

INTRODUÇÃO

Inúmeros mecanismos de busca, armazenamento e distribuição de informações tem se apoiado fortemente na infraestrutura que a internet e o armazenamento em nuvem dispõem. Com o aumento da capilaridade das formas de captura de informação virtualmente, todos os seres humanos e muitos de seus equipamentos podem incluir dados que são tratados, padronizados, consolidados e formam bases de dados de alto valor.

Muitos desses dados têm valor dentro de um contexto específico, como dados de consumo de energia doméstica ou industrial, o tempo dispendido dentro de um meio de transporte, as condições climáticas em uma estação de medida, as compras realizadas por um usuário de cartão de crédito e assim por diante. Porém, conforme os dados se tornam mais confiáveis – em sua origem e na forma de armazenagem –, surgem novas maneiras de utilizá-los, cruzando-os e olhando-os por diferentes dimensões. Dados que antes pareciam ser irrelevantes ou muito localizados passam a ter uma função dentro de um sistema possivelmente globalizado. Só como exemplo, se todos os veículos – carros, ônibus, bicicletas, caminhões, motos etc. – tiverem sensores de posicionamento (GPS), é possível auferir o trânsito e traçar rotas através do uso de programas de computador, e, instantaneamente, devolver esta informação a cada um dos veículos. Com isto, o condutor poderia definir melhor seus caminhos, ou isto poderia auxiliar na condução do veículo sem um motorista.

Uma das vertentes de utilização destes repositórios é a prospecção de novas oportunidades de negócio, seja detectando carências, similaridades, infraestrutura favorável etc. Por exemplo, um empreendedor que queira investir em um novo produto, precisa saber quem compra esse produto (se compra), onde estão os usuários, onde estão os fornecedores de matéria prima, quais as formas de escoar a produção, quais os impactos ambientais e daí por diante.

Ou seja, para um investimento consciente, é preciso fazer uma pesquisa aprofundada e cujo custo, até algum tempo atrás, era significativo ou até proibitivo, e, no final das contas, o empreendedor assumia um risco maior em função de seu desconhecimento ou em função das chamadas “intuições”.

A proposta dos chamados “big data” obedece alguns padrões tais como: o modo como é feita a coleta dos dados, como são analisados e como são processados. E podem ser classificados de acordo com alguns critérios: se tratam de dados analisados em tempo real ou agrupados (temporais), em função da frequência dos dados, do tipo de dados (transacionais, históricos, principais, etc.), da fonte (onde são gerados os dados), do usuário dos dados, dentre outros (IBM, 2014).