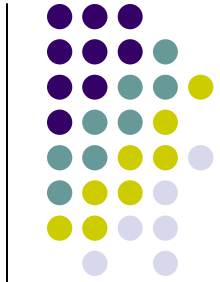


Data Vault Modeling

The Next Generation DW Approach

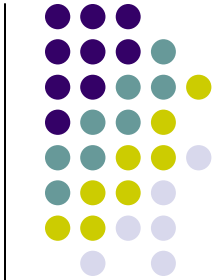
Presented by: Siva Govindarajan

GLOBYTES

HTTP://WWW.GLOBYTES.COM

SIVA02@GLOBYTES.COM

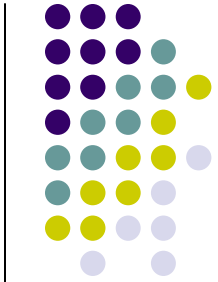
Data Vault Modeling
Agenda



- Introduction
- Data Vault Place in Evolution
- Data Warehousing Architecture
- Data Vault Components
- Case Study
  - Typical 3NF / Star Schema
  - Data Vault Approach
  - Temporal Data + Auditability
- Implementation of Data Vault
- Conclusion

GLOBYTES

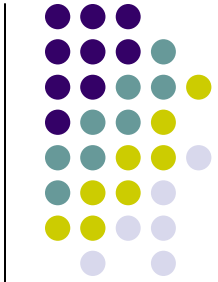
Introduction



- Data Vault concept originally conceived by Dan Linstedt
- Enterprise Data Warehouse Modeling approach.
- Hybrid approach - 3NF & Star Schema
- Adaptable to changes
- Auditable
- Timed right with Technology

GLOBYTES

Data Vault Place in Evolution

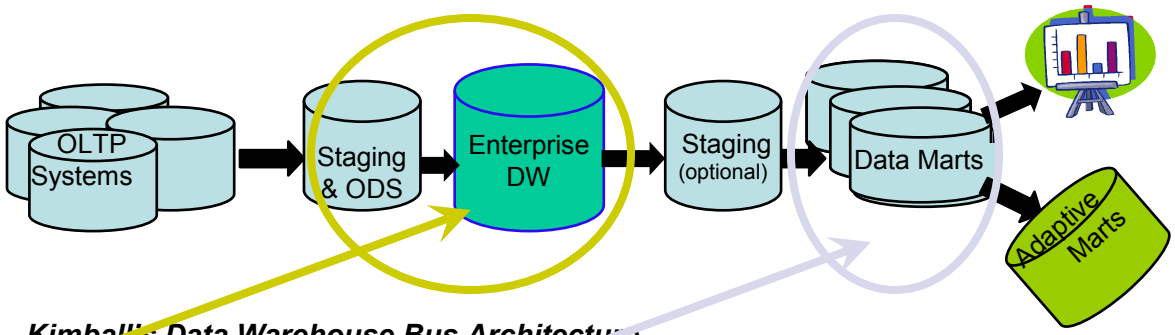


- 1960s - Codd, Date et. al Normal Forms
- 1980s - Normal Forms adapted to DWs
- 1985+ - Star Schema for OLAP
- 1990s - Data Vault concept developed Dan Linstedt
- 2000+ - Data Vault concept published by Dan Linstedt

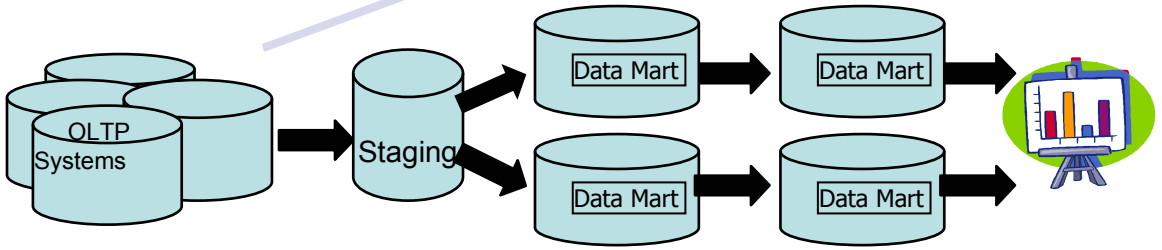
GLOBYTES

Data Warehousing Architecture

Bill Inmon's Data Warehouse Architecture

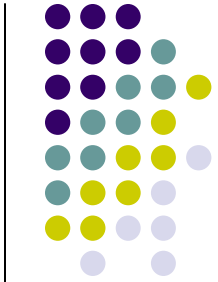


Kimball's Data Warehouse Bus Architecture



GLOBYTES

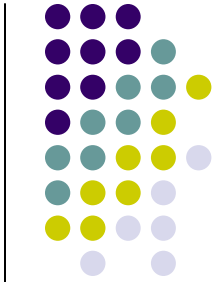
Data Vault Components



- Hub Entities
  - Candidate Keys + Load Time + Source
- Link Entities
  - FKs from Hub + Load Time + Source
- Satellite Entities
  - Descriptive Data + Load Time + Source + End Time
    - Dimension = Hub+ Satellite
    - Fact = Satellite + Link [+ Hub]

