



GOColetivo: um sistema de trânsito baseado em dados públicos de GPS

Tóp. Esp. em BD I (Spark)
Prof. Alexandre Assis
Aluno: Christian Cardozo



Motivação

- Google, Waze, etc. já possuem o serviço
- Utilizam dados dos usuários
- Os resultados são disponíveis apenas em seus mapas e funcionalidades
- Acesso aos dados computados não disponível



Proposta

- Contexto geográfico: Rio de Janeiro/RJ
- Sistema de trânsito e condições das vias
- Utilizar dados públicos de GPS dos ônibus da cidade
- Spark como ferramenta de processamento dos dados



Objetivos

- Sistema de código aberto no GitHub
- Usar Spark para processar o streaming de dados
- Disponibilização dos dados por uma API
- Sistema web com um mapa para visualização

Logradouros

- OpenStreetMap
- Filtro de dados: todos artefatos dentro da cidade do Rio de Janeiro
 - 1.433.415 artefatos
 - Formato OSM (~300MB)
- Extrair ruas do arquivo:
 - Manter Ways do tipo Motorway, trunk, primary, secondary, tertiary, etc. (77285 itens)
 - Manter Nodes (1347498 itens)
 - Excluir Relations
 - Parser feito em Python para converter XML para CSV
 - Importação para PostgreSQL





Logradouros

- Uma rua/avenida/etc. é composta por múltiplos logradouros (segmentos de reta)
- Total de 396.790 logradouros
 - Schema: lid, osm_id, nome, tipo, lat_src, lon_src, lat_dst, lon_dst, limite, ordem
- Mapeamento do limite de velocidade (Baseado no CTB):

OpenStreetMap	CTB	Limite de Velocidade	Total de Ocorrências
trunk, motorway	Vias de trânsito rápido	80	19.941
primary, secondary e unclassified	Vias arteriais	60	47.629
tertiary	Vias coletoras	40	45.860
residential, living_street	Vias locais	30	283.360

API GPS WebService

- Disponível através do portal data.rio
- Dados públicos atualizados a cada 1~2 minutos (em média)
- Exemplo do formato dos dados:

```
{  
  "COLUMNS": [ "DATAHORA", "ORDEM", "LINHA", "LATITUDE", "LONGITUDE", "VELOCIDADE" ],  
  "DATA": [ [ "08-17-2017 00:00:18",  
    "A27574",  
    517.0,  
    -22.916121,  
    -43.247688,  
    18.0  
  ], ...  
]
```





Crawler

- Programa feito em bash
- Execução periódica (a cada 1 minuto, por exemplo)
- Baixa o JSON do data.rio e salva localmente na pasta crawler_data
- Sempre cria um novo arquivo



