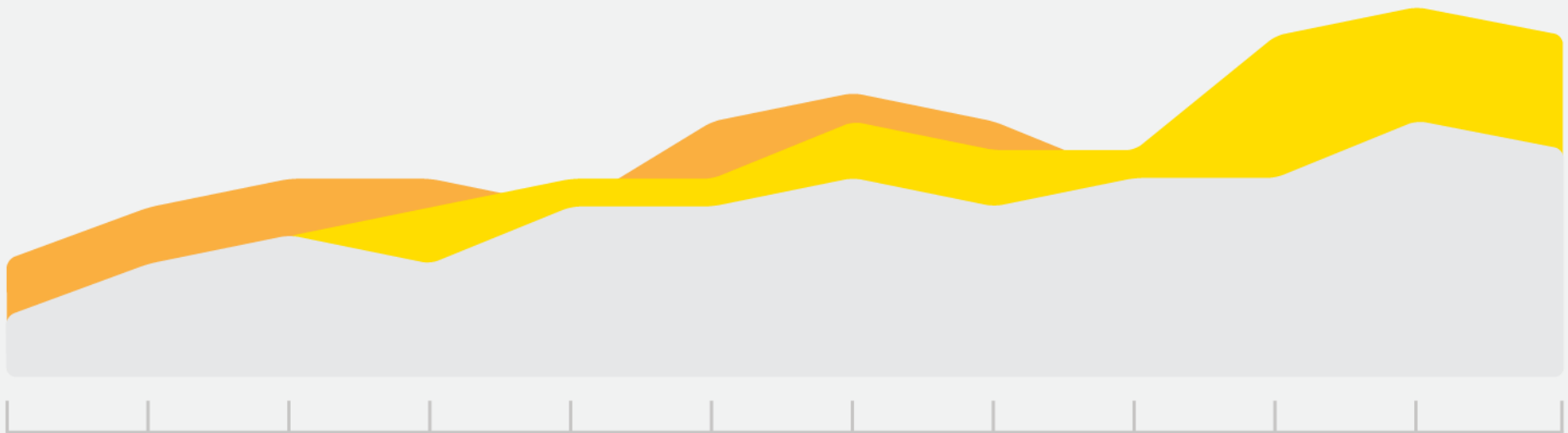




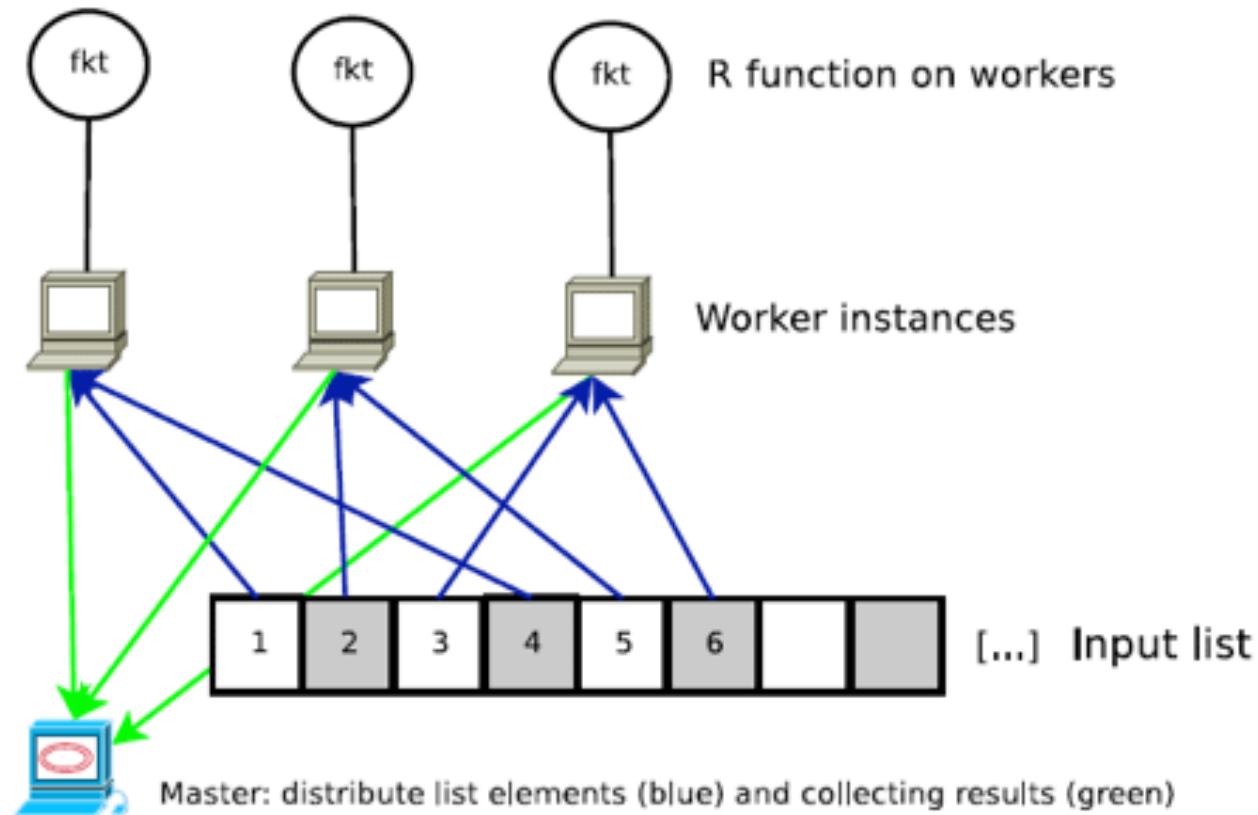
Big Data en R

Procesos Paralelos

El paquete Snow



El paquete “snow” para Computación Paralela



El paquete “snow” para Computación Paralela

- **Snow:** “Simple Network of Workstations” es probablemente el paquete de programación paralela más popular disponible para R.
- Fue escrito por Luke Tierney, AJ Rossini, Na Li, y H. Sevcikova, y es mantenido activamente por Luke Tierney.
- Es un paquete maduro, lanzado por primera vez en el "Comprehensive R Archive Red" (CRAN) en 2003.



El paquete “snow” para Computación Paralela

- El paquete “**snow**” le proporciona a R el soporte necesario para ejecutar funciones en paralelo.
- La mayor parte de las funciones de “**snow**” son variaciones de la función *lapply()* de R.
- Para llevar a cabo estas operaciones en paralelo, “**snow**” utiliza una arquitectura **jefe/peón**, donde el “jefe” envía tareas a los “peones”, luego los “peones” ejecutan las tareas asignadas y devuelven los resultados al “jefe”.



El paquete “snow” para Computación Paralela

- Para ejecutar funciones en paralelo “snow”, primero debe crear un objeto al “Grupo de Peones”. Este objeto se utiliza para interactuar con todos los peones y se pasa como primer argumento a la mayoría de las funciones de “snow”, como sigue:

```
cl <- makeCluster(4, type="SOCK")
```

- El 4 representa el número de “peones”.



El paquete “snow” para Computación Paralela

- El tipo puede ser “SOCK”, “MPI”, “PVM” or “NWS”, por ejemplo:

```
cl <- makeCluster(4, type="MPI")
```

- Exporta las variables a los nodos:

```
clusterExport(cl, list, envir = .GlobalEnv)
```

- Para apagar o terminar la computación paralela ejecute:

```
stopCluster(cl)
```

[<Ver: OtrosPaquetes.html>](#)



Data Analysis in the Distributed World



Parallel R

O'REILLY®

by John McCallum & Stephen Weis



oldemar rodríguez
CONSULTOR en MINERÍA DE DATOS

Gracias ...



oldemar rodíguez
CONSULTOR en MINERÍA DE DATOS