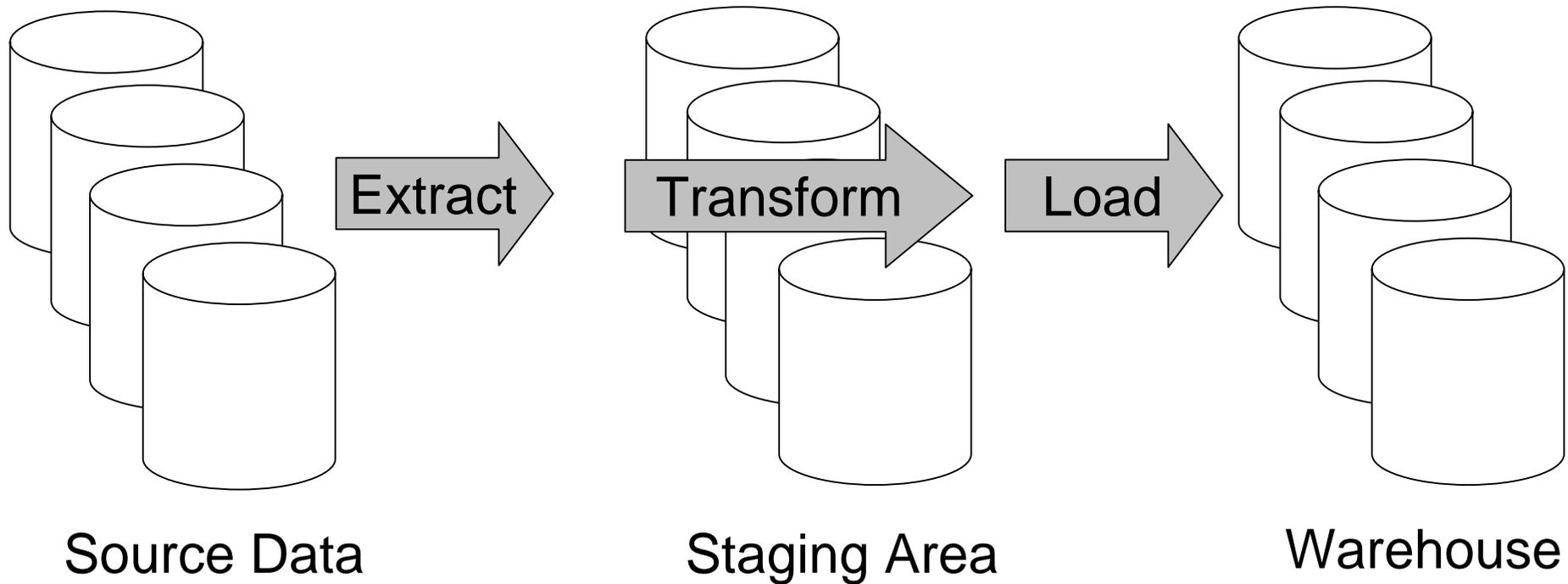


Übung
Data Warehouses

Wintersemester 2000/2001

Extraktion, Transformation, Laden

Datenflußparadigma



Source Data

- ◆ Einbinden von Datenquellen
 - ◆ Relationale Datenbanken (Oracle, SQL-Server...)
 - ◆ Flat Files
- ◆ Builder speichert Definitionen der Datenquellen als „Source Module“

Staging Area

- ◆ Bereich, in dem der Transformationsprozeß vor dem Laden in das Warehouse stattfindet
 - ◆ Bereinigung (cleaning)
 - ◆ Anpassung (editing)
 - ◆ Integration
 - ◆ Berechnung
- ◆ Builder speichert Definitionen und Mappings als „Warehouse Module“

Administration Mode

- ◆ Anlegen, Umbenennen und Löschen von Projekten
- ◆ Ändern der Eigenschaften von Projekten
- ◆ Verwalten von Integratoren (z. Zt. Nur RDBMS und Flat Files)

Project Mode

- ◆ Verwalten von Definitionen für
 - ◆ Data Warehouse Schemata
 - ◆ Datenquellen
 - ◆ Operationen zur Transformation
 - ◆ Mapping der Datenquellen über Transformatoren auf Zieltabellen
 - ◆ Generieren und Ausführen von Skripten