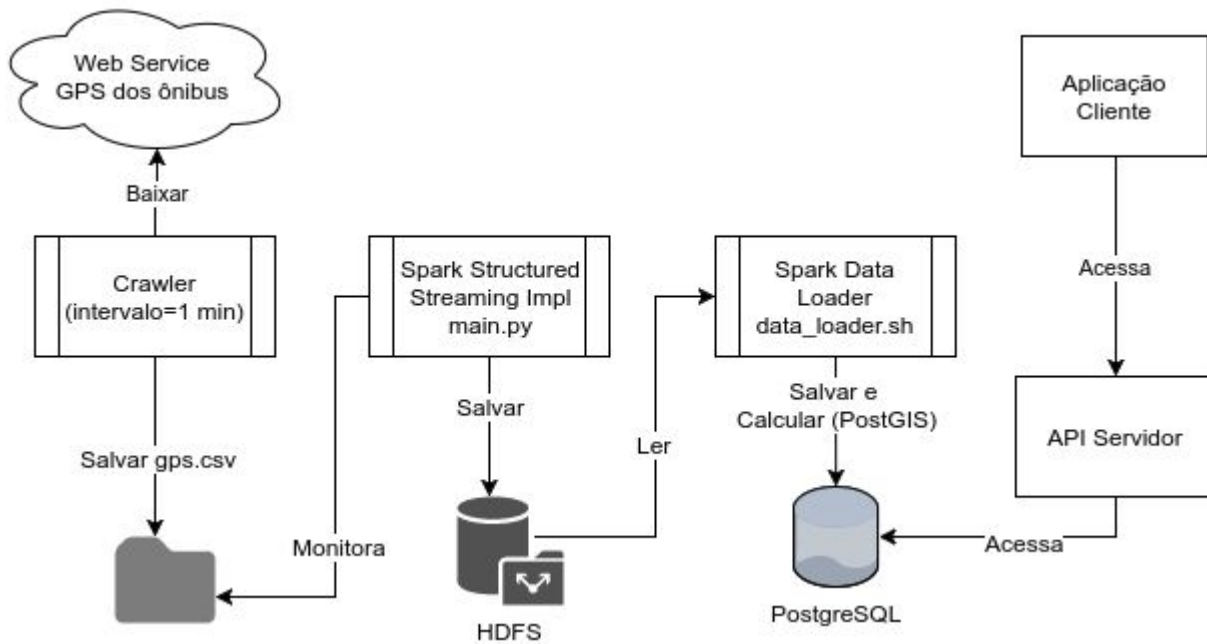
Módulo Structured Streaming - Spark

- Desenvolvido usando a API do Spark para Python
- Monitora por novos arquivos em crawler_data
- Processa um novo arquivo assim que chega
- Carrega o arquivo como um Dataframe com Schema definido
- Usando Spark SQL, filtra por posições válidas:
 - Velocidade maior que zero e menor que 100
 - Data/hora maior que 20 minutos atrás
- Salva em formato CSV no HDFS
- Faz append do dados novos



Módulo Data Loader

- Usando PySpark
 - Carrega os dados do HDFS
 - Filtra novamente por data/hora até 20 minutos atrás
 - Seleciona datahora, latitude, longitude e velocidade
 - Exporta para um arquivo CSV no armazenamento local
 - Assim temos o conjunto das posições válidas nos últimos 20 minutos registrados
- Usando bash
 - Importa CSV exportado para base do PostgreSQL (tabela data_load)
- Usando PostGIS
 - Interpola os dados geoespaciais entre GPS e logradouros fazendo a média de velocidade





API GOColetivo WebService

- Script simples em PHP
- Colocar dentro da pasta do Apache
- Retorna as informações calculadas de todas as vias com velocidade média diferente de nulo (zero)
- Retorno em formato JSON



```
[{"f1":{"lid":163544,"velocidade_media":24,"nome":"Rua Jandyrá Medina","limite":30,"pontos":[{"f1":43.2116476,"f2":22.8054969},{"f1":43.2115843,"f2":22.8052799}]},"f1":{"lid":163545,"velocidade_media":24,"nome":"Rua Jandyrá Medina","limite":30,"pontos":[{"f1":43.2115843,"f2":22.8052799},{"f1":43.2115523,"f2":22.8049246}]},"f1":{"lid":163685,"velocidade_media":42.04,"nome":"Rua Professor Motta Maia","limite":30,"pontos":[{"f1":43.4587479,"f2":23.0207697},{"f1":43.4584386,"f2":23.0206885}]},"f1":{"lid":163741,"velocidade_media":16,"nome":"Rua Barão de Melgaço","limite":40,"pontos":[{"f1":43.3108147,"f2":22.8274725},{"f1":43.3107581,"f2":22.8274485}]},"f1":{"lid":163742,"velocidade_media":16,"nome":"Rua Barão de Melgaço","limite":40,"pontos":[{"f1":43.3107581,"f2":22.8274485},{"f1":43.3088443,"f2":22.8265482}]},"f1":{"lid":163777,"velocidade_media":12.16,"nome":"Auto Estrada Lagoa-Barra","limite":80,"pontos":[{"f1":43.2521975,"f2":22.9926874},{"f1":43.2515859,"f2":22.992561}]},"f1":{"lid":163778,"velocidade_media":19.12,"nome":"Auto Estrada Lagoa-Barra","limite":80,"pontos":[{"f1":43.2515859,"f2":22.992561},{"f1":43.2512555,"f2":22.9925294}]},"f1":{"lid":164195,"velocidade_media":22,"nome":null,"limite":30,"pontos":[{"f1":43.6326059,"f2":22.8743102},{"f1":43.6320431,"f2":22.8753036}]},"f1":{"lid":164322,"velocidade_media":38,"nome":"Rua Doutor Aníbal Moreira","limite":30,"pontos":[{"f1":43.2402545,"f2":22.9257198},{"f1":43.2416345,"f2":22.9285874}]},"f1":{"lid":164390,"velocidade_media":2,"nome":"Avenida Dom Hélder Câmara","limite":60,"pontos":[{"f1":43.2790825,"f2":22.8859084},{"f1":43.2784902,"f2":22.8854104}]},"f1":{"lid":164392,"velocidade_media":20,"nome":"Avenida Dom Hélder Câmara","limite":60,"pontos":[{"f1":43.2784047,"f2":22.8853379},{"f1":43.2779945,"f2":22.884987}]},"f1":{"lid":164393,"velocidade_media":20,"nome":"Avenida Dom Hélder Câmara","limite":60,"pontos":[{"f1":43.2779945,"f2":22.884987},{"f1":43.2778393,"f2":22.8846884}]},"f1":{"lid":164394,"velocidade_media":20,"nome":"Avenida Dom Hélder Câmara","limite":60,"pontos":[{"f1":43.2778393,"f2":22.8846884},{"f1":43.2776872,"f2":22.8847479}]},"f1":
```



