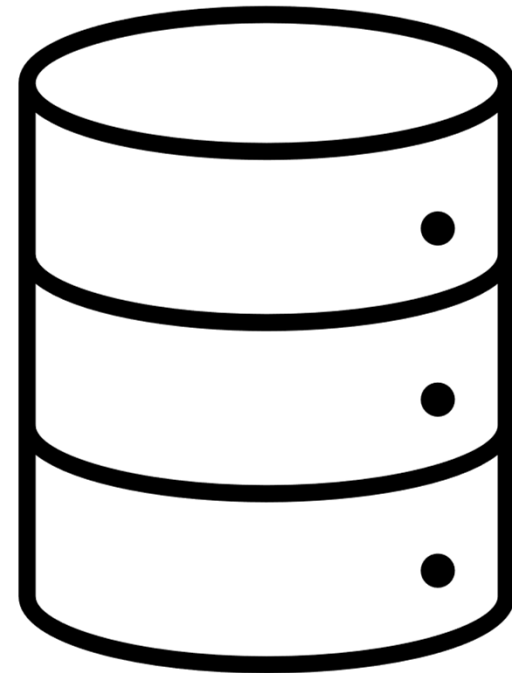


Apache Kafka

Prof. Fernando Amaral

Dados

- Processamos dados que são capazes de produzir informação e conhecimento
- Dado tem um valor
- O valor não está apenas no dado em si, mas em algumas dimensões



