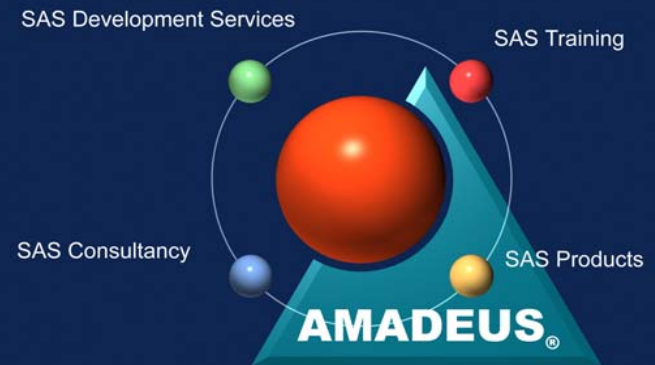


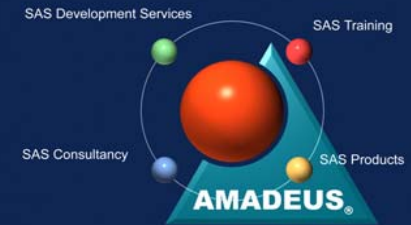


Effective Strategies for Data Warehousing

Katie Essam

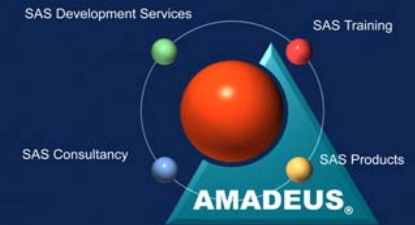
26th September 2001

Overview



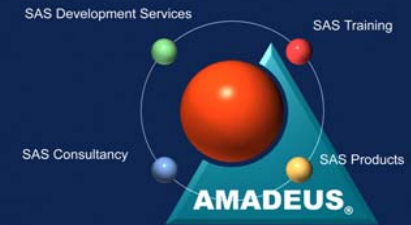
- What is a Data Warehouse?
 - ◆ Collection of data extracted from 1 or more sources for purpose of query and analysis
- Stages of building a Data Warehouse
 - ◆ Design
 - ◆ Implementation
 - ◆ Reporting/Access
 - ◆ Maintenance

Design



- Issues to address
 - ◆ Needs to answer business queries
 - ◆ Must provide efficient access
 - ◆ Ease of maintenance
 - ◆ Scalability

Design - What is a Logical Data Model?



- Logical Data Model combines the two most important components
 - ◆ Business Requirements
 - ◆ Data Structure
- Graphical representation of information required by business area

Design - What is a Logical Data Model?



Example

Customer

Name

Address

Store Card

Sales

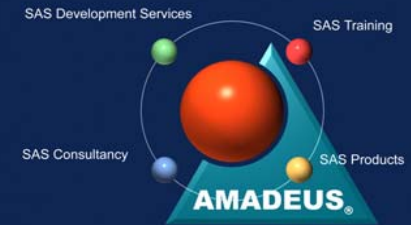
Branch

Product

Quantity

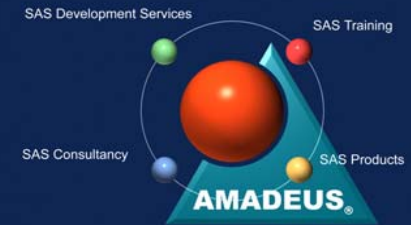
Price

Design - Why should one be used?



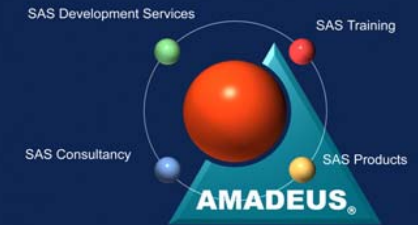
- Independent of technology
 - ◆ Model remains true to business needs regardless of underlying technology used
- Bridge between business users and DBA team
 - ◆ Confirms users understanding of Q's to answer and provides starting point

Design - How does it relate to the data?