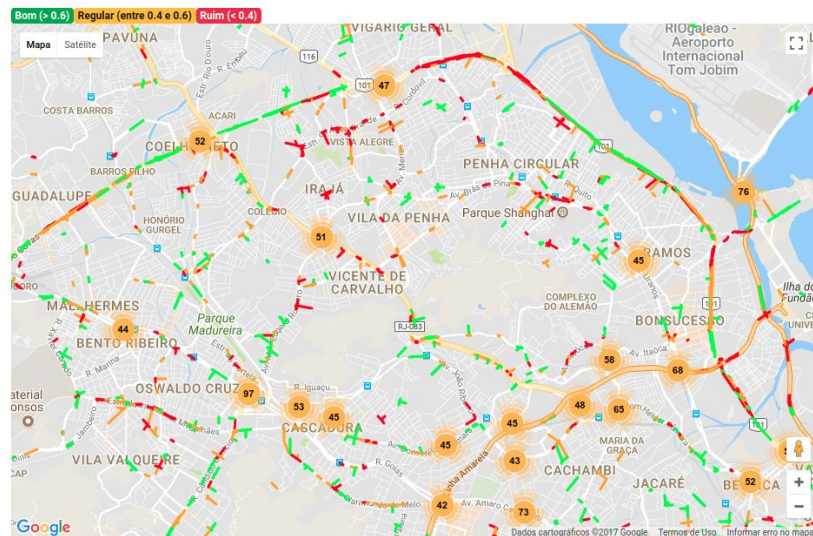
GOColetivo cliente

- Usa os dados retornado da API feita para este trabalho (API GOColetivo Webservice)
- Usa a API do Google Maps
- Exibe um mapa com as condições das vias
- Cor do logradouro
 - Vermelho: trânsito ruim (se velocidade/limite ≤ 0.4)
 - Laranja: trânsito regular (se velocidade/limite entre 0.4 e 0.6)
 - Verde: trânsito bom (se velocidade/limite ≥ 0.6)

GOColetivo - Condições das Vias no Rio de Janeiro

O GOColetivo permite a visualização das condições das vias na cidade do Rio de Janeiro. O cálculo é feito baseado nos dados públicos de GPS da cidade.

Disciplina Tóp. Esp. em BD I (Spark) / Professor: Alexandre Assis
Autor: Christian Cardozo

