



Information Management

Informix Warehouse

neue Möglichkeiten der Informationsanalyse



Michael Köster, IT Specialist
IBM Software Group – Information Management
Michael.Koester@de.ibm.com



Geschäftsabwicklung vs. Geschäftssteuerung?

Geschäftsportal

Geschäftsabwicklung

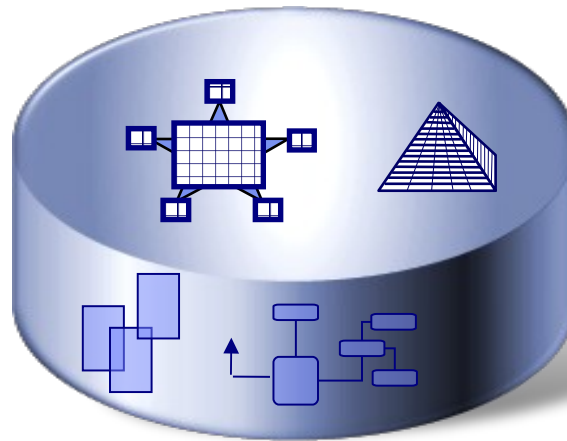
Geschäftssteuerung



Transformation



Datenqualität



Analysis



Reporting



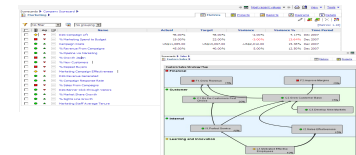
Data Mining



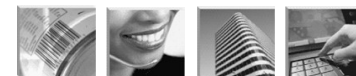
Business Insight



Dashboards



Scorecards, Planning



Operational Applications

Operative Systeme

Datenaufbereitung und -integration

Data Warehouse

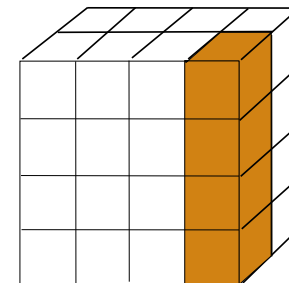
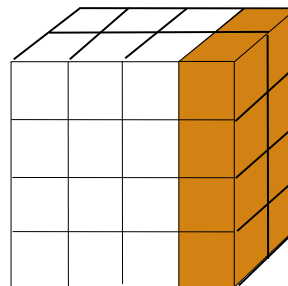
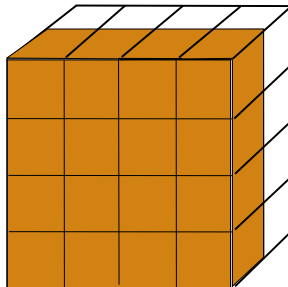
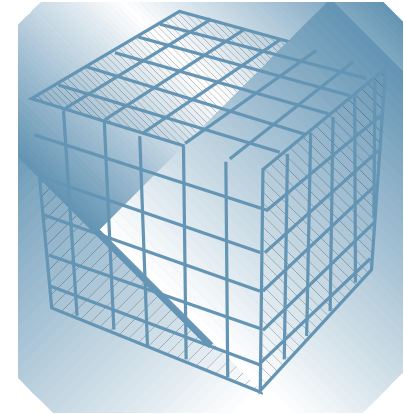
Analytische Systeme

Anforderungen/Herausforderungen an Data Warehouse Systeme

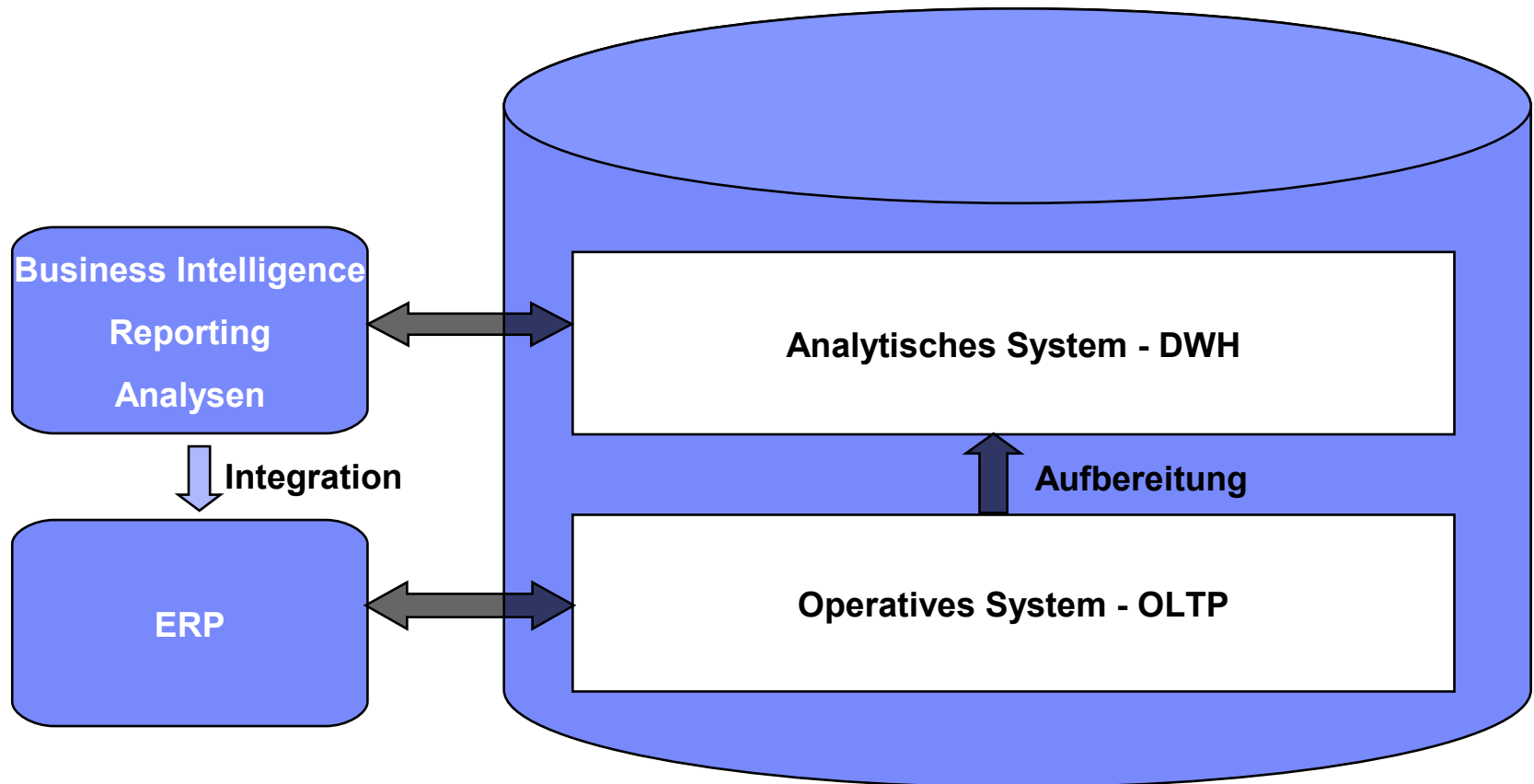
- Datenintegration
 - Integration der verschiedenen operativen Systeme häufig sehr komplex
- Verlässlichkeit der Daten
 - Sicherstellen der Datenqualität der analytischen Daten ist entscheidend für Vertrauen in die darauf basierenden Berichtssysteme
- Performance
 - Informationen müssen zur richtigen Zeit zur Verfügung stehen
- OLAP
 - Für einfache Informationsbereitstellung / Self-Service ist das Vorhandensein einer fachlichen Sicht auf die Daten notwendig

