Overview of
Data Warehousing
and
Real-time
Online Analytical Processing (OLAP)
Systems

by

Greg Cermak, Impartica
in partnership with Axian

Axian
Software Consulting and Training

Recognition of Authors for Content

- Microsoft

- Dennis Kennedy

- Dave Browning

- Joy Mundy

- Michael Haisten

- Justin Langseth

- David Fuller

Axian
Software Consulting and Training

Overview

- Introducing Data Warehousing

- Defining OLAP Solutions

- Understanding Data Warehouse Design

- Understanding OLAP Models

- Applying OLAP Cubes

- Real-time OLAP

Introducing Data Warehousing

- Raw Data vs. Business Information

- OLTP Source Systems

- SQL GROUP BY Clause

- SQL GROUP BY Clause with the CUBE Operator

- Data Warehouse Characteristics

- Data Warehouse System Components

Raw Data vs. Business Information

- Capturing Raw Data

  - Gathering data recorded in everyday operations

- Deriving Business Information

  - Deriving meaningful information from raw data

- Turning Data into Information

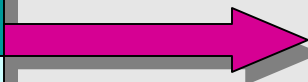 - Implementing a decision support system

OLTP Source Systems

- OLTP System Characteristics

  - Are modeled around documents or transactions

  - Processes real-time transactions of a business

  - Contains data structures optimized for entries and edits

  - Provides limited decision support capabilities
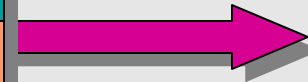
SQL GROUP BY Clause

```
SELECT productid, orderid, quantity
FROM orderhist
```

```
SELECT productid, SUM(quantity)
AS total_quantity
FROM orderhist
GROUP BY productid
```

productid	orderid	quantity
1	1	5
1	1	10
2	1	10
2	2	25
3	1	15
3	2	30

productid	total_quantity
1	15
2	35
3	45

productid	total_quantity
2	35

Only rows that satisfy the WHERE clause are grouped

```
SELECT productid, SUM(quantity)
AS total_quantity
FROM orderhist
WHERE productid = 2
GROUP BY productid
```

Axian
Software Consulting and Training

SQL GROUP BY Clause with the CUBE Operator

```
SELECT productid, orderid, SUM(quantity) AS total_quantity
FROM orderhist
GROUP BY productid, orderid
WITH CUBE
ORDER BY productid, orderid
```

The CUBE operator produces two more summary values than the ROLLUP operator

productid	orderid	total_quantity	Description
NULL	NULL	95	Grand total
NULL	1	30	Summarizes all rows for **orderid 1**
NULL	2	65	Summarizes all rows for **orderid 2**
1	NULL	15	Summarizes only rows for **productid 1**
1	1	5	Detail value for **productid 1**, **orderid 1**
1	2	10	Detail value for **productid 1**, **orderid 2**
2	NULL	35	Summarizes only rows for **productid 2**
2	1	10	Detail value for **productid 2**, **orderid 1**
2	2	25	Detail value for **productid 2**, **orderid 2**
3	NULL	45	Summarizes only rows for **productid 3**
3	1	15	Detail value for **productid 3**, **orderid 1**
3	2	30	Detail value for **productid 3**, **orderid 2**

Axian
Software Consulting and Training

OLTP Source Systems

- OLTP Examples

  - Order tracking

  - Customer service

  - Point-of-sales

  - Service-based sales

  - Banking functions

Data Warehouse Characteristics

- Provides Data for Business Analysis Processes

- Integrates Data from Heterogeneous Source Systems

- Combines Validated Source Data

- Organizes Data into Non-Volatile, Subject-Specific Groups

- Stores Data in Structures that Are Optimized for Extraction and Querying

Data Warehouse System Components
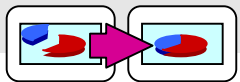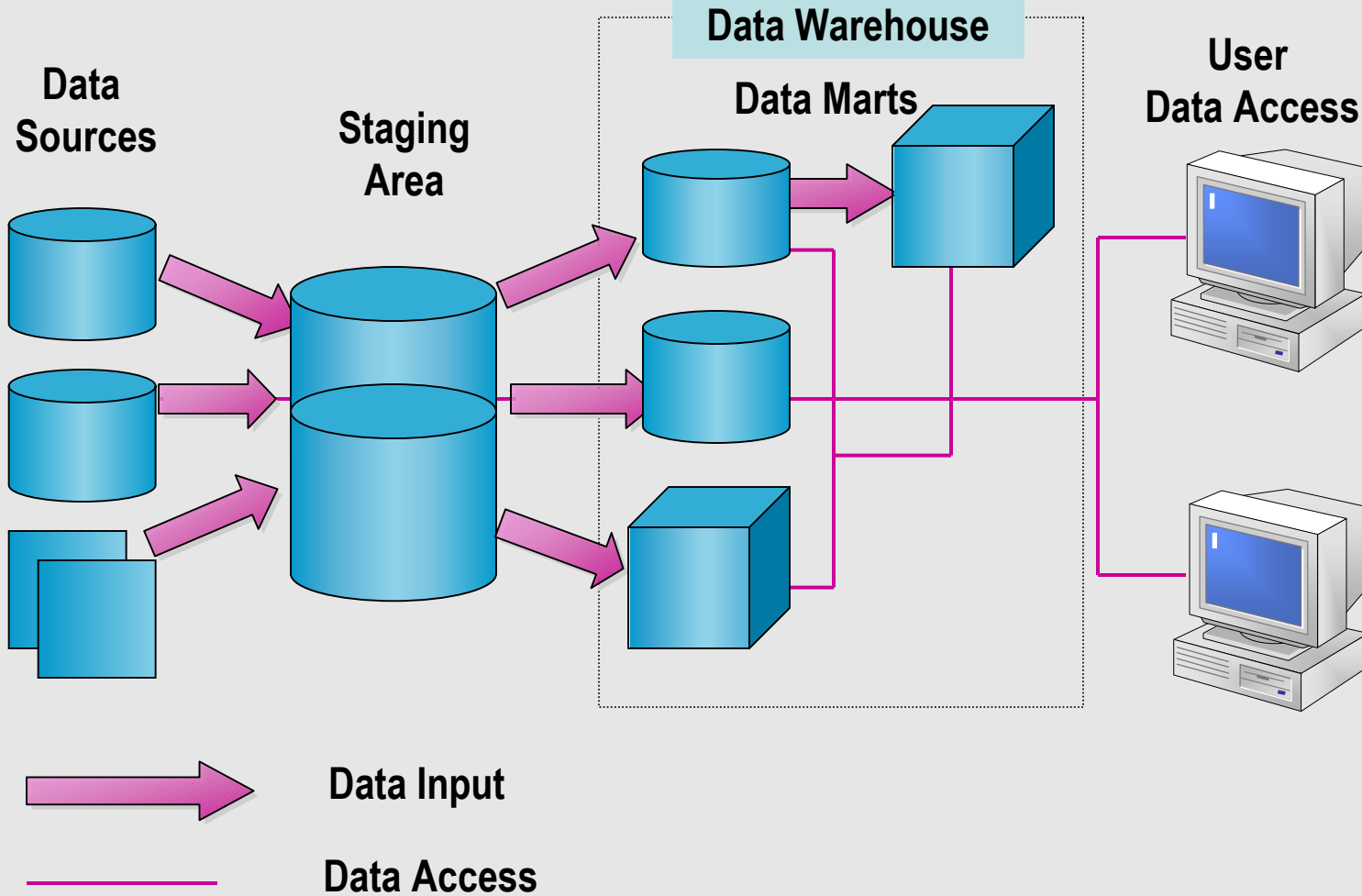
Defining OLAP Solutions

