Aspetos profissionais, legais e éticos na Gestão de Dados em SGBD

João Brito, M9984
Department of Computer Engineering
University of Beira Interior
Covilhã, Portugal
joao.pedro.brito@ubi.pt

Luís Pereira, M10156
Department of Computer Engineering
University of Beira Interior
Covilhã, Portugal
luis.cavaca.pereira@ubi.pt

Carlos Esteves, E10304
Department of Computer Engineering
University of Beira Interior
Covilhã, Portugal
carlos.esteves@ubi.pt

***Abstract*—O presente documento procurará expor a temática da gestão de dados em várias vertentes. Serão referidos aspetos relevantes do ponto de vista profissional, legal e ético, bem como, alguns exemplos reais.**

I. Introdução

A temática da gestão de dados é indubitavelmente abrangente. Do ponto de vista profissional, podem ser considerados aspetos puramente técnicos (i.e. como são, efetivamente, guardados os dados). A título legal, a legislação existente desempenha um papel taxativo na temática aqui abordada. Por fim, existe a ética na gestão de informação, sendo muitas vezes o elemento que norteia os profissionais da área.

II. Gestão de dados: onde começam e acabam os limites?

As aplicações computacionais atuais são cada vez mais capazes e poderosas, no que toca a funcionalidades apelativas ao utilizador. Como tal, pode surgir a necessidade de extrair o máximo de informação dos dados, de modo a oferecer uma experiência mais personalizada. Esta estratégia pode-se revelar adequada (os dados dos utilizadores podem revelar perspetivas inexploradas), mas também recheada de riscos (existe uma fina linha entre a análise a a exploração abusiva).

A. Análise de dados

Nos últimos anos, a preocupação com o manuseamento de dados tem vindo a ganhar cada vez mais atenção. O design de processos e modelos começa a elevar a relação com o utilizador a um caráter inviolável. Se o utilizador confia na empresa/entidade para guardar os seus dados potencialmente sensíveis, por que motivo quebrar tal confiança?A verdade é que, como já foi referido, a aposta parece ser cada vez mais na extração de conhecimento a partir de dados puramente factuais.

Como exemplo, imagine-se que numa Base de Dados (BD), gerida por um Sistema de Gestão de Bases de Dados (SGBD), existem dados geográficos. Tais dados teriam sido recolhidos com o consentimento dos utilizadores, mas, por via da mineração destes dados, são calculados alguns trajetos e rotas que os utilizadores fizeram. Daqui seria possível extrair hábitos de deslocação, avaliar por que lojas ou espaços o utilizadores costumam passar ou quais os horários nos quais não estão em casa.

O exemplo acima concretiza o que tem sido preconizado neste capítulo: a barreira entre o tratamento dos dados e a manipulação excessiva é ténue. Depende da entidade responsável a aplicação de diretrizes fundamentais, como as que se apresentam na secção seguinte.

B. Princípios fundamentais

Na presente secção serão apresentados alguns princípios éticos [1] que qualquer profissional envolvido com a gestão de dados (em qualquer contexto) deve seguir:

- Respeitar a privacidade dos utilizadores, nunca partilhando ou colocando em risco os dados mais sensíveis;
- Preservar a integridade da informação (a alteração e manipulação dos dados é da responsabilidade dos utilizadores) - nos SGBD atuais, *logging* auxilia nesta tarefa;
- Garantir que a organização dos dados beneficia os utilizadores e nunca interesses de terceiros - as tabelas e relações devem ser pensadas do ponto de vista do utilizador e da realidade em que estes se encontram;
- Avaliar constantemente se os resultados que se pretendem atingir, precisam realmente de uma manipulação de dados intensiva. Se é possível obter os mesmos resultados de uma forma menos dependente de dados, essa via deve ser privilegiada;
- Respeitar todos os utilizadores de forma igualitária;
- Considerar cada utilização de dados, de modo a reduzir riscos desnecessários.