OLAP - Exkurs

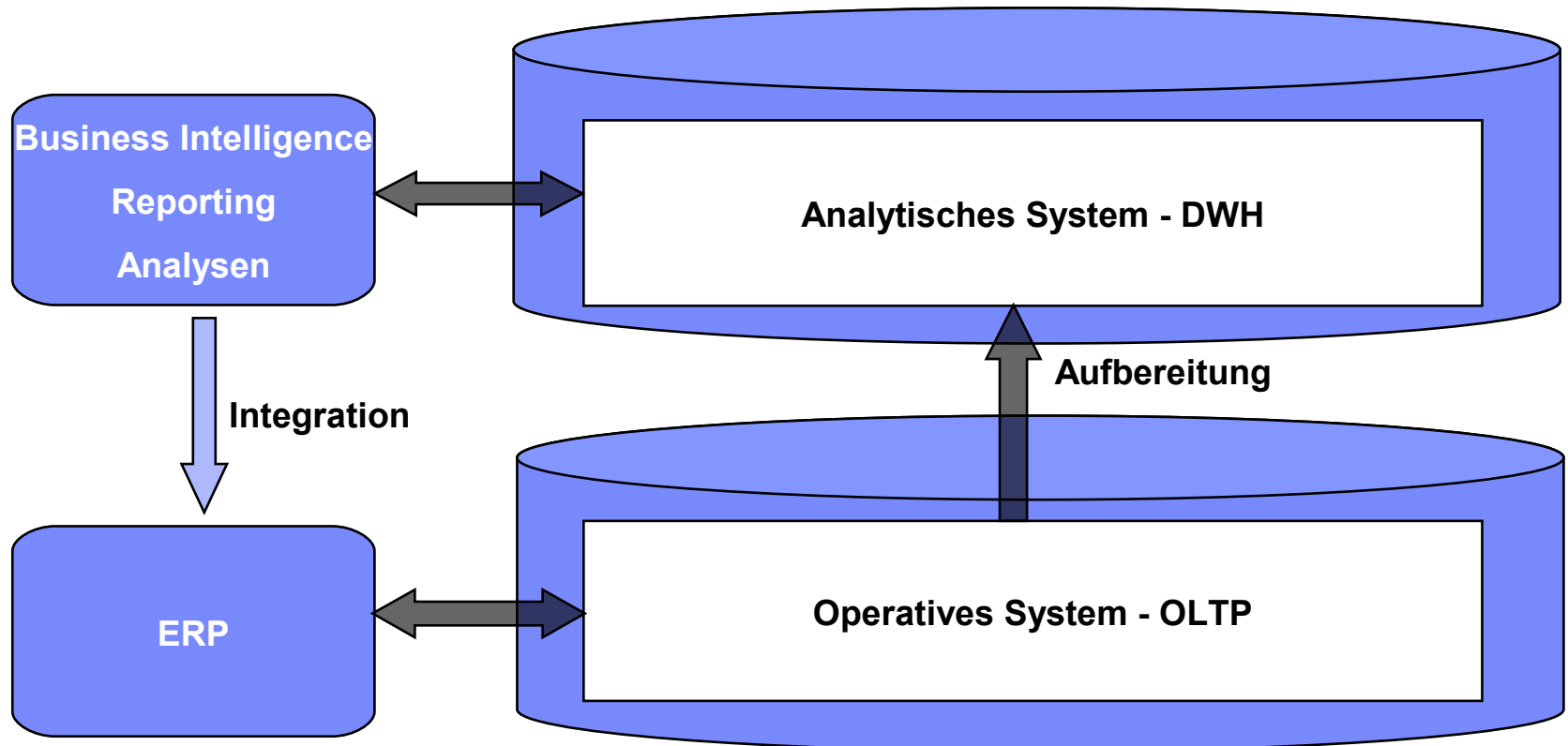
- Fachliche Sicht auf die Daten
- Verschiedene Ausprägungen:
 - ROLAP
 - Berechnung zur Abfragezeit -> Aktualität
 - Sehr flexibel
 - MOLAP
 - Vorberechnet
 - physisch mit eigener Struktur



