

Capítulo 3 - Conceptos y Técnicas de Procesamiento de Big Data

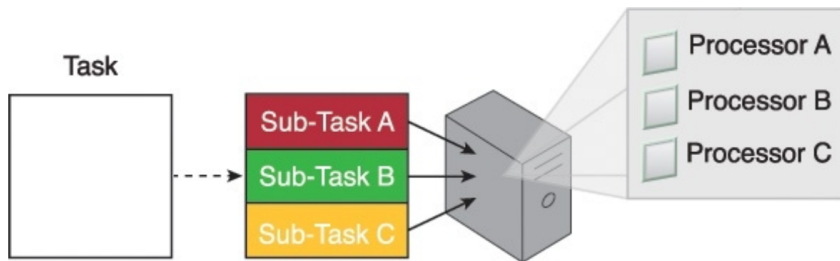
Jaime Veintimilla Reyes

December 1, 2020

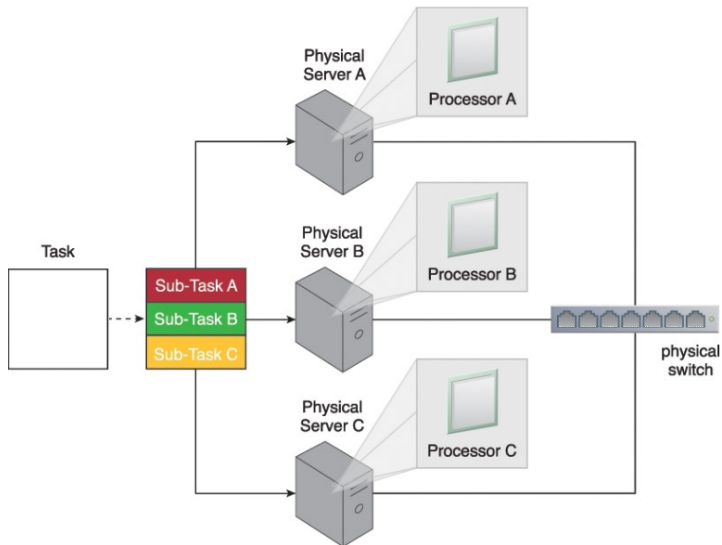
1 Conceptos y Técnicas de Procesamiento de Big Data

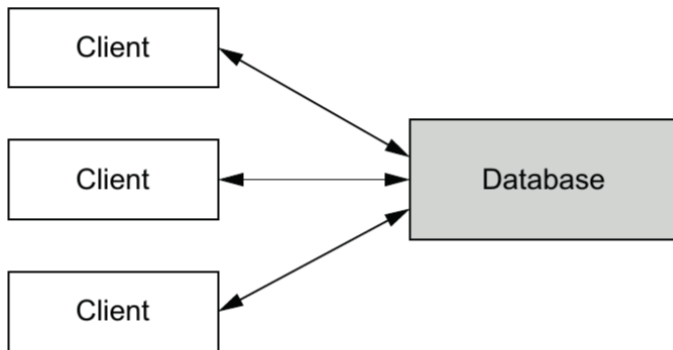
2 Preguntas

Procesamiento Paralelo



Procesamiento Distribuido





With synchronous updates, clients communicate directly with the database and block until the update is completed.

Figure 12.7 A simple speed layer architecture using synchronous updates

Actualización síncrona

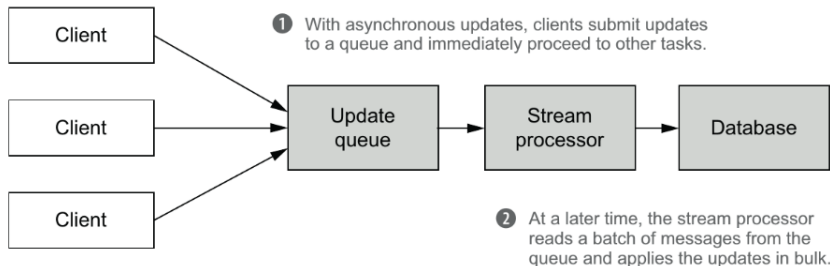


Figure 12.8 Asynchronous updates provide higher throughput and readily handle variable loads.

- Arquitecturas Asíncronas
 - Colas
 - Stream Processing

Colas y Procesamiento de Streams

- Procesamiento sin colas persistentes
 - Dispara y Olvida (Fire and Forget)
 - Tráfico?