C. Conclusão

Em suma, deve reinar o bom senso acima de tudo. Os dados que o utilizadores confiam à empresa ou entidade que os trata são uma prova de confiança, que não deve ser quebrada.

III. Como dar garantias ao utilizador de que a gestão de dados é bem feita?

No capítulo anterior foram abordados princípios e problemáticas no contexto da gestão de dados. Assim, o presente capítulo tratará de apontar várias formas de dar garantias aos utilizadores de que os dados são bem geridos.

A. Procedimentos a adotar

A confiança tem tanto de percetual como de real. Por um lado, no contexto de um SGBD podem ser aplicadas técnicas que inspiram confiança e segurança. Por outro lado, os utilizadores têm de ter, no mínimo, a sensação de que os seus dados estão sob a alçada de profissionais merecedores de confiança. Como tal, alguns procedimentos [2] podem ser incorporados, visando os ideais acima descritos:

- Guardar, manter e distribuir (se necessário) registos de alterações, edições e manipulações em geral dos dados em posse da empresa/entidade em questão;
- Realizar *backups* automáticos e regulares - tomando como exemplo o SGBD PostgreSQL, o comando *pg_dump* é usado com este fim;
- Encriptar dados sensíveis (como *passwords*, por exemplo) e referir na documentação (que tem que existir) as funções criptográficas usadas. Deste modo, são públicas as potenciais debilidades dos algoritmos escolhidos;
- Atribuir privilégios de acesso adequados ao tipo de utilizador;
- Informar os utilizadores de toda e qualquer alteração aos termos de utilização de dados;
- Obter o consentimento dos utilizadores aquando da manipulação de dados;
- Impedir tentativas de ataque ou acesso indevido aos dados - no PostgreSQL a função *pg_escape_string()* é usada para filtrar o input de possíveis carateres mal intencionados e que poderão alterar o sentido das *queries* usadas;
- Otimizar e melhorar regularmente as configurações do SGBD usado - ferramentas como o PGTune são comummente citadas neste contexto.

B. Conclusão

De um modo geral, a postura transparente de qualquer entidade que possua dados privados deve ser visível por todos. Uma documentação rigorosa, aplicação de funcionalidades disponibilizadas pelo SGBD usado e boas práticas da indústria, fazem parte de uma estratégia que se quer respeitosa para com o utilizador final.

IV. Legislação existente

Com o crescimento do sector tecnológico, apareceram problemas éticos e legais referentes à recolha, tratamento e uso dos dados de clientes. O mau uso ou o abuso dos dados dos clientes levaram a uma forte necessidade de regular a politica de dados. Como tal, sistemas de gestão de base de dados não fugiram à regra, isto é, como armazém principal de dados, este necessita uma forte regulação [5].

Ao longo deste capítulo são introduzidas entidades reguladoras e leis existentes criadas para defender a privacidade dos cidadãos e proibir o mau uso de informação.

A. Comissão Nacional de Protecção de Dados

A CNPD é uma entidade administrativa independente com poderes de autoridade. Esta funciona em cooperação com a Assembleia da Republica.

A sua finalidade consiste em controlar e fiscalizar o processamento de dados pessoais, de forma a conferir que estes estão de acordo com os direitos do homem e as liberdades e garantias consagradas na Constituição e na lei.

Dentro da legislação aprovada existem várias subclasses de leis: protecção de dados pessoais, saúde, comunicações electrónicas, videovigilância, trabalho, cartão de cidadão e cibercrime.

Dentro das várias leis, tome-se por exemplo o caso da protecção de dados pessoais. Aqui existem leis que definem como dados pessoais podem ou não ser usados para efeitos de prevenção, detecção, investigação ou qualquer outra tarefa que possa contribuir para a identificação de infracções penais, sujeitas a sanções. De modo a ilustrar o ponto, atente-se no seguinte artigo ”Direito de acesso do titular dos dados aos seus dados pessoais” (Artigo 15º, Lei 59/2019). Isto é, um dado utilizador tem direito a perguntar se os seus dados estão a ser efectivamente utilizados para algum fim. Caso a resposta seja positiva, este tem ainda direito de requerer a remoção destes, as finalidades e fundamentos jurídicos, entre outros.