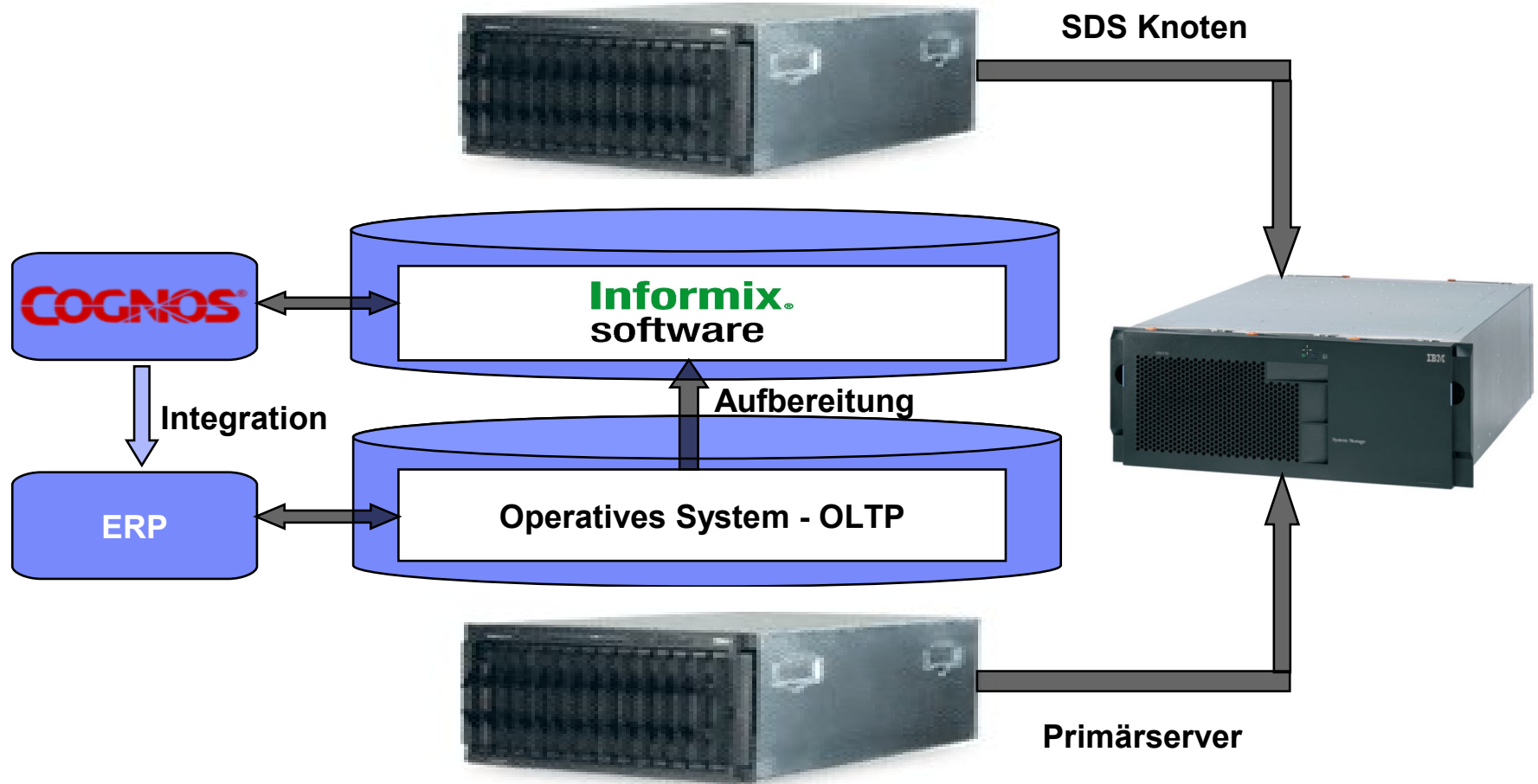
Architektur InfoSysteme



Architektur InfoSysteme



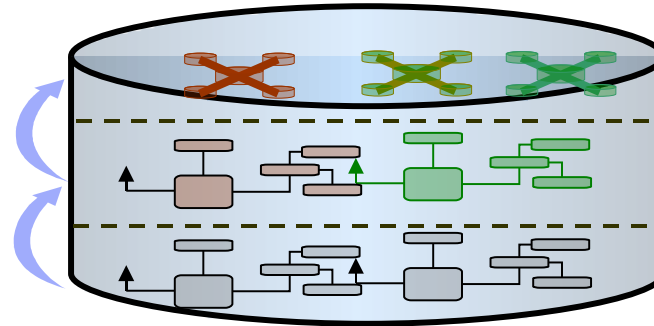
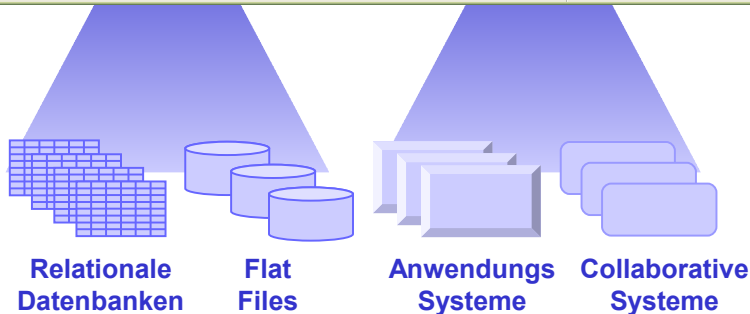
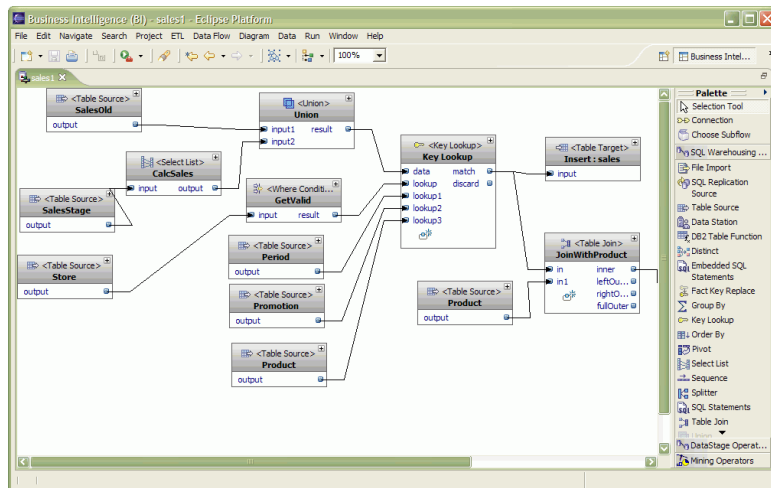
IBM Produkte zur Realisierung von Analytischen Systemen im Informix Umfeld



Informix Warehouse - Übersicht

- Ziel:
 - Warehouse Funktionalitäten für Informix Kunden zur Verfügung stellen, ohne Notwendigkeit einer weiteren Datenbank
- 1. Schritt:
 - SQL Warehousing Tool zur Vereinfachung der Datenintegration

SQL Warehousing Tool – integrierte Bewirtschaftung



- Einheitlicher Zugriff auf alle relevanten Datenquellen
- Einfache Vereinheitlichung der Informationen
- Nutzt die Skalierbarkeit der DB
- Nachvollziehbarkeit der Bewirtschaftung

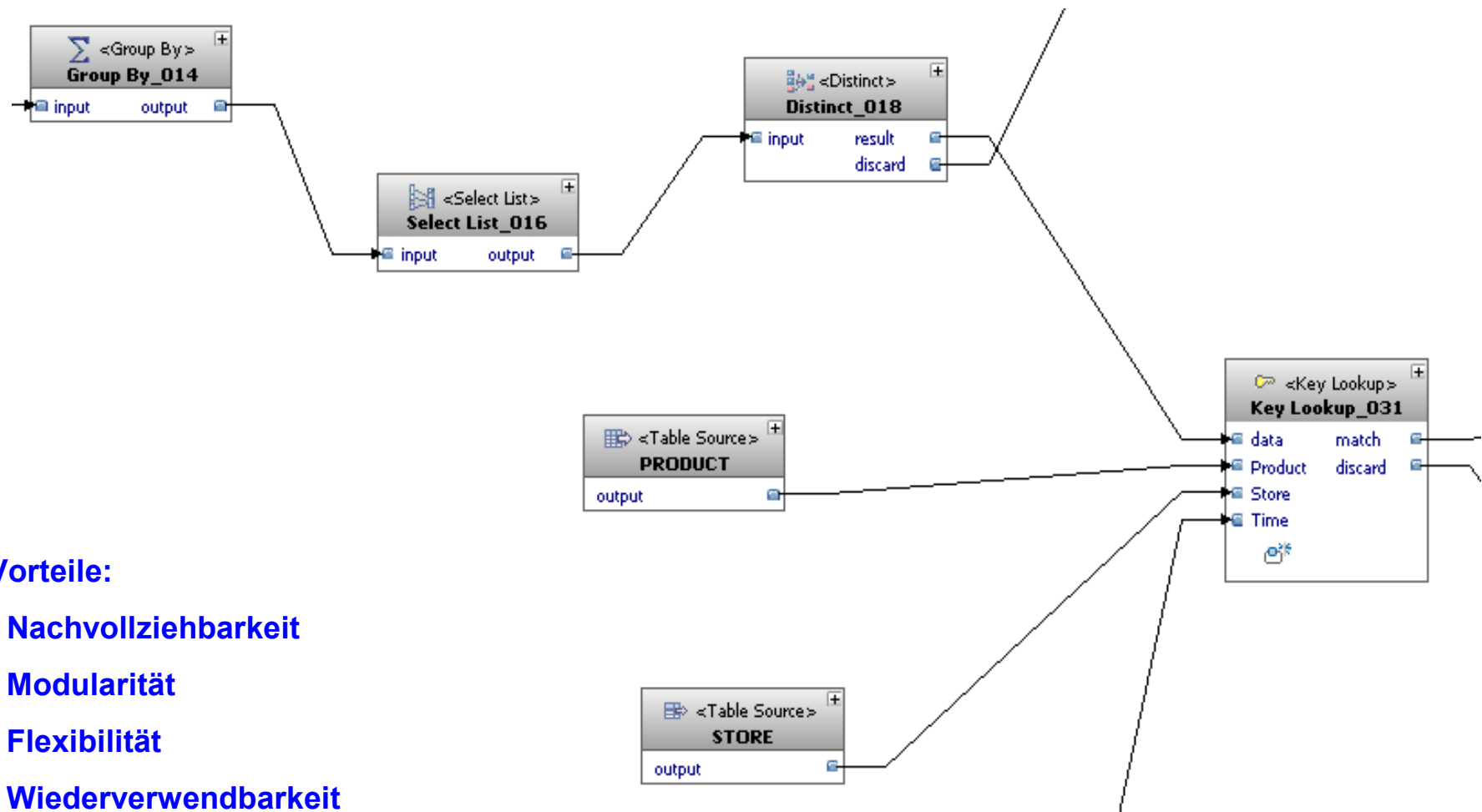
Schreiben Sie lieber SQL Skripte

```
.....
SELECT
SALES.OU_IP_ID AS STR_IP_ID,
SALES.PD_ID AS PD_ID,
SALES.MSR_PRD_ID AS TIME_ID,
SALES.C_D_MKT_BSKT_TXN_ID AS NMBR_OF_MRKT_BSKTS,
SALES.SUM_NBR_ITM AS NUMBER_OF_ITEMS,
CASE
    WHEN SALES.M_BK_PD_SUB_DEPT_NM IN ('BATH AND SHOWER', 'CAMERAS') THEN
        SALES.BKP_SUM_NBR_ITMXPRC * DECIMAL(MARTS.RAND1N(5) + 123) / 100
    ELSE
        SALES.BKP_SUM_NBR_ITMXPRC * DECIMAL(MARTS.RAND1N(5) + 102) / 100
END AS PRDCT_BK_PRC_AMUNT,
CASE
    WHEN
        SALES.MIN_CG_PD_DEPT_NM IN ('TEEN BOYS', 'TEEN BOYS JEANS', 'DRESS FORMAL', 'MEN
        SHOES')
    THEN
        (DECIMAL(68 - MARTS.RAND1N(5)) / 100) * SALES.SUM_CG_NBR_ITMX_PRC
    WHEN
        SALES.MIN_CG_PD_DEPT_NM IN ('ELECTRICAL APPLIANCES', 'ELECTRONICS', 'COLORED
        TELEVISIONS', 'WOMEN SHOES')
    THEN
        (DECIMAL(77 - MARTS.RAND1N(5)) / 100) * SALES.SUM_CG_NBR_ITMX_PRC
    WHEN
        SALES.MIN_CG_PD_DEPT_NM IN ('HEALTH AND BEAUTY')
    THEN
        (DECIMAL(65 - MARTS.RAND1N(5)) / 100) * SALES.SUM_CG_NBR_ITMX_PRC
    ELSE
        (DECIMAL(72 - MARTS.RAND1N(5)) / 100) * SALES.SUM_CG_NBR_ITMX_PRC
END AS CST_OF_GDS_SLD_CGS,
SALES.SUM_NBR_ITMXSTM_PRC AS SALES_AMOUNT
FROM
    SALES) .....
```