GLOBYTES

Case Study

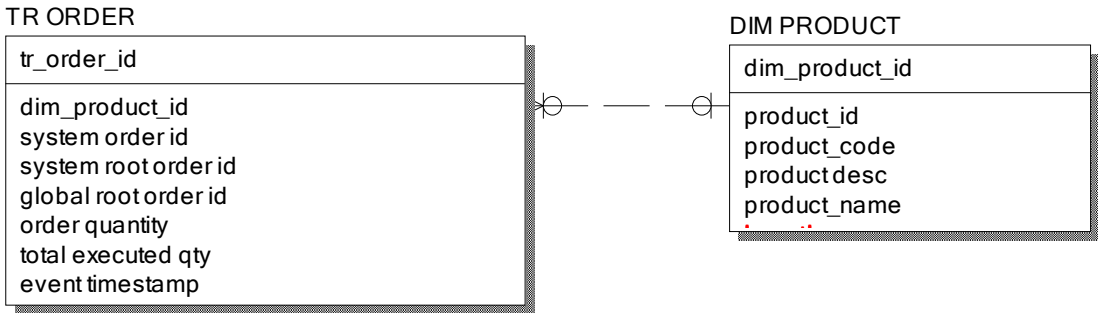


- Modeling Approaches
  - Typical 3NF / Star Schema
  - Data Vault approach
- Scenario
  - Typical Fact and Dimension
  - New Dimension data
  - Reference Data change - 1:M to M:M
  - Additional attributes to Dimension

