

What is I/O?



I/O is an acronym for Input/Output. Input/Output can be best explained using a system that we call a black box. We call it black because we can't see inside the box, making its inner workings hidden. But we can observe and measure how the output changes with respect to the input. If we can accurately model this transformation process to predict the output for any input, then we don't need to know what's happening on the inside.

The examples on the right show four possible transformations, the top transformation shows the output has not changed, the remaining show the outputs color, shape and size changing.

In storage, the black box somehow transports and transforms the input. The transport step happens in all situations while the transform step is optional. In the top example where the output is equal to the input, only transport occurs; for the remaining examples where the output is not equal to the input, both transport and transformation take place.

Operating systems can be divided into two major software subsystems – user space and kernel space. The input usually begins its journey in user space and is transported across user space into kernel space until it eventually arrives at its destination on the storage device. If the input is also transformed, typical transformations are encryption, compression and de-duplication.

The output behavior, no matter the storage technology, we can observe three main phases: (a) a command phase followed by (b) a data phase ending with (c) a response phase. In general, the command phase will consist of a read or write command, the amount of data (i.e., input) to transfer, and the location on the storage device where the data resides; the data phase will be the transmission of the input (transformed or non-transformed) from or to the storage device, and the response phase will be the storage device acknowledging that the data phase was successful or unsuccessful.

One complete command phase, data phase, response phase operation is called a I/O completion. The rate of I/O completions per second are called IOPS. The maximum IOPS a device can achieve is a common performance measurement. Throughput is another common performance measurement and is defined as the number of bytes transferred in a successful I/O completion per second.