Was passiert wenn sich Bezeichner ändern ?

Wie flexibel sind die SQL Skripte (z.B. Unterscheidung Produktion - Testsystem) ?

... oder entwickeln die Integrationsprozesse mit einem intuitiven Werkzeug?



Vorteile:

- Nachvollziehbarkeit
- Modularität
- Flexibilität
- Wiederverwendbarkeit

Design Studio – integrierte Modellierungs- und Entwicklungsumgebung

The screenshot displays the Design Studio environment. The main workspace shows a database model with several tables: **STORE_STG**, **CLASS_STG**, **CLASS**, **PRODUCT**, **SALES**, **PERIOD**, **PRODUCT_STE**, **PARTSFORPRODUCT**, and **PRODUCTPARTSLIS**. Relationships are indicated by lines connecting the tables. The left pane shows the Project Explorer with a tree view of the database structure. The bottom-left pane contains a SQL editor with the following code:

```
CREATE TABLE STORE (  
  STOREKEY INTEGER NOT NULL,  
  STORE_MANAGER CHAR ( 30 ),  
  STORE_TYPE CHAR ( 10 ),  
  STORE_NAME CHAR ( 30 ),  
  STREET CHAR ( 30 ),  
  CITY CHAR ( 20 ),  
  STATE CHAR ( 5 ),  
  ZIP CHAR ( 10 ),  
  EFFECTIVE_DATE DATE,  
  EXPIRE_DATE DATE  
);  
  
CREATE UNIQUE INDEX STORE_PKC  
ON STORE (  
  STOREKEY ASC  
)  
PCTFREE 0;  
  
ALTER TABLE STORE  
ADD CONSTRAINT STORE_PKC PRIMARY  
KEY (  
  STOREKEY
```

The bottom-right pane shows the properties for the **STORE** table, including columns and their primary key status.

Name	Primary Key
STOREKEY	☒
STORE_MANAGER	☐
STORE_TYPE	☐
STORE_NAME	☐
STREET	☐
CITY	☐
STATE	☐
ZIP	☐
EFFECTIVE_DATE	☐
EXPIRE_DATE	☐

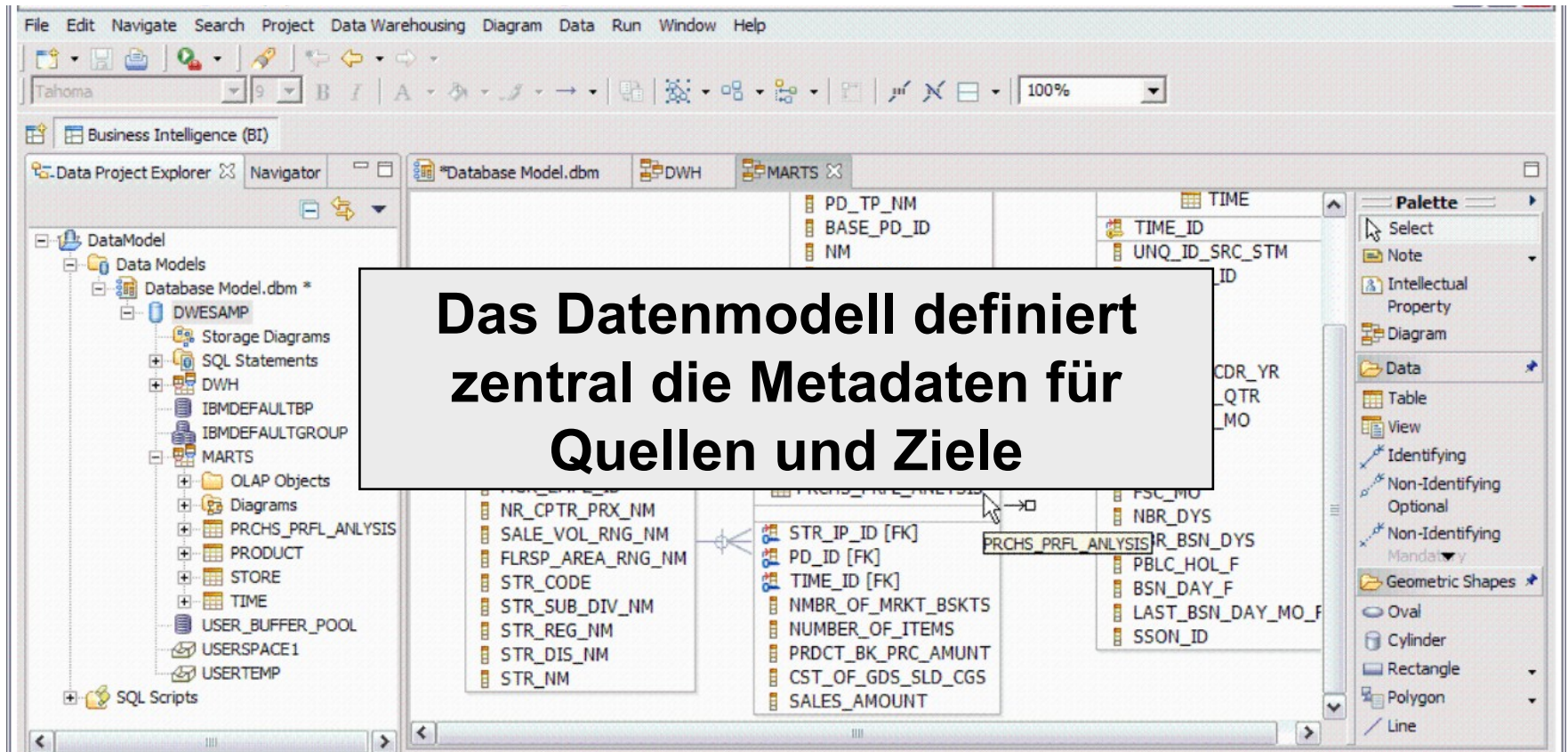
Visuelles Editieren
des physischen DB
Designs, Reverse
Engineering von
DDL &
Datenbanken,
etc...