GLOBYTES

Typical 3NF / Star Schema

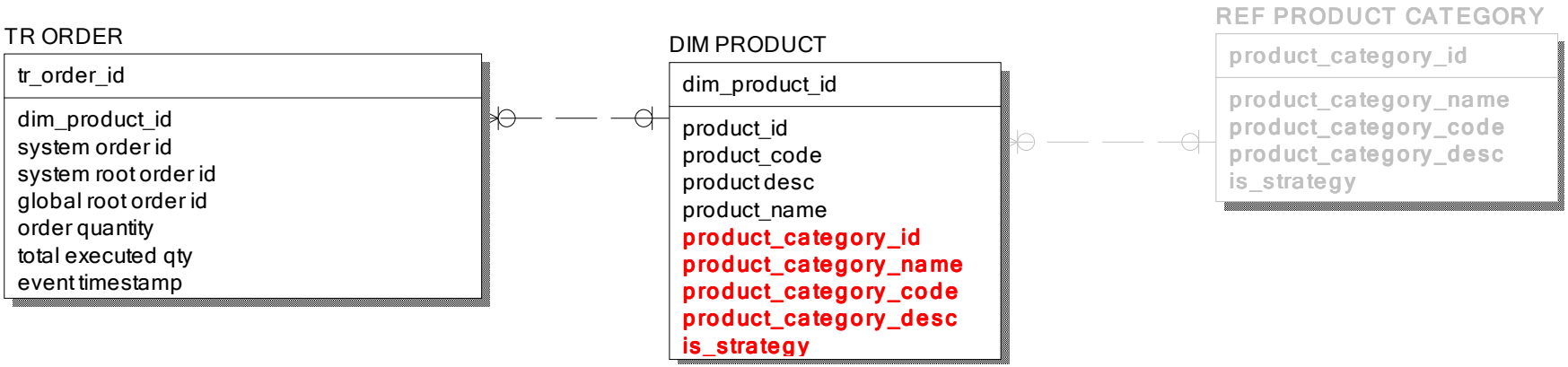
1. Typical Fact and Dimension



GLOBYTES

Typical 3NF / Star Schema

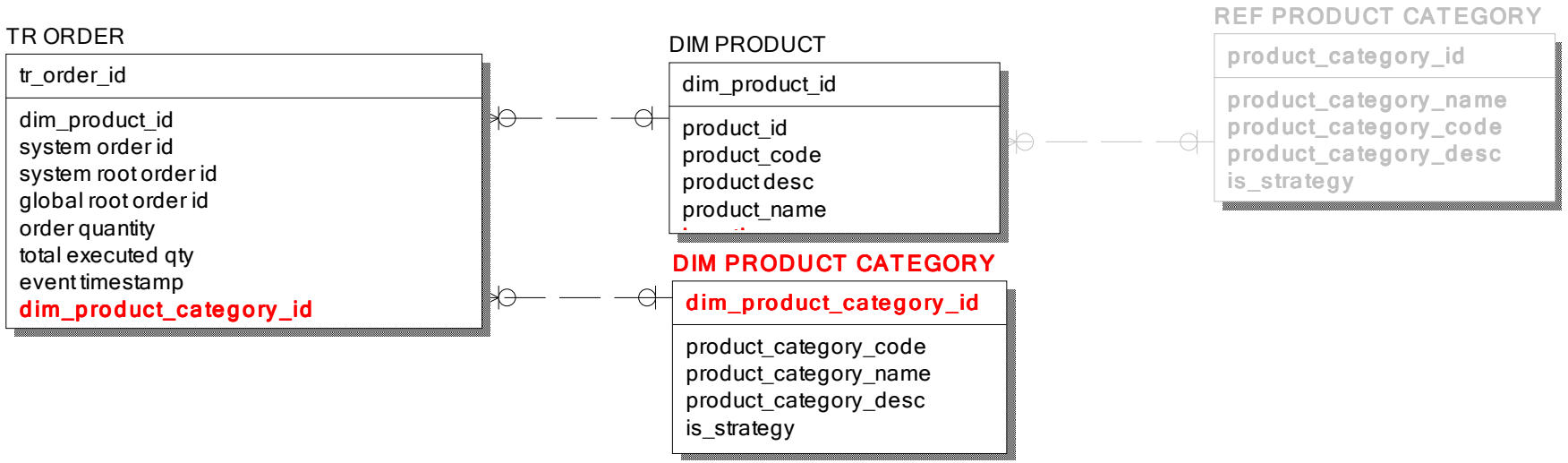
2. New Dimension Data



