KAPITOLA

12

Správa a zabezpečení databáze

Témata kapitoly:

- **Přehled**
- **Správa dat a správa databáze**
- **Zabezpečení databáze**

Přehled

V kapitole 4 jste se seznámili s fázemi životního cyklu vývoje databázových systémů. V této kapitole se budeme zabývat tím, jakou roli hrají administrátor dat (DA) a administrátor databáze (DBA) a vztahem mezi těmito rolemi a fázemi životního cyklu vývoje databázových systémů. Důležitou funkcí DA a DBA je zabezpečení databáze. Budeme se zabývat druhy potenciálních ohrožení pro databázový systém a elektronickými protiopatřeními proti těmto ohrožením.

Studijní cíle

V této kapitole se naučíte:

- Rozdíl mezi administrátorem *dat* a administrátorem *databáze*.
- Účel správy dat a správy databáze a činnosti s nimi spojené.
- Oblast zabezpečení databáze.
- Proč je zabezpečení databáze významným úkolem organizace.
- Druhy rizik, která ohrožují databázový systém.
- Jak chránit databázový systém pomocí počítačově podporované kontroly.

Správa dat a správa databáze

Administrátor dat a *administrátor databáze* odpovídají za správu a kontrolu činností spojených s daty společnosti a s databází společnosti (v tomto pořadí). Správce dat (DA) se více zabývá počátečními fázemi životního cyklu, od pláno-

vání po logický návrh databáze. Naopak správce databáze (DBA) se zabývá více závěrečnými fázemi životního cyklu, od návrhu databáze/aplikací po údržbu systému v provozu. V závislosti na velikosti a složitosti organizace a/nebo databázového systému vykonává funkce DA a DBA jeden nebo více pracovníků. Nejprve pojednáme o účelu a úkolech spojených s rolemi DA a DBA v organizaci.

Správa dat

Administrátor dat odpovídá za data společnosti, což zahrnuje také data mimo počítačový systém, a v praxi se často zabývá správou dat sdílených uživateli nebo provozními aplikačními oblastmi organizace. Navíc se DA musí zabývat otázkou, jak data „vstupují" do organizace a jak data „opouštějí" organizaci. DA má hlavní odpovědnost za konzultace a poradenství vyššímu managementu a zajišťuje, aby aplikace databázových technologií trvale podporovaly cíle společnosti. V některých organizacích správa dat představuje samostatnou provozní oblast, v jiných může být spojena se správou databáze. Typické úlohy spojené se správou dat ukazuje tabulka 12.1.

SPRÁVA DAT
Správa a kontrola dat organizace, včetně plánování databáze, vývoje a údržby databáze a údržby standardů, všeobecných postupů a procedur a logického návrhu databáze.

Tabulka 12.1: Typické úlohy správy dat

Výběr vhodných nástrojů Productivity tools.
Pomoc při vývoji IT/IS systému společnosti a obchodních strategií.
Vypracování studií proveditelnosti a plánování vývoje databáze.
Vytvoření datového modelu společnosti.
Určení datových požadavků organizace.
Stanovení standardů sběru dat a formátů dat.
Určení požadavků na přístup k datům a zajištění dodržování zákonných i etických požadavků organizace.
Spolupráce s pracovníky, kteří provádějí správu databáze a s vývojáři aplikací, aby se zajistilo, že aplikace splňují všechny uvedené požadavky.
Školení uživatelů o standardech dat, zákonné a etické odpovědnosti.
Sledování vývoje v oblasti IT/IS.
Zajištění úplnosti dokumentace, včetně datového modelu organizace, standardů, všeobecných postupů, procedur a kontrol koncových uživatelů.
Spolupráce s koncovými uživateli databáze a administrátory databáze při určení nových požadavků a při řešení problémů s přístupem k datům a s výkonností.
Vývoj bezpečnostních postupů.

Správa databáze

Administrátor databáze je více technicky orientován než DA, musí znát specifické DBMS a prostředí operačních systémů. Primární odpovědností DBA je vývoj a celkové udržování systémů pomocí DBMS softwaru. Typické úlohy správy databáze uvádí tabulka 12.2.

SPRÁVA DATABÁZE
Správa a kontrola fyzické realizace databázového systému organizace, včetně fyzického návrhu databáze a implementace, nastavení zabezpečení a kontroly integrity, monitorování výkonnosti systému a v případě potřeby reorganizace databáze.

Tabulka 12.2: Typické úlohy správy databáze

Vyhodnocení a výběr DBMS produktů.
Provedení fyzického návrhu databáze.
Implementace fyzického návrhu pomocí cílového DBMS.
Odhad objemů dat a jejich růstu.
Určení vzorců a frekvence používání databáze.
Definice zabezpečení a integritních omezení.
Spolupráce s vývojáři databázového systému.
Vývoj testovacích strategií.
Školení uživatelů.
Odpovědnost za „spuštění" implementovaného databázového systému.
Monitorování výkonnosti systému a ladění databáze.
Provádění rutinního zálohování.
Zajištění mechanismů a procedur zotavení.
Zajištění úplnosti dokumentace včetně dokumentace materiálů vytvořených v podniku.
Sledování vývoje v oblasti softwaru, hardwaru a nákladů, a instalace aktualizací v případě potřeby.

Srovnání správy dat a správy databáze

V předcházejících oddílech jsme se zabývali účelem a úlohami spojenými se správou dat a správou databáze. V tomto oddílu se budeme krátce zabývat porovnáním těchto rolí. Tabulka 12.3 shrnuje *hlavní* rozdíly mezi DA a DBA. Snad nejdůležitější rozdíl spočívá v povaze práce, kterou vykonávají. Činnost pracovníků DA je mnohem podobnější manažerské činnosti, zatímco činnost pracovníků DBA je více technická.

Tabulka 12.3: Hlavní rozdíly úkolů DA a DBA

Správa dat	Správa databáze
Zapojení do strategického plánování IS	Vyhodnocuje nové DBMS
Určuje dlouhodobé cíle	Vykonává plány, aby se splnily cíle
Určuje standardy, všeobecné postupy a procedury	Kontroluje dodržování standardů, všeobecných postupů a procedur
Určuje požadavky na data	Implementuje požadavky na data
Vyvíjí logický návrh databáze	Vyvíjí fyzický návrh databáze
Vyvíjí a udržuje datový model organizace	Implementuje fyzický návrh databáze
Koordinuje vývoj databáze	Monitoruje a kontroluje užívání databáze
Manažerská orientace	Technická orientace
Činnost nezávisí na DBMS	Činnost závisí na DBMS

Zabezpečení databáze

ZABEZPEČENÍ DATABÁZE
Mechanismy, které chrání databáze proti záměrným nebo náhodným ohrožením.

Ve zbytku této kapitoly se soustředíme na jednu z klíčových úloh jak správy dat, tak správy databáze, totiž zabezpečení databáze. Popíšeme oblast zabezpečení databáze a budeme se zabývat důvody, proč organizace musí brát potenciální ohrožení svých databázových systémů vážně. Také se budeme zabývat povahou jednotlivých ohrožení a jejich možnými dopady na databázové systémy. Nakonec prozkoumáme mechanismy, které je možné k zabezpečení databáze použít.

