	Lehrveranstaltung	Grundlagen von Datenbanken			WS 2012/13
	Aufgabenzettel	3 (Lösungsvorschläge)			
	Gesamtpunktzahl	40			
	Ausgabe	Fr. 09.11.2012	Abgabe	Do. 22.11.2012	


Aufgabe 1: Konzeptioneller Entwurf

(12 Punkte)

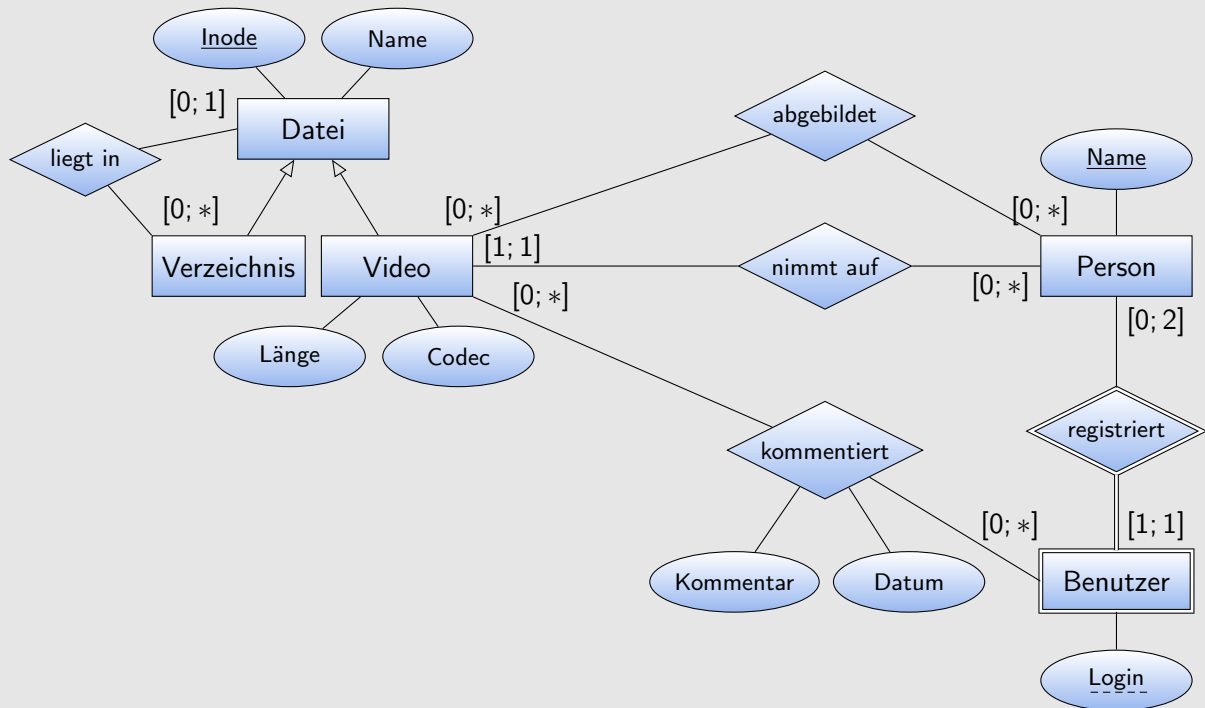
Erfassen Sie die im Folgenden beschriebenen Informationsstrukturen (resultierend aus einer Anforderungsanalyse für eine Anwendung zur Verwaltung und Betrachtung von Video-Dateien) in einem ER-Diagramm. Beziehen Sie sich dabei genau auf die gegebene Beschreibung, ohne weiteres Wissen zu möglicherweise ähnlichen Anwendungsbereichen einfließen zu lassen. Markieren Sie in Ihrem Entwurf Primärschlüssel durch Unterstreichung und notieren Sie die Abbildungstypen in der Form [min;max]. Verwenden Sie unbedingt die aus der Vorlesung (oder Übung) bekannte Notation. Benutzen Sie möglichst wenige Entity-Typen (Ausnahme: Vererbung).


