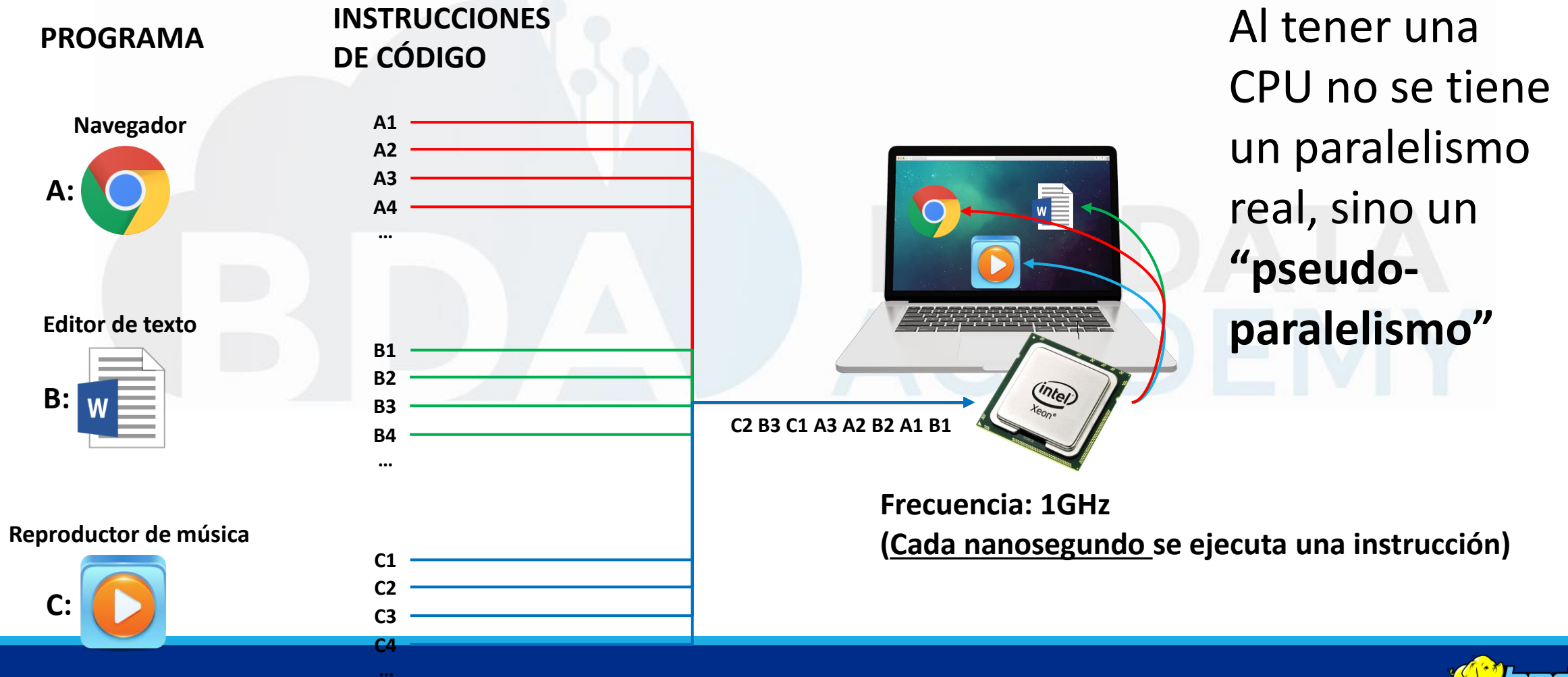


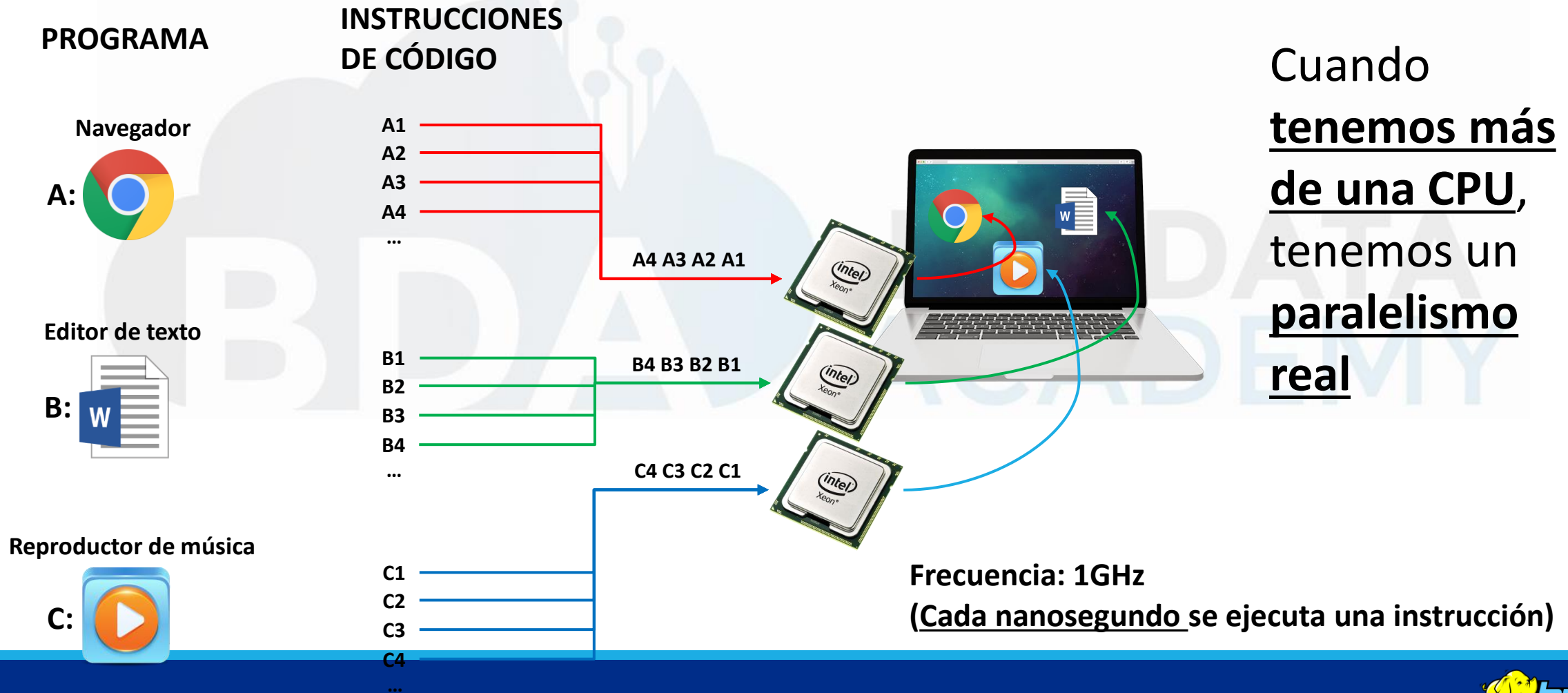
Arquitectura de procesamiento real-time

PARTE DE LA FORMACIÓN BIG DATA ENGINEER DE BIG DATA ACADEMY PERÚ

Funcionamiento de una computadora común



Funcionamiento de una computadora común con más de una CPU



Funcionamiento de un server



CPUs destinadas a servicios para que el servidor funcione (Sistema Operativo, Firewall, Datanode, Worker, etc)

CPUs para el procesamiento de datos

El P3 y el P10 compiten por los tiempos de la misma CPU

Un proceso paralelizable



P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
...