Zajištění bezpečnosti se netýká jen dat uchovávaných v databázi. Narušení bezpečnosti může ovlivnit také jiné části systému, což může mít zpětně dopady na databázi. Proto zabezpečení databáze zahrnuje také hardware, software, pracovníky a data. Efektivní implementace zabezpečení vyžaduje odpovídající kontroly, které jsou specifikovány v dílčích cílech systému. Ačkoli společnosti potřebu zabezpečení v minulosti zanedbávaly nebo přehlížely, v současnosti si uvědomují jeho význam. Důvodem tohoto obratu je skutečnost, že rostoucí podíl klíčových dat společností je uložen na počítačích a ztráta nebo nedostupnost těchto dat může mít katastrofální následky.

Databáze představuje velmi významný zdroj organizace a měla by být řádně zabezpečena pomocí odpovídající kontroly. Zabezpečením se budeme zabývat v souvislosti s následujícími faktory:

- krádež a zpronevěra,
- ztráta utajení (prozrazení),
- ztráta soukromí,
- ztráta integrity,
- ztráta dostupnosti.

V těchto oblastech by organizace měla minimalizovat riziko; to znamená možnost ztráty nebo škody. V některých situacích jsou tyto faktory tak provázané, že ztráta v jedné oblasti může vést také ke ztrátě v jiných oblastech. Navíc platí, že události jako krádež nebo ztráta soukromí vznikají následkem záměrných nebo nezáměrných činů a nemusí mít za následek žádné pozorovatelné změny databáze nebo počítačového systému.

UTAJENÍ
Mechanismus, který zabezpečuje, aby data nebo informace byly přístupné jen osobám, které jsou k přístupu autorizovány.

Krádež a zpronevěra neovlivňují jen databázové prostředí, ale také celou organizaci. Protože těchto činů se dopouštějí lidé, pozornost je třeba zaměřit na redukci příležitostí. Krádež a zpronevěra nemusí změnit data a totéž může platit také pro činnosti, které mají za následek ztrátu utajení nebo soukromí.

Utajení se týká potřeby udržet data v tajnosti, obvykle se týká pouze dat kriticky důležitých pro organizaci, zatímco soukromí se týká nutnosti chránit data o osobách. Narušení bezpečnosti, která mají za následek ztrátu utajení, mohou znamenat ztrátu konkurenceschopnosti, zatímco ztráta soukromí může mít za následek právní akci proti organizaci.

SOUKROMÍ
Schopnost jednotlivce nebo skupiny zabránit zpřístupnění dat nebo informací o sobě jiným osobám, než si sami přejí.

Ztráta integrity dat má za následek, že data jsou neplatná nebo znehodnocená, což může vážně ovlivnit činnost organizace. Mnoho organizací nyní poskytuje nepřetržité virtuální služby, tzv. dostupnost 24/7 (to znamená dostupnost 24 hodin denně 7 dní v týdnu). Ztráta dostupnosti znamená, že data, systém nebo obojí jsou nedostupné, a to může mít vážný vliv na finanční výsledky organizace. V některých případech mohou události, které způsobí nedostupnost systému, také zavinit znehodnocení dat.

V nedávné době počítačová kriminální aktivita dramaticky vzrostla a podle prognóz bude růst i příštích několik let. Například podle US National Fraud Information Center (*www.fraud.com*) se zvýšily osobní ztráty způsobené počítačovou zpronevěrou z 5 787 170 amerických dolarů v roce 2004 na 13 863 003 amerických dolarů v roce 2005. Podle průzkumu počítačové kriminality a zabezpečení provedeného CSI/FBI (Richardson, 2007) přes 46 % organizací, které ve výzkumu odpověděly, zaznamenalo v posledních 12 měsících neautorizovaný přístup do počítačového systému. Průměrná roční ztráta těchto 5 494 společností stoupla z 168 000 dolarů v roce 2006 na 350 424 dolarů v roce 2007. Zpráva dále poznamenává, že nejkritičtější otázkou v oblasti počítačové bezpečnosti pro nejbližší dva roky bude ochrana dat a bezpečnostní zranitelnost aplikačního softwaru.

Zabezpečení databáze má za cíl minimalizovat ztráty způsobené předvídatelnými událostmi, a to efektivně a bez zbytečného omezování uživatelů.

Druhy ohrožení

> OHROŽENÍ
> Situace nebo událost, záměrná i nezáměrná, která může nepříznivě ovlivnit systém a následně organizaci.

Ohrožení může způsobit situace nebo událost zahrnující osoby, činnosti nebo okolnosti, která bude mít pravděpodobně nepříznivé důsledky pro organizaci. Ztráta organizace může být hmotná, například hardware, software, data, nebo nehmotná, například ztráta věrohodnosti nebo důvěry zákazníků. Každá organizace musí identifikovat možná ohrožení. Proto by organizace měly přinejmenším investovat čas a úsilí do identifikace nejvážnějších ohrožení.

V předcházejícím oddílu jsme popsali následky, které mohou vzejít ze záměrných nebo nezáměrných činů. Ohrožení mohou být záměrná i nezáměrná, následky bývají stejné. Záměrná ohrožení se týkají lidí a vznikají od autorizovaných i neautorizovaných uživatelů, z nichž někteří nemusí patřit k organizaci.

Jakékoli ohrožení musí být řešeno jako potenciální narušení bezpečnosti, které v případě úspěchu má určité dopady. Tabulka 12.4 uvádí příklady různých typů ohrožení a možných následků pro organizaci. Například „použití přístupových prostředků jiné osoby" jako ohrožení může mít za následek krádež nebo zpronevěru, ztrátu utajení a ztrátu soukromí pro organizaci.

Rozsah škod utrpěných organizací jako následek ohrožení závisí na řadě faktorů, jako například existence protiopatření a havarijních plánů. Například pokud dojde kvůli selhání hardwaru k poškození vnější paměti, všechny provozní činnosti je třeba přerušit až do vyřešení problému. Zotavení bude závislé na mnoha činitelích, včetně doby posledního zálohování a času potřebného k zotavení systému.

Organizace musí identifikovat druhy ohrožení, kterými může trpět, a iniciovat vypracování odpovídajících plánů a přijetí protiopatření, při zohlednění jejich nákladů. Zjevně by nebylo efektivní věnovat velké množství času, úsilí a finančních prostředků na řešení ohrožení, které může mít za následek jen menší nepohodlí.

