Una cola es una colección ordenada de ítems donde la adición de nuevos ítems tiene lugar en uno de los extremos, denominado “final”, y la remoción de ítems existentes ocurre en el otro extremo, comúnmente llamado “frente”. Un elemento ingresa a la cola por el final y espera hasta el momento que un ítem sea eliminado para avanzar hacia el frente.

Un item más recientemente agregado en la cola debe esperar al final de la colección. El ítem que ha permanecido más tiempo en la colección está en el frente. Este principio de ordenamiento a veces se denomina FIFO (first-in first-out), también conocido como el primero en llegar es el primero en ser atendido.

El ejemplo más simple de una cola es la fila típica en la que todos participamos de vez en cuando. Esperamos en una fila para una película, esperamos en la fila de pago en una tienda de comestibles, y esperamos en la fila de la cafetería (para que podamos extraer de la pila de bandejas). Las filas de buen comportamiento, o colas, son muy restrictivas en el sentido de que sólo tienen un modo de ingresar a ella y una sola salida.

Las ciencias de la computación también tienen ejemplos comunes de colas. Nuestro laboratorio de computadoras tiene 30 computadoras conectadas en red con una sola impresora. Cuando los estudiantes quieren imprimir, sus tareas de impresión “hacen fila” con todas las otras tareas de impresión que están esperando. La primera tarea es la próxima en ser completada. Si usted es el último en la fila, deberá esperar a que todas las otras tareas se impriman por delante de la suya. Vamos a explorar este interesante ejemplo con más detalle más adelante.

Además de las colas de impresión, los sistemas operativos utilizan varias colas diferentes para controlar los procesos dentro de una computadora. La programación de los procesos que se harán a continuación se basa normalmente en un algoritmo de colas que intenta ejecutar programas lo más rápidamente posible y servir a tantos usuarios como pueda. Además, a medida que tecleamos, a veces las pulsaciones de las teclas se adelantan a los caracteres que aparecen en la pantalla. Esto se debe a que la computadora está haciendo otro trabajo en ese momento. Las pulsaciones de teclas se ubican en un búfer de tipo cola para que eventualmente se muestren en la pantalla en el orden correcto.