Privacidade e segurança em big data

Big data tem um enorme potencial e uma infinidade de possibilidades de uso. Hoje em dia podemos ver esse potencial aplicado em várias coisas, como predição do tempo e análise de mercado. Mas todo esse poder pode ser usado tanto de forma benigna como de forma maliciosa, impedir o uso de forma errada sem barrar o avanço que o big data pode trazer à sociedade é um problema atual.

Um desafio para impedir o uso malicioso de big data é a segurança. Um exemplo de problema de segurança foi o incidente ocorrido na universidade de Arkansas em 2014, afetou 50.000 pessoas que tiveram alguns dos seus dados vazados. É um número considerável, mas que se torna irrelevante se comparado a outros incidentes, como o os de 145 milhões de usuários que tiverem seus dados roubados do ebay no mesmo ano.

Uma das melhores estratégias para garantir a segurança dos dados é ter um único ponto de acesso, mas isso vai contra a estruturação do big data que por causa do quantidade e fluxo de dados obriga o uso de uma estrutura distribuída com vários pontos de acesso. E em complemento a isso, vários softwares para estruturação de big data não levam a segurança dos seus sistemas a sério o suficiente. Como é caso do o Apache hadoop, uma coleção de softwares que permite a desenvolvedores criarem soluções para processar grande quantidade de dados em uma pequena estrutura de sistemas distribuída e que, em suas primeiras versões, trabalhava apenas com funcionalidades básicas de segurança e mesmo assim foi usado por grandes empresas em larga escala.

Outro aspecto de segurança a ser levado em conta é a privacidade dos usuários. Isso pode ser garantido por termos e condições e de uso e também por ações tomadas pela empresa, como encriptação dos dados, controle de acesso ou detecção de fraudes. O problema é que algumas medidas de segurança para proteger a privacidade do usuários também podem invadi-la para coletar dados como localização ou histórico de navegação.

Além de coletar dados que não seriam de domínio público dos para proteger a privacidade empresas também podem usar isso para entregar publicidade direcionada, já que com análise de big data esse processo de direcionamento se tornou mais simples e barata. Um exemplo de empresa que usa big data para esse fim é a Target, que faz uso disso para entregar publicidade direcionada para o clientes que a empresa acredita serem futuros pais ou mães.

A target consegue isso guardando um número de registro atrelado a cada cliente e a esse registro um nome ou email, um histórico detalhado de cada compra na loja e toda outra informação

que ela conseguir, sejam essas informações coletadas pela própria empresa ou compradas de outras fontes.

Analisando os dados coletados de compras de alguns produtos como algodão, sabonetes e produtos para higienização das mãos, a empresa consegue prever, até com certa precisão, em que período da gravidez as clientes se encontram. Assim, é possível mandar propagandas de produtos que são mais relevantes baseados no período da gravidez.

Um caso que interessante que aconteceu por causa dessa estratégia da Target, na cidade de Minneapolis nos EUA, um homem entra em uma loja para falar com o gerente, ele reclama que a Target estava estimulando sua filha adolescente a engravidar por enviar à ela emails com propagandas sobre roupas de bebês. O gerente se desculpa com o homem e liga para ele alguns dias depois para se desculpar novamente, mas dessa vez é o homem que se desculpa com o gerente da loja pois sua filha já estava grávida.