Datenschutz und Datensicherheit

Datenschutz und Persönlichkeitsrechte

- Grundrecht auf informationelle Selbstbestimmung
  - “Das Grundrecht gewährleistet insoweit die Befugnis des Einzelnen, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen. Einschränkungen dieses Rechts auf "informationelle Selbstbestimmung" sind nur im überwiegenden Allgemeininteresse zulässig“
  - Aus der Verfassung abgeleitetes Persönlichkeitsrecht, erstmals anerkannt durch Urteil des Bundesverfassungsgerichts vom 15.12.1983 (Volkszählungsurteil)
  - Zusammengefaßt: Soviel Freiheit wie möglich und soviel Bindung wie nötig
- Definition Datenschutz
  - Schutz persönlicher Daten gegen ungesetzlichen Erwerb und ungesetzliche Speicherung und Verbreitung, sowie die Bereitstellung der notwendigen Schutzmaßnahmen gegen die Zerstörung oder Beschädigung legal aufbewahrter Daten

Bundesdatenschutzgesetz (BDSG)

- Historie
  - 1977: Erstes Bundesdatenschutzgesetz
  - 1983: Recht auf informationelle Selbstbestimmung erhält Verfassungsrang (Volkszählungsurteil)
  - 2017: letzte Neufassung im Zusammenhang mit EU-DSGVO
- Geschützte Daten: personenbezogene Daten
  - Daten, die die persönlichen oder sachlichen Verhältnisse einer natürlichen Person beschreiben
  - Person muss nicht namentlich benannt, aber bestimmbar sein, z.B. Telefon-Nummer, E-Mail-Adresse, IP-Adresse beim Surfen
  - Kein Problem bei anonymen Daten oder pseudonymen Daten
  - Besonderer Schutz für sensible Daten (erfordert ausdrückliche Einwilligung des Betroffenen)
    - Rassische und ethnische Herkunft
    - Politische Meinungen, religiöse oder philosophische Überzeugungen
    - Gewerkschaftszugehörigkeit
    - Gesundheit und Sexualleben

BDSG: Grundsätze und Inhalt

- Grundsätze
  - *Verbotsprinzip* mit *Erlaubnisvorbehalt*, d.h. Erhebung und Verarbeitung von personenbezogenen Daten im Prinzip verboten
    - Benötigt klare Rechtsgrundlage oder ausdrückliche Zustimmung des Betroffenen
  - Grundsatz der *Datenvermeidung* und *Datensparsamkeit*
    - Ziel für alle EDV-Systeme: keine oder so wenig wie möglich personenbezogene Daten
    - Möglichkeiten von Anonymisierung und Pseudonymisierung nutzen
- Vorschriften, die Missbrauch der Daten entgegenwirken
  - Festlegung, welche Daten von wem erhoben und gespeichert werden
  - welcher Zugriff auf Daten erlaubt ist,
  - welche Weitergabe der Daten zulässig ist

Datenschutz-Grundverordnung (DSGVO)

- Verordnung der EU zur vereinheitlichten Regeln zum Umgang mit personenbezogene Daten durch private Unternehmen und öffentliche Stellen
- **Gültigkeit**: ab 25.05.2018
- **Grundsätze** der Verarbeitung personenbezogener Daten
  - Rechtmäßigkeit
  - Zweckbindung
  - Datenminimierung (ersetzt Grundsatz der Datensparsamkeit)
  - Richtigkeit
  - Speicherbegrenzung (Daten müssen „in einer Form gespeichert werden, die die Identifizierung der betroffenen Personen nur so lange ermöglicht, wie es […] erforderlich ist“)
  - Integrität und Vertraulichkeit („angemessene Sicherheit der personenbezogenen Daten …“; Schutz der Daten gegen Zerstörung und Verlust)

Datenschutz-Grundverordnung (Forts.)