B. Regulamento Geral sobre a Protecção de Dados

O RGPD é um regulamento do direito europeu sobre privacidade e protecção de dados pessoais aplicável a todo o cidadão na União Europeia e Espaço Económico Europeu. Este foi criado em 2018 e tem como objectivo dar poder aos cidadãos europeus, bem como, métodos para controlar os seus dados pessoais. Adicionalmente, também é regulada a exportação de dados pessoais para fora da União Europeia.

O regulamento procura definir a maneira como são tratadas informações pessoais na União Europeia e é aplicável a toda e qualquer empresa que opera dentro do Espaço Económico Europeu, independentemente do seu país de origem.

De uma forma geral, todo o processo empresarial dentro do Espaço Económico Europeu é obrigado a ser desenhado de raiz e por um conjunto de medidas que respeitem o principio da protecção de dados desde o início da sua conceção, ou seja, os dados usados devem ser guardados segundo métodos de anonimato de forma a evitar que estes sejam usados sem consentimento explícito e que seja impossível identificar o autor, sem informação adicional armazenada em paralelo.

Segundo o RGPD, não é permitido o uso de dados fora do contexto legal especificado no regulamento, excepto quando o proprietário dos dados tenha dado o seu consentimento explicito a quem controla os dados. Note-se, que esta permissão pode ser revogada em qualquer instância pelo proprietário.

O responsável pela recolha de dados tem obrigatoriamente que declarar qualquer recolha de dados, declarar qual o enquadramento jurídico, a finalidade da recolha, qual o prazo durante o qual os dados ficam armazenados, e ainda, se estes serão partilhados para fora da União Europeia. Os utilizadores sujeito à recolha têm o direito de exigir a cópia dos dados recolhidos e exigir a remoção destes em determinadas circunstâncias.

C. Conclusão

De uma forma geral, ao longo dos últimos anos foram criadas legislações e comissões que visam a protecção de dados pessoais. Por norma, empresas que façam recolha de dados estão sujeito a um conjunto de normas, que implica um total respeito pela privacidade do utilizador. Ao longo do presente documento será ainda possível tomar conhecimento de alguns exemplos nos quais não existiu respeito pela privacidade pessoal.

V. Mecanismos de proteção de dados

Nos dias de hoje proteção e segurança de dados é algo extremamente importante para as companhias que processam com regularidade dados pessoais. Estas deveriam integrar a proteção de dados no processo para garantir a conformidade com o RGPD. O RGPD pode ser resumido a um requisito simples: garantir que os dados estejam seguros. Se tal se verificar, não será necessário haver tantas preocupações e todos os problemas que possam surgir tornam-se muito mais fáceis de resolver. Segue-se uma lista dos métodos de proteção de dados mais utilizados, que ajudam a manter as diretivas presentes no RGPD.

A. Avaliação de Risco

Quanto mais arriscados forem os dados, mais proteção estes devem ter. Dados confidenciais devem ser cuidadosamente protegidos, enquanto que dados de baixo risco podem ter menos proteção. O principal motivo destas avaliações é o custo-benefício, pois melhor segurança de dados envolve uma maior despesa. No entanto, é um bom teste determinar quais os dados que precisam de ser melhor protegidos de modo a tornar todo o sistema de processamento de dados mais eficiente.

Existem dois eixos nos quais a avaliação de riscos deve ser baseada:

- A gravidade dos danos no caso de uma violação de dados;
- A probabilidade de uma violação de dados.

Quanto maior for o risco em cada um desses eixos, mais cuidado se deve ter com os dados. Estas avaliações geralmente exigem a assistência de um responsável pela proteção de dados (responsável pela privacidade) que ajuda a estabelecer regras básicas.