GLOBYTES

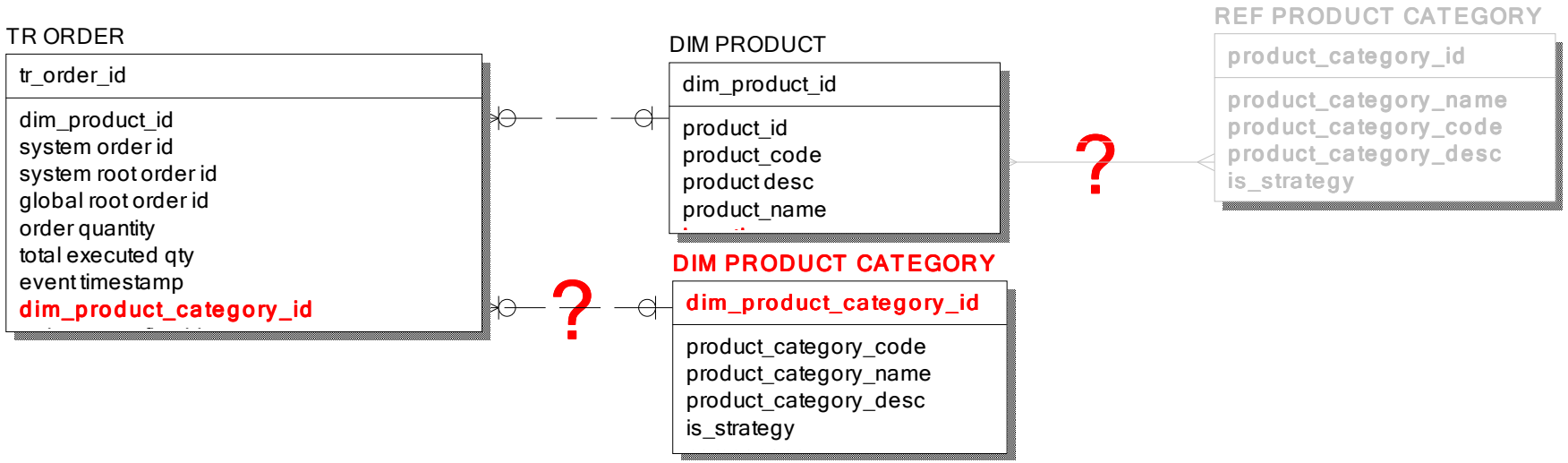
Typical 3NF / Star Schema

2. New Dimension Data



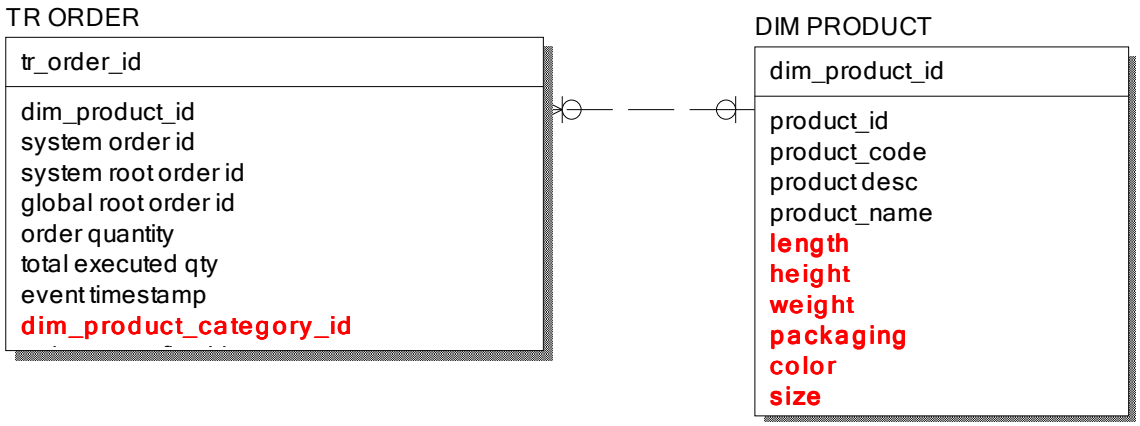
Typical 3NF / Star Schema

3. Reference 1:M to M:M Change



Typical 3NF / Star Schema