- **Transparenz**
  - Recht einer Person auf Auskunft über alle sie betreffenden Daten
  - Auskunft bei einer Datenerhebung u. a. über Zweck, Empfänger und Verantwortliche der Datenverarbeitung, Dauer der Datenspeicherung, …
  - Recht auf Berichtigung von falschen personenbezogenen Daten
  - Recht auf Sperrung der personenbezogenen Daten (wenn Richtigkeit und Grundlage der DV bestritten werden)
- Recht auf **Vergessenwerden**
  - Recht auf Löschen der personenbezogenen Daten, wenn Gründe für Speicherung entfallen
- Recht auf **Datenübertragbarkeit**
  - Recht zum Erhalt der personenbezogenen Daten in einem „strukturierten, gängigen und maschinenlesbaren Format“
- **Privacy by Design**, Privacy by Default
  - Systementwurf mit Voreinstellungen (Defaults), die den Grundsätzen des Datenschutzes entsprechen

Datensicherheit (Security)

***Definition* Datensicherheit:**
Der Schutz einer Datenbank gegen beabsichtigte oder unbeabsichtigte Bedrohungen mit Hilfe von computergestützten und nicht-computergestützten Maßnahmen

Hauptaufgaben / Ziele

- **Vermeidung von Diebstahl und Fälschung:**
  - **Nicht auf die Umgebung der Datenbank begrenzt**
  - **Gefahr besteht in gesamter Organisation**
  - **Daten müssen dabei nicht verändert sein**
- **Geheimhaltung: Vertraulichkeit / Geheimhaltung von Daten, die von entscheidender Bedeutung für das Unternehmen sind**
  - **z.B. bei Verletzung Verlust der Wettbewerbsfähigkeit des Unternehmens**
- **Wahrung der Privatsphäre: persönliche Daten von Individuen dürfen nicht für unbefugte Personen zugänglich sein**
  - **z.B. Ein Student kann nicht die Noten der anderen Studenten sehen.**
- **Integrität: Benutzer sollten Dinge nicht ändern dürfen, wozu sie nicht autorisiert sind.**
  - **z.B. Nur Professoren dürfen Noten eintragen.**
- **Verfügbarkeit:**
  - **Bei kontinuierlicher Arbeitsweise 24x7 Verfügbarkeit erforderlich**
  - **Verlust der Verfügbarkeit → kein Zugriff auf Daten oder aufs ganze System, kann die Existenz des Unternehmens gefährden**

Gegenmaßnahmen

Nicht-Computergestützt

- Sicherheitsrichtlinien und Notfallpläne
- Personalkontrollen
- Sichere Aufstellung von Geräten
- Wartungsverträge
- *Kontrolle des Zutritts*
- *Kontrolle des Zugangs zu DV-Anlagen*

Computergestützt

- *Authentifizierung und Autorisierung (Zugriffskontrolle)*
- *Sichten (Views)*
- Sicherung und Wiederherstellung (Recovery)
- *Verschlüsselung*
- Datenflusskontrolle beim Datentransport

Zutritts- und Zugangskontrolle

- Organisatorische Maßnahmen
  - Pförtner (kontrolliert Zutritt)
  - Bauliche Maßnahmen
- Kryptographische Maßnahmen
  - Verschlüsselung von Daten, Nachrichten und Programmen
- Identitätskontrolle (Authentisierung) bei Zugang zum System
  - Nachweis der Identität des Benutzers gegenüber dem System (Datenbanksystem, Betriebssystem)
  - Verfahrensklassen der Authentisierung:
    - *Was der Benutzer selbst hat:*
      Stimme, Fingerabdruck, Auge, Unterschrift, etc.
    - *Was der Benutzer besitzt:*
      ausgehändigte Gegenständen wie Schlüssel, maschinell lesbarer Ausweise (Badge)
    - *Was der Benutzer weiß:*
      Paßwörter, PINs, Muttis Geburtstag etc.

Zugriffskontrolle (Autorisierung)

- Vergabe von Zugriffsrechten (Lesen, Schreiben, ...) auf DB-Objekten, Programmen etc. *Wer* darf *was womit* machen?
- Subjekte (B)
  - Benutzer
  - Benutzergruppen
  - Terminals / Rechner-Adressen
- Objekte (O)
  - Anwendungs- und Dienstprogramme
  - DB-Objekte: Relationen, Sichte, Attribute, gespeicherte Prozeduren
- Operationen (P)
  - Lesen
  - Ändern
  - Erzeugen
  - Ausführen (bei Programmen)
  - Weitergabe von Zugriffsrechten
- Berechtigungsmatrix