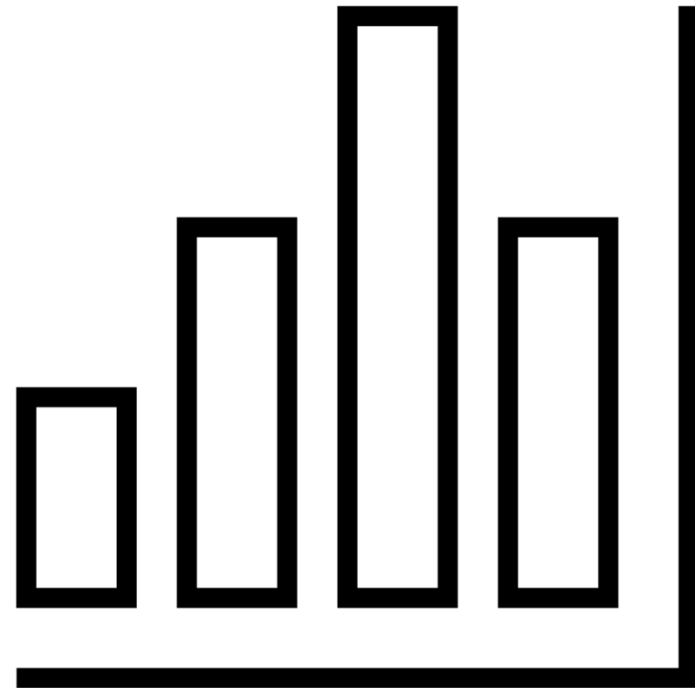
Dimensões do Valor do Dado

- Tempo
 - Identificar fraude
 - 1 minuto depois
 - 1 semana depois



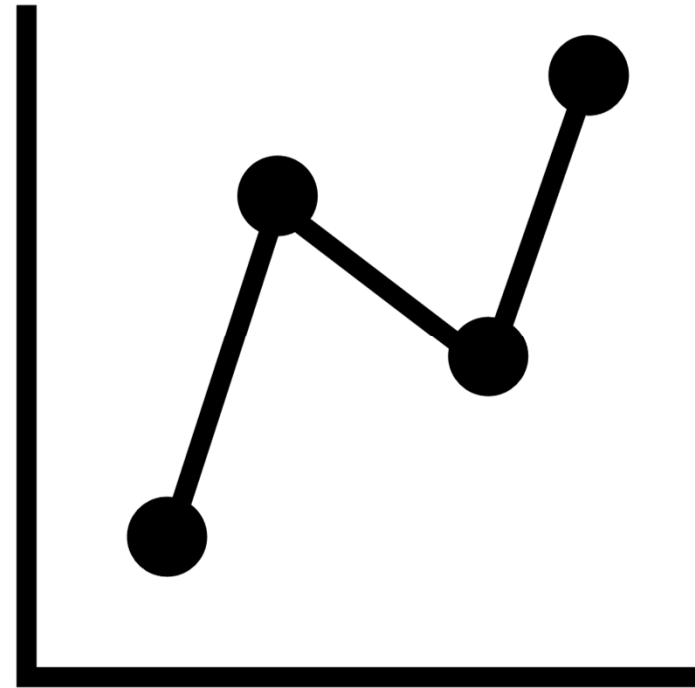
Dimensões do Valor

- Relação
 - Vendas do dia
 - Vendas da semana
 - Vendas do ano



Fato 1

- Dados possuem diferentes características quanto a:
 - Valor
 - Volume
 - Variedade
 - Tempo



Fato 2

- Não existe ferramenta única, universal, que seja capaz de extrair valor de qualquer dado, em qualquer contexto

Apache Kafka

“Apache Kafka é uma plataforma de streaming de eventos distribuída de código aberto usada por milhares de empresas para pipelines de dados de alto desempenho, análise de streaming, integração de dados e aplicações de missão crítica.”

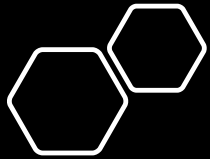
<https://kafka.apache.org/>



Consumidor de Música

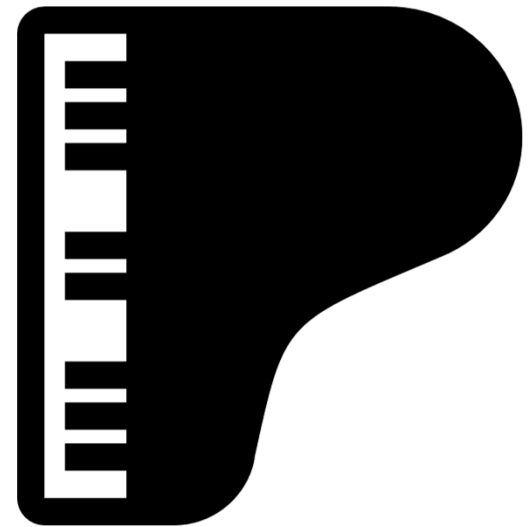
- Você gosta de ouvir música
- Existem vários estilos de música que você gosta de ouvir!





Produtor de Música

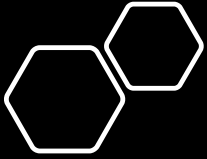
- Você é um músico!
- Você produz música em alguns estilos



Consumidor

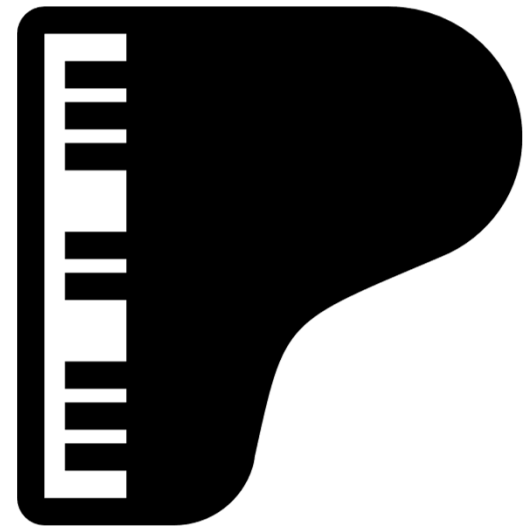
- Você resolve procurar na Internet produtores de musica dos estilos que você gosta
- Você encontra alguns e você consome (ouve) suas músicas