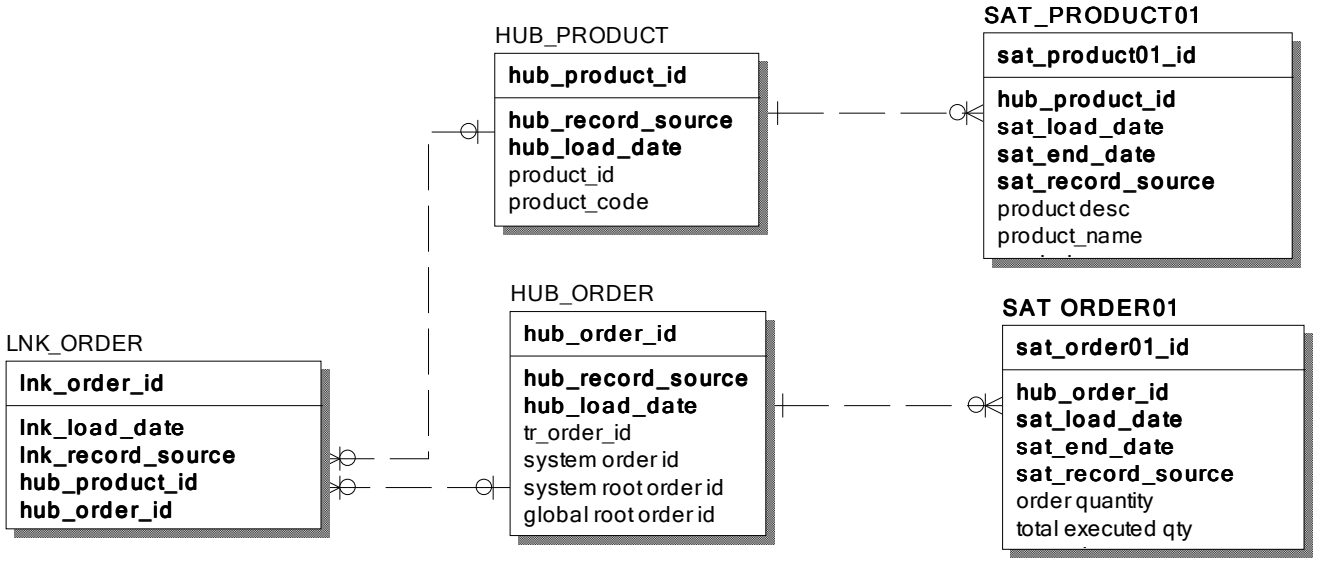
4. Additional attributes to Dimension



Data Vault Approach

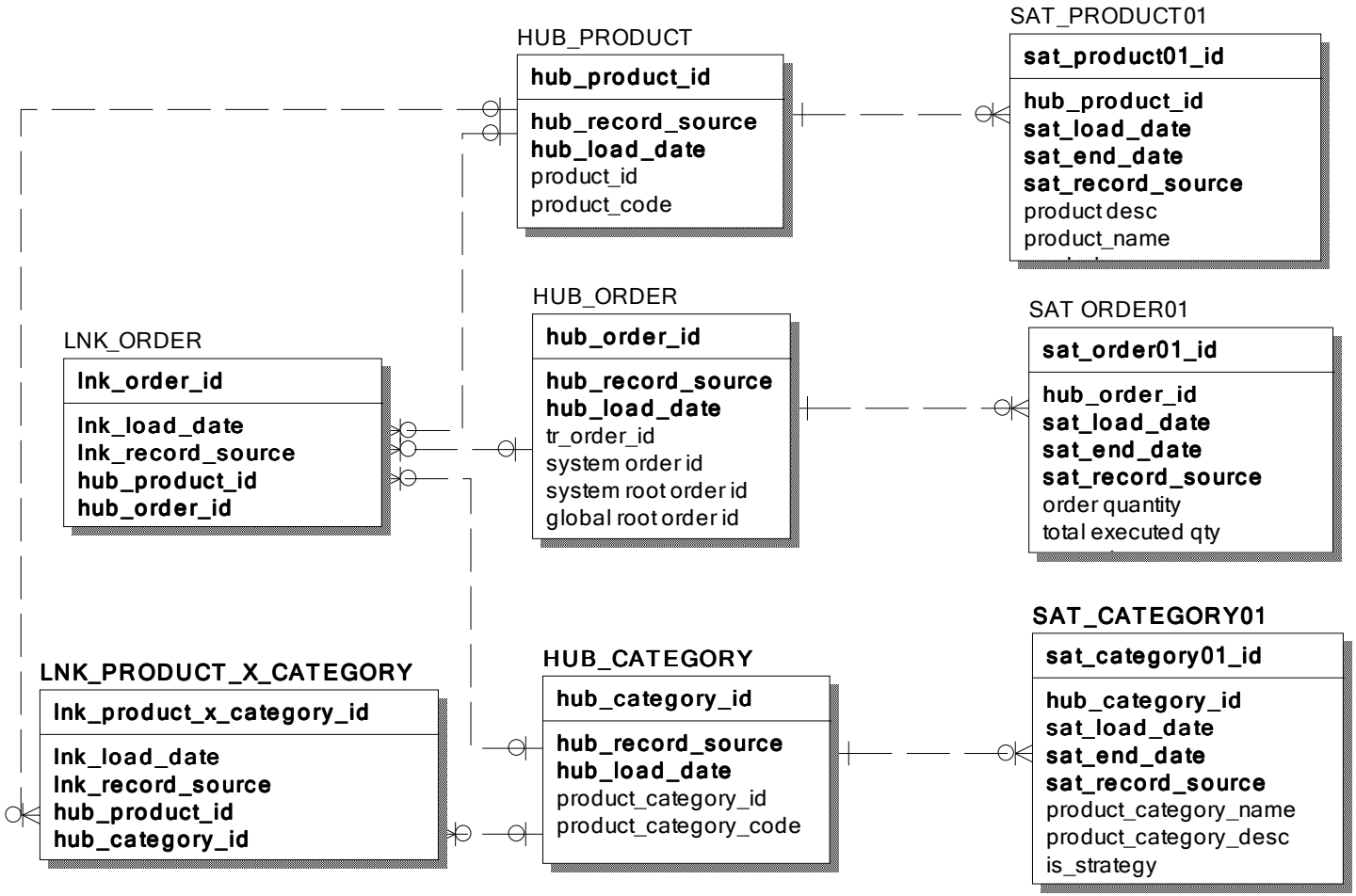
1. Typical Hub, Satellite and Link

(Replacing Fact/Dimension)



