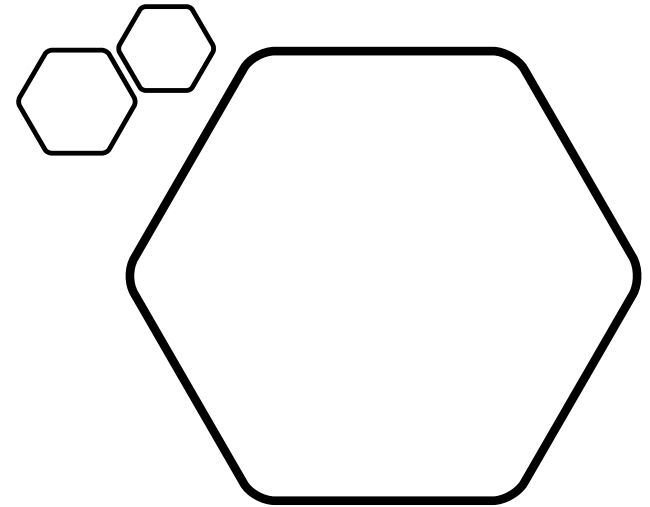
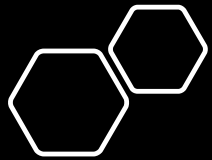


Topics

- “Assunto”
- Imutável
- Semelhante a uma tabela de banco de dados
- Suponde que vamos usar o Kafka para analisar Logs e Web Sites:





Muita coisa
pode ser
publicado!