Alle Dateien werden in einer hierarchischen Struktur verwaltet, wobei Dateien entweder Verzeichnisse oder Videos sind und stets einen Namen und eine eindeutige Inode-Nummer besitzen. Jede Datei liegt in genau einem Verzeichnis, das beliebig viele Dateien aufnehmen kann. Lediglich das Hauptverzeichnis liegt in keinem weiteren Verzeichnis. Videos haben eine bestimmte Länge und ein bestimmtes Codec. Weiterhin sollen Personen verwaltet werden, die durch ihren Namen eindeutig identifizierbar sind. Personen können auf beliebig vielen Videos abgebildet sein und Videos können beliebig viele Personen abbilden. Personen können beliebig viele Videos aufnehmen, aber immer nur genau eine Person ist der Urheber eines Videos.

Jede Person kann bis zu 2 Benutzerkonten registrieren, aber jedes Benutzerkonto ist eindeutig über Login und Name der zugehörigen Person identifizierbar. Benutzer können beliebig viele Videos kommentieren. Videos können von beliebig vielen Nutzern kommentiert werden. Hierbei wird der Kommentartext selbst und das Datum gespeichert, an dem er abgegeben wurde.

	Lehrveranstaltung	Grundlagen von Datenbanken			WS 2012/13
	Aufgabenzettel	3 (Lösungsvorschläge)			
	Gesamtpunktzahl	40			
	Ausgabe	Fr. 09.11.2012	Abgabe	Do. 22.11.2012	

Lösungsvorschlag:



	Lehrveranstaltung	Grundlagen von Datenbanken		WS 2012/13
	Aufgabenzettel	3 (Lösungsvorschläge)		
	Gesamtpunktzahl	40		
	Ausgabe	Fr. 09.11.2012	Abgabe	Do. 22.11.2012

Aufgabe 2: Logischer Entwurf

(12 Punkte)

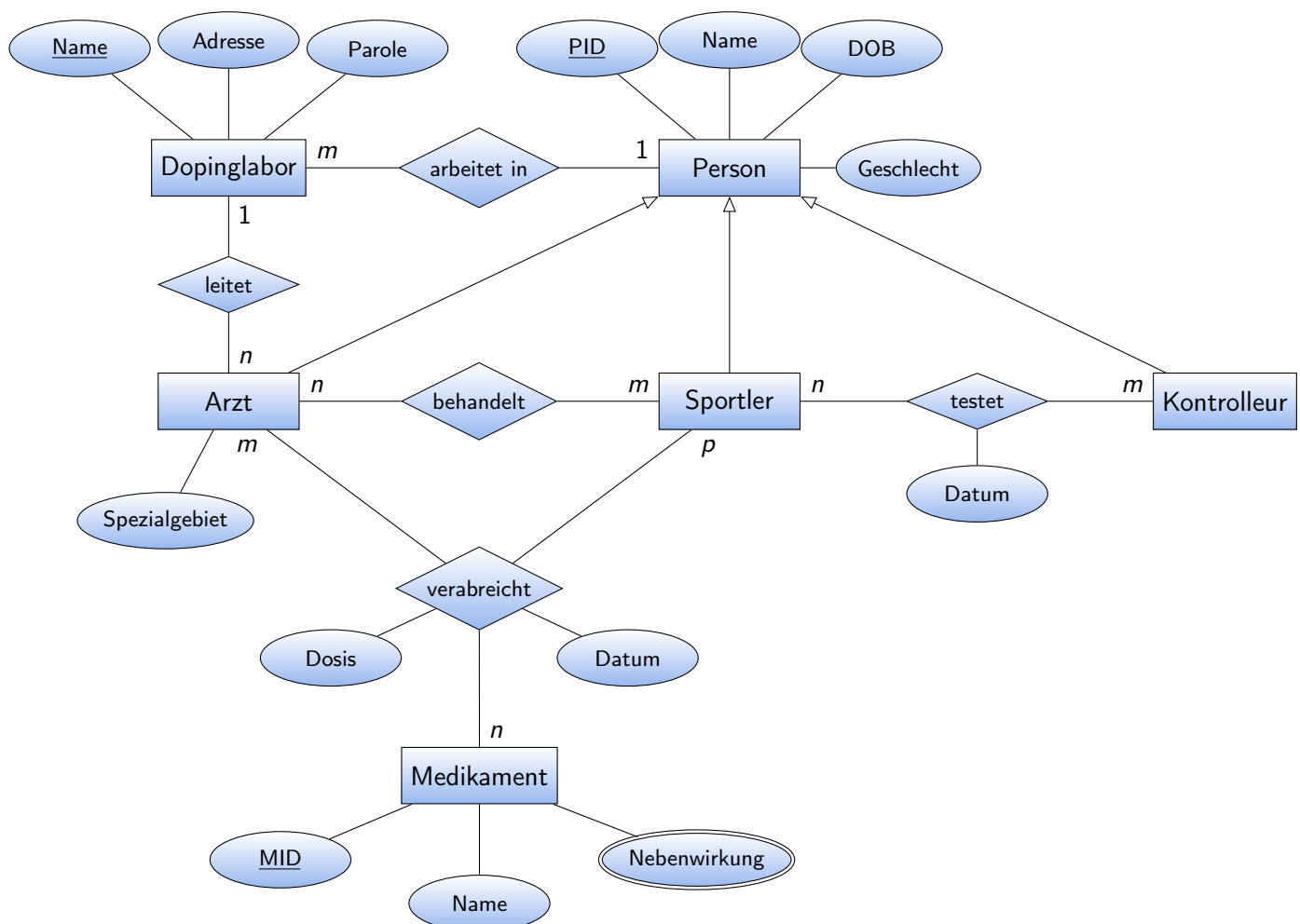
Vollziehen Sie einen logischen Entwurf und transformieren Sie das unten abgebildete ER-Diagramm in ein relationales Datenbankschema. Versuchen Sie das Datenbankschema dabei möglichst minimal zu halten, ohne dadurch Informationseinbußen hinnehmen zu müssen. Verwenden Sie bitte folgende Schreibweise für eine Relation:


Relationenname(Attributname₁, Attributname₂, ..., Attributname_n)

Die Attribute eines Primärschlüssels sind mit einer einzigen durchgezogenen Linie zu unterstreichen. Das Gleiche gilt für zusammengesetzte Fremdschlüssel, in diesem Fall ist jedoch eine gestrichelte Linie zu verwenden. Die Referenzen der Fremdschlüssel sollen wie folgt dargestellt werden:

Attributname_i → Relationenname_j.Attributname_k

Zusätzlich zu den im ER-Diagramm dargestellten Informationen sollen folgende Integritätsbedingungen gelten: „Jede Verabreichung eines Medikaments und jeder Test eines Sportlers ist eindeutig für das jeweilige Datum.“



	Lehrveranstaltung	Grundlagen von Datenbanken			WS 2012/13
	Aufgabenzettel	3 (Lösungsvorschläge)			
	Gesamtpunktzahl	40			
	Ausgabe	Fr. 09.11.2012	Abgabe	Do. 22.11.2012	

Lösungsvorschlag:

Person(PID, Name, DOB, Geschlecht, Arbeitsplatz → Dopinglabor.Name)

Arzt(PID → Person.PID, Spezialgebiet)

Sportler(PID → Person.PID)

Kontrolleur(PID → Person.PID)

Dopinglabor(Name, Adresse, Parole, Leiter → Arzt.PID)

Medikament(MID, Name)

Nebenwirkungen(Medikament → Medikament.MID, Nebenwirkung)

behandelt(Arzt → Arzt.PID, Sportler → Sportler.PID)

testet(Kontrolleur → Kontrolleur.PID, Sportler → Sportler.PID, Datum)

verabreicht(Arzt → Arzt.PID, Medikament → Medikament.MID, Sportler → Sportler.PID, Datum, Dosis)

Aufgabe 3: Relationale Algebra

(12 Punkte)

Betrachten Sie das folgende relationale Datenbankschema.

Personal	<u>PID</u>	Vorname	Nachname	Geburt	Wohnort	<u>Abteilung</u>
	4	Peter	Müller	1962-07-25	Hamburg	2
	8	Bianca	Lohse	1982-01-13	Kiel	4
	11	Murat	Sahir	1990-03-16	Hamburg	2
	21	Frank	Siebenstein	1975-12-02	Norderstedt	1
	22	Bernd	Schmidt	1973-11-26	Norderstedt	1
	24	Ulrike	Müller	1963-10-07	Hamburg	2
	31	Jochen	Fuhrmann	1958-05-09	Stade	2

Abteilung → Abteilungen.AID

Abteilungen	<u>AID</u>	Name	<u>Leiter</u>
	1	Controlling	21
	4	Marketing	8
	2	Einkauf	4