- OLAP Databases

- Common OLAP Applications

- Relational Data Marts and OLAP Cubes

OLAP Databases

- Optimized Schema for Fast User Queries

- Robust Calculation Engine for Numeric Analysis

- Conceptual, Intuitive Data Model

- Multidimensional View of Data

  - Drill down and drill up

  - Pivot views of data

Common OLAP Applications

- Executive Information Systems
  - Performance measures
  - Exception reporting

- Sales/Marketing Applications
  - Booking/Billing
  - Product Analysis
  - Customer Analysis

- Financial Applications
  - Reporting
  - Planning
  - Analysis

- Operations Applications
  - Manufacturing
  - Customer Service
  - Product Cost

Axian
Software Consulting and Training

Relational Data Marts and OLAP Cubes

	Relational Data Mart	OLAP Cube
Data Storage	Relational Data Structure	N-dimensional Data structure
Data Content	Detailed and Summarized Data	Summarized Data
Data Sources	Relational and Non-relational Sources	Relational and Non-relational Sources
Data Retrieval	Fast Performance for Data Extract Queries	Faster Performance for Data Extract Queries

Axian
Software Consulting and Training

Understanding Data Warehouse Design

- The Star Schema

- Fact Table Components

- Dimension Table Characteristics

- The Snowflake Schema

The Star Schema

Employee_Dim

EmployeeKey
EmployeeID ...

Dimension Table

Fact Table

Time_Dim

TimeKey
TheDate ...

Sales_Fact

TimeKey EmployeeKey ProductKey CustomerKey ShipperKey
Sales Amount Unit Sales ...

Product_Dim

ProductKey
ProductID ...

Shipper_Dim

ShipperKey
ShipperID ...

Customer_Dim

CustomerKey
CustomerID ...
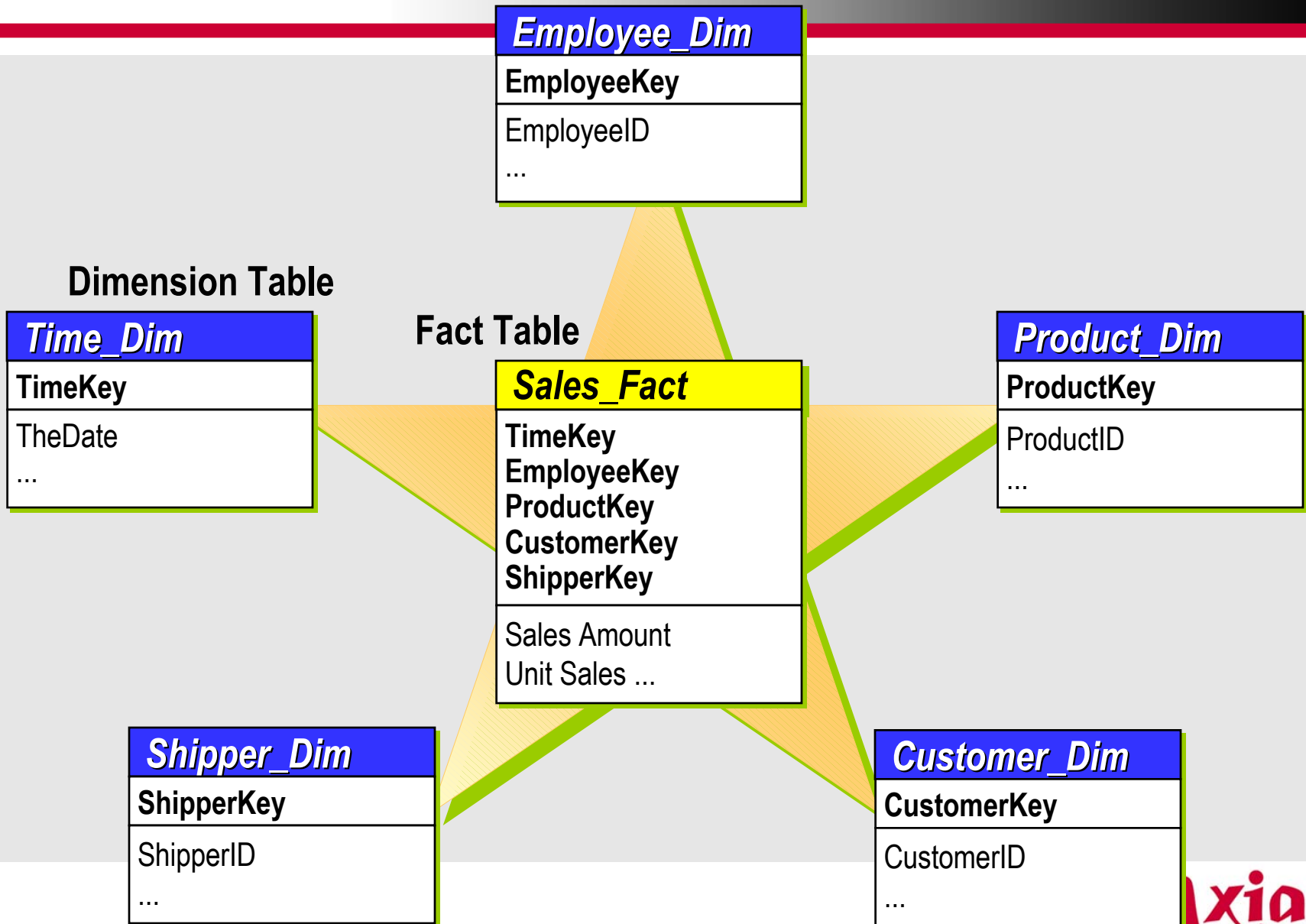
Fact Table Components