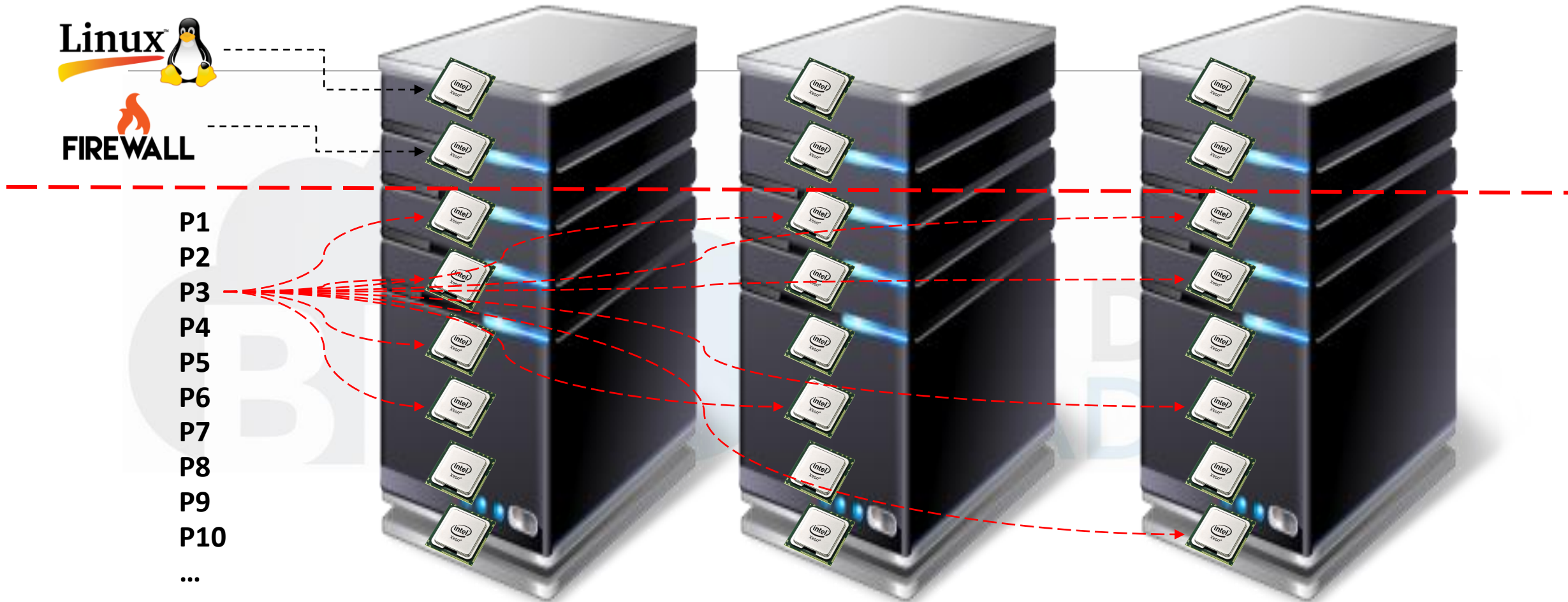


CPU's destinadas a servicios para que el servidor funcione (Sistema Operativo, Firewall, Datanode, Worker, etc)

Si alguno de nuestros procesos es paralelizable, la lógica puede repartirse en más de una CPU, y por lo tanto aumentar en un factor lineal la velocidad de procesamiento (x2, x3, x4)

Mientras más CPUs asignadas al proceso, más factor de paralelización y por lo tanto más rapidez de procesamiento

Un proceso paralelizable sobre un clúster



Un clúster nos permite disponer de muchas más CPUs que un solo servidor y por lo tanto escalar fácilmente los recursos hardware disponibles

Funcionamiento de un proceso BATCH



FIREWALL

P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
...



CPU's destinadas a servicios para que el servidor funcione (Sistema Operativo, Firewall, Datanode, Worker, etc)

CPU's reservadas para el proceso P3:

HORA DE INICIO: 01:00 AM

HORA DE FIN: 02:00 AM

HILOS DE PROCESAMIENTO: 1 x CPU

Al finalizar el procesamiento, las CPU's se liberan y están libres para otros procesos

Funcionamiento de un proceso REAL TIME



P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
...



CPU's destinadas a servicios para que el servidor funcione (Sistema Operativo, Firewall, Datanode, Worker, etc)

CPU's reservadas para el proceso P3:

