

# Cluster Kafka

Broker 0



PARTITION 0

PARTITION 1



Broker 1



PARTITION 1

PARTITION 3



Broker 2



PARTITION 0

PARTITION 3

Broker 3



PARTITION 1

PARTITION 2

PARTITION 3



# Distribuição de Partições

Broker 0



PARTITION 0

PARTITION 3

PARTITION 2

PARTITION 1

Broker 1



PARTITION 1

PARTITION 0

PARTITION 3

PARTITION 2

Broker 2



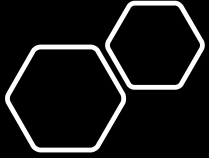
PARTITION 2

PARTITION 1

PARTITION 0

PARTITION 3



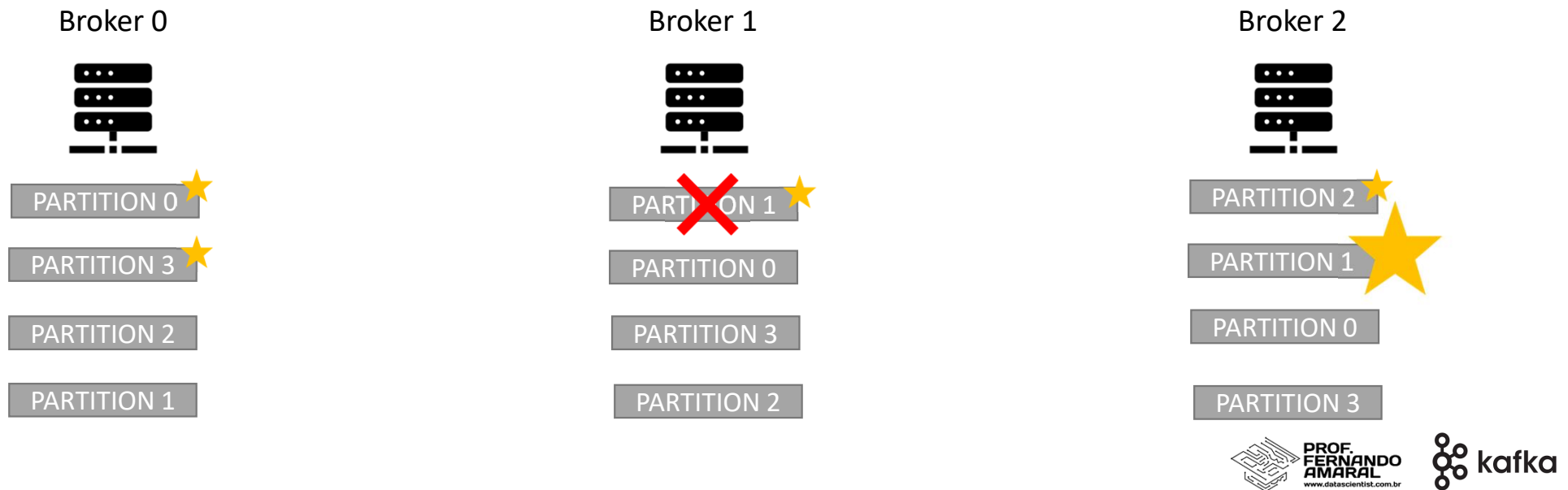


# Followers

- Followers não respondem a requisições!
- Apenas se mantem atualizados com o leader
- Ficam aguardando serem eleitos leaders se o leader falhar
- Followers podem falhar em se manter sincronizados: rede etc.