B. Backups

Os *backups* são um método utilizado para impedir a perda de dados que geralmente pode ocorrer devido a erro do utilizador ou mau funcionamento técnico. Os *backups* devem ser feitos e atualizados regularmente, pois apesar de imporem um custo adicional evitam possíveis interrupções nas operações comerciais normais que custam ainda mais. Os *backups* devem ser executados de acordo com o princípio explicado anteriormente - dados de baixa importância não necessitam de ser copiados com tanta frequência, ao contrário de dados confidenciais. Estes *backups* devem ser armazenados num local seguro e possivelmente encriptados e nunca se deve armazenar dados confidenciais na *cloud*.

C. Encriptação

O dados de alto risco são os principais candidatos ao uso de criptografia em todos os seus passos, ou seja, durante a aquisição (protocolos criptográficos *online*), processamento (criptografia de memória completa) e armazenamento subsequente (RSA ou AES). Dados bem encriptados são inerentemente seguros, mesmo nos casos de violação de dados, estes serão inúteis e irrecuperáveis para os invasores. Por este motivo, a criptografia é mencionada explicitamente como um método de proteção de dados no RGPD, o que significa que o seu uso adequado certamente trará favores aos olhos dos reguladores. Por exemplo, se houver uma violação que afeta os dados cifrados, não será necessário denunciá-los às autoridades de supervisão, pois os dados são considerados adequadamente protegidos. Por este motivo, criptografia deve ser considerada como o método número 1 de segurança de dados.

D. Pseudonimização

A pseudonimização é outro método proposto no RGPD que aumenta a segurança e a privacidade dos dados dos indivíduos. Este funciona bem com conjuntos maiores de dados e consiste em remover informações de identificação de trechos de dados. Por exemplo, substituir os nomes de pessoas por sequências geradas aleatoriamente, a identidade de uma pessoa e os dados que eles forneceram tornam-se impossíveis de vincular. O resultado ainda são dados úteis, mas não contêm mais informações sensíveis identificáveis. Como as pessoas não podem ser identificadas diretamente a partir de dados pseudonimizados, os procedimentos no caso de violação ou perda de dados são muito mais simples e os riscos são bastante reduzidos. O RGPD reconhece isso e os requisitos de notificação foram significativamente reduzidos em caso de violações de dados pseudonimizados. A pseudonimização também é essencial na realização de pesquisas científicas ou estatísticas, pelo que instituições e escolas se deveriam focar na pseudonimização adequada dos seus dados.

E. Controlo de Acesso

A introdução a controlos de acesso ao fluxo de trabalho é um método de redução de risco muito eficiente. Quanto menos pessoas tiverem acesso aos dados, menor o risco de violação ou perda (inadvertida) de dados será. Deve-se garantir o acesso a dados confidenciais apenas a funcionários confiáveis que tenham um motivo válido para aceder a estes. É recomendado que se realize regularmente cursos de educação e atualização de manipulação de dados, principalmente após a contratação de novos funcionários. Também é recomendado elaborar uma politica de proteção de dados clara e concisa, descrevendo os métodos, funções e responsabilidades de cada funcionário (ou grupo de funcionários).

F. Destruição

Pode chegar um momento em que os dados que se possui necessitam de ser destruídos. A destruição de dados pode não parecer um método de proteção à primeira vista, mas na verdade é. Os dados estão a ser protegidos desta maneira

contra recuperação e acesso não autorizado. De acordo com o RGPD, devem-se excluir os dados dos quais não se necessita, sendo que, quanto maior for o grau de confidencialidade, mais eficazes deverão ser os métodos de destruição usados. Os discos rígidos são destruídos com mais frequência através de desmagnetização, enquanto que documentos em papel, CDs e unidades de fita são triturados em pedaços pequenos. A destruição de dados no local é recomendada para dados confidenciais. Os dados encriptados podem ser facilmente excluídos simplesmente destruindo as chaves que permitem a desencriptação, garantindo assim que os dados se mantêm ilegíveis.

