

Database Vocabulary

Is Your Data Set a Dimension (LookUp) Table,
a Fact Table or a Snapshot (Report)?

Ronald J. Fehd
Centers for Disease Control

PharmaSUG 2009

Outline

- 1 Introduction
- 2 Anchors
 - Transaction
 - Location
 - Receipt
- 3 Database Vocabulary
 - Dimension
 - Reports
- 4 Discovery or Data Review

Sound Byte

Information is **the** difference
that makes **a** difference

Anchors

- describe an event
- location
- receipt
- monthly statements
- budget

Transaction: Describe an Event

- who** actors: subject, to other
- what** object
- why** replace with verb
- when** time-stamp
- where** location

Location: Where, GPS

- latitude
- longitude
- elevation
- when: 4th dimension: time

Location: Where, snail mail

- country
- province or state
- city
- street name
- street number
- detail: apt, or floor and room

idea: granularity

Location: When: time

- years
- months
- days
- hours
- minutes
- seconds

idea: granularity

Receipt: Sale or Purchase

where	location
when	time
what	object
how many	number or quantity
price	each or unit
total	units * cost

Receipt is.a Transaction

- where: location:
 - store id: foreign key
 - time
- what: facts
 - product id: foreign key
 - quantity, measurement
 - is.a number, for calculations

action: append

Names of Things

- tables
- columns
- reports

Types of Tables

- dimensions
- facts
- snapshots:
 - periodic
 - accumulating

Dimension: Store

Lookup

- primary key:
 - row number
 - sequence number
- information:
 - address
 - city
 - state
 - zip

Reports: Periodic Snapshots

- entity-id
- date
 - start
 - end
- balance
 - start
 - end

action: append

Reports: Accumulating Snapshots

Yearly Budget

- entity-id
- month.01
- month.02
- ...
- month.12

Process Tracking

- entity-id
- event.1.begin
- event.1.end
- ...
- event.N.begin
- event.N.end

action: update

Cardinality Ratio

$$\text{CR} = \frac{\text{n-levels}}{\text{n-obs}}$$

foreign key 0 <= CR <= 0.5

facts 0.5 <= CR < 1.0

primary key CR = 1.0

see also: SmryEachVar

Cardinality Ratio

$$\text{CR} = \frac{\text{n-levels}}{\text{n-obs}}$$

foreign key 0 <= CR <= 0.5

facts 0.5 <= CR < 1.0

primary key CR = 1.0

see also: SmryEachVar

Conclusion

Database vocabulary necessary for:

- Designing Data Structure
- Proc Transpose
- Summarization
- EG
- SQL

Author Information

Ronald J. Fehd **RJF2@cdc.gov**
Stat Software HelpDesk
SAS Site Rep
Centers for Disease Control
Atlanta, GA, USA

Presentation pdf: \LaTeX Beamer class

References

- paper
 - SmryEachVar data review suite
 - paper: SGF 2008.003
 - page: SAS community wiki
- book
 - Kimball and Ross
 - The Data Warehouse Toolkit