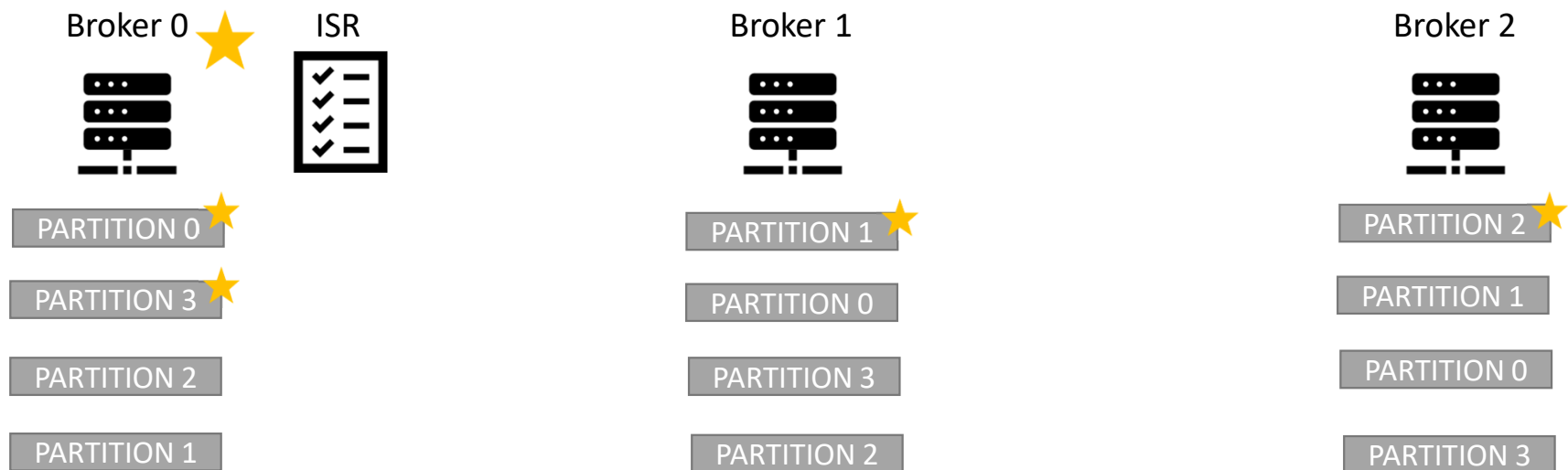
# Partições Podem Falhar!

- O objeto das replicas é disponibilidade
- Um Follower pode se tornar Leader em caso de falha



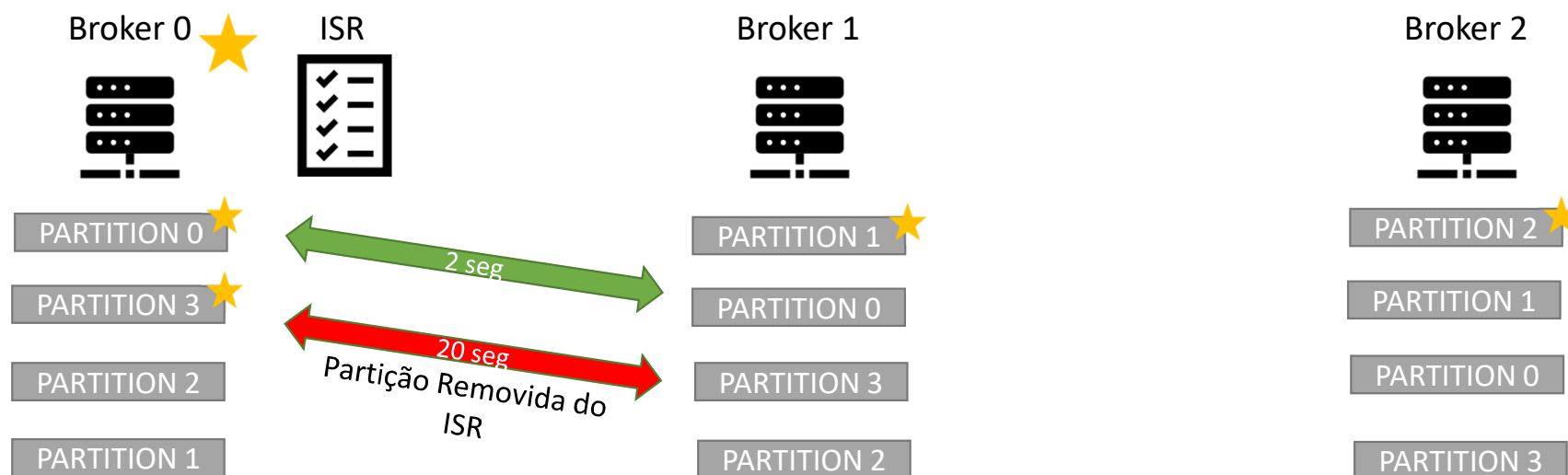
# ISR

- Broker Leader mantem a lista de In-Sync-Replicas (ISR)
- Só fica na lista replicas sincronizadas
- Se estiver sincronizado pode ser eleito leader