• Integrierte Eclipse-basierte
Entwicklungsumgebung

• Nutzung vorhandenen Know-Hows

• erweiterbar

Das Datenmodell als Zentrum des Data Warehouse



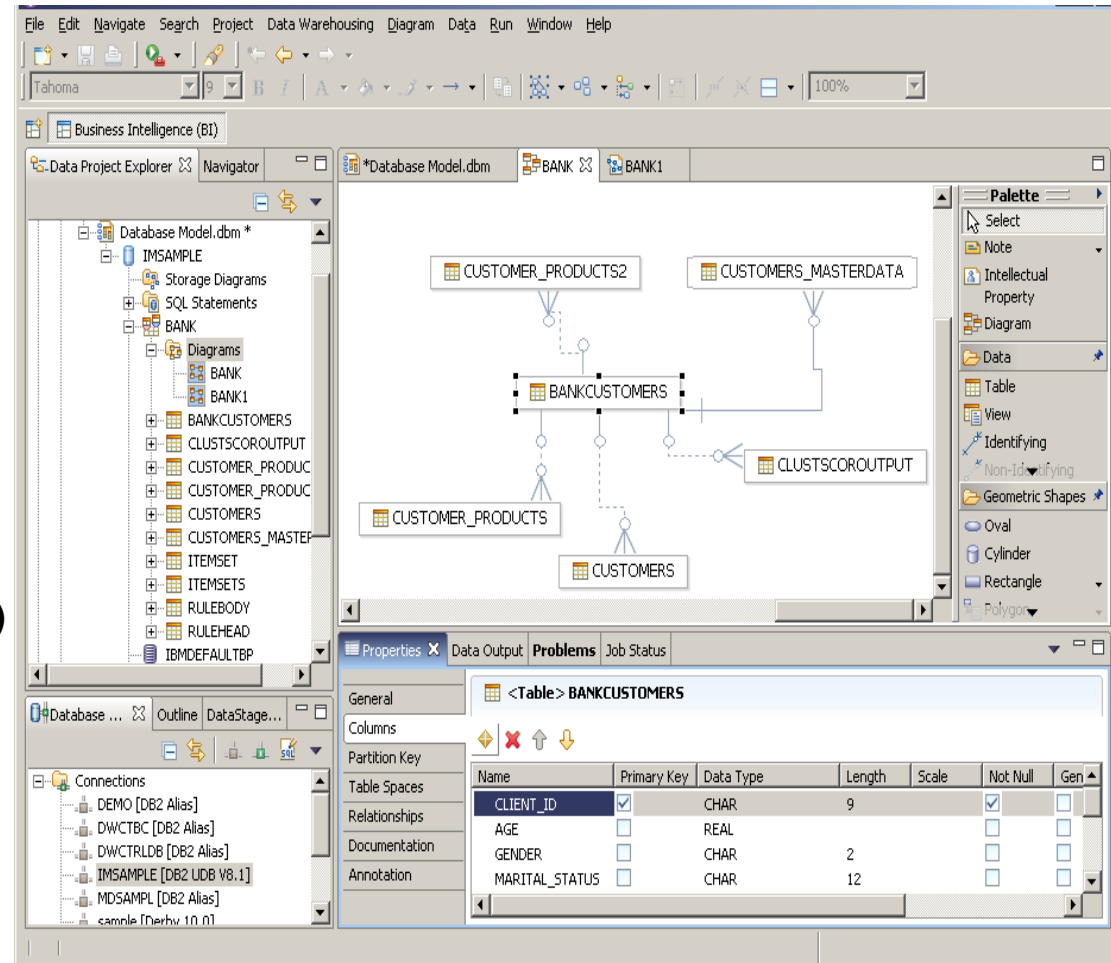
*Erforschen, Modellieren, Visualisieren, Beziehungen aufbauen
für heterogene Landschaften mit einem Tool*

Datenmodellierung Übersicht

- Entwurf und Modifizieren von **physischen** Datenbankmodellen

Key Features:

- Erstellen neuer DB “from scratch”
- Reverse engineer** aus existierender DB oder aus DDL
- Erstellen von Übersichtsdiagrammen
- Graphisches Editieren
- Vergleichen von Datenmodellen oder von Datenmodellen mit DB
- Designanalyse (z.B. Abhängigkeiten)
- Erstellen und Deployen von DDL-Skripten

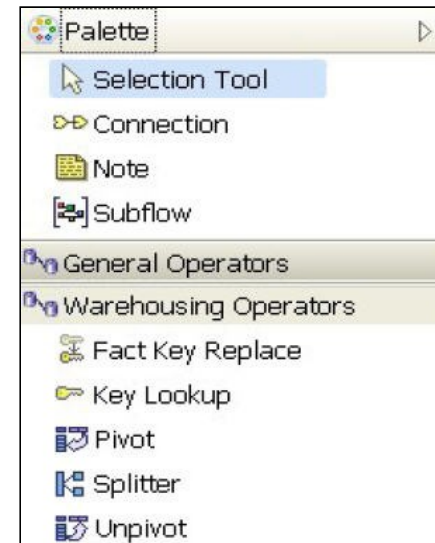
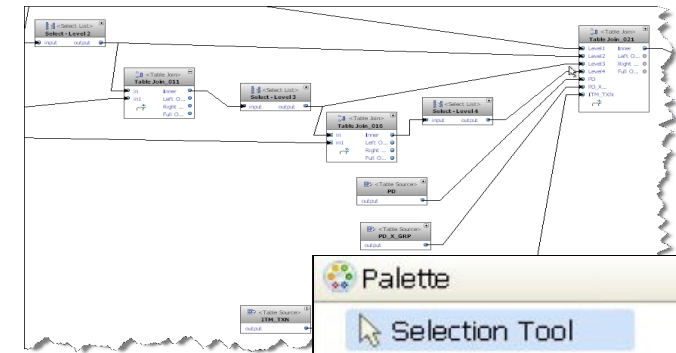


In-database ETL ...

- **SQL wird sehr oft zur Datenaufbereitung verwendet:**
 - Vorteile von SQL Transformationen gegenüber klassischen ETL Werkzeugen:
 - Nutzung vorhandener SQL Kenntnisse
 - Kein unnötiges Verschieben von Daten aus der DB und Reimport
 - Transformationen werden direkt dort verarbeitet, wo sich die Daten befinden
 - Nutzen der Performance und Skalierbarkeit der Datenbank, keine separate Engine notwendig
 - “Hand-Coded” SQL wird häufig für ETL verwendet
-> oftmals hunderte von SQL Skripten
 - SQL ist eine Ergänzung zu traditionellen ETL Werkzeugen in heterogenen Umgebungen
 - Komplexe Datenextraktion (“Konnektoren”) bzw. Datenaufbereitung kann teilweise notwendig sein, da SQL nicht für alle Anwendungszwecke geeignet ist
 - ETL Metadaten Management, Datenqualitätsverfahren, Data Profiling, und spezielle Konnektoren können erforderlich sein. Diese Anforderungen können zum Beispiel mit IBM InfoSphere Information Server (Data Stage) abgedeckt werden

