

1 Introdução

1.1 Motivação

Com o desenvolvimento tecnológico das redes de sensores, muitas indústrias elevaram bastante o nível de controle de sua produção, equipamentos e segurança, dentre outras aplicações. De fato, o aumento no número de processos ou objetos que fazem uso de sensores elevou significativamente a quantidade de dados a serem armazenados e analisados. No estágio atual, a maioria das indústrias ainda não sabe ao certo como usar esta grande massa de dados, o conhecimento está disponível, mas a forma de extraí-lo não é trivial.

Os problemas relacionados a estes dados podem ser divididos em duas categorias básicas: uma trata especificamente da qualidade enquanto a outra está relacionada ao armazenamento eficiente em memória secundária.

Como os sensores podem estar posicionados em ambientes muito hostis e dependem de um meio físico para a transmissão dos dados adquiridos, tais dados estão sujeitos a interferências externas, as quais são conhecidas como “ruídos”. Este fenômeno é tão comum nas redes de sensores, que praticamente inexistem a possibilidade de ignorá-los. Assim sendo, métodos para remoção de ruído em séries históricas de dados adquiridos por sensores são fundamentais para garantir a qualidade dos mesmos e, conseqüentemente, aumentar sua confiabilidade, assim como das decisões tomadas a partir destes dados.

Apesar do custo por megabyte em soluções modernas de armazenamento ter caído abruptamente nos últimos anos, algumas aplicações de redes de sensores podem gerar tantos dados que seria extremamente complicada a manutenção de uma estrutura capaz de armazená-los de forma eficiente. Por esse motivo, a compressão de dados não pode ser desconsiderada quando pensamos em soluções para armazenar dados de sensores.

O problema trazido pelos dados Adquiridos por sensores é ainda mais evidente quando se necessita monitorar e armazenar estes dados por longos períodos de tempo. Podemos ilustrar este problema através de um exemplo real, comum na indústria de Petróleo: o acompanhamento e controle de poços equipados com sensores de fundo.

Um poço de petróleo pode ser entendido como uma interface entre a superfície e a subsuperfície. Através desta interface são feitas as operações de produção que podemos resumir em: trazer o óleo e gás para a superfície e injetar fluidos no reservatório a fim de manter o mesmo pressurizado e os níveis de produção. Para uma análise de qualidade, pode ser necessária a utilização de décadas de dados de centenas de poços.

Seja monitorando a umidade e pressão atmosférica em diversos pontos de uma floresta tropical ou adquirindo dados sobre a temperatura e a vazão da injeção de fluido em um poço de petróleo, o fato é que, isolados, estes dados não geram muito conhecimento sobre o negócio a ser estudado, são “sujos”, cheios de ruídos e numa quantidade tamanha que pode tornar caótica sua atividade de gerenciamento. Apesar de toda esta problemática, existem diversas soluções de armazenamento chamadas de “historiadores”, que funcionam como um grande repositório de dados adquiridos por sensores. Estas soluções permitem que, ao se instalar um sensor em uma válvula ou determinada região, o mesmo seja identificado através de uma “tag” e passe a ter seus valores armazenados neste repositório a uma determinada taxa de frequência. Devido à quantidade de dados, normalmente estas soluções não são implementadas na forma de bancos de dados relacionais, o que naturalmente causa transtornos na hora de relacionar os dados aos objetos de negócio da empresa ou do tema em estudo, normalmente armazenados em bancos de dados relacionais. Além disso, é bastante comum que estas soluções implementem seus bancos de dados de forma proprietária, tornando mais difícil ainda as atividades de análise.

Como os historiadores são otimizados para sua atividade transacional, eles têm um desempenho excelente nas operações de inserção de dados. Em compensação, são bastante ineficientes quando o assunto é a recuperação dos dados. Desempenham muito bem o papel de armazenar séries históricas de dados operacionais brutos. Desta forma, não devem ser levados em consideração como solução

única para o problema enfrentado pelas indústrias em relação aos dados gerados por sensores.

Podemos então afirmar que o problema em questão envolve o tratamento e a remoção de ruídos, a compressão de séries históricas, que torna possível o uso de soluções convencionais de armazenamento e o relacionamento de tais dados com os dados dos objetos de negócio existentes a serem analisados. A existência destes problemas não condena o uso dos historiadores, muito pelo contrário, nos força a pensar em soluções conjuntas melhores, que permitam a análise dos dados contidos neles.

1.2 Abordagem proposta

O objetivo principal deste trabalho é propor uma arquitetura de solução para o problema de tratamento básico e compressão de dados adquiridos por sensores. A arquitetura proposta permite que os dados adquiridos possam ser analisados em conjunto com outros dados do negócio e que os mesmos possuam qualidade suficiente para que as decisões possam ser tomadas. Além disso, este trabalho propõe um mecanismo de compressão de séries históricas baseado em transformadas Wavelet unidimensionais, que além da compressão, oferecem recursos interessantes quanto à remoção de ruídos e a suavização do sinal.

Visto que as análises de dados adquiridos por sensores muitas vezes envolvem a elaboração de gráficos com dados ao longo do tempo, um sistema com a abordagem proposta funcionaria da seguinte maneira. Os sensores instalados estariam associados a um sistema historiador ou a uma outra solução de armazenamento qualquer, de tecnologia proprietária ou não. Estes dados possuem natureza puramente transacional e a solução de armazenamento onde estão disponíveis, na maioria das vezes, oferece desempenho pobre no processamento de consultas. Logo, torna-se necessário que os dados sejam disponibilizados em um banco de dados relacional que possa ter acesso aos demais dados referentes ao negócio.

Uma vez que os dados estejam disponibilizados em um banco de dados relacional, passamos a nos preocupar com a qualidade e o armazenamento de grandes volumes de dados. Um viés da qualidade dos dados está relacionado à integração com os objetos de negócio da empresa. Para isso os dados devem ser confiá-

veis. O outro viés está relacionado à capacidade do SGBD em prover serviços satisfatórios com uma infra-estrutura economicamente viável, o que pode ser extremamente complicado devido à enorme quantidade de dados.

Para testar os conceitos propostos, este trabalho inclui a descrição de um protótipo de toolkit para tratamento básico e compressão de sinais gerados por sensores. O objetivo deste toolkit é dar apoio à construção de aplicações que necessitem destes tipos de serviços. Alguns exemplos de aplicações que poderiam ser construídos com a ajuda deste software, dentre outros, são:

- Ferramentas para análise de sinais em várias resoluções
- Ferramentas para compressão automática de sinais em banco de dados
- Ferramentas para limpeza de sinais contaminados por ruídos
- Aplicações que dão suporte a detecção de padrões de comportamento em séries históricas

Ao longo desta dissertação iremos apresentar as técnicas utilizadas neste toolkit e ao final apresentaremos os resultados obtidos pela metodologia proposta.