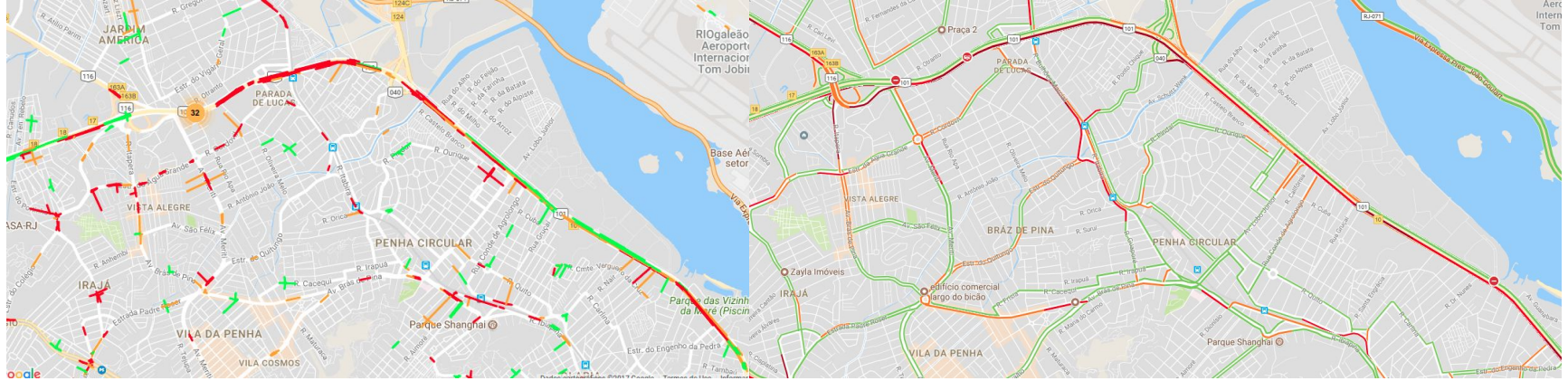


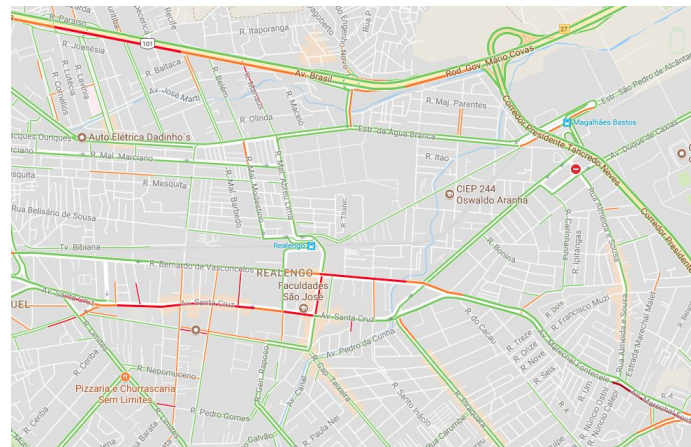
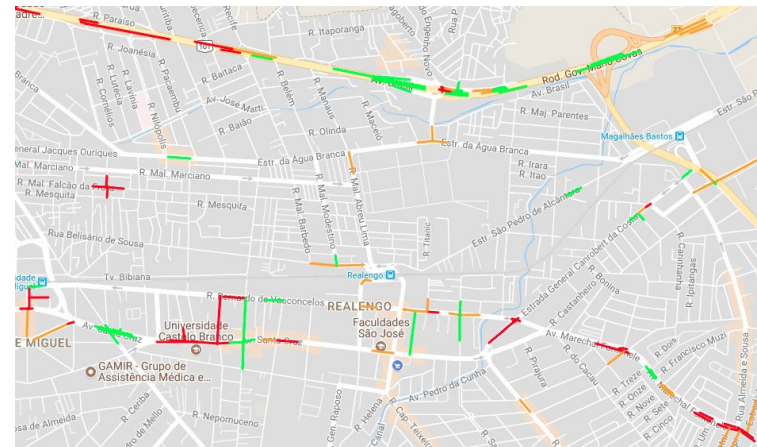
GOColetivo - Condições das Vias no Rio de Janeiro

O GOColetivo permite a visualização das condições das vias na cidade do Rio de Janeiro. O cálculo é feito baseado nos dados públicos de GPS da cidade.

Disciplina Tóp. Esp. em BD I (Spark) / Professor: Alexandre Assis
Autor: Christian Cardozo









Considerações Finais

- Objetivos conquistados:
 - Spark processando o streaming de dados
 - Dados disponibilizados através de uma API retornando um JSON
 - Aplicação de visualização intuitiva com cores dependendo da condição da via
- Structured Streaming bem consolidado: tolerante à falhas e escalável
- Desenvolvimento pronto para rodar em Cluster
- Ah, GitHub com tudo: https://github.com/chriiscardozo/GOC_TransitWithSpark



Referências

- CTB Digital: <http://www.ctbdigital.com.br>
- Google Maps: <https://www.google.com.br/maps>
- Waze: <https://www.waze.com/pt-BR/>
- Apache Spark: <http://spark.apache.org/>
- data.rio: <http://data.rio/>
- OpenStreetMap: <https://www.openstreetmap.org/>
- PostGIS: <http://postgis.net/>
- Google Maps APIs: <https://developers.google.com/maps/>