Dimension Tables

sales_fact Table

customer_dim

201	ALFI	Alfreds

product_dim

25	123	Chai

time_dim

134	1/1/2000

Foreign Keys

Measures

customer_key	product_key	time_key	quantity_sales	amount_sales
201	25	134	400	10,789

The **grain** of the sales_fact table is defined by the lowest level of detail stored in each dimension

Dimension Table Characteristics

customer *

- 🔑 customer_id
- account_num
- lname
- fname
- mi
- address1
- address2
- address3
- address4
- city
- state_province
- postal_code
- country

State *

- 🔑 State_ID
- State_Name
- Region
- Country

time_by_day *

- 🔑 time_id
- the_date
- the_day
- the_month
- the_year
- day_of_month
- week_of_year
- month_of_year
- quarter
- fiscal_period

- Describes Business Entities

- Contains Attributes That Provide Context to Numeric Data

- Presents Data Organized into Hierarchies

The Snowflake Schema
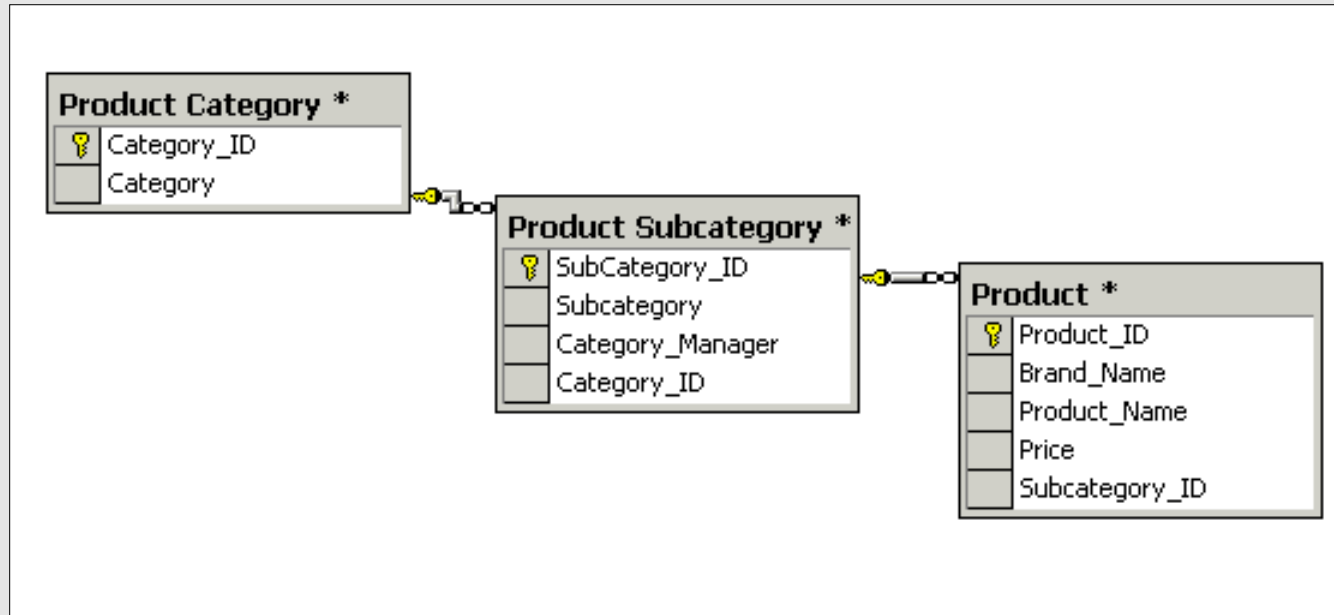
- Defines Hierarchies by Using Multiple Dimension Tables

- Is More Normalized than a Single Table Dimension

- Is Supported within Analysis Services

Understanding OLAP Models

- OLAP Database Components

- OLAP Dimensions vs. Relational Dimensions

- Dimension Fundamentals

- Dimension Family Relationships

- Cube Measures

- Relational Data Sources

OLAP Database Components

- Numeric Measures

- Dimensions

- Cubes

OLAP Dimensions vs. Relational Dimensions

OLAP

REGION
West
   CA
   OR
East
   MA
   NY

Relational

REGION
West
East

STATE	REGION
CA	West
OR	West
MA	East
NY	East

Axian
Software Consulting and Training

Dimension Fundamentals

```
┌─────────────────────────────────┐
│ TIME                            │   ← time dimension
└─────────────────────────────────┘

● Year                               ← year level
    1999                             ← members
    2000
    2001

  ● ● Quarter                        ← quarter level
        Q1                           ← members
        Q2
        Q3
        Q4

    ● ● ● Month                      ← month level
            Jan                      ← members
            Feb
            Mar
            etc.
```

Dimension Family Relationships

USA
  North West
   Oregon
   Washington
  South West
   California

- USA is the **parent** of North West and South West
- North West and South West are **children** of USA
- North West and California are **descendants** of USA
- North West and USA are **ancestors** of Washington
- North West and South West are **siblings**
- Oregon and California are **cousins**
- All are dimension **members**
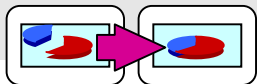
Cube Measures