G. Conclusão

Todos os métodos referidos anteriormente ajudam a manter uma base de dados mais segura e em conformidade com o RGPD, o que deve ser do interesse geral pois ajuda a garantir a segurança, bem como, facilitar a resolução de possíveis problemas.

VI. Falhas e abusos na gestão de dados

Ao longo desta secção serão abordados casos onde o respeito pela privacidade do utilizador foi negligenciado e, mais tarde, os respectivos dados acabaram por ser usados em práticas de exploração.

A. Facebook–Cambridge Analytica [3]

O escândalo *Facebook–Cambridge Analytica* gerou controvérsia no inicio de 2018, quando foi revelado que a empresa *Cambridge Analytica* tinha recolhido informação proveniente do *Facebook* de milhares de utilizadores sem o consentimento destes. Os dados recolhidos foram mais tarde utilizados para propaganda política. O impacto causou uma queda no preço das acções do *Facebook* e pedidos para uma regulamentação mais forte.

De uma forma geral, os dados foram recolhidos através de uma aplicação chamada *"This Is Your Digital Life"*, criada por um dos trabalhadores das empresa. Mais tarde, centenas de milhares de utilizadores da rede preencheram um questionário com fins meramente académicos. No entanto, o design da rede permitiu não só a recolha da informação do questionário, mas também informação pessoal de todos os utilizadores da rede. Desta forma, informação dos utilizadores da redes foi prosperada pela companhia.

A informação prosperada foi, mais tarde, utilizada na campanha do concorrente Ted Cruz para a Casa Branca em 2015/2016 e no mercado publicitário para alterar a tendência de voto.

B. Uber Tracking [4]

Em 2014, a *Uber* era uma das companhias com maior taxa de crescimento. Nesse mesmo ano, um dos funcionários da companhia usou recursos da empresa para detectar o local de um jornalista, que se encontrava atrasado para uma entrevista na empresa. A ferramenta utilizada ia contra a politica privada da companhia, em especifico, funcionários da companhia estavam proibidos de visualizar o histórico das viagens dos clientes, excepto em "casos legítimos".

C. Conclusão

Os exemplos descritos acima, demonstram casos de mau uso de dados. Existem várias fontes para tais situações, desde funcionários, má gestão, entre outros. Porém, o resultado é sempre o mesmo: a violação dos direitos dos utilizadores e, a longo prazo, dados que acabam por ser usados de forma exploratória.

VII. Notas Finais

Ao longo deste documento foram dadas garantias de como a gestão e transparência de dados, a confiança por parte dos utilizadores, o cumprimento das legislações, a proteção de dados e os cuidados com o mau uso de dados são aspetos importantes a ter em consideração num SGBD.
Os SGBD actuais apostam cada vez mais numa gestão de dados realmente bem feita e ao longo dos últimos anos foram criadas legislações e comissões que visam a protecção de dados pessoais assim como um total respeito pela privacidade do utilizador. A proteção de dados também se torna relevante e todos estes métodos apresentados ajudam a manter uma base de dados mais segura e em conformidade com o RGPD, sendo algo do interesse geral.

References

[1] Michael Kassner. 5 ethics principles big data analysts must follow. [Online] https://www.techrepublic.com/article/5-ethics-principles-big-data-analysts-must-follow/

[2] UKEssays. Database Management: Law, Ethics and Security. [Online] https://www.ukessays.com/essays/computer-science/database-management-law-ethics-9552.php?vref=1

[3] Jornal Público. Caso Cambridge Analytica. [Online] https://www.publico.pt/caso-cambridge-analytica

[4] Team ObserveIT. 5 Examples of Data Information Misuse Team ObserveIT - June 25, 2018. [Online] https://www.observeit.com/blog/importance-data-misuse-prevention-and-detection/

[5] Charles Sturt University. Chapter 21 Professional, Legal, and Ethical Issues in Data Management. [Online] https://www.studocu.com/en/document/charles-sturt-university/database-systems/lecture-notes/chapter-21-professional-legal-and-ethical-issues-in-data-management/4296346/view