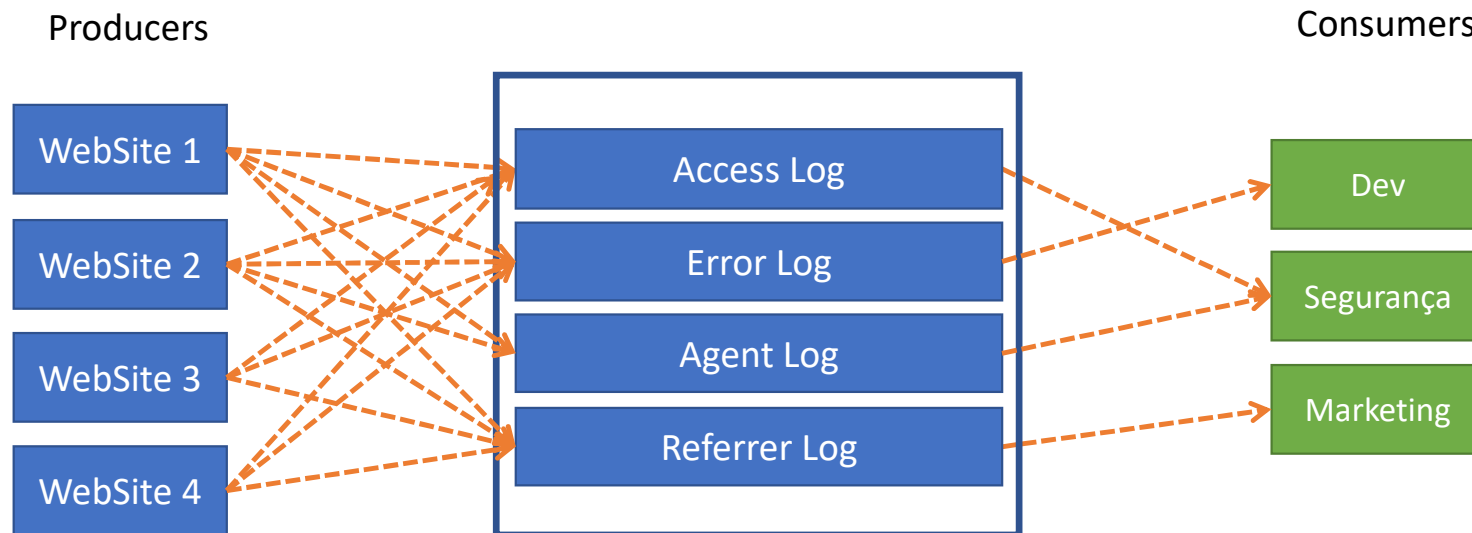
Access Log

Error Log

Agent Log

Referrer Log

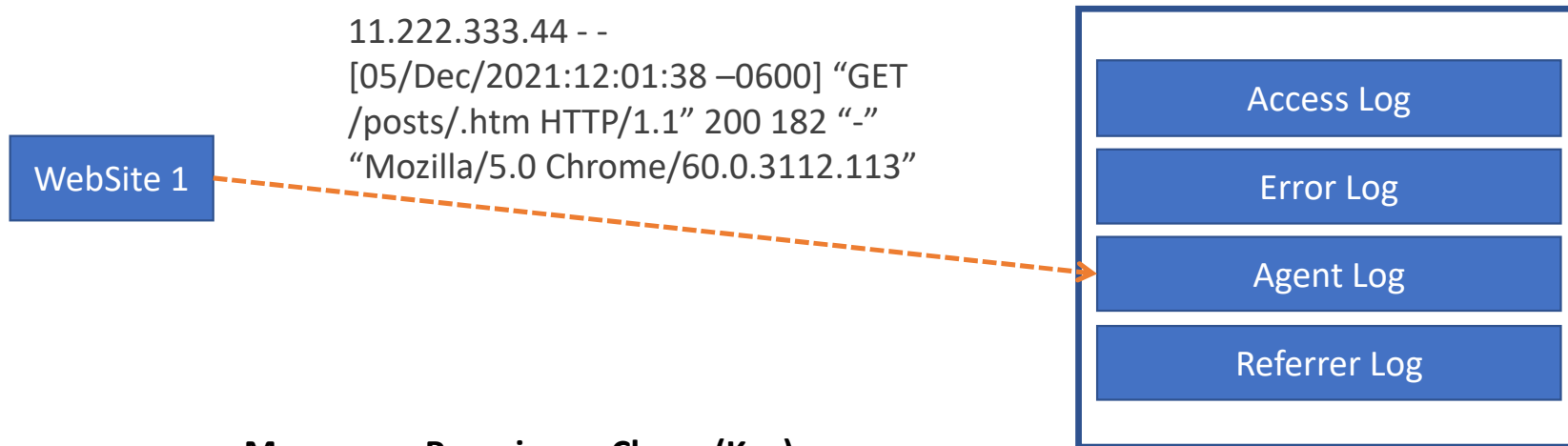
Topics



Imutável!

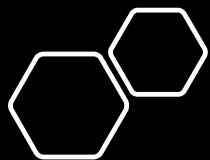


Message (events, records)



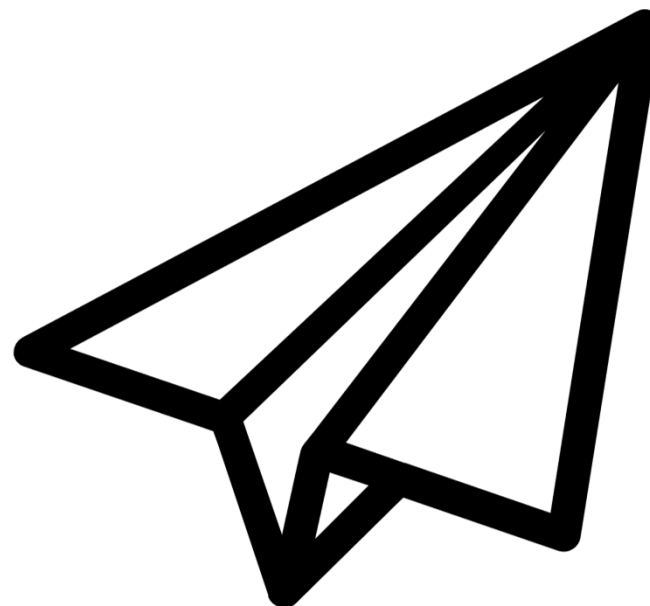
Mensagem Possui uma Chave (Key)



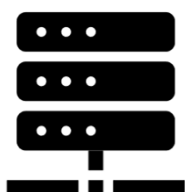


Message

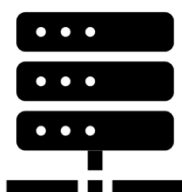
- Retenção
 - Pode ser configurado
 - Por tempo
 - Por tamanho do tópico
 - Pode reter apenas a última mensagem (log compacted)