- Are the Numeric Values of Principle Interest

- Correspond to Fact Table *Facts*

- Intersect All Dimensions at All Levels

- Are Aggregated at All Levels of Detail

- Form a Dimension

Relational Data Sources

- Star and Snowflake Schemas

  - Are required to build a cube with Analysis Services

- Fact Table

  - Contains measures

  - Contains keys that join to dimension tables

- Dimension Tables

  - Must exist in same database as fact table

  - Contain primary keys that identify each member

Applying OLAP Cubes

- Defining a Cube

- Querying a Cube

- Defining a Cube Slice

- Working with Dimensions and Hierarchies

- Visualizing Cube Dimensions

- Connecting to an OLAP Cube

Defining a Cube

Atlanta

Chicago

Denver

Detroit

Market Dimension

Q1 Q2 Q3 Q4

Time Dimension

Grapes
Cherries
Melons
Apples

Products Dimension

Axian
Software Consulting and Training
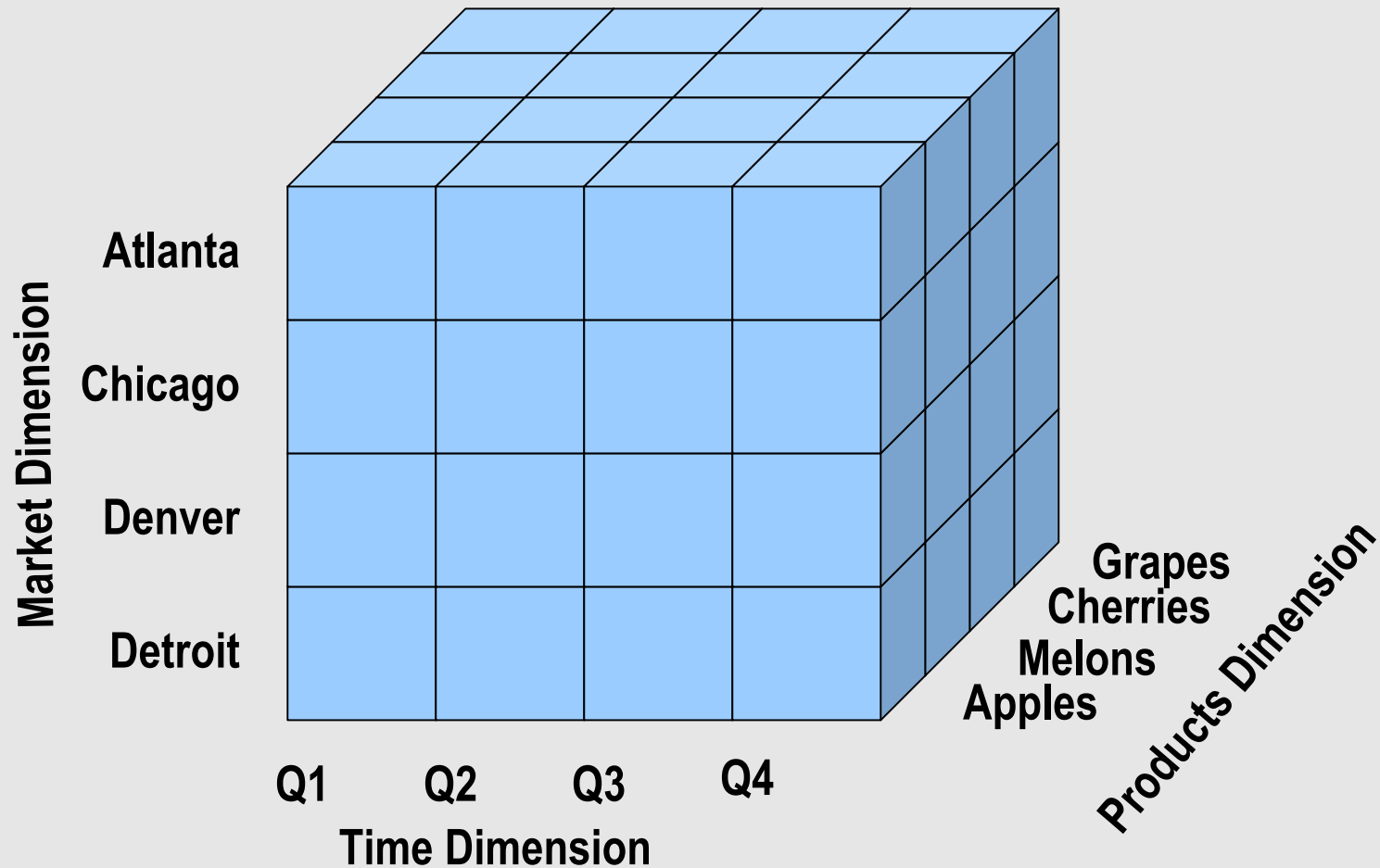
Querying a Cube
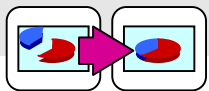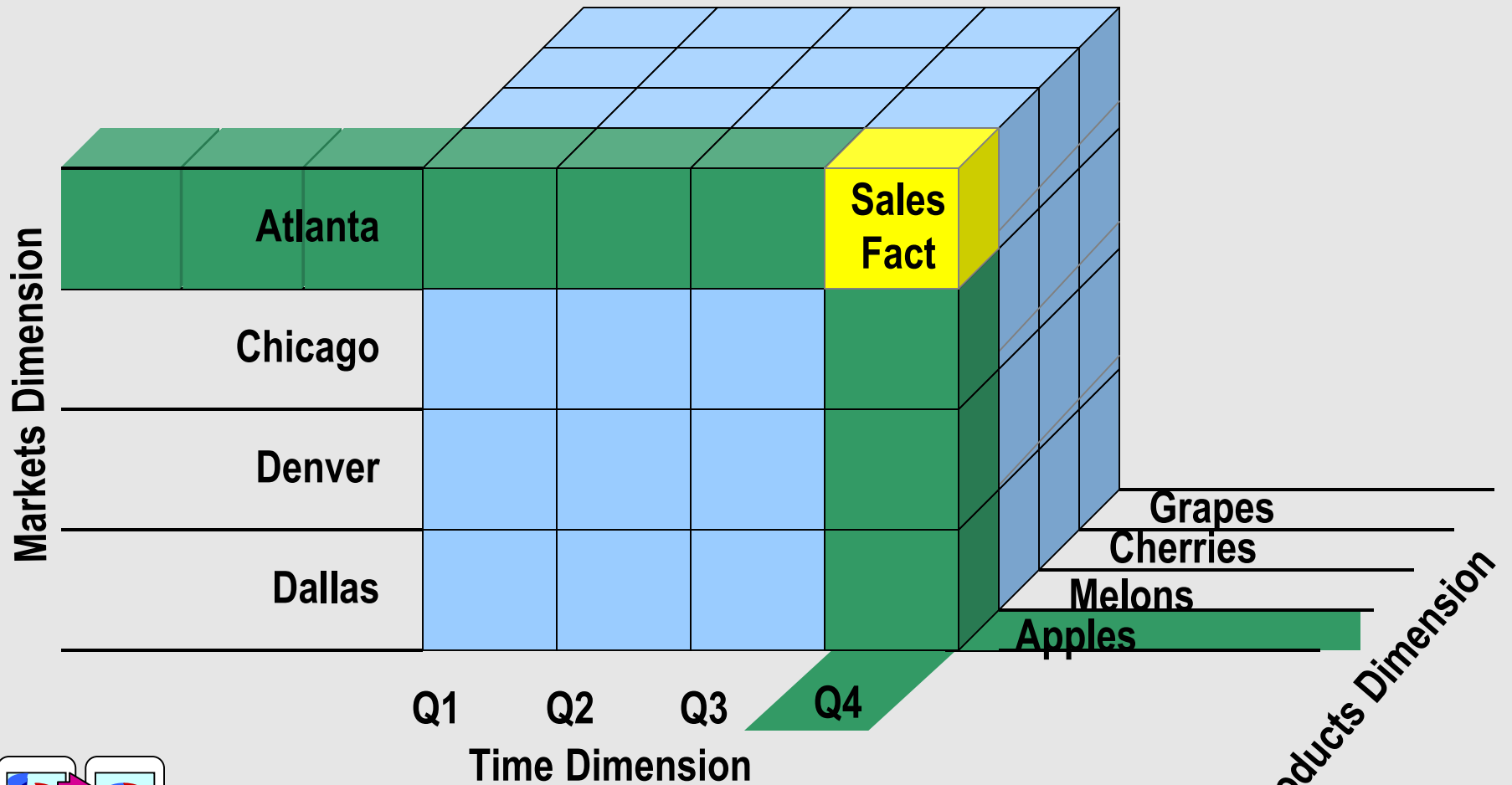
Defining a Cube Slice
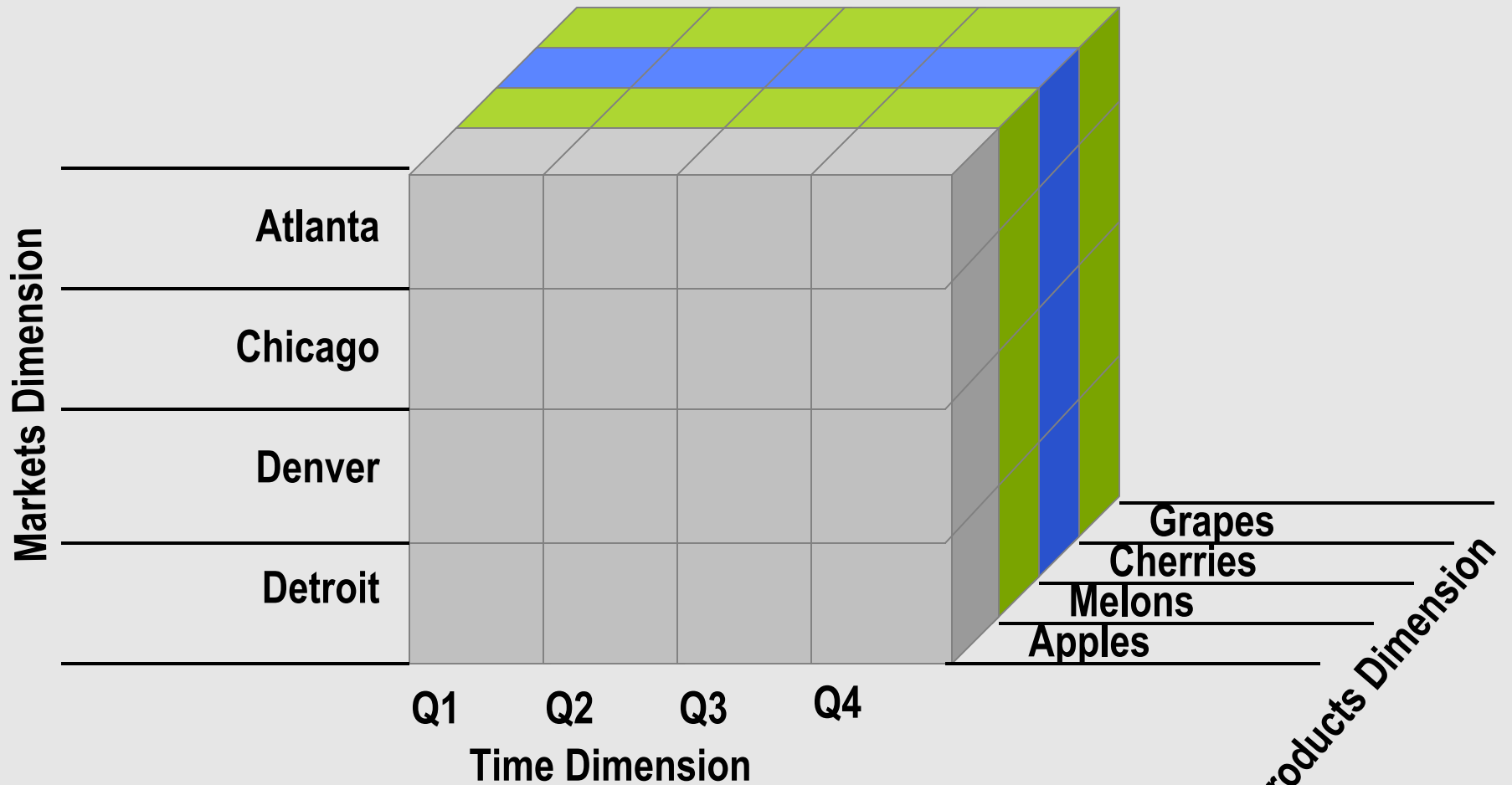
Working with Dimensions and Hierarchies