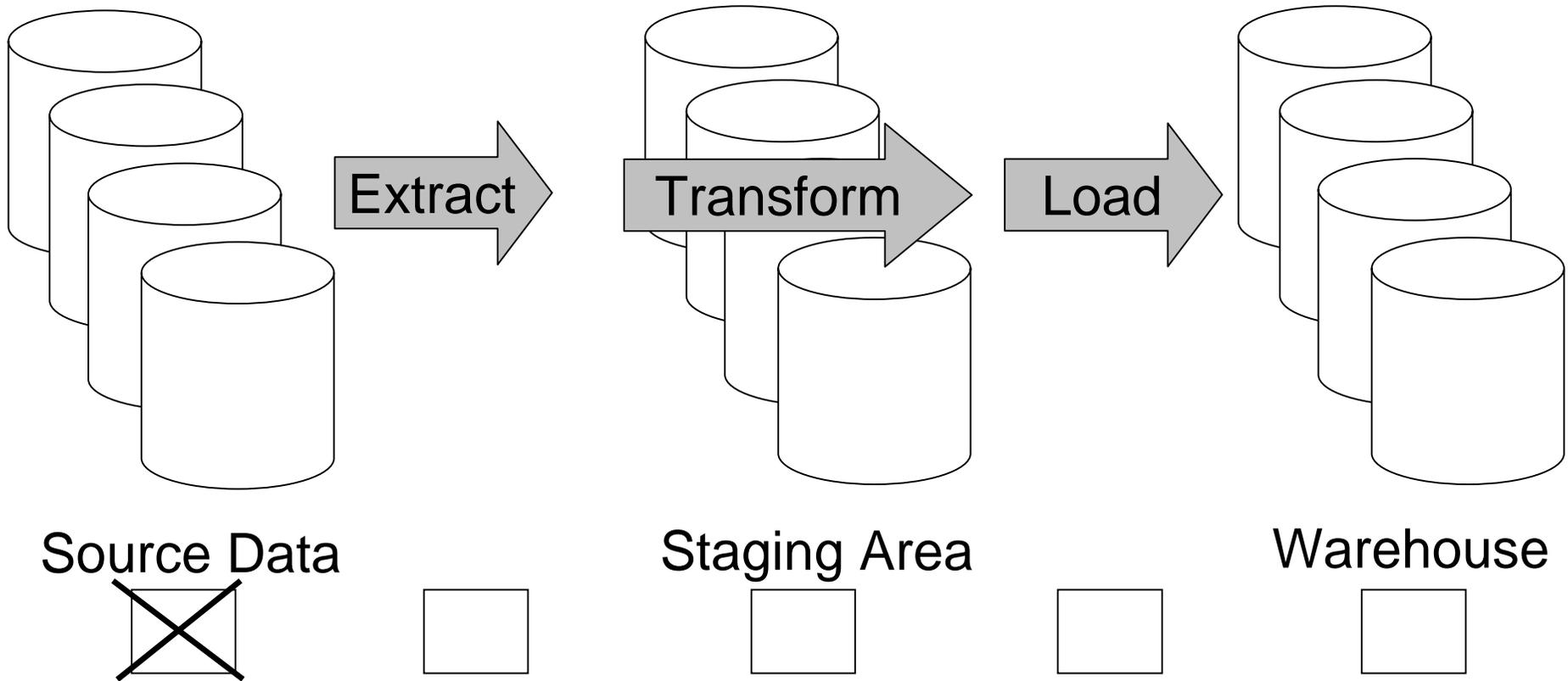
Transformation Mode

- ◆ Anlegen und Ändern von Transformationen
- ◆ Modifizieren der Eigenschaften von Source Modules oder Warehouse Modules

Data Source Module

- ◆ Modul für Oracle-Datenbanken anlegen
 - ◆ Datenbankverknüpfung anlegen
 - ◆ Tabellendefinitionen importieren
- ◆ Modul für Flat-Files anlegen
 - ◆ Verzeichnispfad angeben
 - ◆ Textdateien mit Metadaten Wizard einbinden
- ◆ Für eine graphische Darstellung:
(Source Module Toolbar → View → Schema)