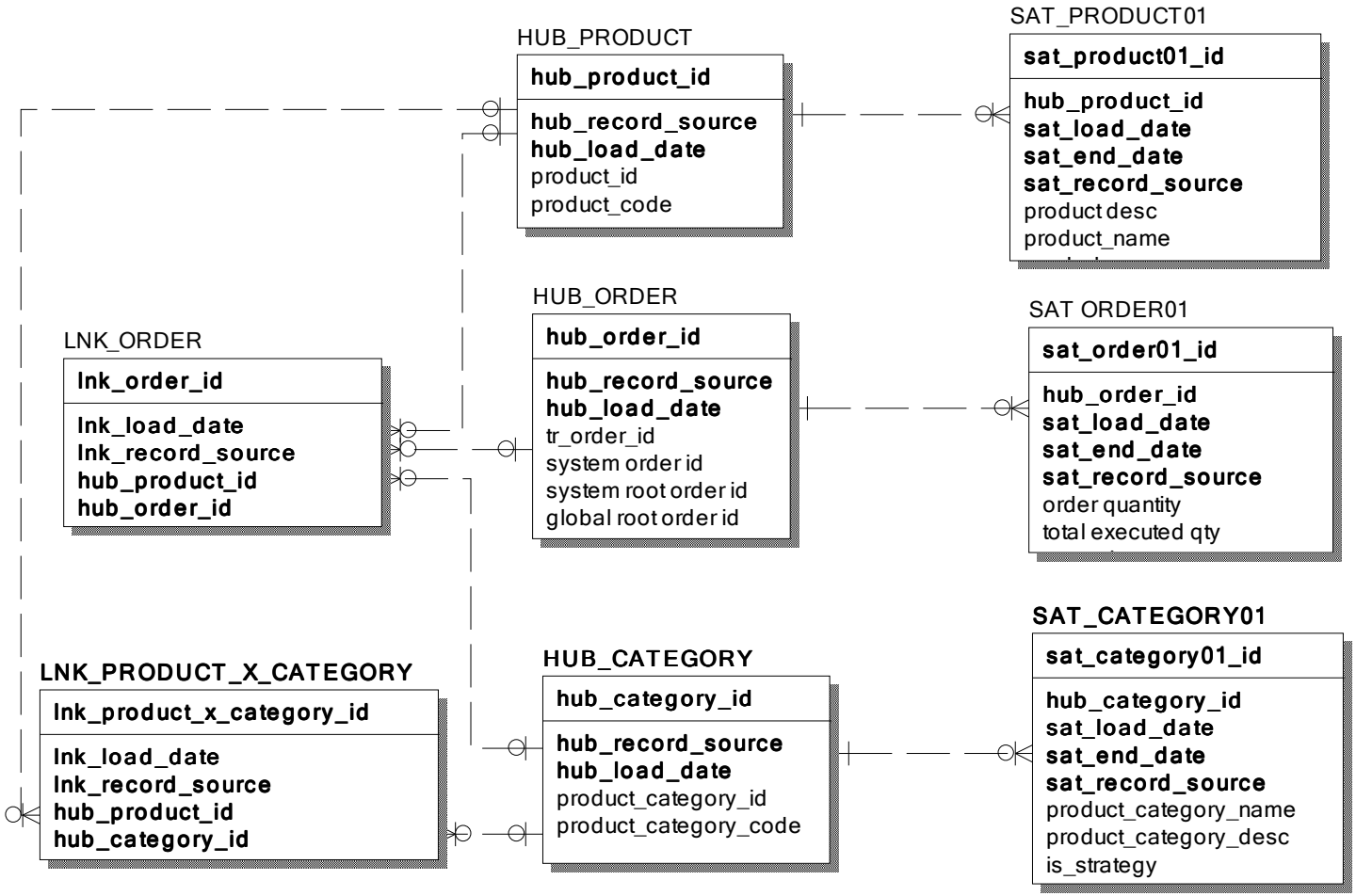
Data Vault Approach

2. New Dimension Data



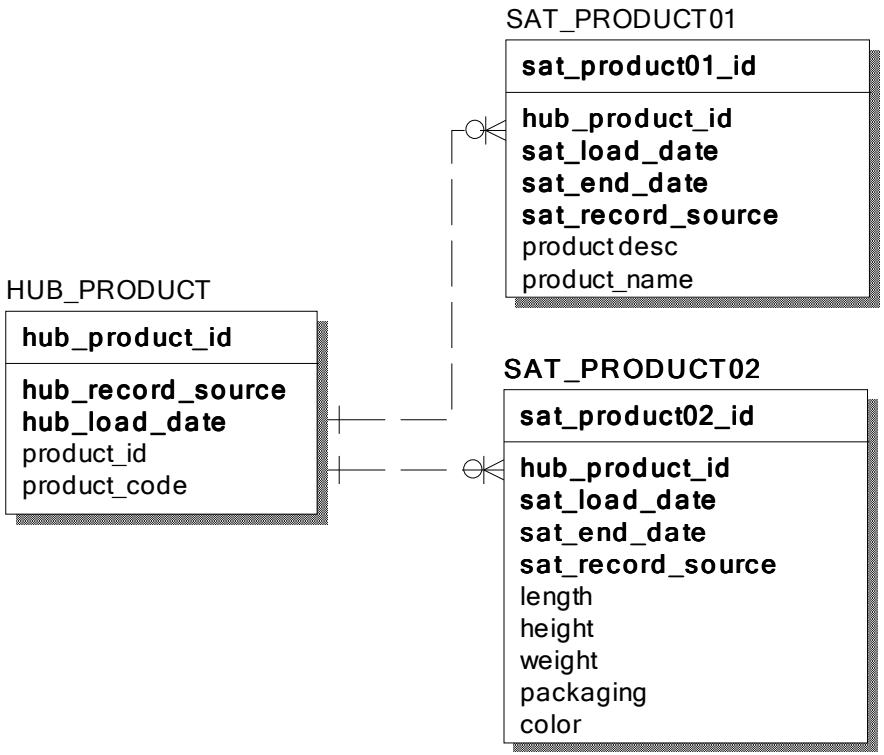
Data Vault Approach

3. Reference 1:M to M:M Change



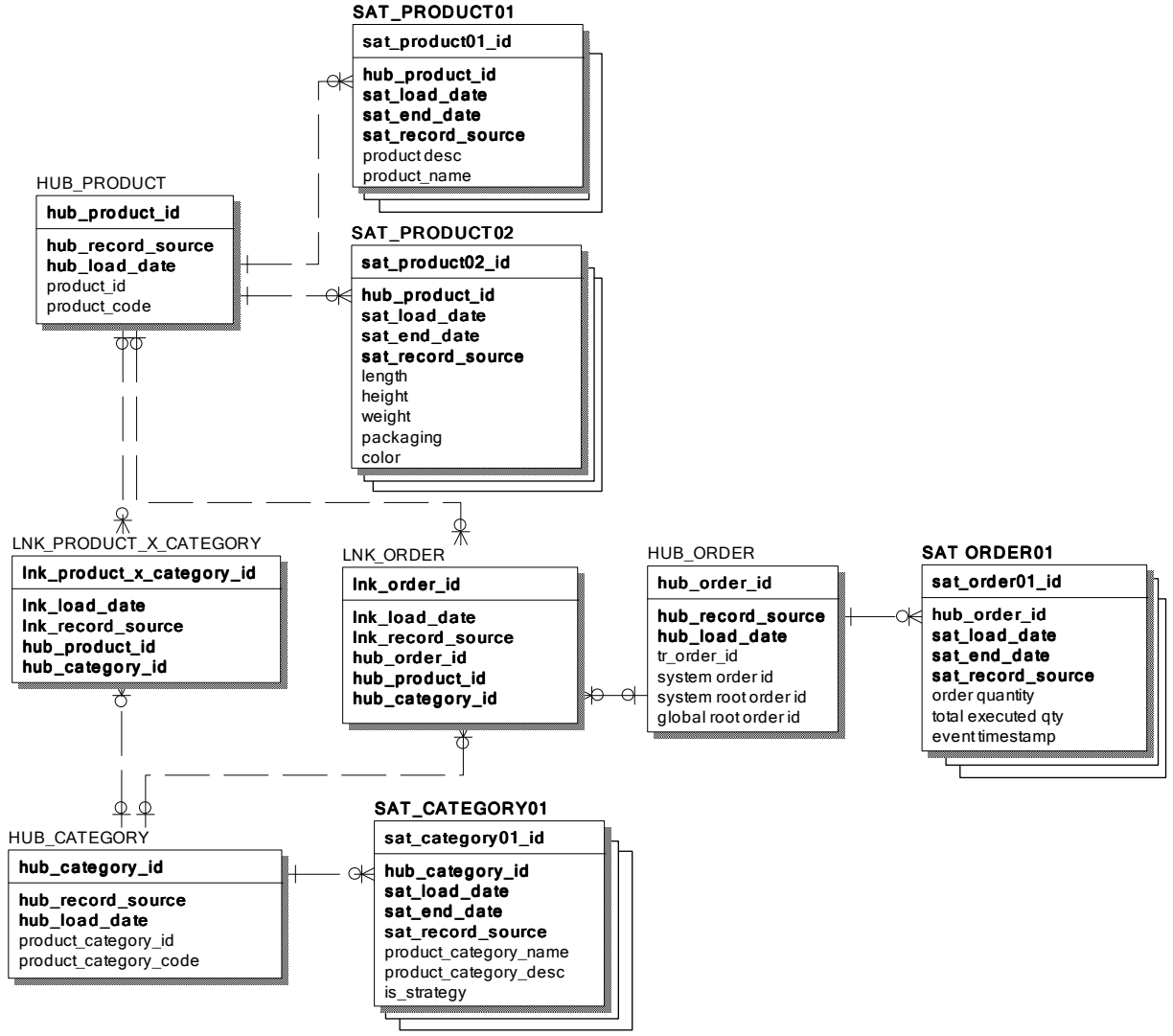
Data Vault Approach

4. Additional attributes to Dimension

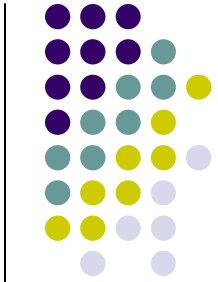


GLOBYTES

Temporal Data + Auditability



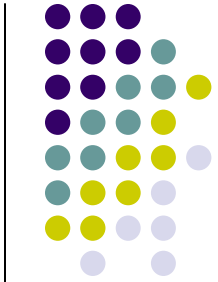
Implementation of Data Vault



- Make it Simple
  - View for Dimension
  - View For Fact
- Data Loading
- Take advantage of Technology
  - DW Appliances
  - High-throughput storage devices
  - RDBMS Features

GLOBYTES

Make it Simple



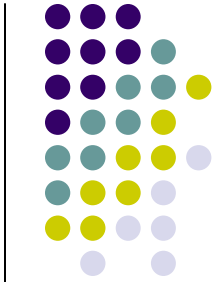
- Create views for Dimension

From our Demo model, Join Product related Hubs and Satellites to build View for Product Dimension as shown below :

- HUB_PRODUCT
- SAT_PRODUCT01
- SAT_PRODUCT02

GLOBYTES

Make it Simple



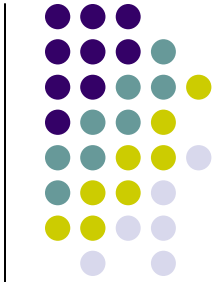
- Create Views For Fact

Join Order related Hubs, Satellites and Links to build View for ORDER Fact using ORDER related Hubs/ Satellites and Links.

- HUB_ORDER
- SAT_ORDER01
- LNK_ORDER

GLOBYTES

Data Loading



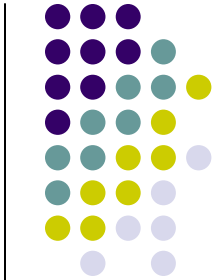
Typically data loading for a Data Vault is in the following sequence.

