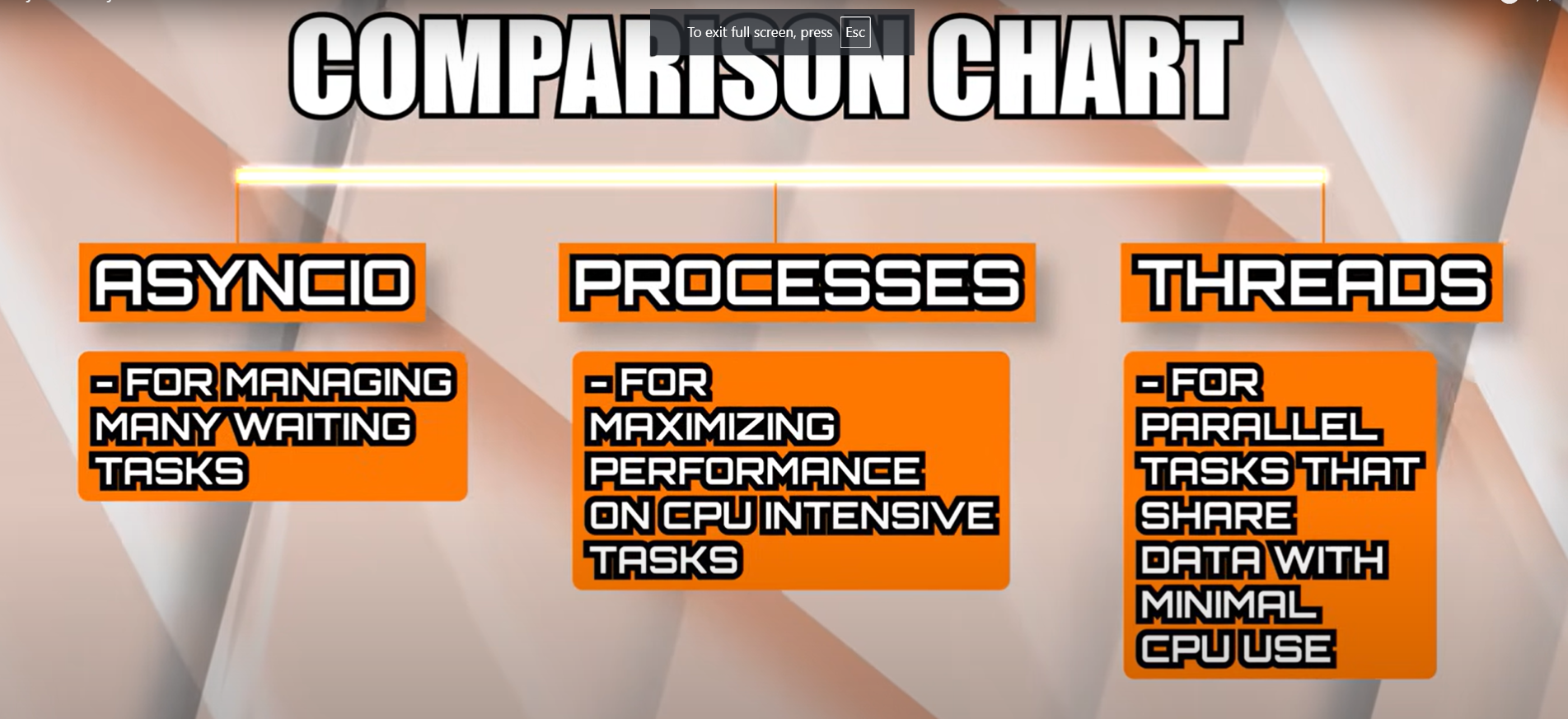
**asyncio**: the Python package that provides a foundation and API for running and managing coroutines.

**Parallelism** consists of performing multiple operations at the same time. **Multiprocessing** is a means to effect parallelism, and it entails spreading tasks over a computer’s central processing units (CPUs, or cores). Multiprocessing is well-suited for CPU-bound tasks: tightly bound [for loops](https://realpython.com/python-for-loop/) and mathematical computations usually fall into this category.

**Concurrency** is a slightly broader term than parallelism. It suggests that multiple tasks have the ability to run in an overlapping manner. (There’s a saying that concurrency does not imply parallelism.)

**Threading** is a concurrent execution model whereby multiple [threads](https://en.wikipedia.org/wiki/Thread_(computing)) take turns executing tasks. One process can contain multiple threads. Python has a complicated relationship with threading thanks to its [GIL](https://realpython.com/python-gil/),



**Coroutines:-**

At the heart of async IO are coroutines. A coroutine is a specialized version of a Python generator function. Let’s start with a baseline definition and then build off of it as you progress here: a coroutine is a function that can suspend its execution before reaching return, and it can indirectly pass control to another coroutine for some time.

­­

1:- **time.sleep()** is like telling your program to pause completely for a few seconds (or however long you specify). During this time, your program can’t do anything else—it’s just sitting there, waiting. This is called "blocking" because nothing else can happen until the wait is over.

2:- **asyncio.sleep()** is different. It still tells your program to wait, but in a way that allows it to do other things during the wait time. This is "non-blocking" because your program can keep working on other tasks while it’s waiting.

**An Intro to Threading in Python**

**Multithreading**is defined as the ability of a processor to execute multiple threads concurrently. In a simple, single-core CPU, it is achieved using frequent switching between threads. This is termed **context switching**. In context switching, the state of a thread is saved and the state of another thread is loaded whenever any interrupt (due to I/O or manually set) takes place. Context switching takes place so frequently that all the threads appear to be running parallelly (this is termed **multitasking**.