Leiter → Personal.PID

ProjektArbeiter	<u>PrID</u>	<u>PID</u>
	15	21
	15	22
	36	8
	36	11
	36	31

PrID → Projekte.PrID, PID → Personal.PID

Projekte	<u>PrID</u>	Name	<u>Leiter</u>	Budget
	15	Prozessoptimierung	22	10.000
	36	B.L.I.C.K.F.A.N.G	8	7.500

Leiter → Personal.PID

	Lehrveranstaltung	Grundlagen von Datenbanken			WS 2012/13
	Aufgabenzettel	3 (Lösungsvorschläge)			
	Gesamtpunktzahl	40			
	Ausgabe	Fr. 09.11.2012	Abgabe	Do. 22.11.2012	

- a) Überprüfen Sie die Korrektheit der folgenden Ausdrücke der relationalen Algebra. Prüfen Sie dabei sowohl die Korrektheit der Syntax als auch der Semantik, d. h. ihre Gültigkeit in Bezug auf das vorliegende Datenbankschema. Begründen Sie jeweils Ihre Antwort. (4 Punkte)

i) $\sigma_{Geburt="1962-07-25"}(Abteilungen \bowtie_{Leiter=PID} (\pi_{PID, Vorname, Nachname}(Personal)))$

Lösungsvorschlag:

Die Selektion anhand der Spalte *Geburt* ist nicht möglich. Diese ist zwar in der Relation *Personal* vorhanden, wurde aber durch die Projektion auf die Spalten *Vorname* und *Nachname* ausgeblendet.

ii) $\pi_{Vorname}(Personal) \cup \pi_{Geburt}(Personal)$

Lösungsvorschlag:

Die Vereinigung ist nicht möglich, da die beiden zu vereinigenden Projektionen nicht vereinbar sind (unterschiedliche Attribute).

iii) $\sigma_{Budget < 1000 \wedge Budget > 5000}(Projekte)$

Lösungsvorschlag:

Dieser Ausdruck liefert zwar in jedem Fall ein leeres Resultat, ist aber dennoch korrekt.

iv) $\pi_{Name}(Abteilungen) \bowtie_{PID=Leiter} \pi_{Vorname, Nachname}(Personal)$

Lösungsvorschlag:

Der Verbund ist nicht möglich, da die Verbundsattribute durch die vorhergehenden Projektionen ausgeblendet wurden.


- b) Übersetzen Sie die folgenden umgangssprachlich formulierten Anfragen in einen zugehörigen Ausdruck der relationalen Algebra, werten Sie die Ausdrücke aus und geben Sie jeweils die Ergebnisrelation an. (4 Punkte)

- i) Namen von Projekten, in denen junge Mitarbeiter (Geboren in oder nach 1990) mitarbeiten

Lösungsvorschlag:

$$\pi_{Name}(\sigma_{Geburt \geq "1990-01-01"}(Personal) \bowtie ProjektArbeiter \bowtie Projekte)$$

$$= \{(B.L.I.C.K.F.A.N.G.)\}$$

	Lehrveranstaltung	Grundlagen von Datenbanken			WS 2012/13
	Aufgabenzettel	3 (Lösungsvorschläge)			
	Gesamtpunktzahl	40			
	Ausgabe	Fr. 09.11.2012	Abgabe	Do. 22.11.2012	

- ii) Persönliche Daten (Vor- und Nachname, Geburtsdatum) aller an dem Projekt *Prozessoptimierung* beteiligten Mitarbeiter.

Lösungsvorschlag:

$\pi_{Vorname, Nachname, Geburt}(\sigma_{Name = \text{"Prozessoptimierung"}}(Projekte) \bowtie ProjektArbeiter \bowtie Personal)$
 $= \{(Frank, Siebenstein, 1975-12-02), (Bernd, Schmidt, 1973-11-26)\}$

- iii) Mitarbeiter, die an keinem Projekt mitarbeiten.

Lösungsvorschlag:

$Personal - \pi_{PID, Vorname, Nachname, Geburt, Wohnort, Abteilung}(Personal \bowtie ProjektArbeiter)$
 $= \{(4, Peter, Müller, 1962-07-25, 2), (24, Ulrike, Müller, 1963-10-07, 2)\}$

- iv) Vor- und Nachnamen der Abteilungskollegen von Jochen Fuhrmann (PID=31).

