Renato Malafaia: O monstro do data lake

Reza a lenda que próximo à cidade de Inverness, na Escócia, vive uma criatura que remonta à época dos dinossauros. Naquela imensidão aquática de 37 km de extensão e 230 metros de profundidade, viveria uma espécie de animal que teria focinho de jacaré, corpo de dinossauro e cauda de serpente e assombraria os humanos locais.

Algumas expedições enviadas ao local já no novo milênio por gigantes como Google e BBC confirmaram não passar de fruto da imaginação humana, uma vez que não haveria nas profundezas daquele lago nenhuma espécie com essas características. Mas o que não aparece nos noticiários é que os profissionais da privacidade e da proteção de dados podem ter encontrado a tal criatura escondida. Onde?

No *data lake*.

Bastante popular e ganhando cada vez mais mercado nos últimos anos (ou devemos dizer nas últimas décadas?), as ERPs (*Enterprise Resource Planning*) propõem a solução para o tratamento de dados — pessoais e comuns — de forma completamente disruptiva, por centralizar e organizar o local onde os dados são armazenados. Isso garante uma facilidade operacional gigantesca, já que todos os demais projetos e aplicações da companhia podem ser desenhados para se alimentar de dados de uma única fonte, cujos códigos e APIs já são mais familiarizados.

Parece maravilhoso, certo? Permitir que todo o setor de *Analytics* e de negócios possam consumir dados numa base única, segura, eficaz e que ainda economize tempo de trabalho parece um sonho. Porém pode existir um monstro escondido nessa imensidão de dados, pronto para abocanhar a primeira vítima.

Escondido bem debaixo dos nossos narizes e ausente em parte dos debates e preocupações na adequação à LGPD, o artigo 6º, II, desmascara-o e nos faz propor um debate imprescindível: estão os *data lakes* preocupados com o *princípio da adequação*?

A lógica por trás dessa pergunta olha bem para a estrutura de negócios mais comum: empresas estão antecipando a vigência da LGPD e tomando a cautela de adaptar todos os seus processos de tratamento de dados às novas diretrizes, para controlar ao máximo os riscos durante todo o ciclo de vida do dado pessoal.

Portanto, a preocupação do momento é garantir que a coleta não é intrusiva, a finalidade é totalmente legítima, foi dada transparência ao usuário, o tratamento está minimizado, os acessos estão restringidos e tudo isso se encaixa em uma base legal. Ufa!

Ufa?

A base de dados a qual o dado é incorporado (após a coleta) está interligada a outras aplicações ou a outros procedimentos? Quais? O que acontece com o dado após ele virar mais uma gota do lago?

Segundo o princípio da adequação, o dado pessoal pode ser tratado de forma compatível com as finalidades informadas ao titular, segundo o contexto do tratamento. Por exemplo, se o dado foi coletado por um portal de *e-commerce*, pode fazer sentido, se dentro do legítimo interesse, tratá-lo para identificar produtos similares nos quais o consumidor possa estar interessado.

Você já entendeu aonde quero chegar. Após a internalização do dado ao *data lake*, temos a certeza de que, pela integração a outras aplicações, esse dado não esteja sendo usado para outras finalidades (e expondo a empresa a um risco que talvez não tenha sido mapeado)?

Se não tiver, podemos ter um grande ponto de atenção aqui, pelo seguinte:

1) Será que as outras finalidades empregadas no tratamento de dados foram informadas ao titular?

2) Há compatibilidade com o contexto original que justificou a coleta do dado?

Sem dúvidas já há lanternas ligadas e olhares curiosos em direção a esse lago, porém a visibilidade dessa luz, em muitos casos, ainda parece superficial e não permite enxergar os riscos que estão no fundo desse lago. Uma boa recomendação de antemão é a adoção de medidas de segregação de acesso de acordo com cada finalidade de armazenamento dos dados. Com isso, evita-se que o dado, uma vez mergulhado no *data lake*, seja atacado pelo monstro, também conhecido como princípio da adequação.

Date Created
08/06/2020