- Dimensions Allow You to
  - Slice
  - Dice
- Hierarchies Allow You To
  - Drill Down
  - Drill Up

Category ▾	
Bread	Sales Units
Dairy	Sales Units
Meat	Sales Units
Grand Total	Sales Units

Country ▾

Ca

Category ▾	Subcategory	Product Name	Sales Units
⊟ Bread	⊟ Bagels	Colony Bagels	264
		Fantastic Bagels	299
		Great Bagels	334
		Modell Bagels	304
		Sphinx Bagels	340
		Total	1541
	⊞ Muffins		6445
	⊞ Sliced Bread		6546
	Total		14532

Visualizing Cube Dimensions

Connecting to an OLAP Cube

Real-time OLAP

- Real-time Data Warehousing Defined
- Buzzwords
- Traditional Data Warehousing as Archeology
- Change Data Capture
- Time Compression
- Transitory States
- RT DW Components - Capture
- RT DW Components - Delivery Mechanism
- RT DW Components - Transformation Engine
- RT DW Components - Data Warehouse
- RT DW Components - Incremental Aggregator
- Capacity Management RT DW for BPA TBL

Axian
Software Consulting and Training

Real-time Data Warehousing Defined

- Real-time BI enables business users to react rapidly to changing business conditions

- Reduces the latency between a business event and the time it takes to react to the event

- Ideally, latency should be zero, but realistically > 0

- How much more than 0?