Datenflüsse für in-database ETL

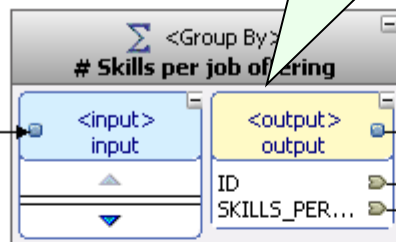
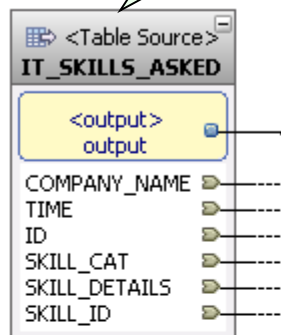
- **Eclipse Tool generiert SQL, um Daten innerhalb der Datenbank zu transformieren und zu bewegen**
 - Einfach zu benutzen per Drag & Drop, dadurch schneller DWH Entwicklungszyklus
 - Integriert mit physischer Modellierung für Quellen und Ziele sowie andere DB Objekte
 - Vielfältige Auswahl an Operatoren für Transformationsprozesse
- **DWH Erstellung in homogener DB Umgebung**
 - Verwendet existierende DB Infrastruktur
 - Wiederverwendbarkeit von SQL Code, Nutzung vorhandener Kompetenzen
 - Integration für schnelle BI Entwicklung
 - Datenbank Modellierung und Administration
 - Datenaufbereitung
 - Abhängigkeitsanalysen



Ein einfacher Datenfluss

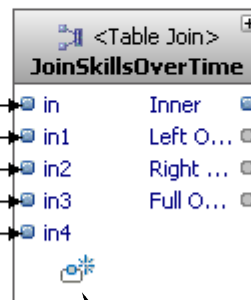
IT_SKILLS_ASKED
Table as a source

Aggregation: count of
number of skills per
job posting

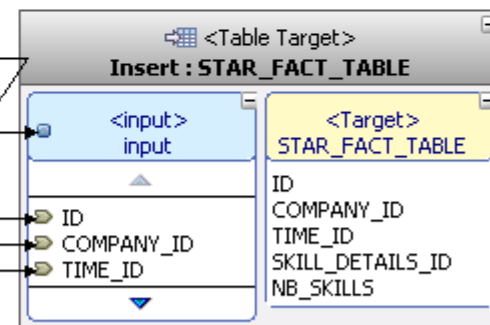


My dimension table
sources

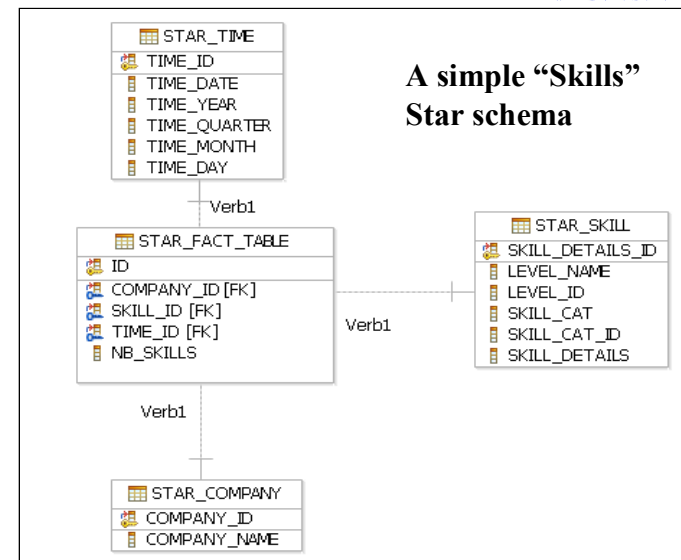
"SQL Execution DB"
= where SQL is
executed.
(usually your WH DB)



Lookup dimension
table (surrogate)
keys, calculate
'NB_SKILLS' weight



Insert new fact rows
into the fact table

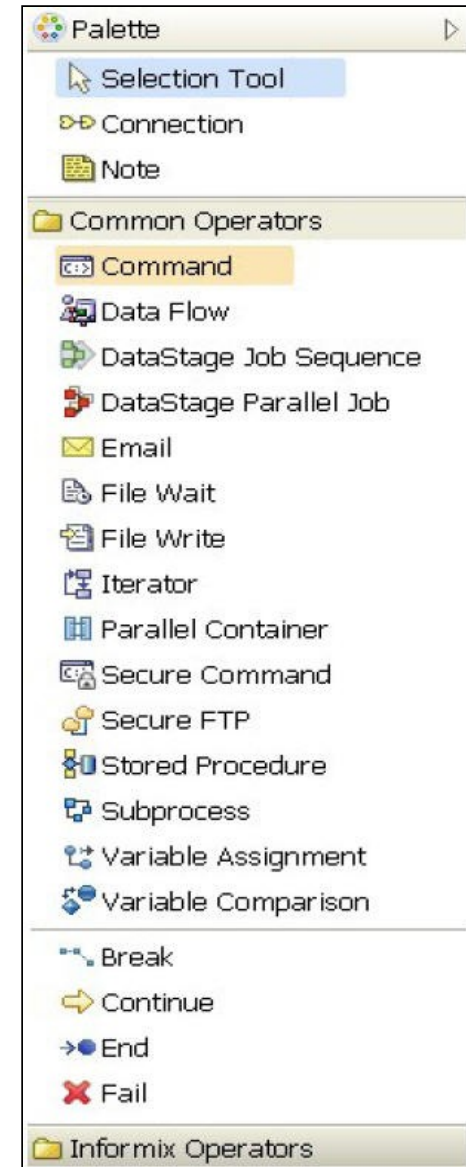
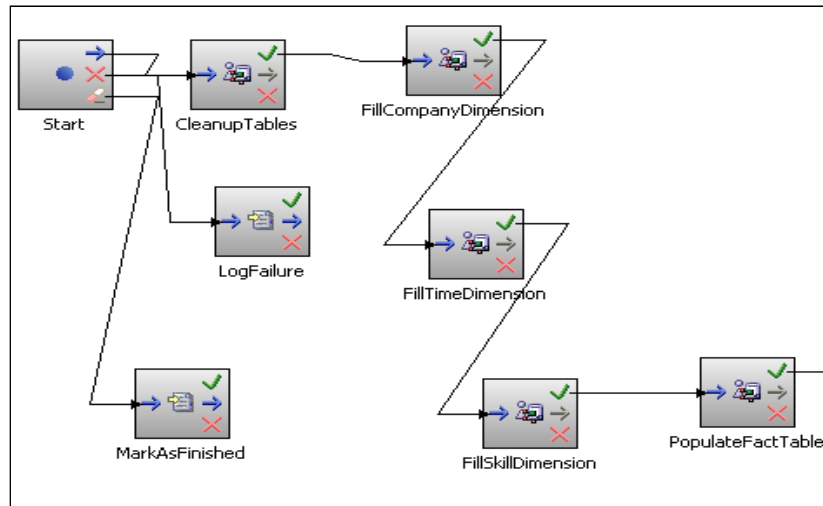


