UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA DE COMPUTAÇÃO E AUTOMAÇÃO

**Roteiro de Aula Prática – Programação Spark Dataframes**

DISCIPLINA: DCA0132 – ENGENHARIA DE DADOS

PROFESSOR: Carlos Manuel Dias Viegas

Esta prática consiste em criar aplicações com Apache Spark programando com Dataframes na linguagem Python.

- Os arquivos com o código fonte (.py) ou (.ipynb) devem ser submetidos em uma tarefa específica no SIGAA.

- Esta prática pode ser realizada em duplas.

**Tarefas**

* Desenvolver uma aplicação utilizando os conceitos de Spark Dataframes que exiba as informações das linhas de ônibus de uma determinada cidade.

Pretende-se que a aplicação seja capaz de associar as informações de três arquivos que compõem o conjunto de dados.

Entrada do usuário:

- Número da linha de ônibus.

Saída:

- O número e o nome da linha de ônibus;

- Os horários e o nome de cada ponto em que essa linha passa.

Conjunto de dados disponível em: <https://goo.gl/A3MhFS>

Datasets > onibus.tar

Fonte dos dados:

Dados abertos da UFPR: <http://dadosabertos.c3sl.ufpr.br/curitibaurbs/>