Tabulka 12.4: Příklady ohrožení a možných následků

Ohrožení	Krádež a zpronevěra	Ztráta utajení	Ztráta soukromí	Ztráta integrity	Ztráta dostupnosti
Použití přístupových prostředků jiné osoby	✓	✓	✓		
Neautorizovaná úprava nebo kopírování dat	✓			✓	
Pozměnění programu	✓			✓	✓
Neadekvátní všeobecné postupy a procedury vedoucí k promíchání tajných a normálních výstupů	✓	✓	✓		
Zachycování přenášených zpráv	✓	✓	✓		

Ohrožení	Krádež a zpronevěra	Ztráta utajení	Ztráta soukromí	Ztráta integrity	Ztráta dostupnosti
Ilegální průnik hackera	✓	✓	✓		
Vydírání	✓	✓	✓		
Vytvoření „zadních dvířek" do systému	✓	✓	✓		
Krádež dat, programů a vybavení	✓	✓	✓		✓
Poskytování rozsáhlejšího přístupu než normálně v důsledku selhání bezpečnostních mechanismů	✓	✓	✓		
Nedostatek pracovníků nebo stávky				✓	✓
Neodpovídající školení zaměstnanců		✓	✓	✓	✓
Prohlížení a prozrazení neautorizovaných dat	✓	✓	✓		
Elektromagnetické rušení a vyzařování				✓	✓
Poškození dat kvůli přerušení dodávky elektrické energie				✓	✓
Požár (elektrické závady, úder blesku, žhářství), povodeň, hurikán, bombový útok				✓	✓
Fyzické poškození zařízení				✓	✓
Přerušení nebo vypojení kabelů				✓	✓
Zhroucení softwaru (DBMS) a operačního systému				✓	✓
Působení počítačových virů				✓	✓

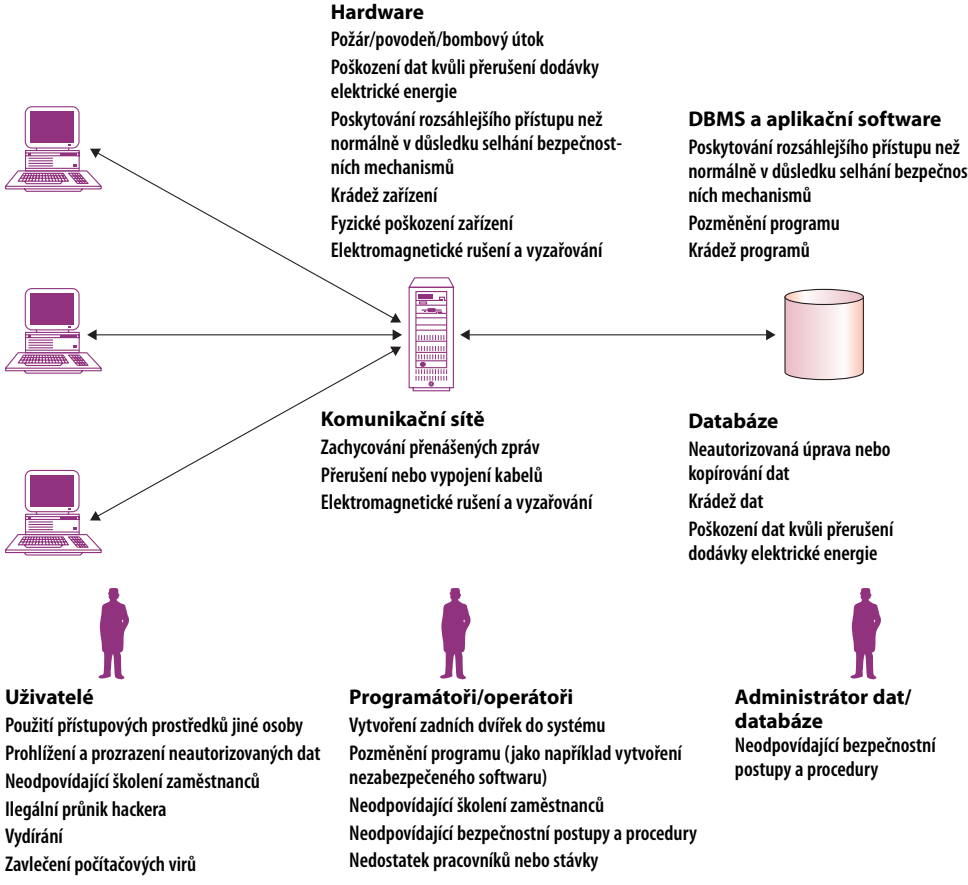
Provoz organizace může být hodnocenými ohroženími také ovlivněn, některá z těchto ohrožení se vyskytují jen vzácně. Vzácně se vyskytující události je však třeba také vzít v úvahu, zejména pokud by jejich dopady mohly být významné. Shrnutí možných ohrožení počítačových systémů uvádí tabulka 12.1.

Protiopatření – počítačová kontrola

Druhy protiopatření proti ohrožením databázových systémů sahají od fyzické kontroly po administrativní procedury. Přes velký rozsah počítačových protiopatření, která jsou k dispozi-

ci, je třeba si uvědomit, že zabezpečení DBMS je pouze tak dobré, jak dobré je zajištění operačního systému, a to kvůli jejich úzké provázanosti.

Typické víceuživatelské prostředí výpočetního systému ukazuje obrázek 12.2. V tomto oddílu se zaměříme na následující počítačově založená bezpečnostní opatření pro víceuživatelské prostředí (některá z nich nemusí být dostupná v prostředí PC):



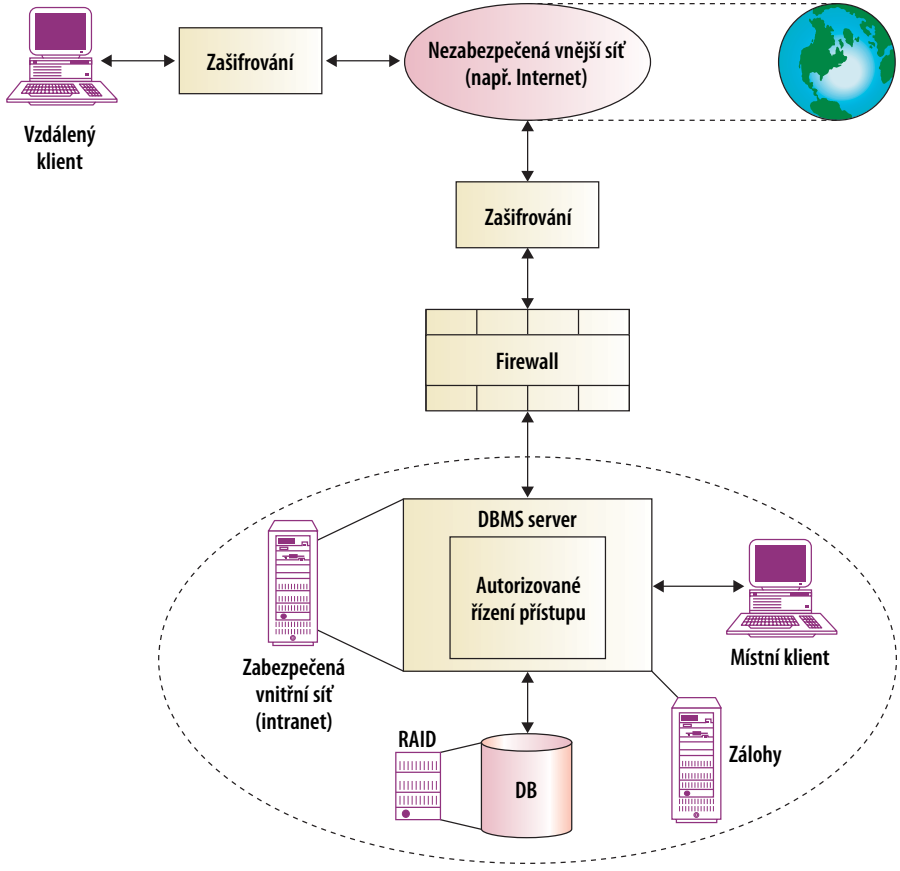
Obrázek 12.1: Shrnutí potenciálních ohrožení počítačových systémů

- autorizace,
- pohledy,
- zálohování a zotavení,
- integrita,
- zašifrování,
- redundantní pole nezávislých disků (RAID).