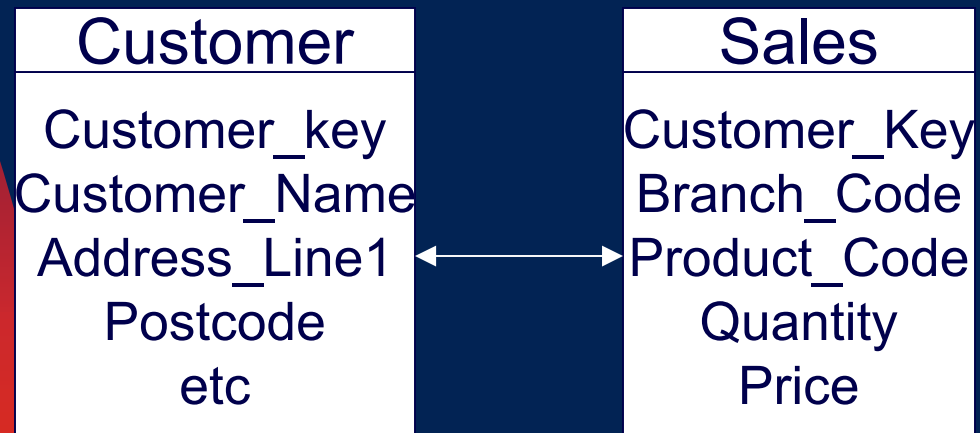


- Physical model is built based on the logical model
- Used to define appropriate
 - ◆ Keys
 - ◆ Indices
 - ◆ Integrity checks
- Based on technology available
- Considerations – Storage, maintenance

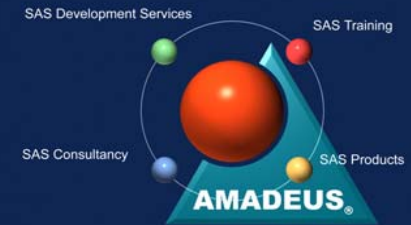
Design - How does it relate to the data?



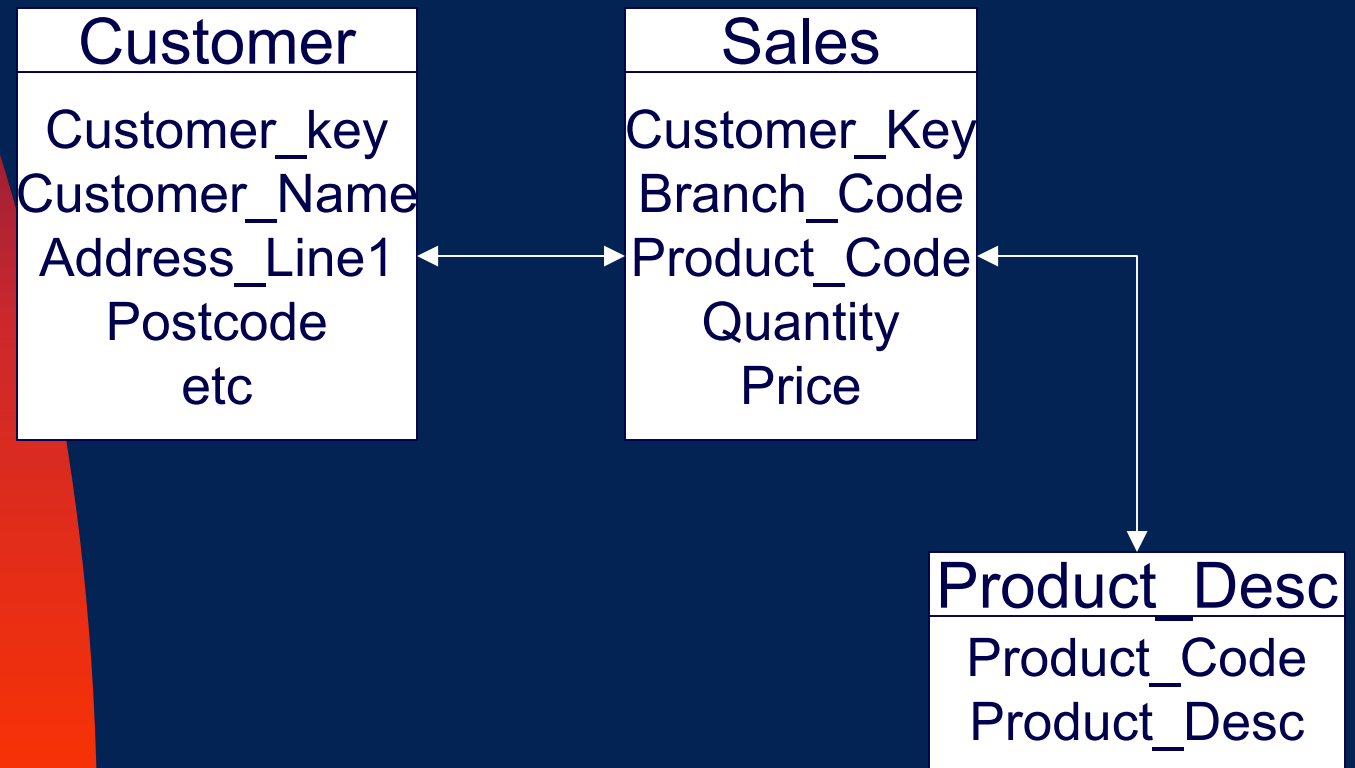
Example of relationship



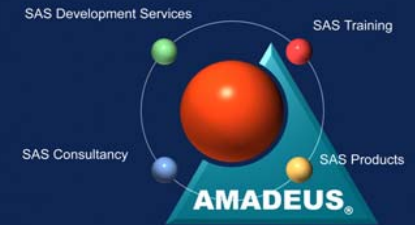
Design - How does it relate to the data?



