

Trabalho avaliativo da disciplina DM-109

Alunos: Isabel Mendes

Renan Ribeiro

Questão 1

Caso British Airway:

Um dos maiores desafios do Big Data é o armazenamento. A quantidade de dados gerados por ele pode chegar a inúmeros gigabytes, necessitando de cada vez mais espaço.

Por isso, é extremamente importante contar com uma ferramenta de Big Data que consiga armazenar esses dados e o Apache Hadoop é uma das mais eficientes nessa tarefa.

Um caso de uso do Apache Hadoop é da empresa de aviação aérea British Airways, uma gigante no setor que coleta dados de seus clientes e analisa seu comportamento para estar à frente de seus concorrentes.

Em 2018, a gigante já possuía pentabytes de dados referente ao seu programa de fidelidade e ao seu website. Com esses dados a empresa reconhece seus clientes fiéis, oferece atendimento e ofertas personalizadas.

O que é Apache Hadoop?

É um dos principais framework utilizado no processamento e armazenamento de grandes conjuntos de dados. Inicialmente inspirado no MapReduce e no GoogleFS, o Apache Hadoop é um software Open-Source, construído em Java e mantido pela Apache Foundation.

Os principais benefícios são:

- Custo Baixo
- Escalável
- Flexível
- Open-Source
- Tolerante a falhas.

Questão 3

Numa rede capaz de receber e processar milhares de dados há a necessidade de coordenação, centralização e sincronização dos serviços que fluem por ela. Os dados devem ser escoados de forma segura, ordenada e sem sobrecarga para os servidores. Um dos softwares responsáveis por essa coordenação em data centers é o Apache Zookeeper.

Zookeeper é um serviço de código aberto criado pelo Yahoo, que se tornou padrão para organizar os serviços no Hadoop, Kafka e outros.

Suas principais funções são:

- Coordenação de rede centralizada, capaz de manter as informações de configuração atualizadas;
- Encontrar um nó da rede capaz de escoar os dados (chamado de nó mestre);
- Permitir que um nó encontre outro nó dentro de milhares de máquinas;
- Resolver problemas de sincronização utilizando blocos, barreiras e filas;
- Expõe uma API simples, permitindo aos desenvolvedores implementar tarefas comuns como gerenciamento de metadata.

Dois exemplos do uso do Zookeeper são:

- Apache Solr: Em sua forma distribuída (SolrCloud), usa o Zookeeper para armazenar e coordenar as atualizações de metadados sobre o cluster.
- Yahoo!: Usa o Zookeeper para tarefas como eleição do nó mestre, detecção e falhas e armazenamento de metadados.

Um dos concorrentes do Zookeeper é o software etcd, também de código aberto e disponível no GitHub. Ele é capaz de manipular as eleições mestras durante as partições de rede, descoberta de serviços e tolerância a falhas.

O etcd é utilizado pelo npm registry, Bandai Namco (jogos), Giant Swarm entre outros.

Referências

<http://igti.com.br/blog/apache-hadoop-casos-de-sucesso-no-uso-da-tecnologia/>. (Julho de 2019).

<https://medium.com/@joseantonio11/bigdata-40-zookeeper-6cedff6e4cc0>. (Julho de 2019).

<https://medium.com/develbyte/introduction-to-zookeeper-bcda7ef136cd>. (Julho de 2019).

<https://pt.slideshare.net/CloudxLab/apache-zookeeper-97027256>. (Julho de 2019).

https://pt.slideshare.net/Hadoop_Summit/making-the-case-for-hadoop-in-a-large-enterprisebritish-airways. (Julho de 2019).

<https://stackshare.io/ampstories/etcd>. (Julho de 2019).

<https://www.cetax.com.br/apache-hadoop-tudo-o-que-voce-precisa-saber/>. (Julho de 2019).

<https://zookeeper.apache.org/>. (julho de 2019).