AUTORIZACE
Udělení přístupových práv, která subjektu umožňují oprávněný přístup do databázového systému nebo k objektu databázového systému.

Autorizace

Kontrolu autorizace je možné zabudovat do softwaru, a tak kontrolovat nejen to, k jakým objektům databázového systému mají uživatelé přístup, nýbrž také co s nimi mohou dělat.



Obrázek 12.2: Typické víceuživatelské prostředí výpočetního systému

AUTENTIZACE
Mechanismus, který kontroluje, zda uživatel je skutečně tím, za koho se vydává.

Z tohoto důvodu se řízení autorizace někdy označuje jako *řízení přístupu* (viz krok 6 v kapitole 11). Proces autorizace zahrnuje autentizaci subjektů požadujících přístup k objektu; v tomto případě „subjekt" představuje uživatele nebo program a „objekt" představuje tabulku databáze, pohled, proceduru, spoušť nebo jiný objekt existující v databázovém systému.

Administrátor systému obvykle odpovídá za povolení přístupu uživatelů k výpočetnímu systému prostřednictvím individuálních uživatelských účtů. Každý uživatel získá jedinečný identifikátor, který operační systém používá k určení jeho identity. S každým identifikátorem je spojeno heslo, zvolené uživatelem a známé operačnímu systému, které musí být zadáno, aby umožnilo operačnímu systému autentizaci (nebo *verifikaci*) uživatelů.

Tento postup umožňuje autorizované používání výpočetního systému, ale neznamená nutně autorizaci přístupu k DBMS a souvisejícím databázovým aplikacím. Oddělená, samostatná procedura může být spojena s přístupem uživatelů k DBMS. Odpovědnost za autorizované používání DBMS má obvykle administrátor databáze (DBA), který musí také zřídit jednotlivé uživatelské účty a hesla pomocí funkcí poskytovaných DBMS.

Některé DBMS udržují seznam platných uživatelů a jejich hesel, který může být odlišný od seznamu operačního systému. Jiné DBMS však položky seznamu ověřují podle seznamu operačního systému na základě přihlašovacích údajů uživatele. Tím se zabrání tomu, aby se uživatel přihlásil k DBMS s odlišným jménem, než pod jakým se přihlásil k operačnímu systému.

Přístupová práva

> PŘÍSTUPOVÁ PRÁVA
> Práva, přidělená uživatelem jinému uživateli nebo skupině uživatelů, přistupovat k databázovému systému nebo objektu v databázovém systému.

Jakmile má uživatel přístup do DBMS, může obdržet na základě toho automaticky také různá přístupová práva. Přístupová práva se mohou týkat například práva přístupu k databázovým objektům nebo práva vytváření databázových objektů jako tabulek, pohledů a indexů nebo spouštění různých utilit DBMS. Přístupová práva uživatelé získávají, aby mohli plnit úkoly spojené se svou pracovní činností. Nadměrné přidělování zbytečných přístupových práv může ohrozit bezpečnost, přístupová práva je třeba přidělovat jen uživatelům, kteří je nezbytně potřebují k pracovní činnosti.

Některé DBMS pracují jako *uzavřené systémy*, takže uživatelé kromě autorizace při přístupu do DBMS jsou autorizováni také při přístupu k určitým objektům. Tuto autorizaci jim poskytuje administrátor databáze nebo vlastník daného objektu. *Otevřený systém* implicitně dovoluje uživatelům úplný přístup ke všem objektům v databázi a pro kontrolu přístupu je nutné přístupová práva uživatelům explicitně odebrat.

Vlastnictví objektů a přístupová práva

Některé objekty v DBMS vlastní přímo DBMS samotný, obvykle jako specifický superuživatel, například DBA. Vlastnictví objektu dává vlastníkovi všechna odpovídající přístupová práva k danému objektu. Stejná situace se týká také autorizovaných uživatelů, pokud vlastní nějaké objekty. Tvůrce objektu vytvořený objekt vlastní a může k němu udělovat přístupová práva. Tato přístupová práva jsou předávána jiným autorizovaným uživatelům. Například vlastník několika tabulek může autorizovat jiné uživatele k dotazování těchto tabulek, ale nikoli k provádění jejich aktualizace.

Pokud DBMS podporuje několik různých typů autorizačních identifikátorů, s každým typem je možné spojit jinou prioritu. Například DBMS může povolovat vytvoření identifikátorů individuálních uživatelů i identifikátorů skupin a identifikátory pro jednotlivé uživatele mohou mít

vyšší prioritu než identifikátory skupin. V takových DBMS mohou být definovány identifikátory uživatelů a skupin tak, jak ukazuje Obrázek 12.3.

Obrázek 12.3 (a) ukazuje všechny uživatele systému spolu s typem uživatele, což odlišuje individuální uživatele od uživatelských skupin. Obrázek 12.3 (b) ukazuje všechny skupiny a uživatele, kteří jsou členy těchto skupin. Některá přístupová práva je možné spojit se specifickými identifikátory, aby se určilo, jaká přístupová práva (například Select, Update, Insert, Delete nebo All) jsou přidělena k určitým objektům databáze.

V některých DBMS musí uživatel systému sdělit, pod jakým identifikátorem pracuje, zejména pokud je členem více skupin. Je důležité seznámit se s dostupnými mechanismy autorizace a dalšími kontrolními mechanismy konkrétního DBMS, zejména pokud je možné přidělit odlišným autorizačním identifikátorům různá přístupová práva. Tak se umožní stanovit typy přístupových práv podle potřeb uživatelů a databázových aplikací, které tito uživatelé používají.

(a) Identifikátory uživatelů

Identifikátor uživatele	Typ
S0099	Uživatel
S2345	Uživatel
S1500	Uživatel
Prodej	Skupina

(b) Identifikátory skupin

Skupina	Identifikátor skupiny
Prodej	S0099
Prodej	S2345

Obrázek 12.3: Identifikátory uživatelů a identifikátory skupin

Přístupovými právy jsme se zabývali v kroku 6 metodologie fyzického návrhu databáze v kapitole 11.

Pohledy

> POHLED
> Virtuální tabulka, která nemusí existovat v databázi, ale může být vytvořena na základě požadavku konkrétního uživatele v době vznesení požadavku.

