Podcast

Disciplina: **Linguagens de programação para ciência de dados**
Título do tema: Real Time Analytics com Python e Spark
Autoria: Yuri Vasconcelos de Almeida Sá
Leitura crítica: Henrique Salustiano Silva

Abertura:

Olá! No podcast de hoje vamos falar sobre o grande campo da aquisição e transformação de dados em tempo real.

O *Spark* tem uma grande qualidade que é processar os dados em tempo real. Isto é uma característica que realmente destaca o framework entre os outros que processam BigData.

Porém isso gera uma necessidade que é aquisição de dados em tempo real. Tudo isto é fantástico, porém se não tivermos um fluxo intenso de dados o framework não se justifica, a infraestrutura de processamento em memória que o spark oferece só é válido se tivermos dados para o processo ao vivo. Na hora, para já.

Uma grande área de trabalho com BigData é justamente isso, o que achar e onde achar. Estes conectores fazem sua aplicação funcionar. O exemplo mais clássico hoje são as redes sociais. Elas geralmente fornecem acesso à APIs em formato uniforme para acesso ao vasto acervo de publicações por segundo.

Em tempos de internet a oportunidade de aquisição de dados em larga escala é enorme. No princípio destas análises, a filtragem de spam em servidores de e-mail era um clássico. Uma vez que era muito fácil enviar mensagens de propaganda não solicitadas e o volume chegava a ser 90% das mensagens enviadas.

Existe ainda uma outra frente de dados que são os dados de acesso de qualquer site, estas análises podem gerar relatórios em tempo real. E qualquer site é capaz de gerar um volume de dados interessante para a produção da análise.

No entanto, um exemplo que somente os grandes estão utilizando no momento é o processamento de mídias vivas, em áudio e vídeo.

Esta análise amplia nosso campo de estudo e podemos utilizar técnicas mais avançadas de análise, como *Text-To-Speech*, localização de objetos em imagens dinâmicas e criar metadados e até mesmo transcrições em tempo real.

Em um contexto de empresas pequenas e médias, podemos criar transcrições de todas as ligações da empresa, por exemplo, podendo criar um modelo em tempo real e identificar possíveis leads para o time de vendas, ou então alertar a supervisão de alguma postura para intervir.

Sempre lembrando que todos os tipos de formatos digitais podem ser convertidos em números e dimensões, qualquer imagem é uma tabela de dados e vídeos são uma série de imagens. Qualquer arquivo de áudio pode ser resumido em cadeias de números, a onda de áudio é sempre uma tabela, indexada pelo tempo.

Essa habilidade pode até gerar dados para e se comunicar com sistemas de CRM e gerar eventos e oportunidades de relacionamento.

Este tipo de conector se torna muito importante neste contexto, e você pode incluir ele já no código para ser processado e distribuído junto, distribuindo a carga até mesmo da aquisição.

O mundo é nossa ostra, e o mundo digital é nossa Matrix. Quando a gente começa a pensar desse jeito, a gente pensa em extrair dados em tempo real de todos os lugares. Aplicações industriais, sensores de IoT, qualquer lugar oferece uma oportunidade de dados em tempo real.

Hoje existem relógios esportivos que monitoram nosso batimento cardíaco, pressão sanguínea, oxigenação e seus passos. Tudo em tempo real, e conectado na internet. Hoje em dia você mesmo pode ser uma fonte de dados em tempo real.

Dados! Dados em todos os lugares!

Fechamento:

Este foi nosso podcast de hoje! Até a próxima!