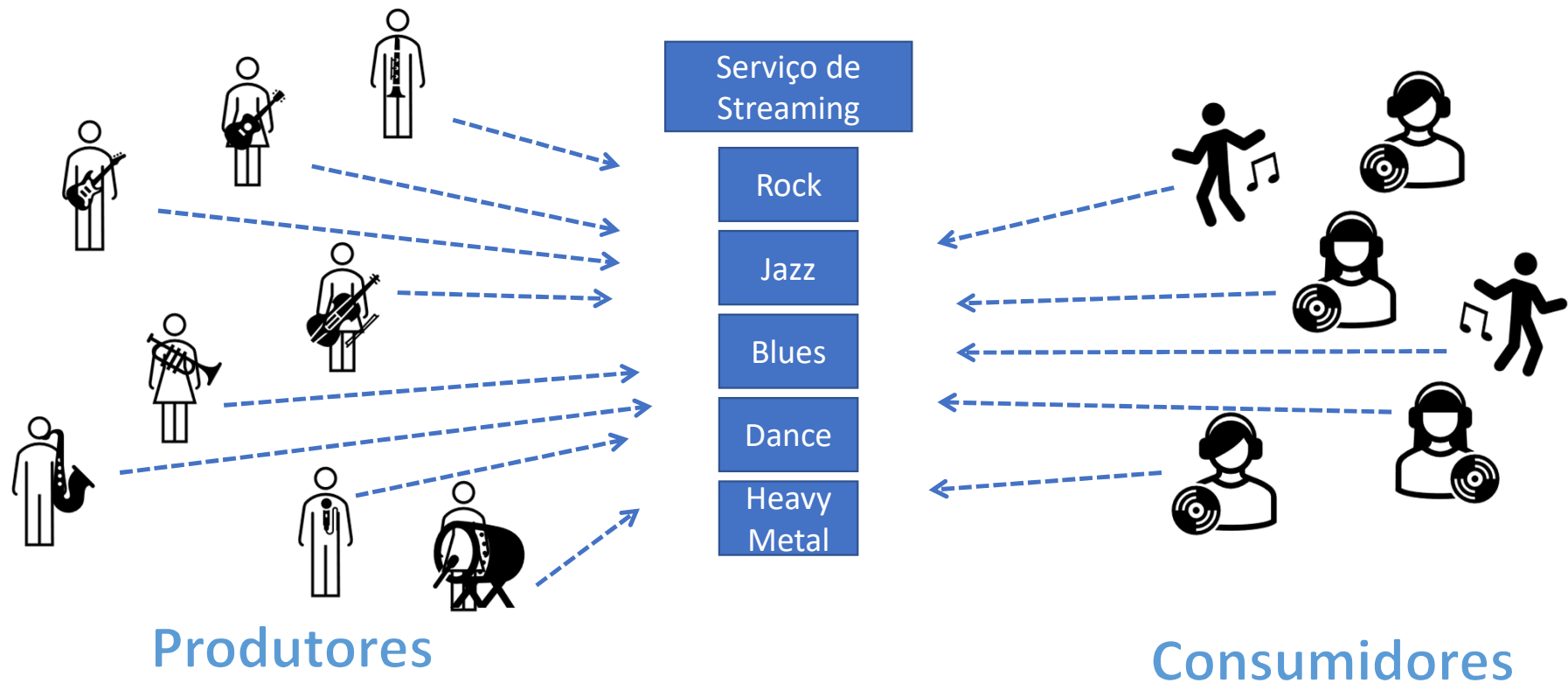


Produtor

- Produtores querem publicar suas músicas para que todos os consumidores que gostam do mesmo estilo, possam consumir suas músicas
- Alguns consumidores encontram suas músicas na Internet