Koncept pohledů jsme představili na str. 39. Mechanismus pohledů poskytuje mocný a flexibilní bezpečnostní mechanismus díky tomu, že před určitými uživateli skrývá části databáze. Uživatelé si nejsou vědomi existence sloupců nebo řádků, které pohled neobsahuje. Pohled může být definován nad několika tabulkami a uživatel musí získat odpovídající přístupová práva k použití pohledu, ale nikoli k použití podkladových tabulek. Díky tomu může být použití pohledů bezpečnější než přidělení přístupových práv k tabulkám.

Zálohování a zotavení

> ZÁLOHOVÁNÍ
> Proces periodického kopírování databáze a log souboru (případně také programů) na offline paměťová zařízení.

DBMS může poskytovat možnosti zálohování, aby se usnadnilo zotavení databáze po selhání. DBMS zaznamenává probíhající transakce pomocí log souboru (neboli žurnálního souboru),

který obsahuje informace o všech aktualizacích databáze. Doporučuje se zálohovat databázi a log soubor v pravidelných intervalech a zajistit bezpečné uložení záložních kopií. V případě selhání, které způsobí nepoužitelnost databáze, záložní kopie a údaje v log souboru zajistí obnovení databáze v nejaktuálnějším možném konzistentním stavu.

> ŽURNÁLOVÁNÍ
> Proces udržování a správy log souboru (neboli žurnálu) obsahujícího všechny změny učiněné v databázi, aby se umožnilo efektivní zotavení po selhání.

DBMS by měl poskytovat možnosti logování, někdy označovaného jako žurnálování, tedy udržovat informace o aktuálním stavu transakcí a změn v databázi, aby se usnadnilo zotavení po případném selhání. Výhoda žurnálování spočívá v tom, že v případě selhání je možné databázi obnovit do posledního známého konzistentního stavu pomocí záložní kopie databáze a informací obsažených v log souboru. Pokud není žurnálování zapnuto a systém havaruje, jedinou možností zotavení je obnovení databáze pomocí nejaktuálnější záložní kopie databáze. Ale bez log souboru budou ztraceny všechny změny, k nimž došlo po posledním zálohování. Podrobněji se budeme zabývat použitím log souboru pro zotavení na str. 342.

Možnosti zálohování je možné kategorizovat podle těchto voleb:

- Zda databáze v době zálohování běží a zpracovává transakce („horké“ zálohování) nebo je vypnutá („studené“ zálohování).
- Zda se zálohuje celá databáze (všechna data) (plné zálohování) nebo jen část databáze (inkrementální nebo diferenciální zálohování).

V závislosti na těchto volbách může zálohování trvat několik málo minut až několik hodin. Nejnovější zálohovací zařízení jako například jednotky SDLT 600 a LTO-4 mohou kopírovat data z disku na magnetickou pásku rychlostí přibližně 120 megabytů za sekundu (MB/s). Zálohování veškerých dat v databázi o velikosti 400 GB zabere za těchto podmínek necelou jednu hodinu času. Protože náklady na diskový prostor se snižují, je v mnoha případech možné také zálohovat data jednoho disku na jiný disk místo na magnetickou pásku.

> ČASOVÉ OKNO ZÁLOHOVÁNÍ
> Doba, během níž je možné provést zálohu databáze.

Doba, kdy se může provádět zálohování, se nazývá časové okno zálohování. Časové okno zálohování závisí na potřebách organizace. Například organizace, která vyvíjí činnost jen od 9:00 ráno do 5:00 odpoledne, má k dispozici větší časové okno pro zálohování. Podnik provozující e-komerci, který musí být dostupný kdykoli, má jen malé časové okno, během něhož musí dojít k zálohování.

Zálohování databáze během provozu a zpracování transakcí může způsobit problémy s integritou dat. Například předpokládejme, že během dvou hodin potřebných k provedení zálohování dojde k mnoha novým transakcím, které přidávají, mění nebo odstraňují data poté, co byla tabulka zkopírována na magnetickou pásku. Při obnovení ze zálohy by data byla nekonzistentní, některá data by chyběla, jiná by byla přítomna, ačkoli by měla být již vymazána. Jednoduchým řešením této situace by bylo vypnout databázi, provést zálohu dat a pak databázi znovu spustit. Pokud je k dispozici dostatečně dlouhé časové okno zálohování, je to jedno-

značně nejlepší řešení. Ale jak jsme už zmínili, v provozu organizace se vyskytují situace, kdy vypnutí databáze není možné.

Následující možnosti se liší podle toho, jak velké množství dat se zálohuje během jednoho zálohování. Existují tři obecně používané možnosti:

1) Plné zálohování – zálohování všech dat v databázi.
2) Inkrementální zálohování – zálohování všech dat, která se změnila od posledního zálohování.
3) Diferenciální zálohování – zálohování všech dat, která se změnila od posledního plného zálohování.

Abychom tyto možnosti blíže vysvětlili, vezmeme si jako příklad databázi společnosti *StayHome*:

- Plné zálohování se provádí vždy v neděli v noci.
- V pondělí bylo přidáno 10 nových zákazníků, bylo zpracováno 50 nových pronájmů a přidána 3 nová DVD.
- V úterý bylo přidáno 5 nových zákazníků, bylo zpracováno 20 nových pronájmů a přidána 4 nová DVD.

Plné zálohování prováděné každou noc by znamenalo zálohovat všechna data v databázi. Každý den by velikost zálohy vzrůstala podle toho, kolik nových dat by se každý den v databázi objevilo. Obnovení dat by bylo jen otázkou volby nejaktuálnější zálohy a kopírování dat z magnetické pásky na pevný disk. Například selhání disku ve středu dopoledne by se zvládlo obnovením ze zálohy provedené v úterý v noci.

Inkrementální zálohování by začalo s plným zálohováním v neděli a poté by se každou noc zálohovala jen ta data, která by byla mezitím přidána nebo změněna. Pro pondělí by to znamenalo jen 10 nových záznamů o zákaznících, 50 záznamů o pronájmu a 3 záznamy o DVD. Pro úterý by to znamenalo jen 5 nových záznamů o zákaznících, 20 záznamů o pronájmu a 4 záznamy o DVD. Zotavení po selhání disku ve středu dopoledne by vyžadovalo nejprve obnovení zálohy z nedělní noci, pak obnovení inkrementální zálohy z pondělí a nakonec obnovení inkrementální zálohy z úterý.

Diferenciální zálohování by také začalo s plným zálohováním v neděli a poté by se každou noc zálohovala jen data, která by byla přidána nebo změněna od doby plného zálohování. Pro pondělí by to znamenalo jen 10 nových záznamů o zákaznících, 50 záznamů o pronájmu a 3 záznamy o DVD. Pro pondělí by to znamenalo 15 záznamů o zákaznících, 70 záznamů o pronájmu a 7 záznamů o DVD. Zotavení po selhání disku ve středu dopoledne by vyžadovalo nejprve obnovení zálohy z nedělní noci a pak obnovení diferenciální zálohy z úterý.

