

# Trabalho DM109 - Big Data e Análise de Dados

Trabalho entregue como forma de avaliação da Disciplina DM109 do curso de especialização em Desenvolvimento de Aplicações para Dispositivos Móveis e Cloud Computing no Inatel.

**Alunos:**

- Rafael William da Silva
- Thiago Scodeler

## Questão 1

- Faça uma aplicação para processamento de streams em tempo real utilizando o Apache Flink. Utilize o TORCS para gerar os dados de telemetria e o Apache Kafka como broker. Considere utilizar o conteúdo da aula de Big Data para te orientar. A aplicação deverá imprimir na tela, para cada carro, quantas vezes houve troca de marcha, dentro de um intervalo de 3 segundos. Utilize a aplicação "TV Dashboard", onde calculamos a velocidade média de cada carro, como ponto de partida. Se necessário, o código fonte desta aplicação está em <https://github.com/dmazzzer/kafka-flink-101>

Resposta: <https://github.com/ThiagoScodeler/DM109>

## Questão 2

- Explique a funcionalidade principal do software Zookeeper.

Resposta: ZooKeeper é um serviço de coordenação que oferece um conjunto de ferramentas para ajudar a gerenciar aplicações distribuídas. Construir aplicações distribuídas tem as dificuldades que são intrínsecas às aplicações distribuídas, o que inclui o mantimento de informações de configuração, grupos, nomeação e sincronização. ZooKeeper permite que desenvolvedores lidem com essas dificuldades para que criem aplicações distribuídas robustas. ZooKeeper vem com um conjunto de garantias: consistência sequencial, atomicidade, confiabilidade e pontualidade.

O Apache ZooKeeper permite que os processos distribuídos em sistemas de grande porte sincronizem informações um com o outro sem falha, de modo que todos os clientes que fazem solicitações recebam dados consistentes.

- Explique a importância do Zookeeper para aplicações no contexto de Big Data.

Resposta: O Hadoop, principal plataforma para computação distribuída voltada para clusters e processamento de grandes volumes de dados (Big Data), se baseia no ZooKeeper. O Hadoop utiliza o ZooKeeper para gerenciamento de configuração e coordenação. Zookeeper fornece uma infra-estrutura centralizada com serviços que permitem coordenação e sincronização dentro de um Hadoop cluster. As aplicações utilizam estes serviços para coordenar o processamento distribuído em grandes clusters.

Ao projetar um sistema distribuído, geralmente há necessidade de projetar e desenvolver alguns serviços de coordenação: sincronização de recursos, nomeação (name service), controle de acesso a recursos, gerenciamento de configuração, auto eleição para nós de cluster, entre outros. Embora seja possível projetar e implementar esses serviços a partir do zero, isso requer um esforço muito grande, além de termos que levar em conta aspectos como confiabilidade, escalabilidade e replicação. Para resolver todas essas questões, o Zookeeper surge como uma ótima opção.

**- Explique onde o Zookeeper foi utilizado na aplicação fictícia de telemetria para Formula1, utilizada em sala de aula.**

Resposta: O Zookeeper é inicializado junto com o Kafka e coordena o serviço que atua como "consumer". Esse serviço assina o tópico do Kafka (flink-demo) para receber os dados de telemetria provenientes do TORCS. O ZooKeeper atua como um componente mandatório para o Kafka, sendo assim os dois precisam estar rodando no momento dos testes.

### **Questão 3**

**- Crie um programa em Python para imprimir alguns gráficos. Utilize um notebook Jupyter e importe o arquivo disponível no endereço <https://goo.gl/RSTes9>. Crie os seguintes gráficos utilizando a biblioteca matplotlib:**

- Comparação da velocidade de dois carros qualquer;
- Um gráfico de linha mostrando a velocidade de um carro sobreposto ao RPM;
- Um gráfico a escolha do aluno (use a criatividade).

**Nota: Para obter a velocidade em Km/h multiplicar o atributo "Speed" por 35/9.**

**A entrega deste exercício deverá ser o link para a biblioteca do aluno, no serviço Microsoft Azure Notebooks.**

Resposta:

<https://notebooks.azure.com/rafaelws/libraries/inatel-dm109-big-data/html/TrabFinal.ipynb>