Axian
Software Consulting and Training

Buzzwords

- Active Data Warehouse

- Zero Latency Enterprise (ZLC)

- Real-time Data Warehousing

- Real-time Analytics

- Business Activity Monitoring

- Real-time Personalization

Traditional Data Warehousing as Archeology

- Incomplete record. A flood may wash away a whole layer of the fossil record

- Similar to a purge or selective but undocumented periodic record deletions

- Geological intrusion can replace something without leaving evidence of what was there before

- Operational systems allow update of information that destroys historical record

- Items that survive are those…

  - built to last

  - based on preferences of that time

- Expensive to store things to last

Time Compression

- As pace of business increases, processing cycles compressed

- Puts strain on distinctions between operational and analytic info services

- On-the-fly analysis

- Trend line confirmation

  - Any time sensitive activity can be compared to a baseline at close to real-time

Change Data Capture

- Process of recording incremental net changes which allows recording history accurately

- Point-in-time currency

- Difficult, if not impossible to do by reading OLTP source data in batches

- Real-time data warehousing eliminates the dig

Transitory States

- Short duration activities that may end in some form of reversal that leaves no evidence behind

- Example: a credit hold is a temporary condition in an OLTP system…

  - For an approved order, may not know it was on hold at all

  - Disapproved order recorded as a credit rejection

  - Orders per rep decline because of hold delays

  - Disapproved orders may increase, but does not account for lower numbers of orders per rep
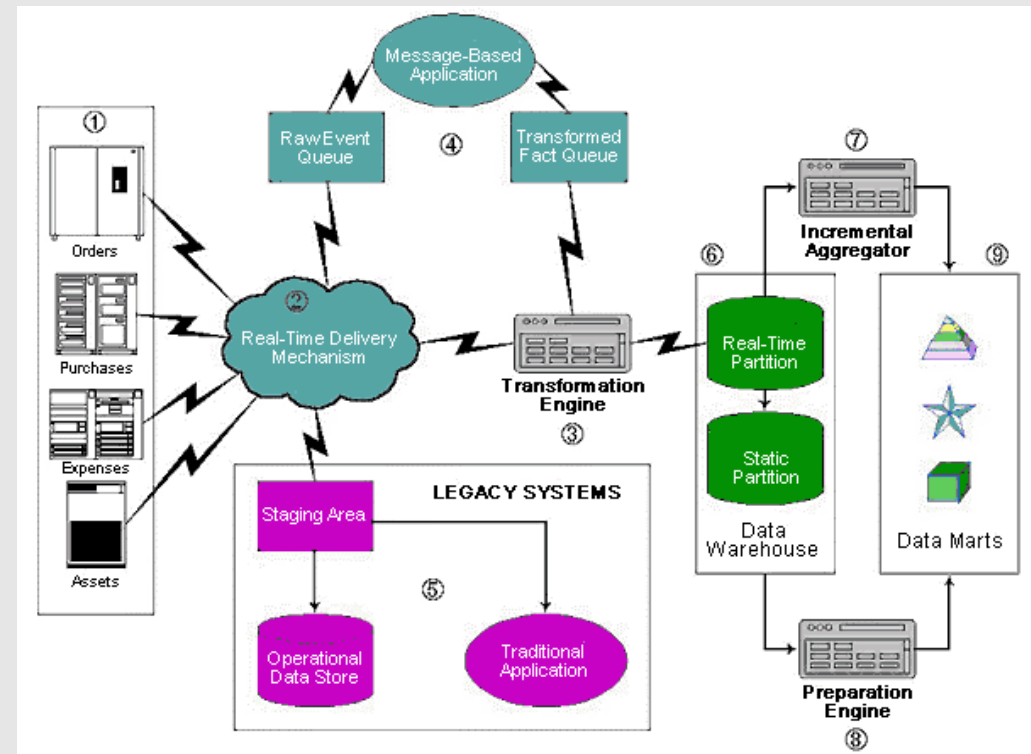
RT DW Components - Capture

- Data event monitoring

  - Database triggers

  - Replication

  - Recovery Log Processing

- Application event monitoring

  - Message queues
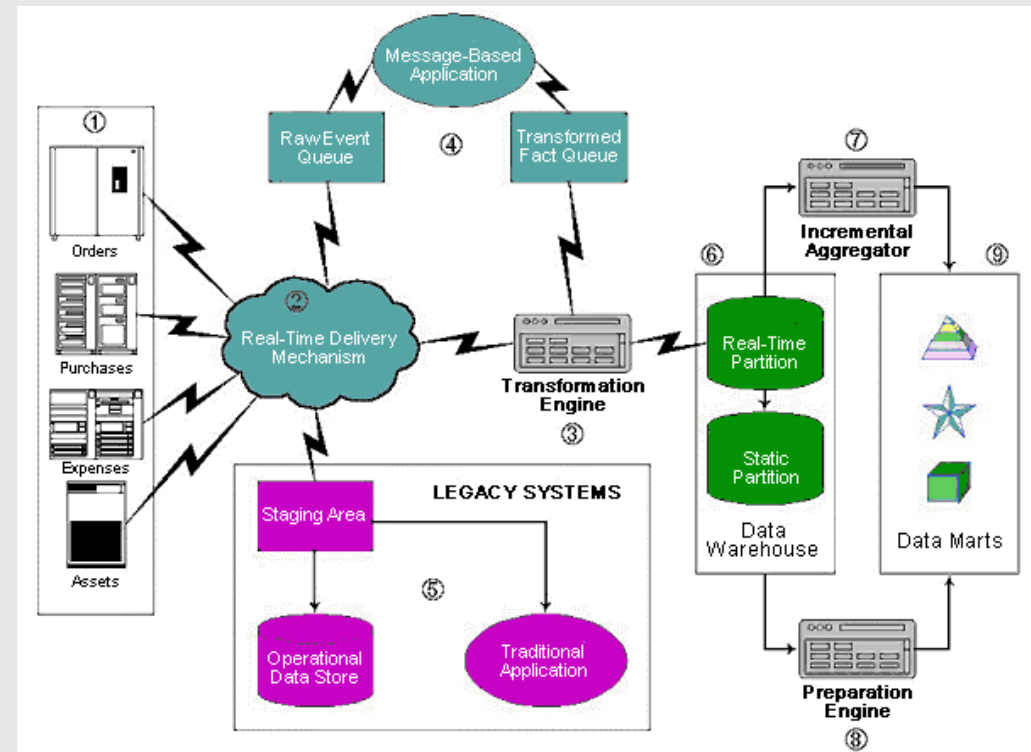
  - Enterprise Integration Apps (EAI)
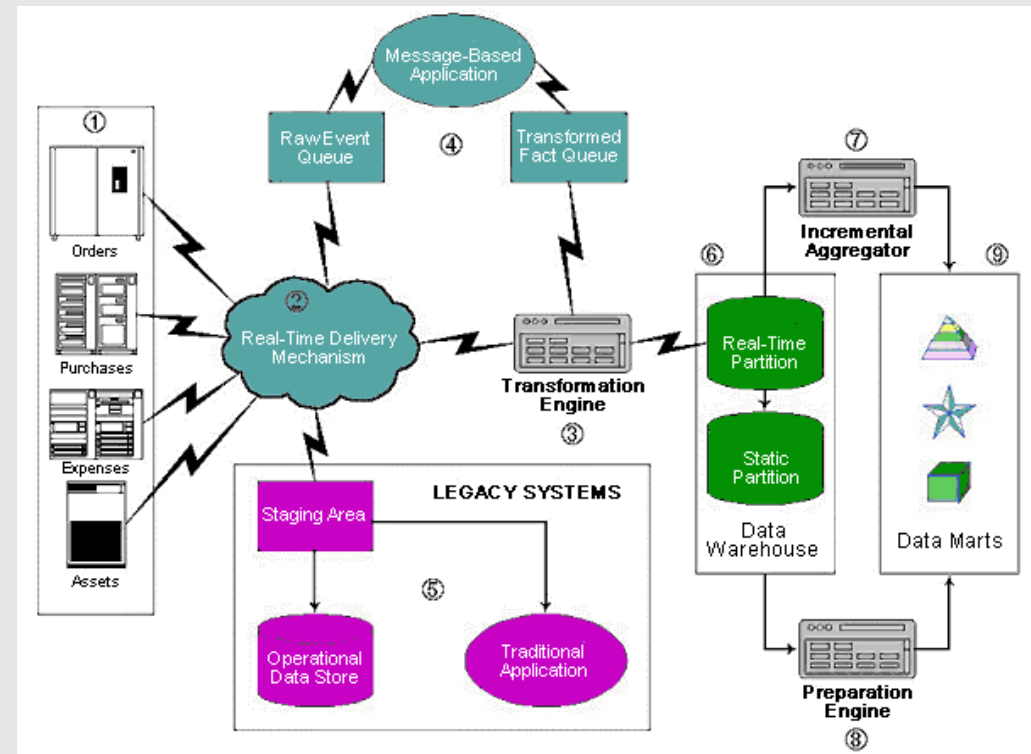
RT DW Components - Delivery Mechanism