  $<B_i, O_j,$ Liste von P‘s$>$

Klassifikation der Autorisierung

- Dezentrale Vergabe von Zugriffsrechten durch Eigentümer (*Discretionary Access Control*)
  - Zugriffsrechte oder Privilegien für DB-Objekte (Tabellen und Sichten)
  - Mechanismen zum Gewähren und Entziehen von Privilegien an bestimmte Benutzer
  - Eigentümer eines DB-Objekts erhält automatisch alle Privilegien darauf
  - DBMS kontrolliert den Erwerb und Verlust von Privilegien und sichert, daß nur Anfragen von Benutzer mit den notwendigen Privilegien erlaubt sind (zur Zeit der Anfrage); dafür verantwortlich: Zugriffskontrollsystem als eine Komponente des DBMS
- Alternativer Ansatz: systemweite Vergabe von Rechten (*Mandatory Access Control*), siehe später
- Wertabhängige vs. wertunabhängige Festlegung von Privilegien
  - View-Konzept: Untermengenbildung einzelner Relationen

GRANT-Kommando

GRANT privileges ON object TO users [WITH GRANT OPTION]

- Definition von Privilegien (für Objekte):
  - SELECT: Alle Spalten können gelesen werden (einschließlich derer, die später mittels ALTER TABLE hinzugefügt werden)
  - INSERT (col-name): Es können Tupel mit Werten (ungleich NULL oder ungleich Default) in dieser Spalte eingefügt werden
    - INSERT bedeutet das gleiche Recht bezogen auf alle Spalten
  - UPDATE: Analog UPDATE, mit Spalteneinschränkung möglich
  - DELETE: Tupel können gelöscht werden
  - REFERENCES (col-name): Erlaubt die Definition von Fremdschlüsseln (Foreign Keys) in anderen Tabellen, die diese Spalte referenzieren. Ohne Spaltenname gilt dieses Recht für das gesamte Tupel
  - ALL: Gewährung aller oben genannten Rechte
  - EXECUTE: erlaubt die Ausführung von gespeicherten Funktionen und Prozeduren
- dynamische Weitergabe von Zugriffsrechten mit der GRANT OPTION, an andere Benutzer (dezentrale Autorisierung)
- Nur der Eigentümer eines Schemas kann DB-Objekte erzeugen, verändern oder löschen: Kommandos CREATE, ALTER, DROP
- Empfänger: Liste von Benutzern (users) oder PUBLIC

Rücknahme von Zugriffsrechten

```
REVOKE [GRANT OPTION FOR] privileges
ON object FROM users {RESTRICT | CASCADE}
```

- Zurücknahme eines Privilegs (SELECT, UPDATE etc.) oder nur der GRANT OPTION auf einem Privileg
- Probleme: Rechteempfang aus verschiedenen Quellen möglich
- CASCADE: Rücknahme aller Privilegien oder Grant Options vom Benutzer, sofern dieser diese Rechte ausschließlich vom Aufrufer X erhalten hat (entweder direkt oder indirekt über andere Benutzer)
- RESTRICT: Rücknahme der Privilegien nur dann möglich, wenn keine weiteren Abhängigkeiten bestehen
- Hilfsmittel: Führen der Abhängigkeiten in einem *Autorisierungs-graphen* mit Knoten für die Benutzer und Kanten für die Privilegien

GRANT und REVOKE Beispiele

- GRANT INSERT, SELECT ON Sailors TO Horatio
  - Horatio kann SELECT-Abfragen auf Sailors oder Tupel einfügen
- GRANT DELETE ON Sailors TO Yuppy WITH GRANT OPTION
  - Yuppy kann Tupel löschen und auch andere Benutzer dazu autorisieren
- GRANT UPDATE (*rating*) ON Sailors TO Dustin
  - Dustin kann ändern, aber nur die Spalte *rating* in Tabelle Sailors
- GRANT SELECT ON ActiveSailors TO Guppy, Yuppy
  - Dies gestattet den ‘uppies nicht, direkt in Sailors zu lesen!
- Kaskadierende Rücknahme von Rechten (Beispiel):

  GRANT SELECT ON Sailors TO Art WITH GRANT OPTION *(Joe)*

  GRANT SELECT ON Sailors TO Bob WITH GRANT OPTION *(Joe)*

  GRANT SELECT ON Sailors TO Bob WITH GRANT OPTION *(Art)*

  REVOKE SELECT ON Sailors FROM Art CASCADE *(Joe)*

  Art verliert seine SELECT-Rechte. Bob hat seine Rechte von Joe und von Art. Somit behält Bob seine Rechte.