Při porovnání těchto přístupů máme k dispozici volbu mezi minimalizací času potřebného pro zálohování a minimalizací času a úsilí potřebného pro obnovení dat v případě selhání disku. Plné zálohování vyžaduje při provádění nejdelší dobu, ale umožňuje nejrychlejší zotavení. Inkrementální zálohování vyžaduje při provádění nejkratší dobu, ale také vyžaduje nejdelší dobu pro obnovení dat. Diferenciální zálohování je uprostřed podle obou kritérií.

Komerční DBMS, jako například Oracle a Microsoft SQL Server, poskytují řadu metod pro zálohování dat jak s vypnutou databází, tak během provozu databáze. Oba systémy nabízí podporu pro „horké" zálohování udržováním záznamů o transakcích, které se provádějí během zálohování. Například u Oracle se tato možnost nazývá Archived REDO log.

Integrita

Integritní omezení také přispívají k udržování databáze v bezpečném stavu, a to tím, že brání znehodnocení dat a z toho plynoucím zavádějícím nebo nesprávným výsledkům. Integritní omezení jsme představili na str. 56 a podrobně probrali v metodologii návrhu databáze (viz krok 2.4 v kapitole 10 a krok 3 v kapitole 11).

Zašifrování

> ZAŠIFROVÁNÍ
> Zakódování pomocí speciálního algoritmu, které způsobí, že bez dešifrovacího klíče jsou data pro jakýkoli program nečitelná.

Pokud databáze obsahuje velmi citlivá data, může být nezbytné je zašifrovat, což je preventivní opatření proti možným vnějším ohrožením nebo pokusům o neautorizovaný přístup k datům. Například citlivá data jako čísla sociálního pojištění, čísla řidičských průkazů, čísla kreditních karet, čísla bankovních účtů, hesla atd. jsou na paměťovém zařízení typicky zašifrována, aby se zabránilo přímému čtení těchto dat. Některé DBMS nabízejí funkce pro tento účel. DBMS může k datům přistupovat (poté, co je dešifruje), ale dochází k degradaci výkonnosti kvůli času potřebnému k dešifrování. Zašifrování také chrání data přenášená po komunikačních linkách. Pro zašifrování dat existuje velké množství technik, některé se označují jako reverzibilní, jiné jako ireverzibilní. *Ireverzibilní techniky*, jak plyne z názvu, neumožňují poznat původní data. Data je však možné použít k získání validních statistických informací. *Reverzibilní techniky* jsou používány více. Bezpečný přenos dat přes nezabezpečenou síť vyžaduje použití šifrovacího systému, který obsahuje:

- Šifrovací klíč pro zašifrování dat (čitelného textu).
- Šifrovací algoritmus, který spolu s šifrovacím klíčem převede čitelný text na zašifrovaný text.
- Dešifrovací klíč pro rozšifrování zašifrovaného textu.
- Dešifrovací algoritmus, který spolu s dešifrovacím klíčem transformuje zašifrovaný text na čitelný text.

Jedna technika, *symetrické šifrování*, používá stejný klíč pro zašifrování i dešifrování a při výměně klíčů mezi účastníky komunikace je nezbytné použít zabezpečené komunikační cesty. Protože většina uživatelů nemá přístup k zabezpečené komunikační cestě, musí být klíč, aby byl bezpečný, stejně dlouhý jako samotná zpráva. Ale většina šifrovacích systémů je založena na použití klíče kratšího než vlastní zpráva. Jedním ze schémat pro šifrování je Data Encryption Standard (Standard pro šifrování dat, DES), jenž používá stejný klíč pro zašifrování i dešifrování, a tento klíč je nutno utajit, i když algoritmus nikoli. Algoritmus transformuje 64bitové bloky textu pomocí 56bitového klíče. DES není obecně považován za příliš bezpečný a podle některých stanovisek by bylo třeba používat delší klíč. Například Triple DES používá posloupnost tří šifrovacích fází a v každé fázi se používá odlišný 56bitový klíč a schéma PGP (Pretty Good Privacy) používá 128bitový symetrický algoritmus pro zašifrování odesílaných dat.

Klíče dlouhé 64 bitů lze v současnosti prolomit i při použití běžného hardwaru. Předpokládá se, že klíče dlouhé 80 bitů bude v blízké budoucnosti také možné prolomit, klíče dlouhé 128 bitů zůstanou pravděpodobně v dohledné budoucnosti neprolomitelné. Termíny „silná autentizace" a „slabá autentizace" se někdy používají k odlišení mezi algoritmy, které podle součas-

ných poznatků není možné prolomit pomocí známých technologií a postupů, a těmi, u nichž to možné je.

Jiný typ šifrovacích systémů používá odlišné klíče pro zašifrování a dešifrování. Označuje se jako *asymetrické šifrování*. Jedním příkladem tohoto systému je systém s *veřejným klíčem*, který používá dva klíče, z nichž jeden je veřejný a druhý tajný. Šifrovací algoritmus může být také veřejný, takže každý, kdo chce poslat uživateli zprávu, může použít veřejně známý klíč a algoritmus pro její zašifrování. Ale pouze vlastník tajného klíče může provést dešifrování zprávy. Šifrovací systém s veřejným klíčem se také používá k posílání „digitálního podpisu" se zprávou, což slouží jako důkaz, že zpráva pochází od osoby s danou identitou. Nejznámějším systémem s asymetrickým šifrováním je RSA (jméno je odvozeno od iniciál jeho návrhářů: Rivest, Shamir a Adelman).

Obecně jsou symetrické algoritmy při zpracování na počítači mnohem rychlejší než asymetrické. V praxi se však často používají společně; veřejný algoritmus se použije k zašifrování náhodně generovaného klíče a tento náhodný klíč se použije k zašifrování skutečné zprávy pomocí symetrického algoritmu. V systémech e-komerce (o nichž pojednáme v kapitole 15) se používá protokol Secure Socket Layer (SSL), který je založen na tomto algoritmu. Webové stránky udržují pár veřejný/tajný klíč a „zveřejňují" veřejný klíč v „Certifikátech". Když prohlížeč požaduje zabezpečené spojení, webový server pošle kopii certifikátu a prohlížeč použije veřejný klíč, který je ve zprávě obsažen, pro přenos tajného (symetrického) klíče, který bude používán pro zabezpečení následujících přenosů.

RAID

> RAID
> Skupina nebo pole fyzických diskových jednotek, které se uživatelům databáze (a také programům) jeví jako jedna velká logická paměťová jednotka.

Diskové vstupy a výstupy (I/O) prošly revoluční změnou díky zavedení technologie RAID. Zkratka RAID původně pocházela z výrazu *Redundant Array of Inexpensive Disks* (redundantní pole levných disků), ale později začalo I ve zkratce znamenat *Independent* (nezávislý). RAID pracuje na principu velkých diskových polí obsahujících několik nezávislých disků, které jsou organizovány tak, aby se zvýšila výkonnost a současně spolehlivost.