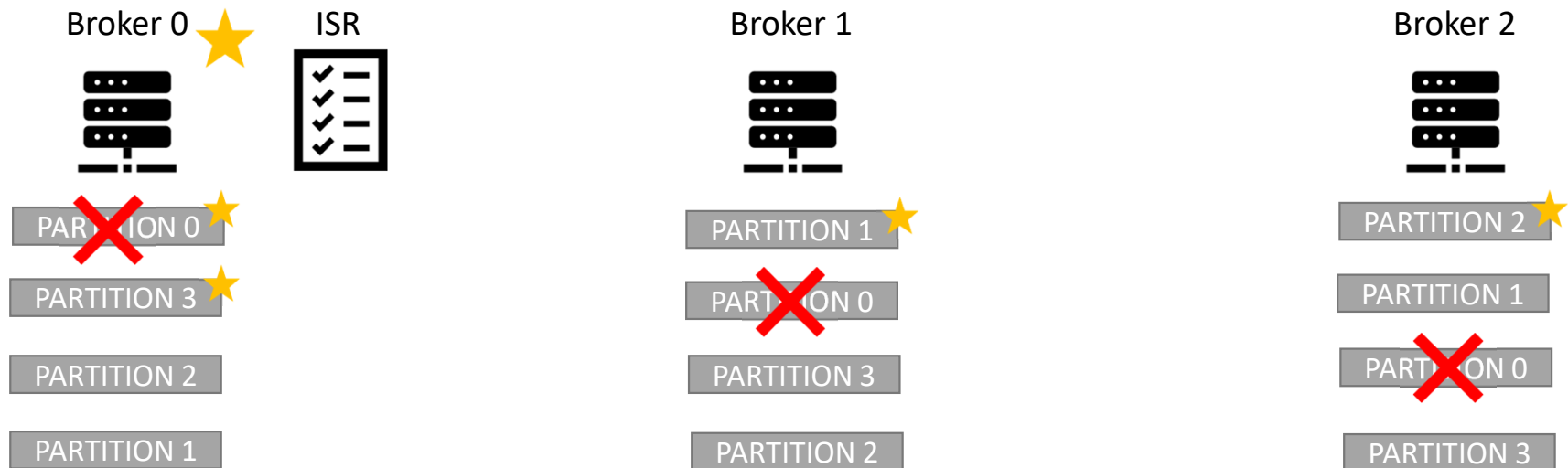
# ISR

- Como se mantem atualizados?
  - Pelos offsets (números sequenciais) solicitados pelas replicas
  - A tolerância por padrão é de 10 segundos: se a replica solicitou as mensagens mais recentes nos últimos 10 segundos, ela se mantém na lista