Views und Security

- Views (Sichten) können genutzt werden, um benötigte Informationen darzustellen (oder eine aggregierte Form davon); Details werde in zugrundeliegender Relation versteckt
  - Beispiel: Anlegen einer View ActiveSailors (alle Segler, die mindestens ein Boot reserviert haben), ohne die *bid*s (Boot-Identifier)
- Erzeuger einer View hat Rechte auf der View sofern er auch die Rechte auf zugrundeliegender Basis-Relation hat
- Wenn der Erzeuger einer View das SELECT-Recht auf der Basistabelle verliert: View wird gelöscht
- Wenn der Erzeuger einer View ein mit GRANT gewährtes Recht auf der Basistabelle verliert: Recht geht auch auf der View verloren
- Zusammen mit GRANT/REVOKE sind Views ein sehr mächtiges Konzept der Zugriffskontrolle

Rollenbasierte Autorisierung

- In SQL-92 werden Rechte intern an Autorisierungs-Identifikatoren zugewiesen, welche einen einzelnen Benutzer oder eine Gruppe von Benutzern repräsentieren
- In SQL:1999 (und in vielen aktuellen Systemen): Zuweisung von Privilegien an Rollen:
  - Rollen umfassen eine Menge von Rechten und können Benutzern oder auch anderen Rollen zugewiesen werden
  - Rollen entsprechen einem bestimmtes Profil (z.B. Admin, Developer, Student) und reflektieren die Organisation des Unternehmens