- Cross platform message queuing

- Events dispatched to subscribers using a publish and subscribe

- Data transformed in memory instance by instance

- Provides physical format conversion, calculations and derivations, code translation, table lookups

- ETL batch tools will migrate to real-time

RT DW Components - Data Warehouse

- Collect data continuously in real-time partition

- Sweep a consistent snapshot into static partition

- Host for incremental aggregation

- Separate databases or separate tables

- Horizontal table partitioning provides single logical view while providing physical isolation between real-time and static partitions

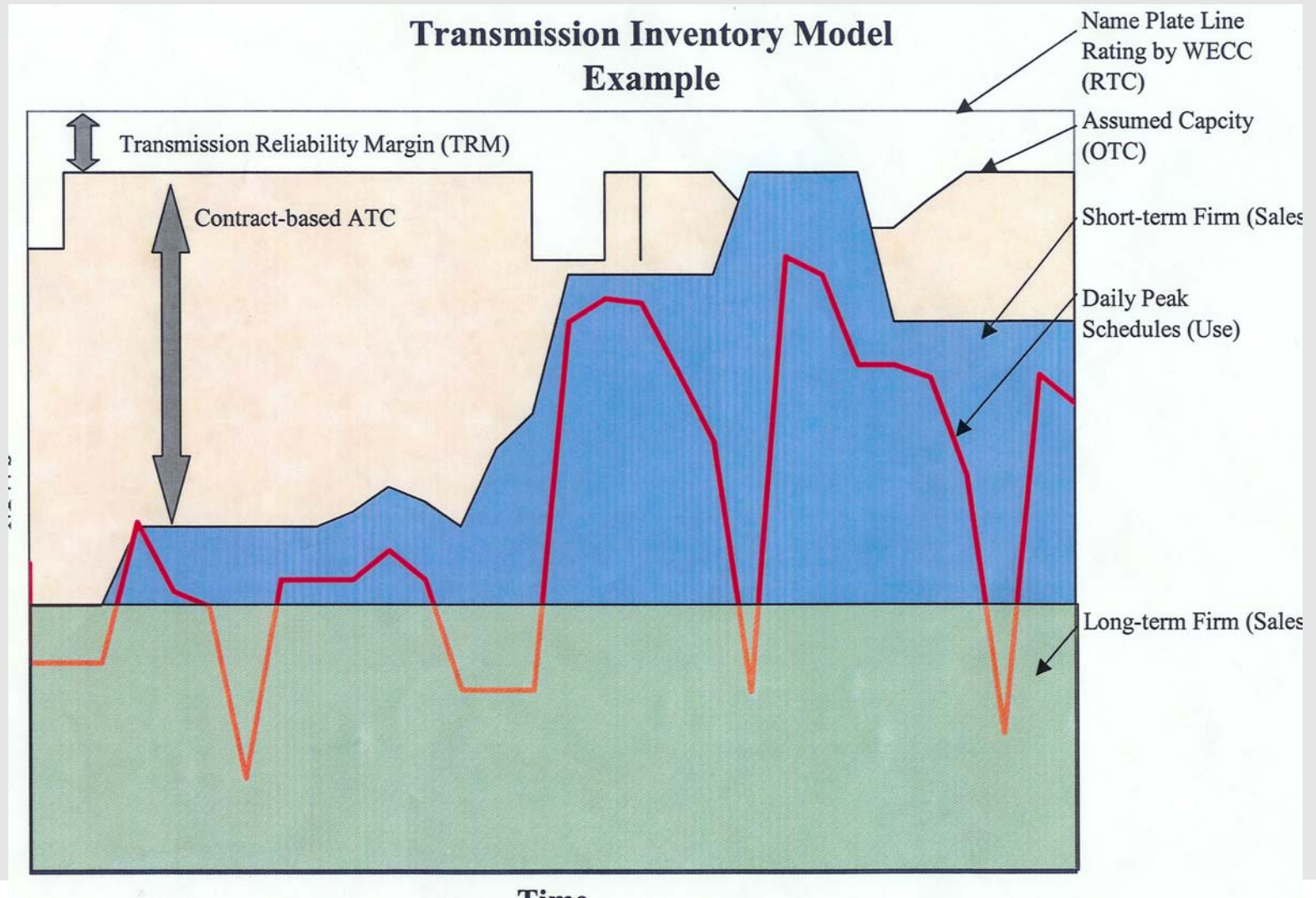Axian
Software Consulting and Training
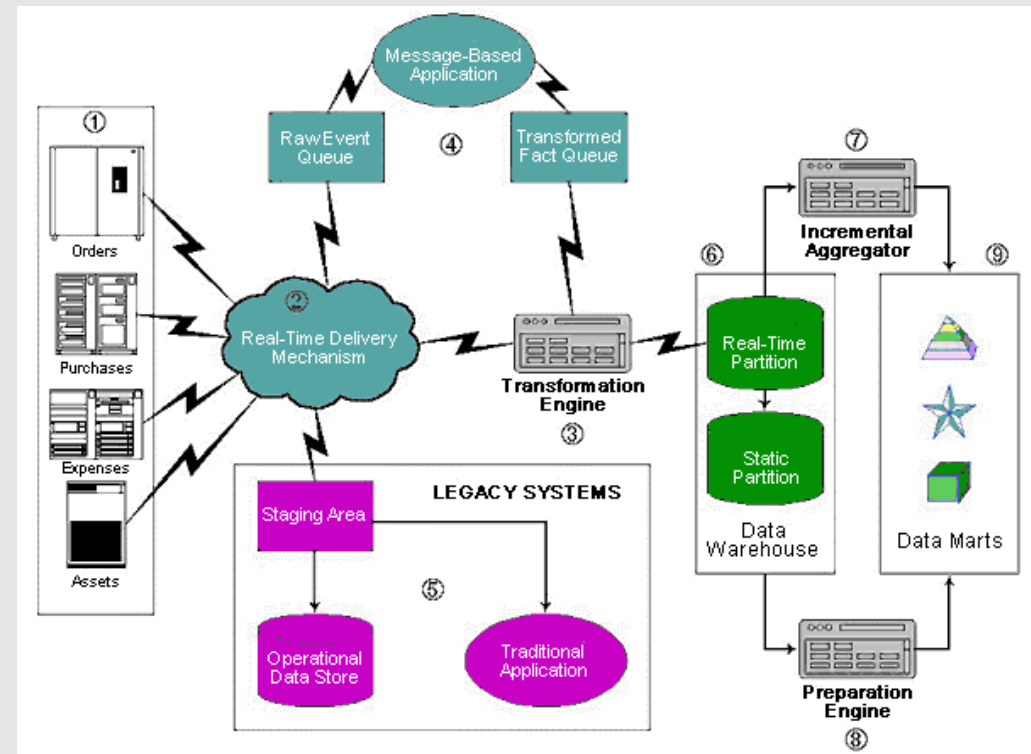
RT DW Components – Incremental Aggregator