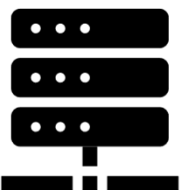
Topic Partitions



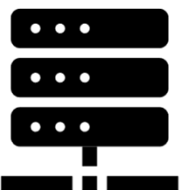
Access Log
Partition 0



Access Log
Partition 1



Access Log
Partition 2



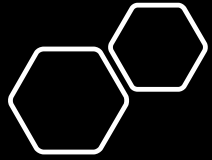
Access Log
Partition 3

- Partições são uma parte menor de um tópico
- Armazenado em um único nó do cluster!
- Mensagens são distribuídas nas partições – as partições não precisam ter o mesmo número de tópicos
- Definido pelo administrador do cluster
- Não pode ser dividido posteriormente!
- **Só pode haver um Consumer por Partition**
- **Não podem haver mais Consumers do que Partition**

Replication Factor

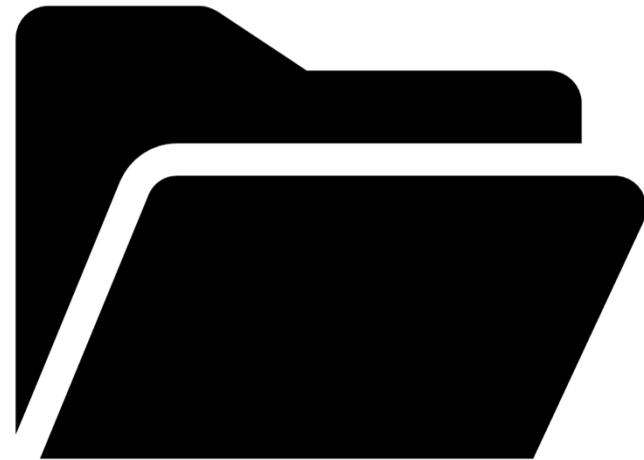
- Uma mesma partição pode estar replicada em mais de um broker: redundância
- = alta disponibilidade!

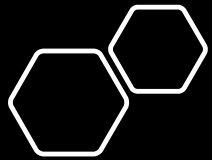




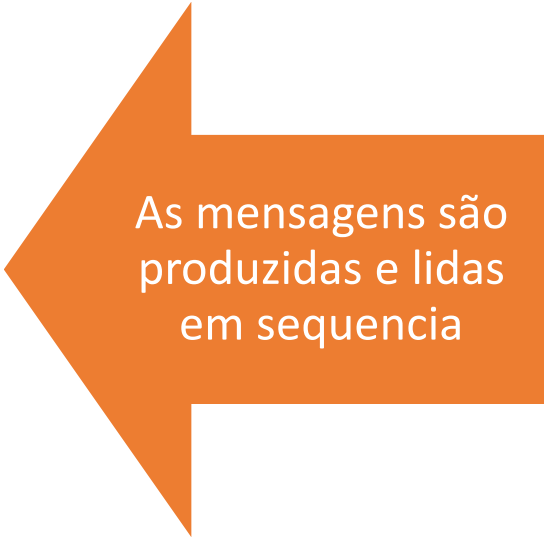
Topic Partitions

- Partições nada mais são do que Diretórios usados para dividir os dados
- Podem estar no mesmo servidor, ou em servidores diferentes