- Hubs for Dimensions
- Links for Dimensions
- Satellites for Dimensions
- Hubs for Fact (if any)
- Links for Fact
- Satellites for Fact

Refer to Data Vault Series article 5 of Linstedt for further details

GLOBYTES

DW Appliances

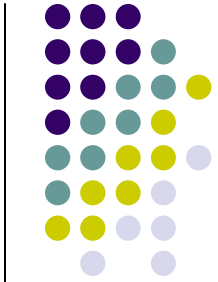


Next wave of IT Data Warehouse hardware solutions are the Data Warehouse Appliances. Take advantage of the technology where possible. Few DW Appliances in the market :

- Oracle Exadata
- Netezza
- Teradata
- RedBrick
- GreenPlum

GLOBYTES

High-throughput storage devices

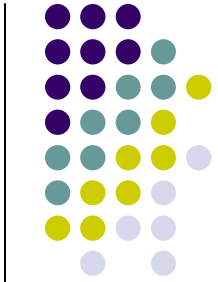


Utilize Storage devices designed for DW / OLAP applications. Following are some examples only and would change over time. Please do your research based on your sizing requirements.

- EMC CLARiiON Storage family
- Texas Memory systems RamSan – Solid State devices
- Hitachi USP / VSP Storage family
- Other Hybrid solutions with High performance storage and Solid State devices

GLOBYTES

RDBMS Features



Some Oracle Features catering DW / OLAP applications:

- Exadata Smart Scans
- OLAP Based Materialized views
- Partitioning Reference partition
- Advanced data compression
- Automatic degree of parallelism (ADOP).
- Star query optimization
- Oracle OLAP/DWH Features.