HORA DE INICIO: **PERMANENTE**

HORA DE FIN: **PERMANENTE**

HILOS DE PROCESAMIENTO: **¿? X CPU**

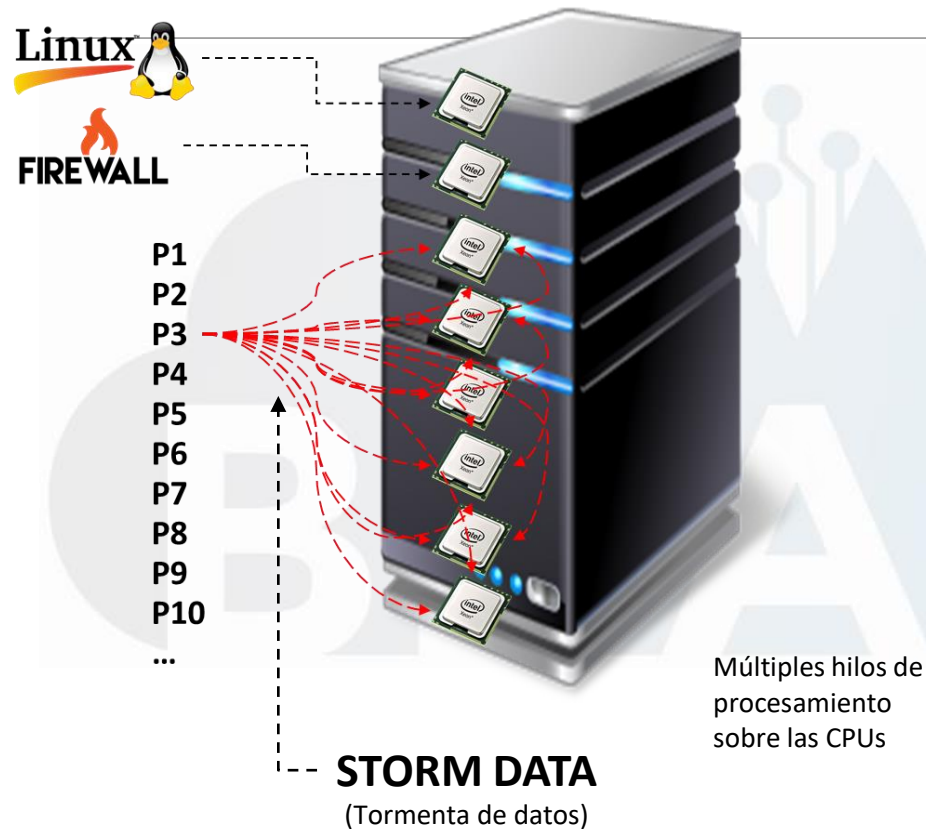
Existe un gran riesgo de que el proceso monopolice los recursos de CPU

ANTI-PATRÓN: MONOPOLIZACIÓN REAL TIME

Las peticiones de procesamiento se atienden como van llegando

Micro-batch

REAL TIME



Cada hilo se atiende como va llegando (0.001 segundos)

MICRO-BATCH



Los hilos se acumulan y se atienden por lote cada cierto tiempo (1 segundo)

Funcionamiento de un server para un proceso micro-batch



FIREWALL

P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
...

encolamiento



CPU's destinadas a servicios para que el servidor funcione (Sistema Operativo, Firewall, Datanode, Worker, etc)

CPU's reservadas para el proceso P3:

HORA DE INICIO: **PERMANENTE**

HORA DE FIN: **PERMANENTE**

HILOS DE PROCESAMIENTO: **1 X CPU**

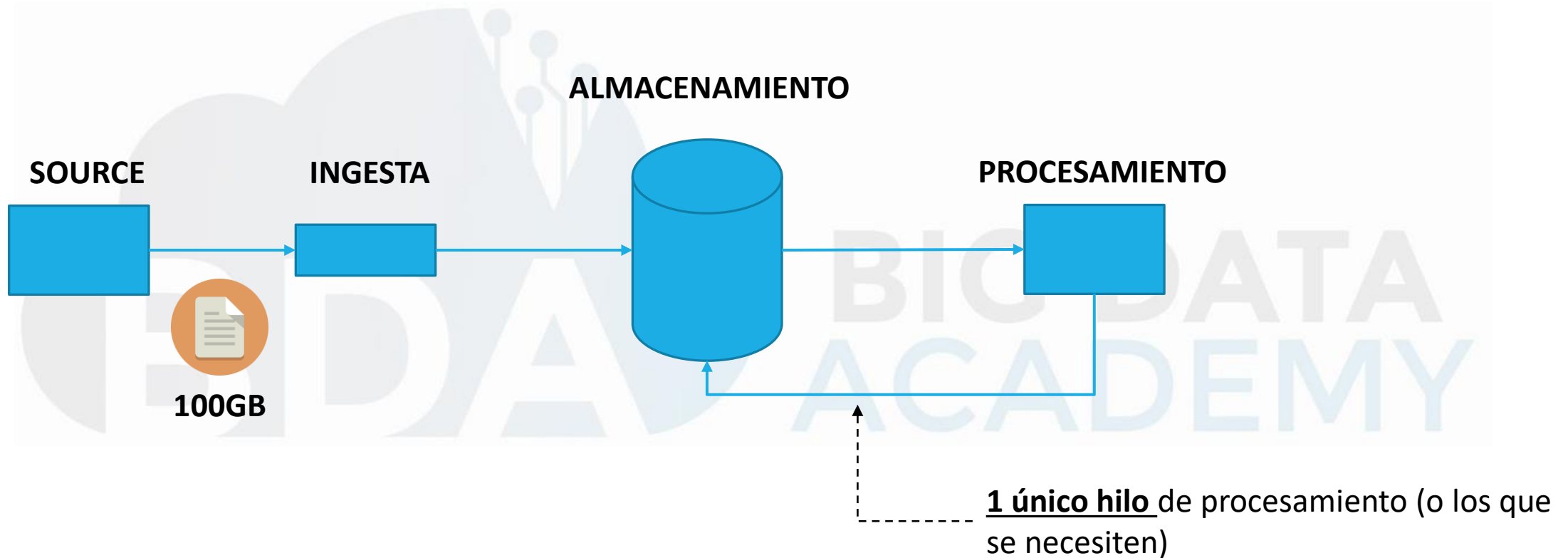
TIEMPO DE ENCOLAMIENTO: 1 segundo

Existe un gran riesgo de que el proceso monopolice los recursos de CPU