1. Erzeugung einer Rolle

```
CREATE ROLE rolle
```

2. Spezifikation der Rolle

```
GRANT privileges ON object TO rolle
```

3. Zuweisung der Rolle an Benutzer

```
GRANT rolle TO users
```

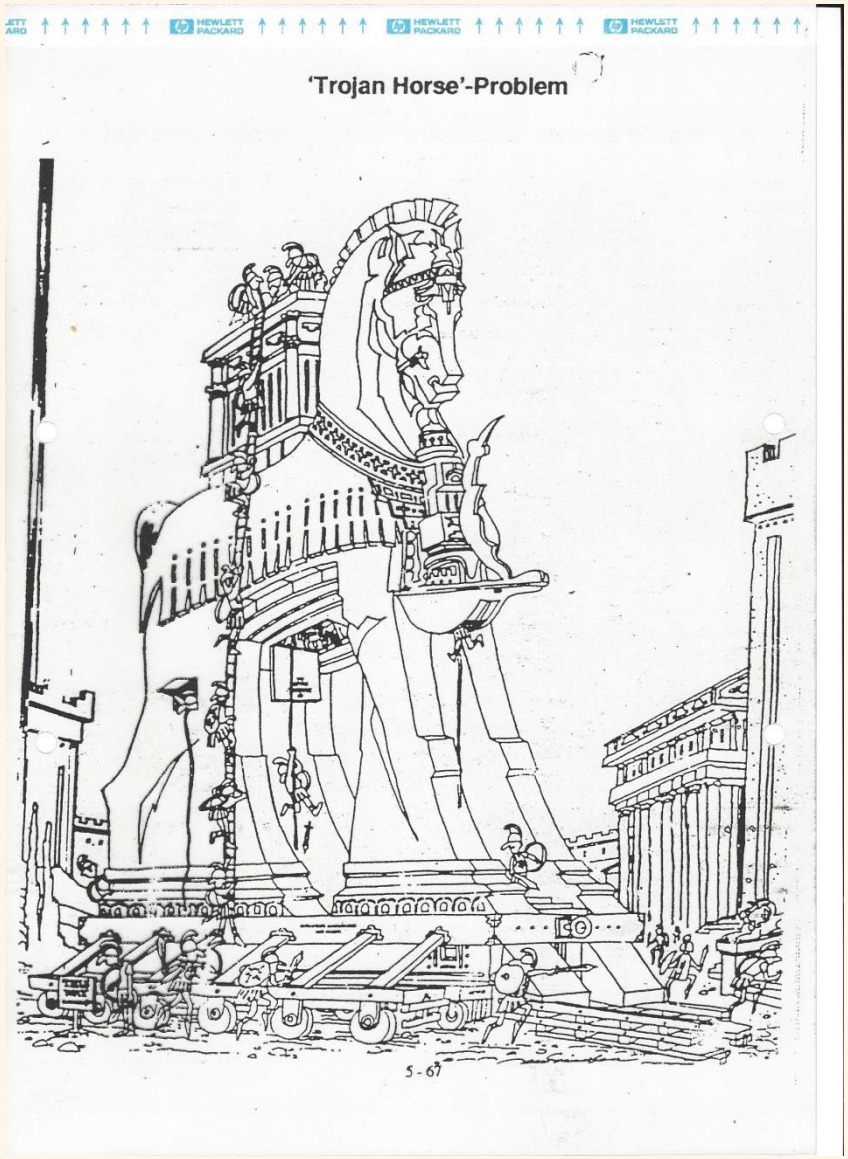
Mandatory Access Control

- Beruht auf systemweiter Policy, die nicht durch einzelne Benutzer verändert werden kann
  - Jedes DB-Objekt wird einer Security-Klasse zugewiesen
  - Jedes Subjekt (User oder User Program) wird einer Clearance für eine bestimmte Security-Klasse zugewiesen
  - Regeln, die auf Security-Klassen und Clearances beruhen, steuern die Zulässigkeit von Read/Write-Operationen auf den Objekten
- Die meisten kommerziellen Systeme unterstützen Mandatory Access Control nicht (mit Ausnahme einiger Versionen)
- Einsatz von Mandatory Access Control:
  - Dort, wo erhöhte Sicherheitsanforderungen existieren, z.B. militärische Anwendungen

Warum Mandatory Access Control?

- Dezentrale Vergabe von Zugriffsrechten hat einige Schwachstellen, z.B. Einbau *"Trojanischer Pferde"* ins System möglich:
- Beispiel-Szenario "Trojanisches Pferd"
  - Eindringling Dick erzeugt die Tabelle Horsie und gibt INSERT-Rechte an Justin (der das nicht weiß)
  - Dick modifiziert den Code eines Applikationsprogramms, das von Justin genutzt wird, um zusätzlich einige geheime Daten in die Tabelle Horsie zu schreiben
  - Jetzt kann Justin die geheimen Informationen sehen
  - Anschließend Modifikation des Anwendungsprogramms wieder beseitigen
- Die Modifikation des Codes ist außerhalb der Kontrolle des DBMS, aber es kann versuchen, den Gebrauch der Datenbank als Kanal für geheime Informationen zu verhindern

Das Trojanische Pferd



Bell-La Padula Modell

- Objekte (z.B. Tabellen, Views, Tupel)
- Subjekte (z.B. User, User Programs)
- Security-Klassen:
  - Top Secret (TS), Secret (S), Confidential (C), Unclassified (U): TS > S> C > U
- Jedes Objekt und jedes Subjekt wird einer dieser Klassen zugewiesen.
  - Subject S kann Objekt O lesen nur wenn gilt:
    class(S) >= class(O) (Simple Security Property)

    Beispiel: Ein Benutzer mit TS-Clearance darf ein S-klassifiziertes Objekt lesen. Ein Benutzer mit C-Clearance darf nicht ein TS-klassifiziertes Objekt lesen
  - Subject S kann Objekt O schreiben nur wenn gilt:
    class(S) <= class(O) (*-Property)

    Beispiel: Ein Benutzer mit S-Clearance kann nur Objekte schreiben, die mit S oder TS klassifiziert sind.

Grundidee zentraler Zugriffskontrolle

- Idee:
  - Garantiere, dass Informationen niemals von einer höheren zu einer niedrigeren Sicherheitsstufe fließen
- Beispiel:
  Dick hat Security-Klasse C, Justin hat Klasse S, und die geheime Tabelle hat Klasse S:
  - Dicks Tabelle, Horsie, hat ebenfalls Dicks Clearance, C.
  - Justins Anwendungsprogramm hat seine Clearance, S.
  - Somit kann das Programm nicht in die Tabelle Horsie schreiben.
- Mandatory Access Control Regeln werden zusätzlich angewandt zu den Zugriffsrechten, die von den Eigentümern der Objekte bereits definiert wurden

Multilevel-Relationen

bid	bname	color	class
101	Salsa	Red	S
102	Pinto	Brown	C

- Benutzer mit S und TS clearance kann beide Tupel sehen; ein C-Benutzer sieht nur das 2. Tupel; ein U-Benutzer sieht gar nichts!
- Wenn ein C-Benutzer versucht, ein Tupel einzufügen <101,Pasta,Blue,C>:
  - Erfolgreiches Einfügen verletzt Key Constraint
    Schlüssel muß eindeutig sein!
  - Verbieten der Einfügung gibt dem Benutzer die Information, daß es bereits ein anderes Objekt mit Schlüssel 101 gibt, das eine Klasse > C hat!
  - Lösung des Problems: Behandle Spalte *class* als Teil des Schlüssels