Výkonnost se zvýší pomocí technologie *data striping*: data jsou segmentována na jednotky stejné velikosti (*stripovací jednotky*), které jsou transparentně distribuovány na více disků. Diskové pole funguje stejně, jako by šlo o jediný velký a velmi rychlý disk, ačkoli jsou data ve skutečnosti rozdělena na několik malých disků. Stripování dat zlepšuje všeobecnou výkonnost I/O operací tím, že několik I/O operací je možné provádět paralelně. Stripování dat současně také rovnoměrně rozděluje zátěž mezi disky. Spolehlivost se zvýší díky ukládání redundantních informací na více diskových jednotek pomocí schématu *parity* nebo schématu *opravy chyb*. V případě havárie disku je možné redundantní informace použít k rekonstrukci obsahu disku postiženého selháním.

Existuje několik různých konfigurací disků, které se označují jako úrovně RAID. Každá z nich poskytuje poněkud odlišné možnosti vyvážení výkonnosti a spolehlivosti.

Existují následující úrovně RAID:

- RAID 0 – bez redundance: Na této úrovni se neuchovávají redundantní data a díky neexistenci replikací při aktualizaci má nejlepší výkon při zápisu. Stripování dat se provádí na úrovni bloků.
- RAID 1 – zrcadlení: na této úrovni se udržují (zrcadlí) dvě identické kopie dat na různých discích. Aby se udržela konzistence v případě havárie disku, zápis se nesmí provádět simultánně. Jde o nejdražší paměťové řešení.
- RAID 1+0 (někdy označovaná RAID 10) – bez redundance a se zrcadlením: tato úroveň spojuje stripování a zrcadlení.
- RAID 2 – opravné kódy: na této úrovni je stripovací jednotkou jediný bit a jako schéma redundance se používají opravné kódy.
- RAID 3 – bitově prokládaná parita. Informace o paritě se ukládají na jednom disku z pole. Informace o paritě na tomto disku slouží k obnově dat na jiných discích, pokud havarují. Tato úroveň vyžaduje méně diskového prostoru než RAID 1, ale slabinou může být disk s paritami.
- RAID 4 – blokově prokládaná parita: Na této úrovni je stripovací jednotkou diskový blok. Parita diskového bloku se udržuje na odděleném disku spolu s paritou řady odpovídajících bloků z dalších disků. Pokud jeden z těchto disků havaruje, paritní blok spolu s odpovídajícími bloky dalších disků je možné použít k obnovení bloků havarovaného disku.
- RAID 5 – blokově prokládaná parita rozložená po discích. Na této úrovni se používají paritní data podobně jako v RAID 3, ale jsou rozdělena na všechny disky podobným způsobem jako zdrojová data. To zlepšuje situaci s úzkým místem, jímž je paritní disk.
- RAID 6 – P + Q redundance: tato úroveň je podobná RAID 5, ale uchovávají se další redundantní data pro případ havárie více disků současně. Místo parity se používají opravné kódy.

Pro většinu databázových aplikací se zpravidla vybírají úrovně RAID 1, RAID 1+0 nebo RAID 5. Například Oracle doporučuje použití RAID 1 pro redo log soubory. Pro databázové soubory ORACLE doporučuje použití RAID 5, pokud jsou přijatelné další náklady na správu, jinak RAID 1 nebo RAID 1+0.

Protiopatření – zabezpečení sítě

Předcházející protiopatření se zaměřovala na interní kontrolu v rámci DBMS a databáze. Moderní databázové systémy jsou však jen zřídka implementovány izolovaně. Místo toho se vyskytují dvouvrstvé a třívrstvé architektury, a zejména třívrstvé architektury implementované s webovými servery (podrobně popisujeme v kapitole 15), které zahrnují síť jako propojení mezi různými vrstvami systému. Například v typických architekturách e-komerce pracuje uživatel s prohlížečem, který je připojen k webovému serveru společnosti přes různé sítě Internetu. Tento webový server je zase připojen k internímu databázovému serveru společnosti (kde se nachází DBMS a databáze) pomocí privátní sítě, která patří společnosti. Zabezpečení sítě je proto důležitým aspektem zabezpečení databázových zdrojů.

ZABEZPEČENÍ SÍTĚ
Ochrana serverů před neoprávněným vniknutím.

Základní architekturu zabezpečení sítě v případě třívrstvého databázového systému ukazuje obrázek 12.4. Upozorňujeme, že podrobná diskuse o otázkách zabezpečení sítě a alternativních architekturách zabezpečení sítě přesahuje rámec této knihy; uvedeme však některé obecné definice.

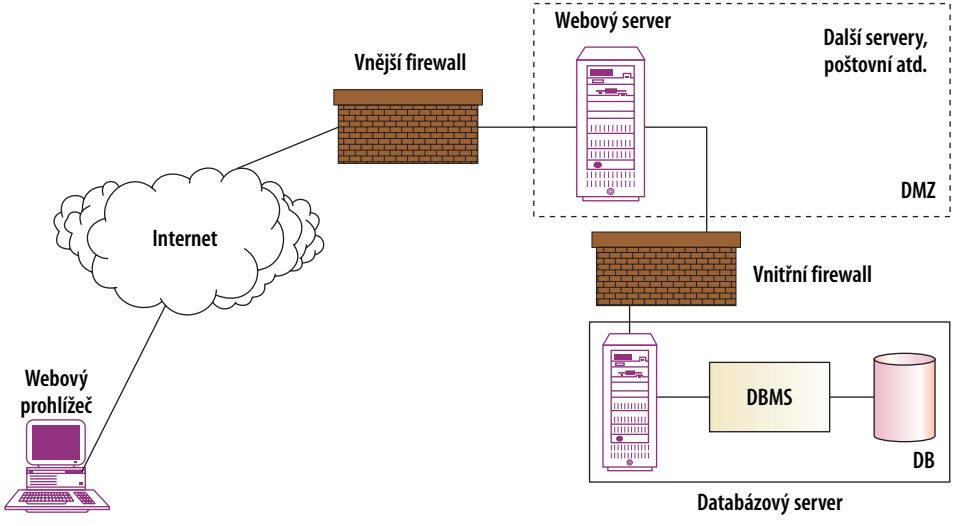
FIREWALL
Server nebo router s dvěma nebo více síťovými rozhraními a speciálním softwarem, který filtruje nebo selektivně blokuje zprávy procházející mezi těmito sítěmi.

Firewall je možné konfigurovat tak, aby propouštěl mezi různými sítěmi pouze zprávy určitých druhů. Na obrázku 12.4 například je možné vnější firewall nakonfigurovat tak, aby blokoval všechny zprávy přicházející z Internetu s výjimkou požadavků na webové stránky a přicházejících e-mailů. Tímto způsobem je možné dosáhnout toho, aby zákazníci mohli prohlížet webové stránky na webovém serveru společnosti a posílat společnosti e-mailové zprávy. Všechny ostatní zprávy, které by se případně mohly snažit využít slabin v dalších službách, jsou blokovány „na hranicích" organizace. Vnitřní firewall na obrázku 12.4 je ještě restriktivnější a je nakonfigurován tak, že dovoluje jen spojení mezi webovým serverem a databází pro účely běhu transakcí. Tato interakce je podrobněji vysvětlena na str. 362.