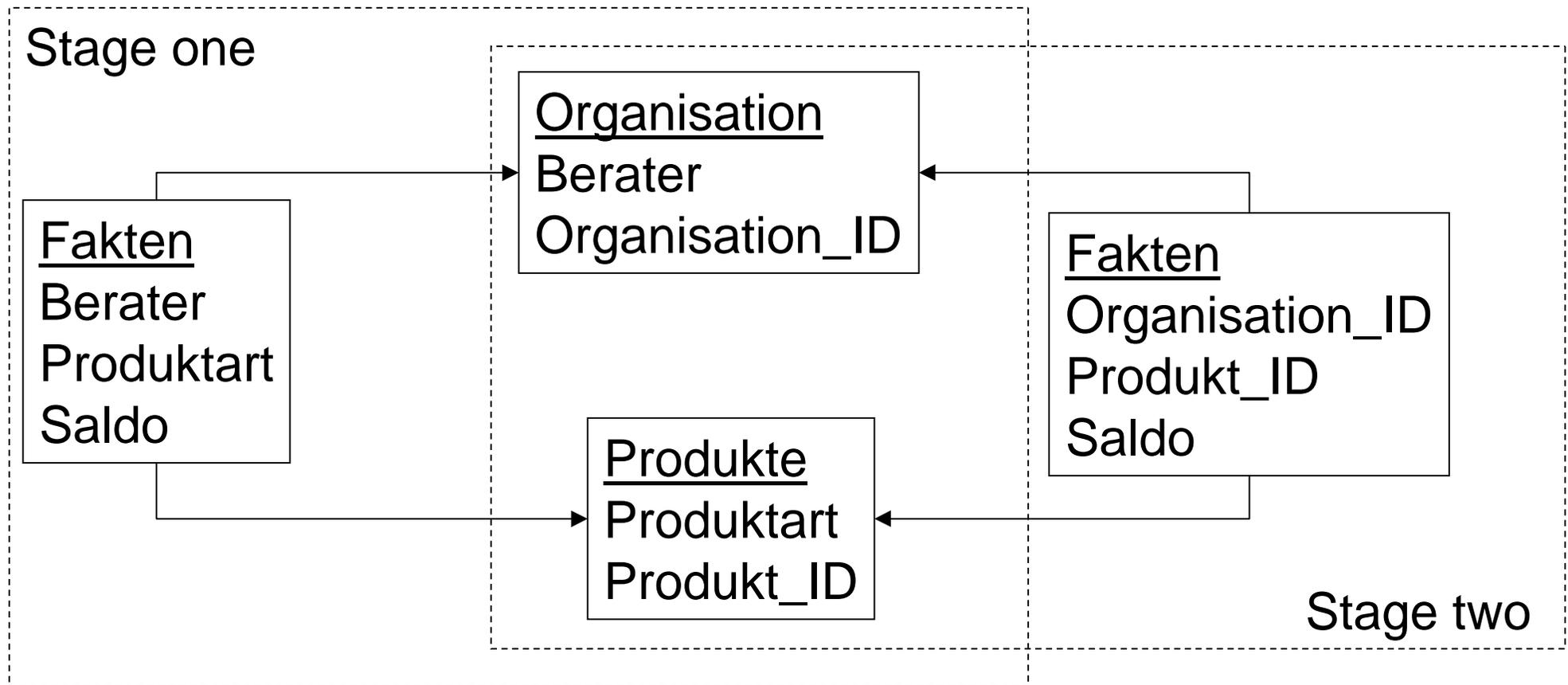
Entwicklungsstand



Schema der Staging Area

- ◆ Ähnelt oft den Datenquellen
- ◆ Erzeugung der künstlichen Schlüssel für die Dimensionen
 - ◆ Künstliche Schlüssel sind eindeutig (meist Sequenz)
 - ◆ Ursprünglicher Primärschlüssel bleibt erhalten
- ◆ Stage One: Faktentabelle mit Referenz auf die ursprünglichen Primärschlüssel
- ◆ Stage Two: Faktentabelle, erzeugt durch Star-Join bei Projektion auf die Kennzahlen und künstlichen Schlüssel

Beispiel einer Staging Area



Anlegen der Staging Area

- ◆ Staging Area ist ein Transformationsziel ⇒ Speicherung im Warehouse Module
- ◆ Möglichkeiten der Speicherung
 - ◆ Zieldatenbank (im eigenen oder allg. Schema)
 - ◆ Zwischendatenbank (bspw. aus Performancegründen lokal)
 - ◆ System einer der Datenquellen (unüblich)

Mapping von Datenquellen auf die Staging Area

- ◆ Auswählen des Warehouse Moduls der Staging Area
→ Mappings → Create Mapping
- ◆ Öffnen des Mappings, Hineinziehen der Datenquelle
(select from existing tables)
- ◆ Mit Shift-Drag in das Zielschema kopieren → Create-Like wizard

Fertigstellen des Mappings

- ◆ Künstl. Schlüssel (Warehouse Key) hinzufügen
 - ◆ Columns → Add (NUMBER)
 - ◆ Constraints → Add → Primary Key
- ◆ Vervollständigen der Tabellendefinition
 - ◆ Eindeutiger natürlicher Schlüssel (s.o., aber Unique Key)
 - ◆ Filtern von Spalten

Erzeugen des Warehouse Keys

- ◆ Bei relationalen Quellen
 - ◆ Create Sequence
 - ◆ Drop Secondary Source auf Mapping Definition
 - ◆ Drop NEXTVAL der Sequence auf Warehouse Key
- ◆ Bei Flat Files
 - ◆ Create Local Variable
 - ◆ Variablenname muß SEQ sein, Syntax
SEQUENCE(MAX, 1)