GLOBYTES

Conclusion

In this presentation we have addressed high level overview of Data Vault Architecture. Topics discussed are

- Data Vault concept and architecture
- Data Vault Components such as Hubs, Satellites and Link tables
- Typical modeling challenges with traditional modeling approaches
- How those challenges could be handled using Data Vault Modeling Approach.
- Auditing and Temporal data capture using DV Approach.
- And finally, some implementation details

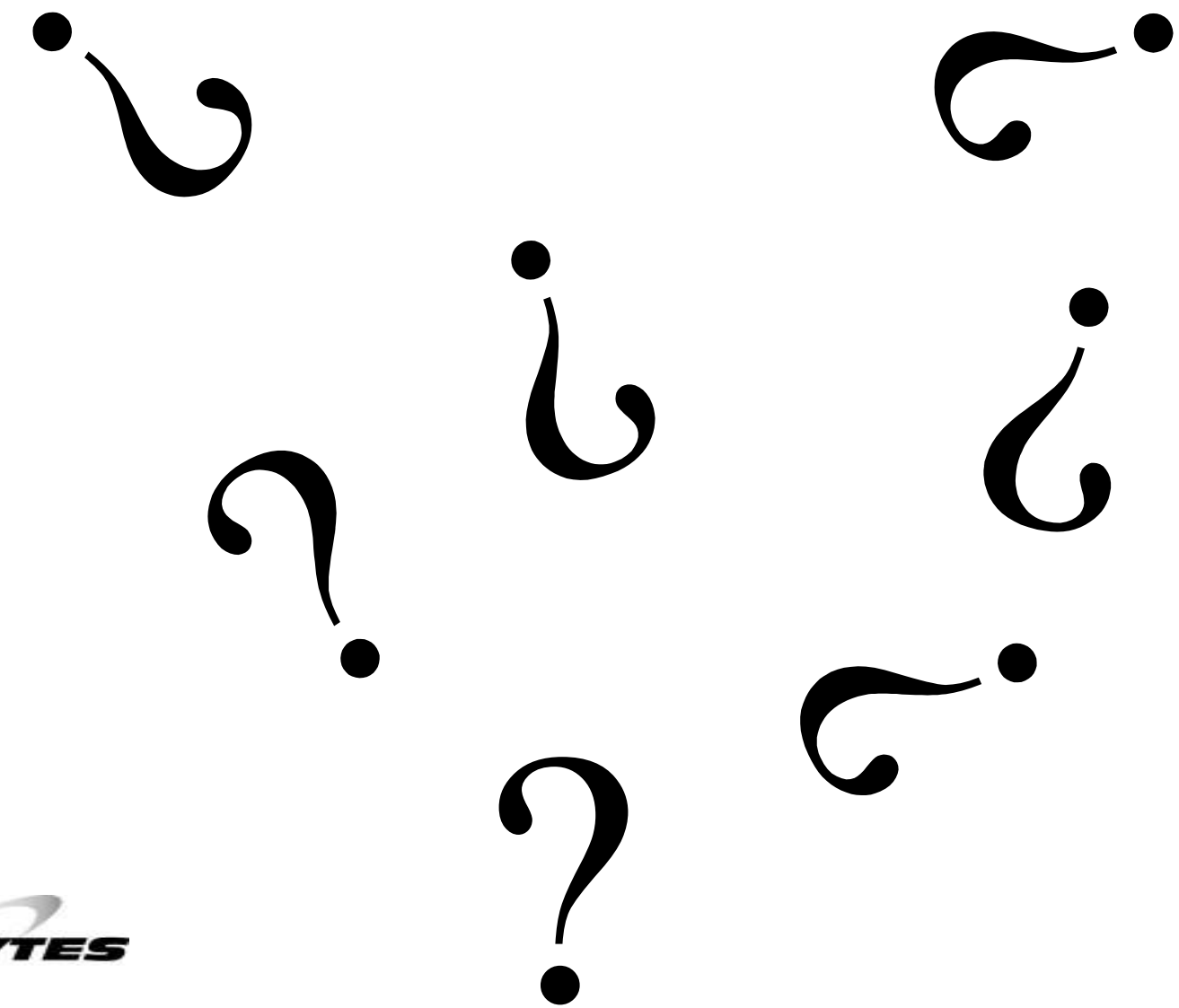
If interested in learning more, try

Dan Linstedt's special coaching area at:

http://danLinstedt.com/my-coach.

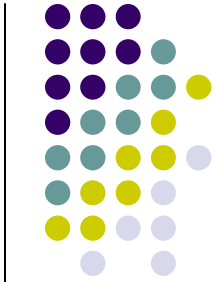
GLOBYTES

Questions ?



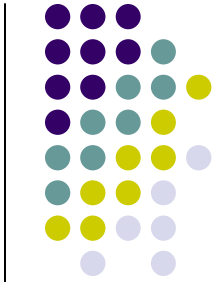
GLOBYTES

References



- Data Vault Series articles by Dan E. Linstedt -- http://www.tdan.com
- Referred few articles from Genesee Academy -- http://geneseeacademy.com
- Articles and books by Ralph Kimball and Bill Inmon from multiple sources.
- Technical Documents from http://www.oracle.com/technetwork

GLOBYTES

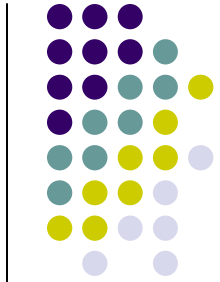


Special Thanks to

- Dan Linstedt for review and corrections to the article.
- My colleague Mark Bruscke for his assistance in preparing the article.

GLOBYTES

The End



Contact Email :

siva02@globytes.com

GLOBYTES