Topic



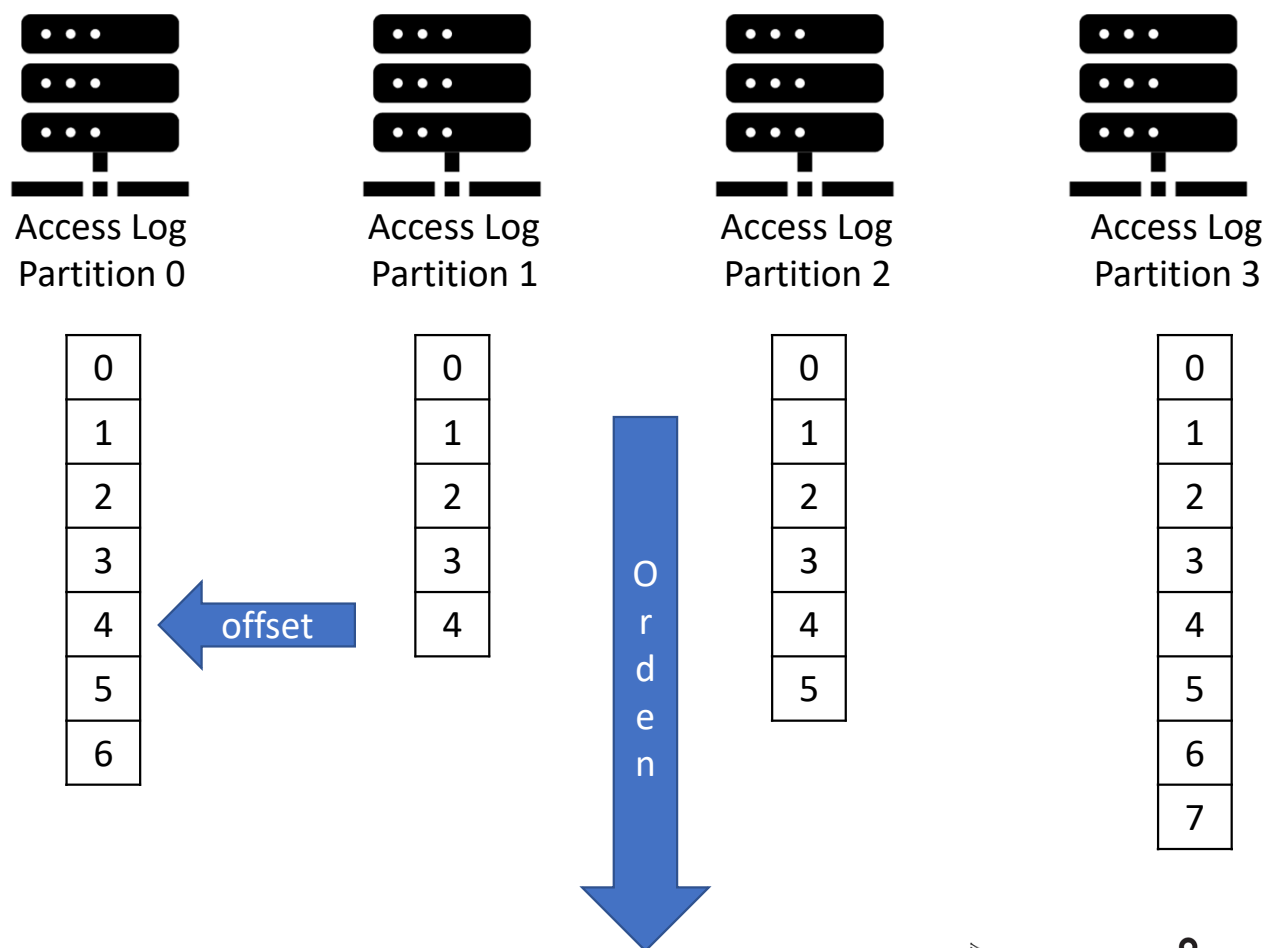
As mensagens são
produzidas e lidas
em sequencia

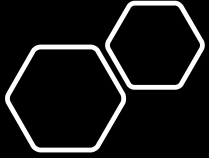


“Ordem”

Partition offset

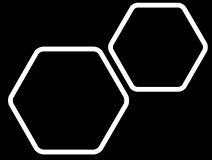
- ✓ Identificador único de uma mensagem na partição
- ✓ Gerado automaticamente
- ✓ Imutável
- ✓ Em ordem de chegada
- ✓ A ordem só existe na partição, mesmo em um mesmo Topic





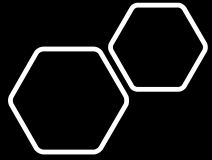
offset

- Offset são comitados num tópico chamado `__consumer_offsets`
- Um consumidor pode ler novamente a partir do número do offset
- Commit é o ponto de referência



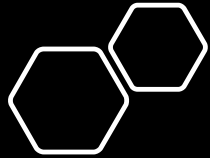
Identificar Mensagem de Forma Única

- Topic
- Partição (número)
- E offset



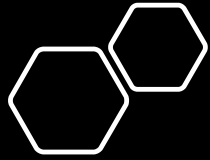
Como funciona

- O Consumer requisita e processa mensagens em intervalos de offset



Elementos de uma Mensagem

- Argumentos obrigatórios
 - Topic
 - Mensagem
- Opcional
 - Partition:
 - Podemos definir com base na estratégia de particionamento
 - Default: hash key ou rodizio
 - Timestamp
 - Pode ser definido, mas sempre haverá um
 - Pode ser a hora de criação ou hora de log
 - Message Key : usado para particionamento, agrupamento, joins etc.



Serialização

- Dados são serializados para transferir pela rede
- Podem ser usados serializadores como Avro