Lösungsvorschlag:

$\pi_{Vorname, Nachname}(\sigma_{PID \neq 31}(Personal \bowtie \pi_{Abteilung}(\sigma_{PID=31}(Personal))))$
 $= \{(Peter, Müller), (Murat, Sahir), (Ulrike, Müller)\}$

- c) Interpretieren Sie die folgenden relationalen Ausdrücke, indem Sie eine umgangssprachliche Beschreibung sowie die Ergebnisrelation angeben. (4 Punkte)

- i) $\pi_{Vorname, Nachname}(\sigma_{Budget > 8000}(Projekte) \underset{Leiter=PID}{\bowtie} Personal)$


Lösungsvorschlag:

Vor- und Nachnamen der Projektleiter, die über ein Budget von mehr als 8.000 EUR verfügen.
 $= \{(Bernd, Schmidt)\}$

- ii) $\pi_{PID, Vorname, Nachname, Geburt, Wohnort, Abteilung}(Personal \underset{PID=Leiter}{\bowtie} Projekte)$
 $\cap \pi_{PID, Vorname, Nachname, Geburt, Wohnort, Abteilung}(Personal \underset{PID=Leiter}{\bowtie} Abteilungen)$

Lösungsvorschlag:

Mitarbeiter, die gleichzeitig Projektleiter und Abteilungsleiter sind.
 $= \{(8, Bianca, Lohse, 1982-01-13, Kiel, 4)\}$

	Lehrveranstaltung	Grundlagen von Datenbanken			WS 2012/13
	Aufgabenzettel	3 (Lösungsvorschläge)			
	Gesamtpunktzahl	40			
	Ausgabe	Fr. 09.11.2012	Abgabe	Do. 22.11.2012	

iii) $\pi_{Name}(\pi_{Abteilung}(\sigma_{Name="B.L.I.C.K.F.A.N.G."}((Projekte) \bowtie ProjektArbeiter \bowtie Personal) \bowtie_{Abteilung=AID} Abteilungen)$

Lösungsvorschlag:

Die Namen aller an dem Projekt *B.L.I.C.K.F.A.N.G.* beteiligten Abteilungen.
 $= \{(Marketing), (Einkauf)\}$

iv) $\pi_{PrID}(\sigma_{Name="Controlling"}((Abteilungen) \bowtie_{AID=Abteilung} Personal \bowtie ProjektArbeiter) \bowtie Projekte)$

Lösungsvorschlag:

Projekt-ID, Name, Leiter-PID und Budget aller Projekte, an denen Mitarbeiter der Abteilung *Controlling* mitwirken.
 $= \{(15, Prozessoptimierung, 22, 10.000)\}$

Aufgabe 4: Algebraische Optimierung


(4 Punkte)

Betrachten Sie erneut das Datenbankschema aus Aufgabe 3. In der folgenden Aufgabe sind zwei relationale Ausdrücke angegeben. Beide Ausdrücke liefern dasselbe Ergebnis zurück und sind daher semantisch äquivalent, unterscheiden sich jedoch in ihrem Optimierungsgrad. Zeichnen Sie zu jedem relationalen Ausdruck einen Operatorbaum und bestimmen Sie, welcher der zwei Operatorbäume den höchsten Optimierungsgrad besitzt. Begründen Sie Ihre Entscheidung mit Hilfe der in der Vorlesung behandelten Optimierungsheuristiken I-VII (Folien 62-71, Kapitel 4.).

Für die zugehörige Datenbank werden in dieser Aufgabe jedoch folgende Kardinalitäten angenommen:

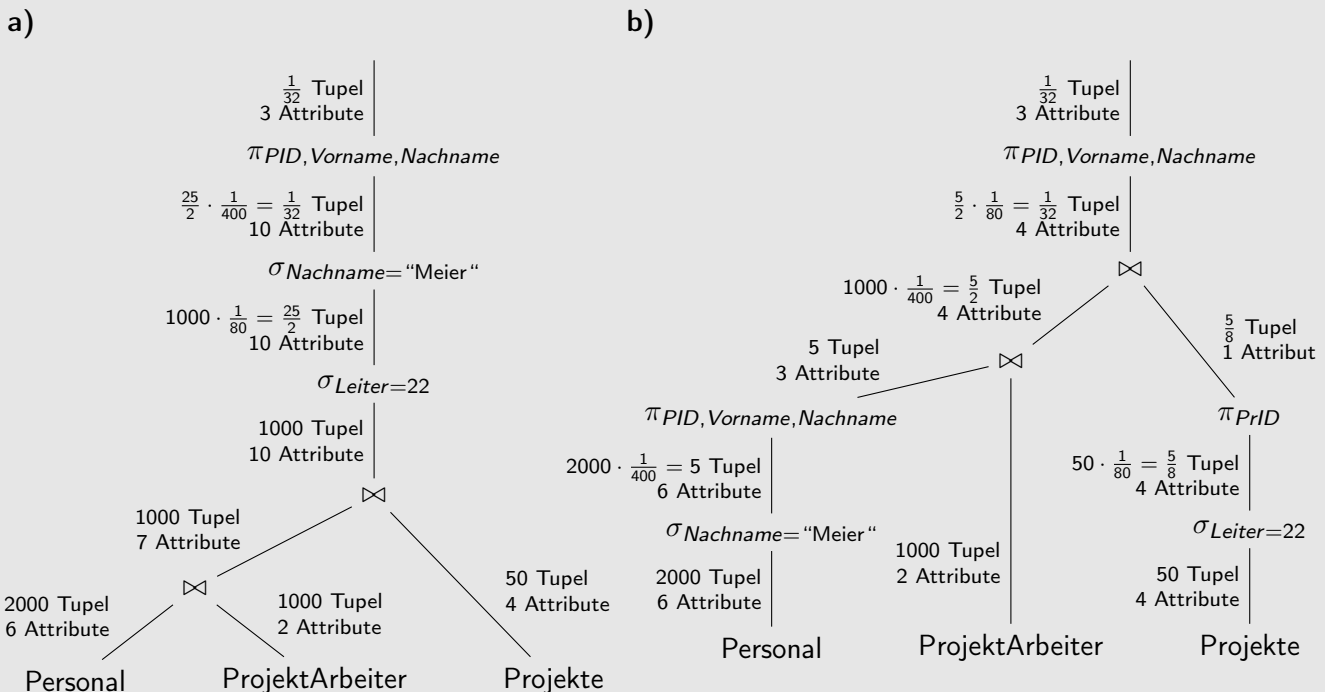
$Card(Personal) = 2000$, $Card(Abteilungen) = 40$, $Card(Projekte) = 50$, $Card(ProjektArbeiter) = 1000$

Es gibt 50 verschiedene Werte für das Attribut Wohnort und je 400 verschiedene Vor- und Nachnamen. Jeder 25. Personalangestellte ist ein Projektleiter. Die Namen von Abteilungen sind eindeutig.

	Lehrveranstaltung	Grundlagen von Datenbanken			WS 2012/13
	Aufgabenzettel	3 (Lösungsvorschläge)			
	Gesamtpunktzahl	40			
	Ausgabe	Fr. 09.11.2012	Abgabe	Do. 22.11.2012	

- a) $\pi_{PID, Vorname, Nachname}(\sigma_{Nachname=Meier}(\sigma_{Leiter=22}((Personal \bowtie ProjektArbeiter) \bowtie Projekte)))$
- b) $\pi_{PID, Vorname, Nachname}(((\pi_{PID, Vorname, Nachname}(\sigma_{Nachname="Meier"}(Personal)) \bowtie ProjektArbeiter) \bowtie (\pi_{PrID}(\sigma_{Leiter=22}Projekte)))$

Lösungsvorschlag:



Der zweite Operatorbaum besitzt den höchsten Optimierungsgrad, da auch hier Projektionen und Selektionen so früh wie möglich ausgeführt werden. Der Verbund zwischen den Relationen Personal und ProjektArbeiter führt zu einem Zwischenergebnis von etwa $\frac{5}{2}$ Tupeln mit 4 Attributen (vgl. 1000 Tupel mit 7 Attributen) und der nachfolgende Verbund mit der Projekte-Relation liefert $\frac{1}{32}$ Tupel mit 4 Attributen (vgl. 1000 Tupel mit 10 Attributen).