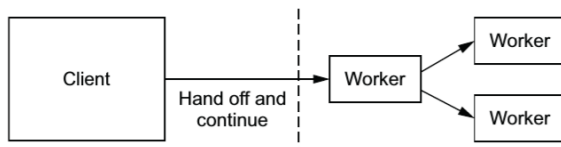


Figure 14.1 To implement asynchronous processing without queues, a client submits an event without monitoring whether its processing is successful.

(Single-consumer queue server)

- Los mensajes se eliminan de la cola cuando son confirmados
- Múltiples aplicaciones consumiendo los mismos eventos?
- El problema es que la cola controla que fue consumido y que no

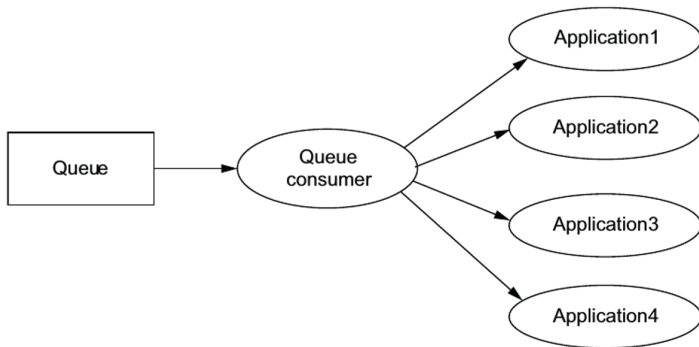


Figure 14.2 Multiple applications sharing a single queue consumer

(Multi-consumer queues)

- Pasar el control de los evento consumidos a la aplicación

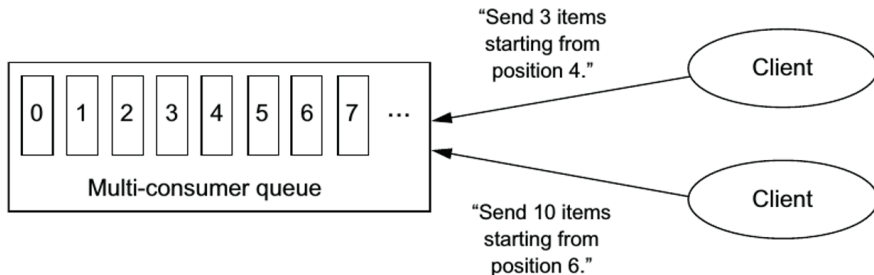


Figure 14.3 With a multi-consumer queue, applications request specific items from the queue and are responsible for tracking the successful processing of each event.

Stream processing

- Procesar los eventos y actualizar las vistas en tiempo real
 - Uno a la vez
 - Micro Batches

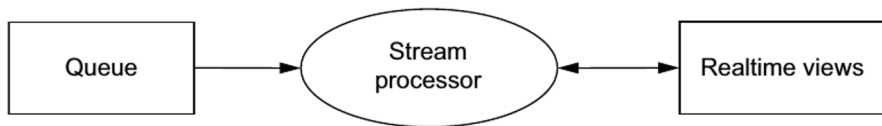


Figure 14.4 Stream processing

Stream processing

	One-at-a-time	Micro-batched
Lower latency	✓	
Higher throughput		✓
At-least-once semantics	✓	✓
Exactly-once semantics	In some cases	✓
Simpler programming model	✓	

Colas y Procesadores(Workers)

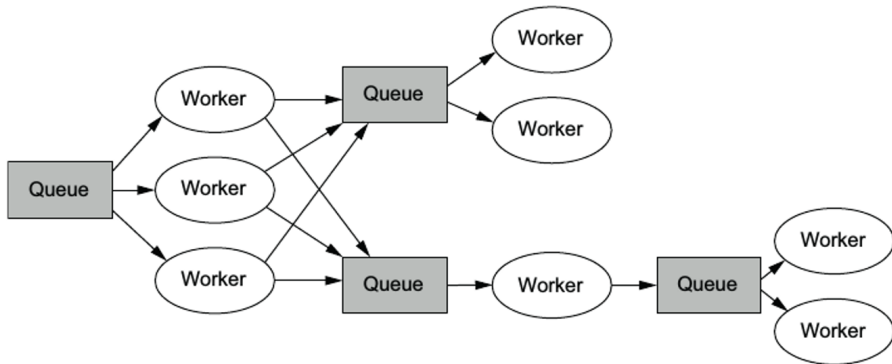


Figure 14.6 A representative system using a queues-and-workers architecture. The queues in the diagram could potentially be distributed queues as well.

1 Conceptos y Técnicas de Procesamiento de Big Data

2 Preguntas

- ¿Alguna Pregunta?.