SQL Warehousing Überblick: Prozesse

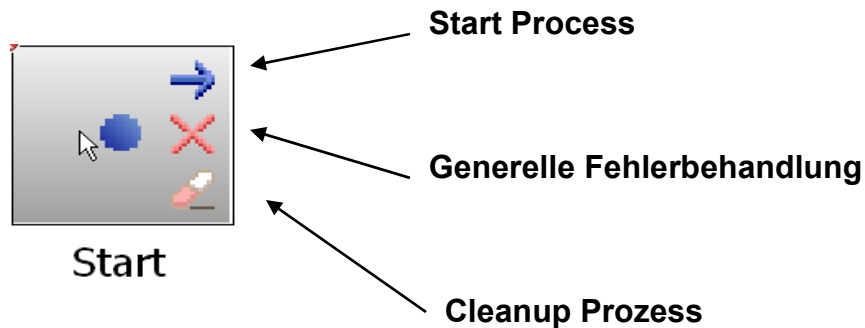
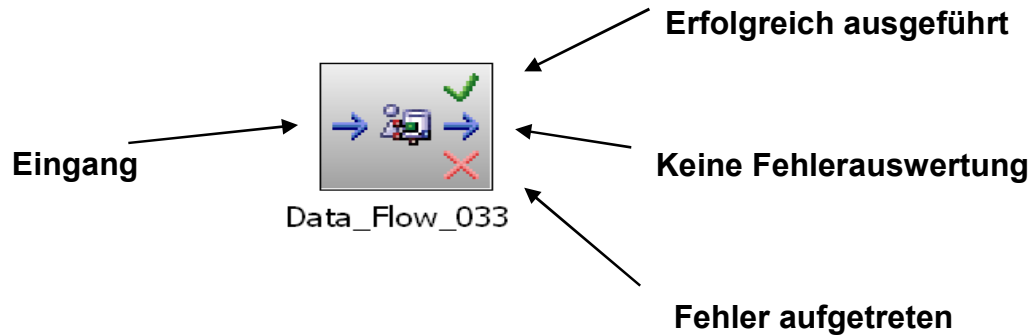
■ Steuerflüsse als Definition von ETL Prozessen

Key Features:

- **Control flows** koordinieren die Ausführung von Datenflüssen und anderen Aktivitäten
- Unterstützen Ausführungsbedingungen
- Unterstützung für parallele Ausführung
- Modularisierung durch Subprozesse
- Parametrisierung durch Verwendung von Variablen



Steuerflüsse – Definieren von Prozessen



Wird ausgeführt falls ein Fehler aufgetreten ist

Cleanup-Prozess

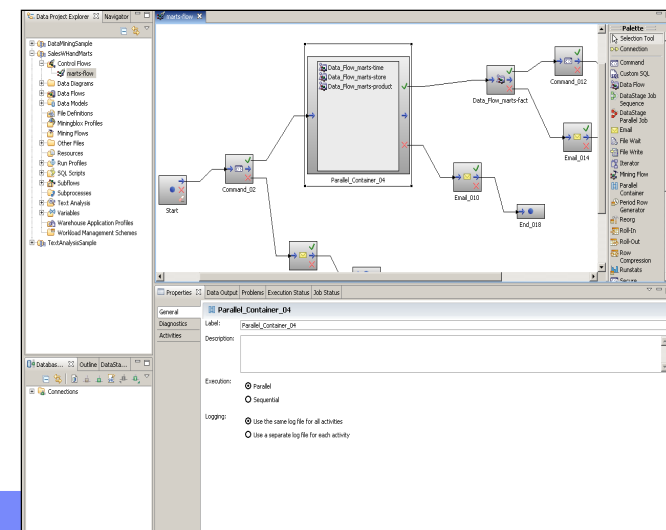
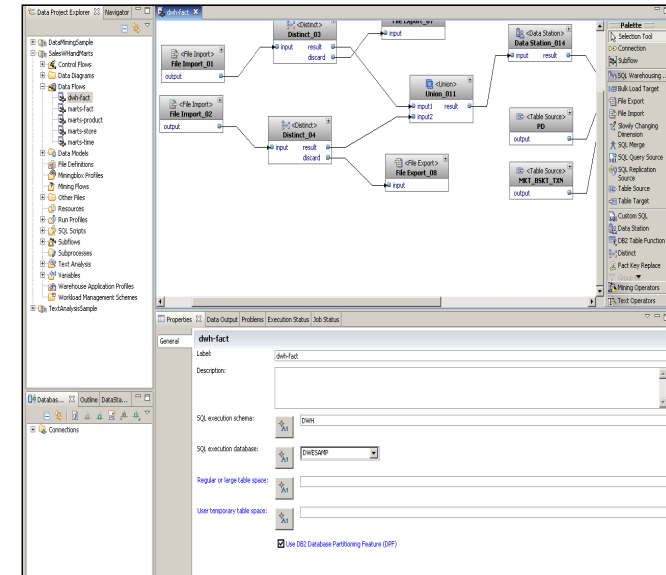
Zusammenfassung – Vorteile von SQW

Ohne SQW:

- Handgeschriebene SQL-Skripte oder Notwendigkeit externer Tools
- Komplexe IT-Landschaft, Wartung/Anpassung teilweise sehr umfangreich
- Umfangreiche Entwicklungszyklen
- Nutzung selbstprogrammierter Logiken

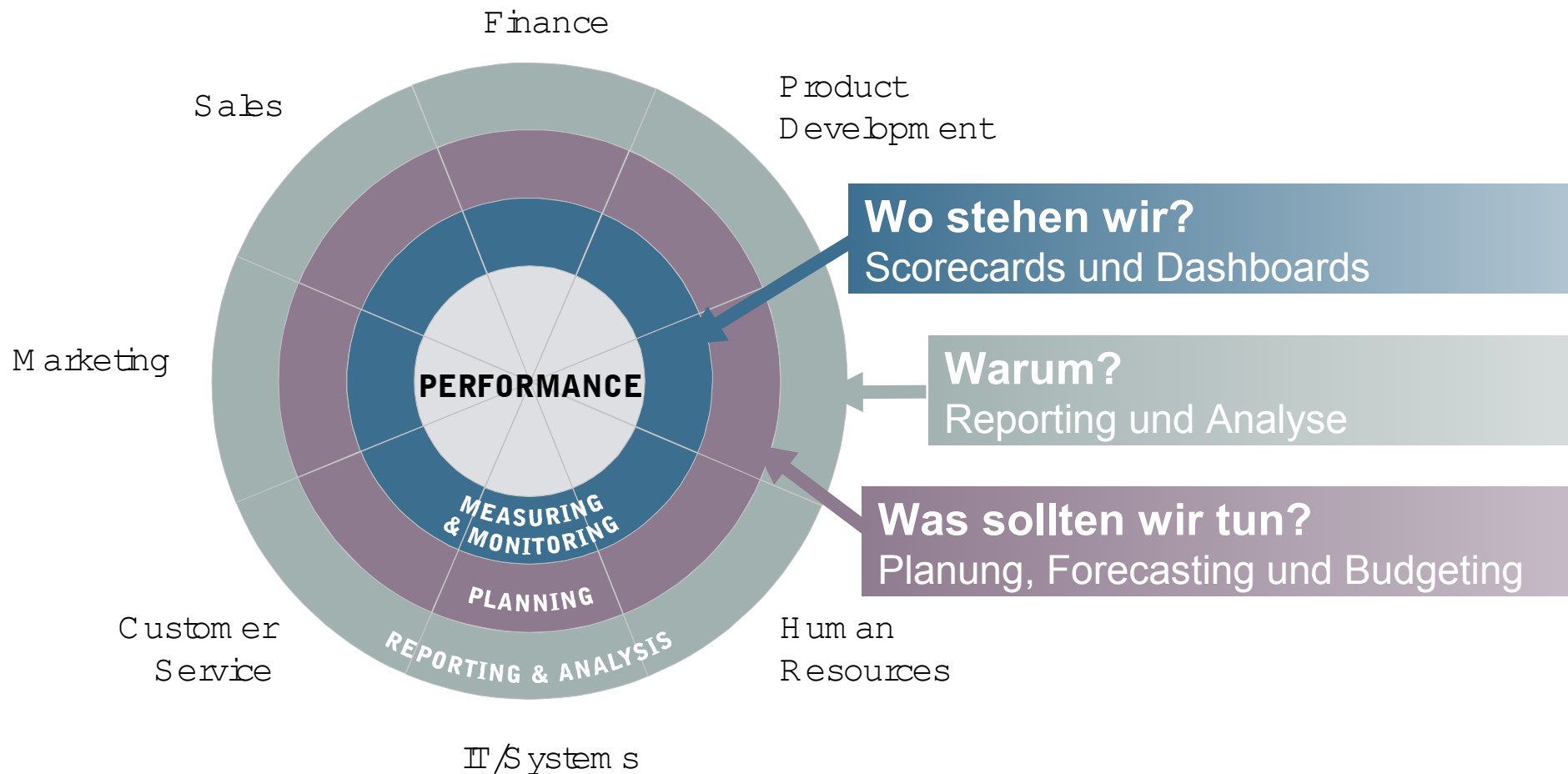
Mit SQW:

- Visualisierte Transformationen, leicht zu erstellen und leicht nachvollziehbar
- Integriertes Metadatensystem
- Automatische SQL Generierung
- Integrierte Abbildung der Prozesse, integrierter Scheduler
- Nutzung vorhandener Infrastruktur, homogene IT-Landschaft



Performance Management

Antworten auf drei Fragen, die die Performance beeinflussen



Breites Angebot an BI-Funktionen

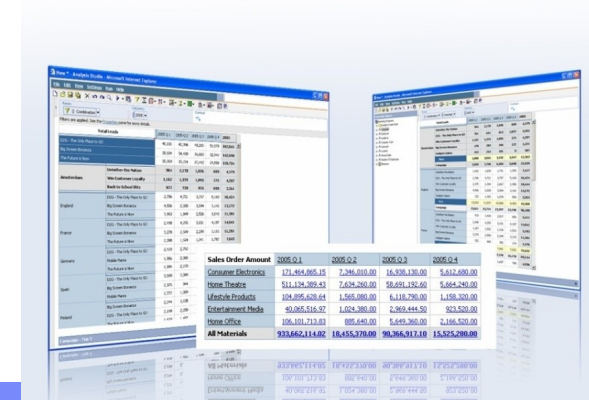
Reporting

- **Volle Bandbreite aller Reporttypen**
- **Konsistente Informationen über alle Arten von Ausgabeformaten**
- **Berichte können personalisiert und zielgerichtet verteilt werden**
- **Ermöglicht die enge Zusammenarbeit zwischen Benutzern und der IT**
- **Ermöglicht den Zugriff via E-Mail, Portal, Microsoft Office, Such-Portale und mobile Endgeräte**



Analyse

- **Geführte multidimensionale Erkundung von Daten**
- **Liefert komplexe Analysen und Szenarien Modellierung einfach und schnell**
- **Liefert das “warum” hinter den Trends**
- **Müheleses Wechseln zwischen Aggregationsebenen auf Detailebenen (DrillDown / DrillUp)**



11 2 1 0 1 1

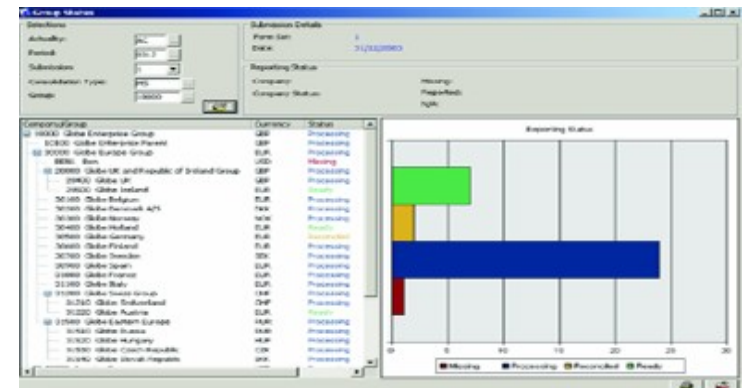
Planung und Forecasting

- **Verbindung von Top-Down- und Bottom-Up-Planung, Gegenstromverfahren**
- **Verknüpfung von operativen Plänen und der Finanzplanung**
- **Planungsworkflow zur Prozesskontrolle**
- **Verlässlichkeit und Akzeptanz der Planung erhöhen**



Konsolidierung

- Einheitliche Sicht auf alle unternehmensweiten Finanzdaten
- Vordefinierte Konsolidierungslogiken und Konsolidierungsfunktionen
- Unterstützt verschiedene Rechnungslegungsstandards
- Audit-trail



Breites Angebot an BI-Funktionen

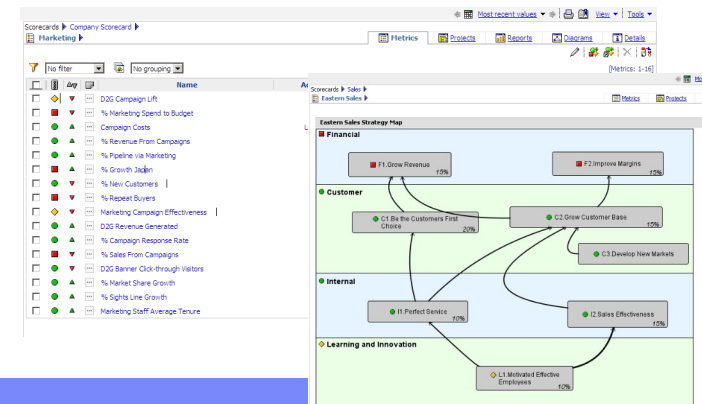
Dashboards

- Liefert wirkungsvolle Sichten auf komplexe Informationen
- Hilfestellung bei der Fokussierung auf kritische Angelegenheiten, die Ihrer Handlung und Aufmerksamkeit bedürfen
- Die Dashboards sind sehr visuell und intuitiv
- Vereint Informationen über verschiedenste Quellen



Balanced Scorecards St. Galler Modell

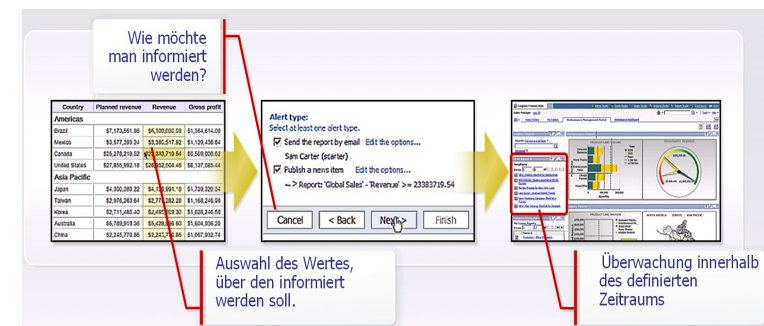
- Regelmäßige Überwachung von Kennzahlen im Bezug zu Zielwerten und Benchmarks
- Richtet Entscheidungen und Taktiken an strategischen Initiativen aus
- Unterstützt die Methodik der Balanced Scorecard
- Berücksichtigt Eigentümerschaft und Verantwortlichkeit



Überwachungselemente

Event Management

- Anwender können einfach automatisch geplante Reports überwachen
- Anwender können ohne die IT bestimmen, wann und wie sie informiert werden wollen.
- Ereignisbasiertes Reporting
- Schwellwertüberwachung
- Prozessautomation





Information Management

Informix Warehouse

neue Möglichkeiten der Informationsanalyse



Michael Köster, IT Specialist
IBM Software Group – Information Management
Michael.Koester@de.ibm.com