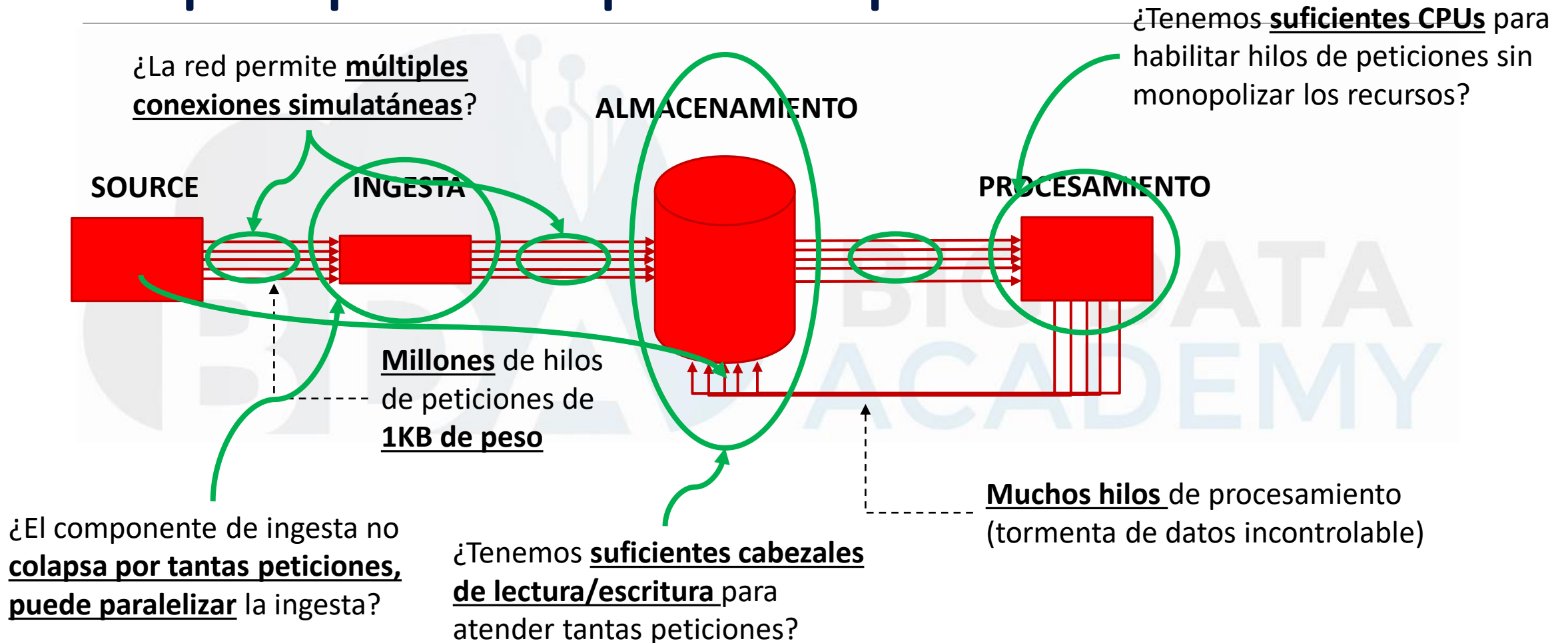
PATRÓN: MICRO-BATCH COMO REAL TIME

Las peticiones de procesamiento se atienden por lotes

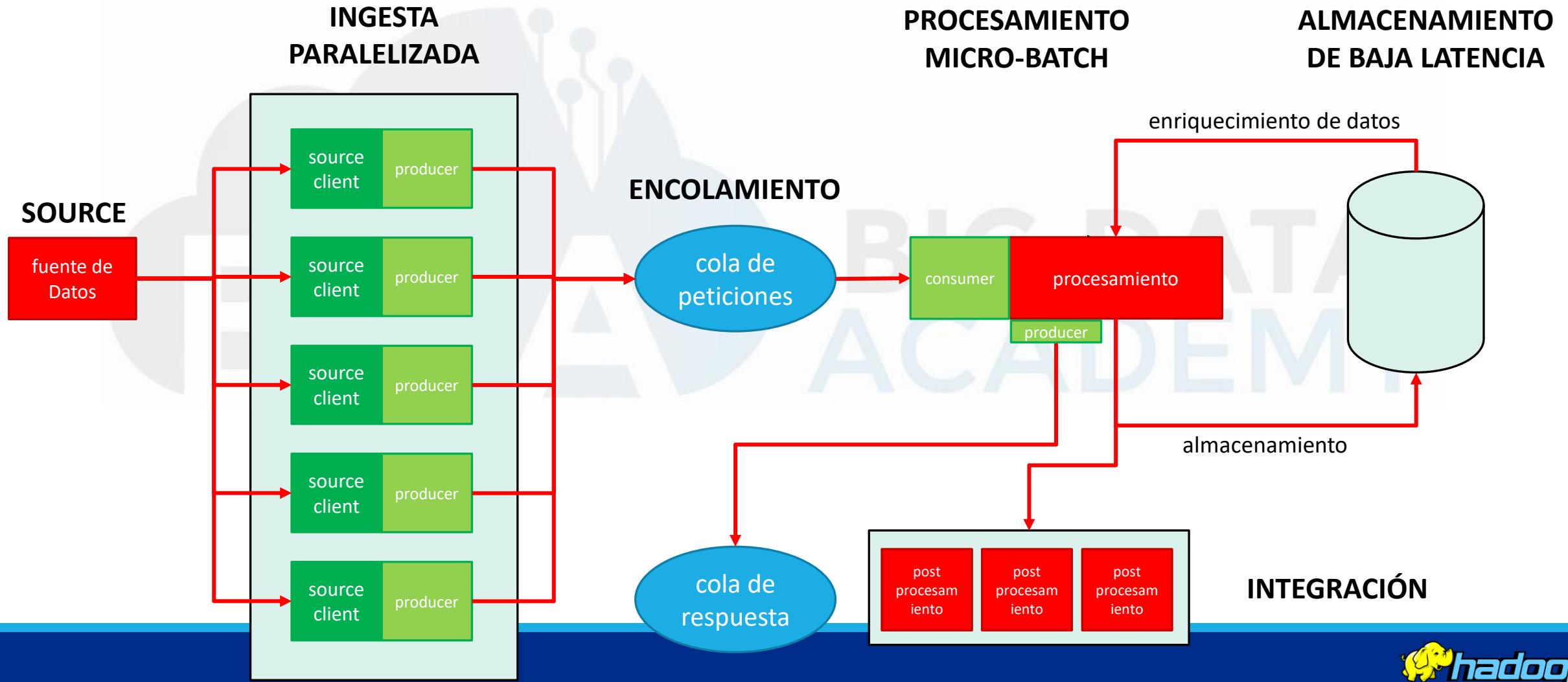
Arquetipo base para un proceso batch