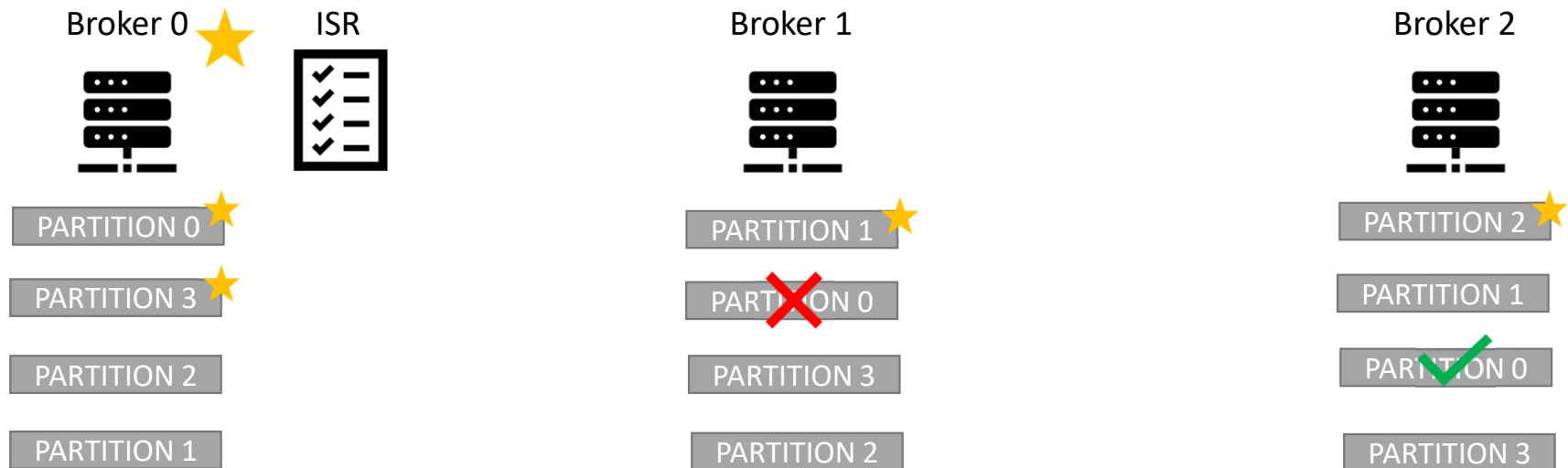
# Exemplo: Partição Zero

- Se todos os followers ficarem de fora da lista e o leader falhar?
- Como escolher um novo líder?

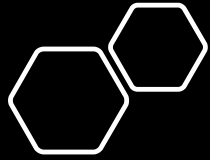


# Committed vs Uncommitted

- Podemos configurar o leader para não considerar a mensagem confirmada até que seja copiada a todas as replicas da lista ISR
- Dessa forma, o leader pode ter mensagens confirmadas e não confirmadas
- Se perdemos o leader perdemos só as mensagens não confirmadas
- Elas devem ser reenviadas pelos produtores, que ficam aguardando um sinal de recebido (acknowledgment)







# Minimum ISR List

- Podemos definir um número mínimo de replicas na lista
- Porém, se houver um problema, o broker não aceitará mensagens informado que não há replicas suficientes
- O leader se torna “Read Only”

# Replication Factor

- Quantas cópias haverão de cada partição?
- As replicas são criadas em outros brokers, que normalmente estarão em outros servidores
- Se definimos 4 partições com 3 réplicas, temos 12 partições
- Elas vão estar distribuídas nos Brokers disponíveis
- Obs: Tudo faz parte do mesmo Topic

Broker 0



PARTITION 0

PARTITION 3

PARTITION 2

PARTITION 1

Broker 1



PARTITION 1

PARTITION 0

PARTITION 3

PARTITION 2

Broker 2



PARTITION 2

PARTITION 1

PARTITION 0

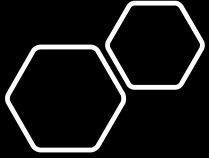
PARTITION 3





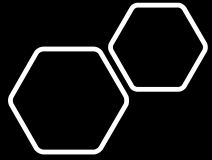
# Segments

- Dados ficam em arquivos físicos: diretórios nomeados “logs”
- Arquivos de partições são divididos em arquivos menores: chamados de segmentos
- Os dados vão para o primeiro segmento, até o limite, então ele inicia um novo arquivo
- Tamanho máximo é um 1 GB ou 7 dias
- Isso pode ser configurado



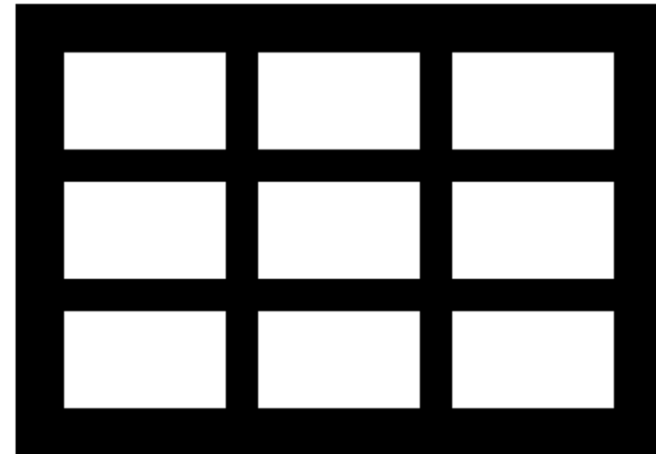
# offset

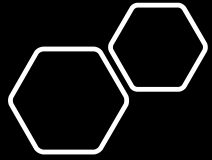
- O número de offset não reinicia em um novo segmento
- O numero do arquivo do segmento possui o numero do primeiro offset do segmento
- Off set não é único no tópico, e sim na partição!



# Partition offset

- Permite manter o estado
- Reiniciar
- Identificar mensagem de forma única





# Arquivo timeindex

- Para buscar mensagem baseadas em intervalos de tempo