Verschlüsselung

- Definition: Codierung von Daten durch einen speziellen Algorithmus, der sie in eine Form überführt, die kein Programm lesen kann, das den Schlüssel nicht kennt.
- Verschlüsselung wird von vielen DBMS angeboten
- Codierung reversibel (meistens) oder irreversibel (für statistische Auswertungen)
- Verschlüsselungssystem
  - Verschlüsselungsschlüssel + Verschlüsselungsalgorithmus: Klartext → Schlüsseltext
  - Entschlüsselungsschlüssel + Entschlüsselungsalgorithmus: Schlüsseltext → Klartext
- Symmetrisch: gleicher Schlüssel für Ver- und Entschlüsselung
  - DES (Data Encryption Standard) von IBM („schwach“)
  - PGP (Pretty Good Privacy), 128-Bit Schlüssel („stark“)
- Asymmetrisch: unterschiedl. Schlüssel für Ver- und Entschlüsselung
  - Public Key Verfahren: ein öffentlicher und ein geheimer Schlüssel
  - Anwendung: digitale Unterschrift
  - bekanntestes Verfahren: RSA

Sicherheitsprobleme in statistischen DB

- Statistische Datenbank (z.B. Data Warehouse / Data Mart)
  - enthält personenbezogene Daten
  - erlaubt keinen Zugriff auf personenbezogene Daten (z.B. “das Alter von Joe“)
  - erlaubt Verwendung von statistischen Funktionen wie AVG, MIN, MAX, COUNT (z.B. “ermittle das Durchschnittsalter“)
- Neues Problem: Einzelwerte sind ableitbar bei
  - selektiven Anfragen (kleine Treffermengen)
  - Ergebnisverknüpfung mehrerer Anfragen
- Beispiel:
  - Statistische Anfragen auf PERS ohne Attribute PNR und NAME
  - Wissen über bestimmte Personen (z.B. Alter, Beruf Familienstand) kann leicht für gezielte Anfragen genutzt werden:

    SELECT COUNT(*), AVG(GEHALT)

    FROM PERS

    WHERE ALTER=51 AND BERUF=‘Programmierer‘ (dies ist Hugo!)

    Ermittlung des Gehalts möglich (wenn nur ein Treffer)

    bei mehr als einem Treffer kann Treffermenge durch weitere Bedingungen reduziert werden

Abhilfemöglichkeiten

- Idee: Antwortausgabe nur, wenn Treffermenge über festgelegtem Grenzwert N liegt
  - Selbst das kann aber umgangen werden, zum Beispiel:
  - Anfrage stellen "Wie viele Programmierer über X", solange bis das System die Antwort verweigert (da Ergebnismenge < N). Somit erhält man N Programmierer einschließlich Hugo
  - Nächster Schritt: "Wie hoch ist die Gehaltssumme dieser Programmierer älter als X?". Das Resultat sei S1.
  - Nächster Schritt: "Wie hoch ist die Gehaltssumme der Programmierer älter als X außer Hugo plus mein Gehalt". Das Resultat sei S2.
  - Gesamtergebnis: S1-S2+MeinGehalt = Gehalt von Hugo
- Überprüfung, ob mehrere Anfragen aufeinander aufbauen
- Gezielte Einstreuung von kleineren Ungenauigkeiten

Zusammenfassung Sicherheit

- Datensicherheit (Security) = Schutz der Datenbank gegen Bedrohungen
- Datenschutz = Schutz personenbezogener Daten entsprechend gesetzlicher Vorschriften (organisatorisch / technisch)
- DB-Administrator insgesamt für Security verantwortlich:
  - Entwirft Security Policy
  - Bewirtschaftet Audit Trail bzw. die Historie aller DB-Zugriffe durch die Benutzer
- Zwei Hauptansätze für Zugriffskontrolle in einem DBMS:
  - Discretionary Control: beruht auf dem Konzept der Zugriffsrechte (muß mit beim Entwurf definiert werden)
  - Mandatory Control: beruht auf dem Konzept der Security-Klassen
- Statistische Datenbanken (Inferenzkontrolle)
  - versuchen den Schutz persönlicher Daten durch Unterstützung von aggregierten Anfragen, aber weisen oft Sicherheitslücken auf, so daß persönliche Daten leicht abgeleitet werden können