Example of relationship

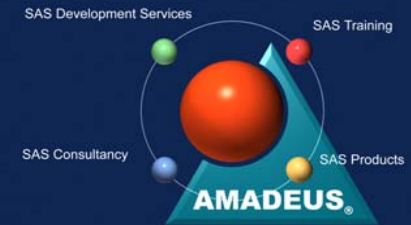


Design - How does it relate to the data?



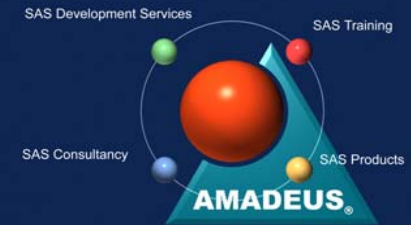
- Example of maintenance considerations
 - ◆ Customer and sales are stored separately as need to be refreshed at different intervals

Design - How should the data be structured?



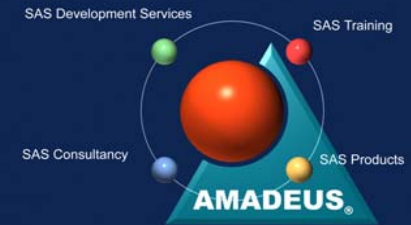
- Design must be twofold
 - ◆ Efficient access to answer business questions
 - ◆ Ease of maintenance
- May involve
 - ◆ De-normalisation
 - ◆ Pre-summarisation
- These processes may help provide first, too extreme will hinder second

How should the data be structured?



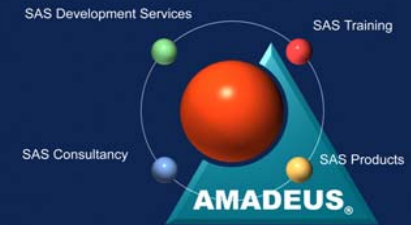
- Where do the feeds for the warehouse come from?
- What software is used to store the warehouse?
- How are the users to access the warehouse?
 - ◆ Directly/via results of manipulation

How should the data be stored?



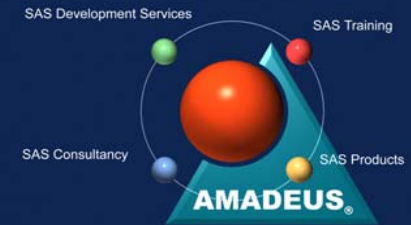
- Depends on the answers to the previous questions
 - ◆ SAS/ACCESS to underlying data source
 - ◆ SAS Tables – summarised
 - ◆ SAS/MDDB
 - ◆ HOLAP Groups
 - ◆ SPD Server

Implementing the Warehouse



- Pilot
 - ◆ Develop data mart for one area
 - ◆ Reduced data
 - ◆ Deployed to few users
- 'Proof of Concept'
- Scalable
 - ◆ Volumes of data
 - ◆ Additional data marts
 - ◆ 1,000's of users

Accessing the Warehouse



■ Tools

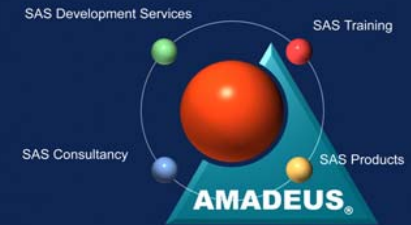
- ◆ Base SAS Software
- ◆ Client/Server
- ◆ SAS/AF, SAS/EIS
- ◆ Internet
 - ★ SAS/IntrNet
 - ★ HTML, Javascript, Java
 - ★ Web/AF, Integration Technologies

Maintaining the Warehouse

- What, when and where?
- Warehouse Administrator



The Future of the Warehouse



- Increasing Volumes
 - ◆ Data
 - ◆ Departments
 - ◆ Users
- E-Business
- CWM (Common Warehouse Metamodel) inc XML

Summary

- Communication
 - ◆ Business users
 - ★ Logical Model
 - Business Model
- Accessibility
- Scalability



Contact information

Katie Essam
Amadeus Software Ltd
Orchard Farm
Leafield
Oxon
OX29 9PG

www.amadeus.co.uk