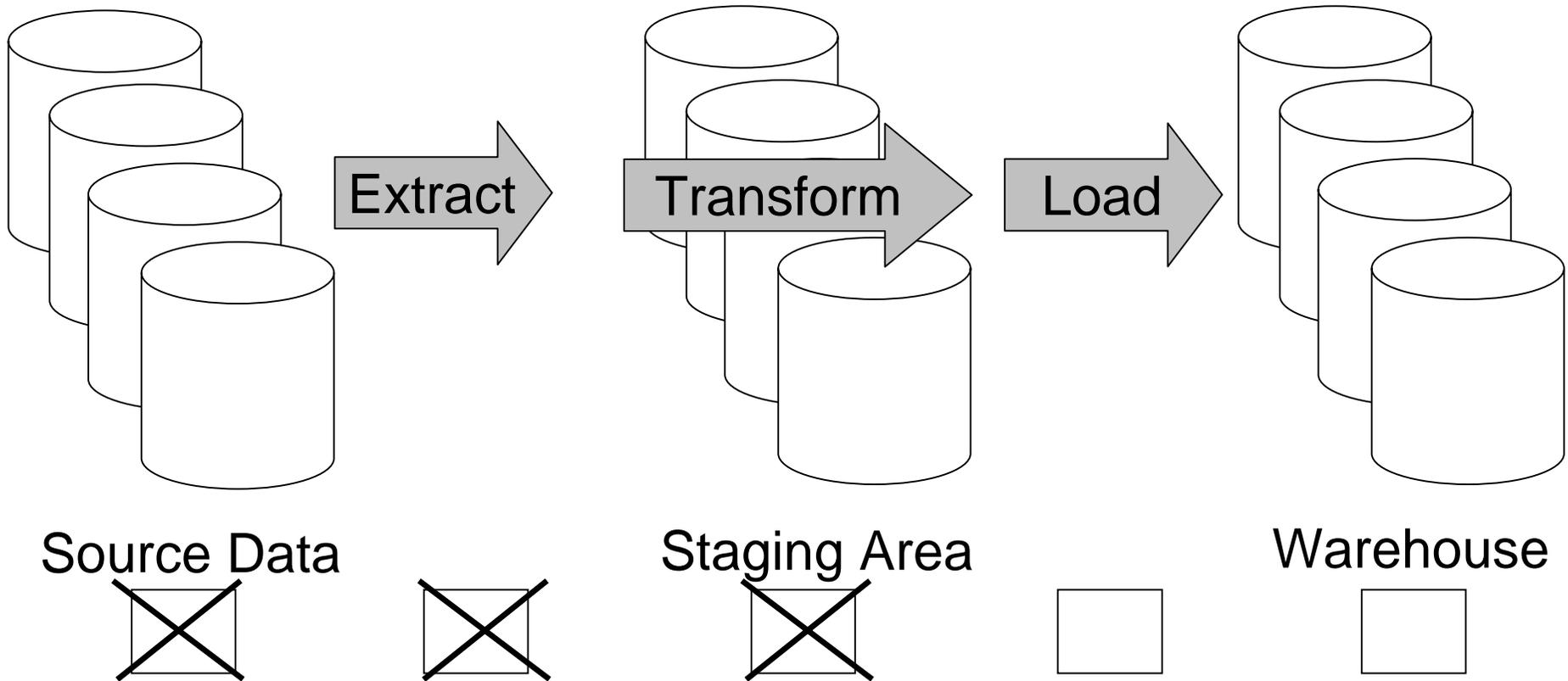
Erzeugen des Mappings für die Faktentabelle (Stage 1)

- ◆ Create-Like Wizard aufrufen und Tabelle mit Transaktionsdaten in die Staging Area kopieren
- ◆ Spalten für die natürlichen Schlüssel aller Dimensionstabellen der Staging Area anlegen
- ◆ Fremdschlüssel für die natürlichen Schlüssel der Dimensionen Eindeutigkeit für den eigenen definieren.
- ◆ Überflüssige Spalten entfernen
- ◆ Bei mehreren Quellrelationen über Secondary Source einbinden (Foreign Keys notwendig)

Erzeugen des Mappings für die Faktentabelle (Stage 2)

- Tables → Create Table (Faktentabelle Stage 2)
- Anfügen der Spalten
- Mapping anlegen
 - Keine Source Modules
 - Add Table (from existing tables) (Fakten 1 und 2)
 - Drop Fakten 1 auf Fakten 2 (rechte Maustaste)
 - Dimensionstabellen als Secondary Sources einbinden

Entwicklungsstand



Warehouse Module

- ◆ Neues Projekt anlegen und in dieses wechseln
- ◆ Neues Warehouse Module zur Abbildung des Warehouse-Modells anlegen und Eigenschaften einstellen
- ◆ Bestandteile definieren:
 - ◆ Dimensionen (Sonderfall: Zeitdimension)
 - ◆ Faktentabelle(n)
 - ◆ Später: (Materialized) Views

Entwicklungsstand