DEMILITARIZOVANÁ ZÓNA (DMZ)
Speciální, omezená síť zřízená mezi dvěma firewally.

Na obrázku 12.4 vymezuje demilitarizovanou zónu (DMZ) přerušovaná čára mezi dvěma firewally. Záměrem je silně omezit funkce serverů v DMZ, protože jsou částečně „vystavené" Internetu.

Ve větších organizacích spolupracuje administrátor databáze s pracovníky zabezpečení sítě, aby se zajistilo, že databáze nebude zranitelná přímými útoky z Internetu. V menších organizacích může tuto odpovědnost nést přímo administrátor databáze.



Obrázek 12.4: Základní architektura zabezpečení sítě v případě třívrstvého databázového systému

Znovu zdůrazňujeme, že obrázek ukazuje jen jednu z mnoha možných architektur zabezpečení sítě.

Shrnutí kapitoly

- **Správa dat** je správa a kontrola dat organizace, včetně plánování databáze, vývoje a údržby databáze a údržby standardů, všeobecných postupů, procedur a logického návrhu databáze.
- **Správa databáze** je správa a kontrola fyzické realizace databázového systému organizace, včetně fyzického návrhu databáze a implementace, nastavení zabezpečení a kontroly integrity, monitorování výkonnosti systému a v případě potřeby reorganizace databáze.
- **Zabezpečení databáze** se zabývá ochranou před následky těchto událostí: krádež a zpronevěra, ztráta utajení (prozrazení), ztráta soukromí, ztráta integrity, ztráta dostupnosti.
- **Utajení** se týká potřeby udržet data nepřístupná, obvykle se vztahuje jen na data kriticky důležitá pro organizaci, zatímco **soukromí** se týká potřeby chránit osobní data.
- **Ohrožení** je situace nebo událost, záměrná i nezáměrná, která může nepříznivě ovlivnit systém a následně organizaci.
- **Počítačově založená bezpečnostní opatření pro víceuživatelské prostředí** se týkají autorizace, pohledů, zálohování a zotavení, žurnálování, integrity, šifrování a RAID.
- **Autorizace** je udělení přístupových práv, která subjektu umožňují oprávněný přístup do databázového systému nebo k objektu databázového systému.
- **Autentizace** je mechanismus, který určuje, zda uživatel skutečně je tím, za koho se vydává.
- **Pohled** je virtuální tabulka, která nemusí existovat v databázi, ale může být vytvořena na základě požadavku konkrétního uživatele v době vznesení požadavku.
- **Zálohování** je proces periodického kopírování databáze a log souboru (případně také programů) na offline paměťová zařízení.
- **Časové okno zálohování** je doba, během níž je možné provést zálohu databáze.
- **Žurnálování** je proces udržování a správy log souboru (neboli žurnálu) obsahujícího všechny změny učiněné v databázi, aby se umožnilo efektivní zotavení po selhání.
- **Integritní omezení** také přispívají k udržování databáze v bezpečném stavu, a to tím, že brání znehodnocení dat a z toho plynoucím zavádějícím nebo nesprávným výsledkům.
- **Zašifrování** je zakódování dat pomocí speciálního algoritmu, které způsobí, že bez dešifrovacího klíče jsou data nečitelná pro jakýkoli program.
- **RAID** je skupina nebo pole fyzických diskových jednotek, které se uživatelům databáze (a také programům) jeví jako jedna velká logická paměťová jednotka.
- **Zabezpečení sítě** se týká ochrany serverů před vniknutím pomocí implementace architektury zabezpečení sítě.

Kontrolní otázky

12.1 Definujte účel a úkoly spojené se správou dat a správou databáze.

12.2 Srovnejte úkoly správce dat a úkoly správce databáze a uveďte rozdíly mezi nimi.

12.3 Vysvětlete účel zabezpečení databáze a popište oblasti, kterých se týká.

12.4 Jaký je rozdíl mezi utajením a soukromím?

12.5 Uveďte hlavní druhy ohrožení, které se týkají databázových systémů, a ke každému druhu uveďte možná protiopatření.

12.6 Vysvětlete následující termíny z oblasti zabezpečení databází:

(a) autorizace a autentizace

(b) pohledy

(c) zálohování a zotavení

(d) integrita

(e) šifrování

(f) RAID

12.7 Jaký je rozdíl mezi plným zálohováním, inkrementálním zálohováním a diferenciálním zálohováním?

12.8 Jaký je rozdíl mezi symetrickým a asymetrickým šifrováním?

12.9 Co je firewall a jak je možné ho použít k ochraně databázového systému?

Cvičení

12.10 Navštivte webové stránky Computer Security Institute (Institutu počítačové bezpečnosti) (*http://www.gosci.org*) a stáhněte si nejnovější CSI/FBI Computer Crime and Security Surfy (Průzkum počítačové kriminality a bezpečnosti). Diskutujte o současných trendech v druzích útoků nebo zneužití, které jsou nejkritičtější pro zabezpečení a integritu dat.

12.11 U libovolného DBMS, k němuž máte přístup, prozkoumejte, jak podporuje následující funkčnost:

(a) autentizaci

(b) vlastnictví objektů a přístupová práva

(c) zálohování a obnovení

(d) šifrování

12.12 Vezměte si případovou studii *StayHome Online Rentals* popsanou v kapitole 5. Vytvořte zprávu pro manažera společnosti s nástinem ohrožení, která se mohou týkat databázového systému společnosti, a podejte soubor doporučení, jak tato ohrožení minimalizovat.

12.13 Vezměte si případovou studii *Perfectpets* popsanou v příloze B. Vytvořte zprávu pro manažera společnosti s nástinem ohrožení, která se mohou týkat databázového systému společnosti, a podejte soubor doporučení, jak tato ohrožení minimalizovat.

12.14 Předpokládejme, že databáze o velikosti přibližně 500 gigabytů (GB) dat vyžaduje každodenní zálohování.

(a) Pokud je k dispozici jako zálohovací zařízení magnetická páska založená na technologii LTO-2 (objem dat na pásku může činit až 200 GB, s přenosovou rychlostí 24 megabytů za sekundu (MB/s)), jak dlouho bude trvat plné zálohování?

(b) Zjistěte formát nejnovější magnetické pásky LTO-4 (dva možní výrobci jsou Hewlett-Packard a Tandberg Data) a popište rozdíl v době zálohování při použití této nové technologie.

(c) Současné disky typu Serial Attached SCSI (SAS) a Serial ATA (SATA) mají přenosovou rychlost až 300 MB/s. Popište rozdíl v době zálohování při použití této technologie pro zálohování „z disku na disk".

12.15 Napište technickou zprávu s porovnáním a popisem rozdílů mezi asymetrickým šifrovacím algoritmem a symetrickým šifrovacím algoritmem.

12.16 Prozkoumejte dva rozdílné přístupy k zajištění bezpečnosti webového přístupu k databázi při použití firewallu. Připravte zprávu o tom, která alternativa je podle vás nejbezpečnější.