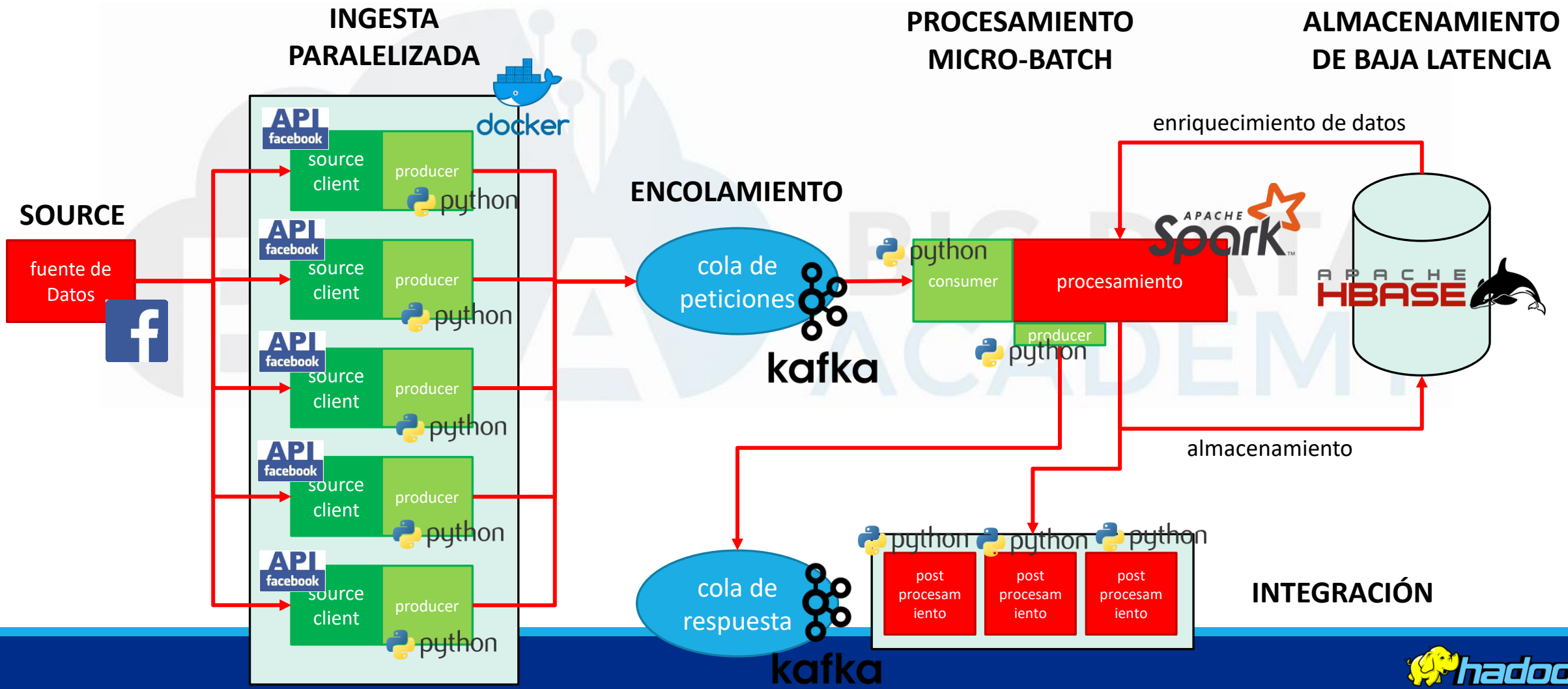
Arquetipo base para un proceso real time



Arquitectura base para un proceso micro-batch como Real Time



Arquitectura aplicada: Scrapping de Facebook



Arquitectura aplicada: Geolocalización móvil