- At regular intervals, aggregator engine runs against real-time partition to create intermediate totals and statistics

- Aggregates cascaded into static partition

- Aggregation done more continuously to support higher volume than bulk processing

- 15 min for securities, 30-60 minutes for product orders

Axian
Software Consulting and Training

Transmission Inventory Model Example
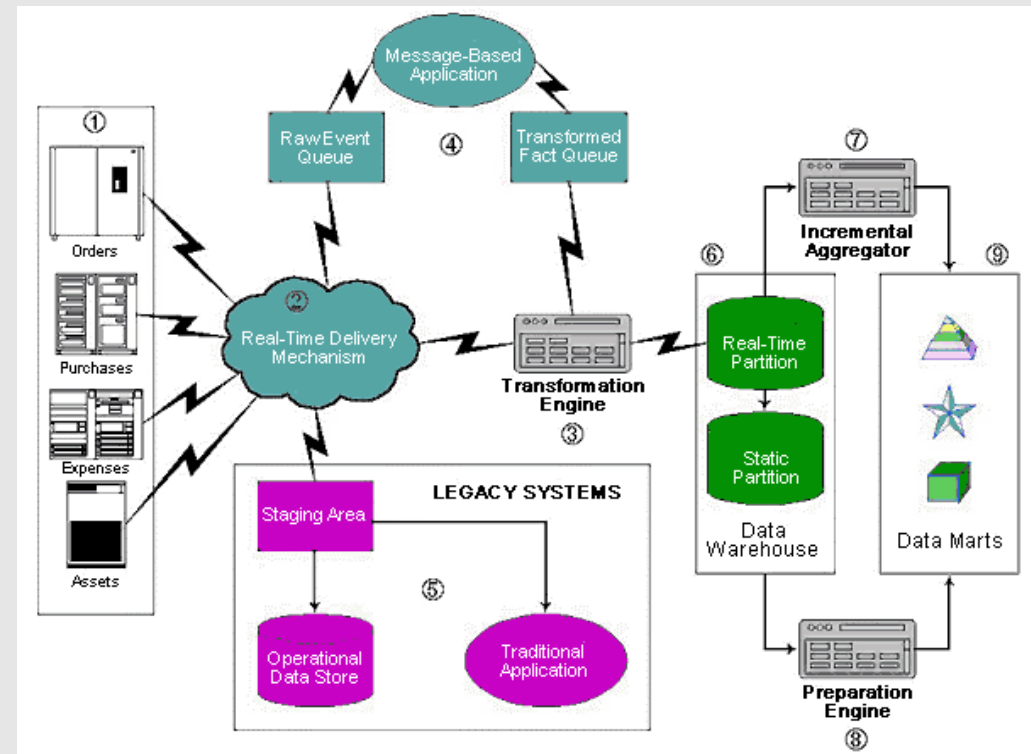
Review

- Introducing Data Warehousing

- Defining OLAP Solutions

- Understanding Data Warehouse Design

- Understanding OLAP Models

- Applying OLAP Cubes

- Real-Time OLAP

More Info

- www.axian.com

- www.datawarehouse.com

- www.dw-institute.com

- Data Warehouse Design Considerations
  (http://msdn.microsoft.com/library/default.asp?url=/library/en-
  us/dnsql2k/html/sql_dwdesign.asp)

- Real-time Data Warehouse: Real-time Data Warehousing Defined
  (http://www.dmreview.com/master.cfm?NavID=198&EdID=1522)

- Real-time Data Warehousing: Challenges & Solutions
  Whitepaper (http://www.claraview.com/claraview-rtw.pdf)

- PowerPoint slides from this presentation
  (http://www.impartica.com/rtdw.pdf)

- Real Time Data Warehouse Challenges and Solutions
  Table(http://www.impartica.com/rtdwsoln.pdf)

Axian
Software Consulting and Training