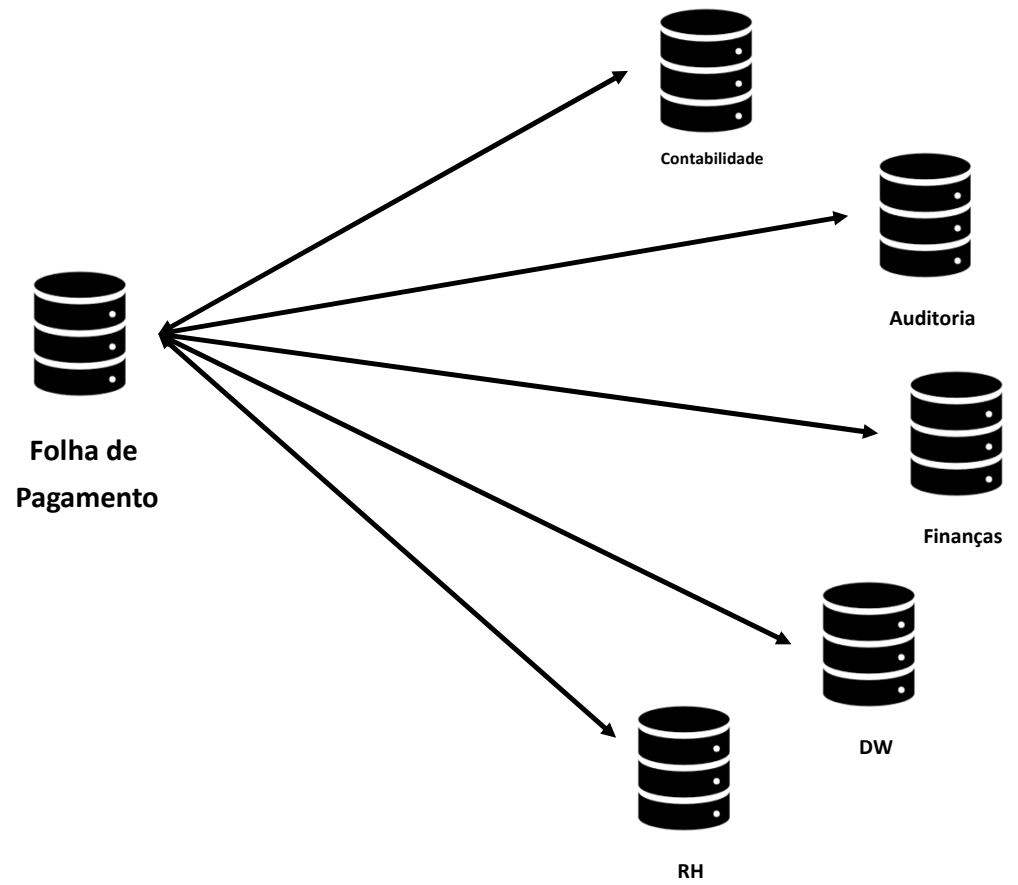
Serviço de Streaming de Música



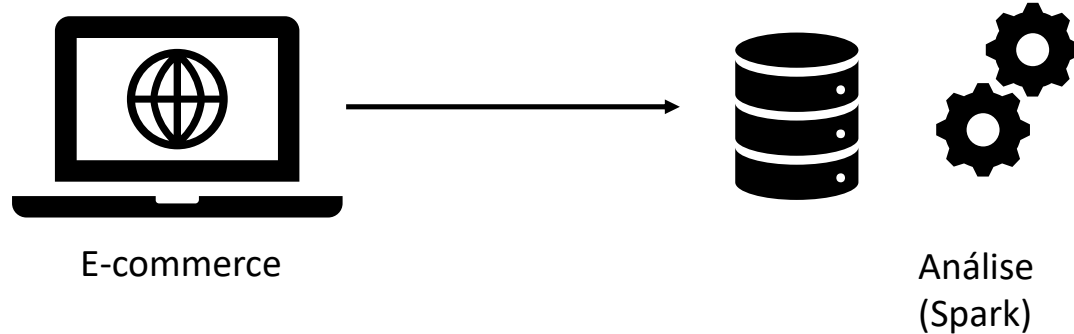
Vantagens

1. Produtor publica música em um único lugar
2. Consumidor encontra música em um único lugar
3. Uma vez publicada, a música está disponível quando o consumidor quiser ouvir
4. A mesma música é compartilhada para todos os interessados

Dados



Logs



- ✓ E se a demanda de logs aumentar além da capacidade do banco de dados (promoção)?
- ✓ Se o banco de dados ficar indisponível?
- ✓ Se por um erro parte dos logs se perder (falha em partição)?

Problemas Identificados

- Mesma informação compartilhada
- Sobrecarga
 - Diferentes “ritmos”
- Indisponibilidade
- Perda de Dados ou Reprocessamento

Apache Kafka



- ✓ Os dados são publicados um única vez
- ✓ Quem tem Interesse nos dados “assina” e consome só o que interessa
- ✓ Produtores e Consumidores desacoplados
- ✓ Podem trabalhar em ritmos diferentes
- ✓ Consumidores podem ler dados mais de uma vez
- ✓ Indisponibilidade do Produtor não afeta o processo
- ✓ Além disso, Kafka traz alto disponibilidade e capacidade com recursos de cluster e particionamento