Week 11 Topic: Dimensional Modeling: More

Dimension Patterns and Considerations

Dealing with NULL

➤ We will cover

Junk Dimension

Outriggers

IIT School of Applied Technology

ILLINOIS INSTITUTE OF TECHNOLOGY

information technology & management

526 Data Warehousing

Week 11 Presentation April 2, 2019

Dimension Value Band

Bridge Tables

Monster Dimension

ILLINOIS INSTITUTE OF TECHNOLOGY

School of Applied Technology

Dealing with Nulls

- ➤ NULL dimension attributes
- unexpected query results (e.g. Strongly discouraged to avoid invalidating index strategy)
- Unknown, Invalid, To be determined,... ■ Use default values instead − N/A
- ➤ NULL facts
- Use ONLY IF it truly means N/A, Unknown, and Invalid, not zero

School of Applied Technology Operations with Nulls Dealing with Nulls: LLINOIS INSTITUTE OF TECHNOLOGY

> Operations with Null can be tricky sometimes

FROM dual; -- null SELECT 1 + null col

SELECT CASE WHEN null = null THEN 1 ELSE 2 END AS col FROM dual; -- 2 SELECT CASE WHEN null is null THEN 1 ELSE 2 END AS col SELECT CASE WHEN 1 IS null THEN 1 ELSE 2 END AS col FROM dual; -- 1

SELECT DECODE (null, null, 1, 2, 3) as col FROM dual; -- 1

FROM dual; -- 2

SELECT DECODE (1, null, 1, 2, 3) as col FROM dual; -- null

School of Applied Technology

Dealing with Nulls:

Operations with Nulls (Cont'd)

➤ Operations with Null can result in missing information

SELECT deptid, empid, month_pay*12+bonus AS annual_salary SELECT deptid, SUM(annual_salary) AS annual_salary_by_dept SELECT 'd01' AS deptid, 'e0001' AS empid, SELECT 'd01' AS deptid, 'e0002' AS empid, 100000 AS month_pay, null AS bonus 10000 AS month_pay, 1000 AS bonus 121000 -- \$1,200,000 missing! FROM dual UNION ALL DEPTID ANNUAL_SALARY_BY_DEPT FROM dual GROUP BY deptid d01

ILLINOIS INSTITUTE OF TECHNOLOGY

School of Applied Technology

Rule of Thumb in Data Warehousing Dealing with Nulls:

> NULL fact table foreign keys

- No NULL is allowed as it breaks referential integrity
- Substitute key to special dimension row (a.k.a. dummy dimension row)

NULL dimension attributes A

- Strongly discouraged to avoid unexpected query results (e.g. invalidating index strategy)
- Use default values instead N/A, Unknown, Invalid, To be determined,...

NULL facts A

 Use ONLY IF it truly means N/A, Unknown, and Invalid, not zero

ILLINOIS INSTITUTE OF TECHNOLOGY

School of Applied Technology

Operations with Nulls (Cont'd) Dealing with Nulls:

Aggregate functions handle Null gracefully

SELECT 'rm2001' AS room_no, 's0001' AS student_id SELECT 'rm2001' AS room_no, 's0002' AS student_id, SELECT room_no, AVG(math) AS avg_math_by_room, AVG(writing) AS SELECT room_no, student_id, math, writing 90 writing 100 AS math, null writing ROOM_NO AVG_MATH_BY_ROOM AVG_WRITING_BY_DEPT FROM dual UNION ALL 80 AS math, FROM dual 90 avg_writing_by_dept GROUP BY room_no; rm2001

LLINOIS INSTITUTE OF TECHNOLOGY

School of Applied Technology

ø

Avoid Too Many Dimension "Centipede" Fact Tables

Promotion Media Type Dimension Promotion Reduction Type Dimension Package Type Dimension Department Dimension **Brand Dimension** POS Retail Sales Transaction Fact Extended Discount Dollar Amoun Promotion Meda Type Key (FK) Extended Cost Dollar Amount Store Region Key (FK) Store Floor Plan Key (FK) Category Key (FK)
Department Key (FK)
Package Type key (FK)
Store Key (FK) Store District Key (FK) Figure 3.17 Centipede fact table with too many norm Year Key (FK) Fiscal Year Key(FK) iscal Month Key (FK) Store State Key (FK) Date Key (FK) Week Key (FK) Month Key (FK) Quarter Key (FK) Product Key (FK) Brand Key (FK) Store State Dimension Store County Dimension Store District Dimension Store Region Dimension Fiscal Year Dimension Quarter Dimension Year Dimension

ILLINOIS INSTITUTE OF TECHNOLOGY

Junk Dimensions (cont'd)

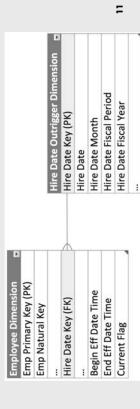
- flags/indicators into junk dimension ➤ Combine miscellaneous transaction
- SCD Type 2 is not necessary
- > Potentially less desirable alternative:
- Multiple fact table keys to low-cardinality dimensions
- ➤ Undesirable alternatives:
- Place flags/indicators directly in fact table as text facts or DDs

ILLINOIS INSTITUTE OF TECHNOLOGY

School of Applied Technology

Outriggers

- ➤ Dimension tables joined to other dimension tables
- outrigger to the employee dimension via role-➤ In this case, a date dimension serves as an playing
- Outriggers are acceptable in moderation but should be viewed as the exception rather than the rule



Customer Key (FK) Product Key (FK)

Payment Option Dim Payment Option Key (PK) Weather Type Key (PK) Payment Option Code Payment Option Desc Weather Type Code Weather Type Desc

More Foreign Keys... (FK)

Measures...

Payment Option Key (FK)

Weather Type Key (FK)

Weather Type Dim

More Foreign Keys... (FK) Junk Type Key (FK) Customer Key (FK) Product Key (FK) Measures...

Weather Type Code

Junk Type Key (PK) Junk Type Dim

Payment Option Code Payment Option Desc Weather Type Desc

2

School of Applied Technology ILLINOIS INSTITUTE OF TECHNOLOGY

Dealing with Rapidly Changing Monster Dimensions: Monster Dimensions

customer dimension (e.g. 30 million) with rapidly ➤ Imagine an insurance company with a big changing demographics (in green)

CUSTOMER DIMENSION Employment Risk Score Payment History Score Number Dependents Home Owner Status Customer Key (PK) **Customer Address Customer Name** Number Autos Age Group Transaction Date Key (FK) Transaction Type Key (FK) Insured Item Key (FK) **Transaction Amount** INSURANCE FACT Customer Key (FK) Coverage Key (FK)

The dimension table size can be easily doubled within a short period making this a rapidly changing monster dimension A

Dealing with Rapidly Changing Monster Dimensions:

Monster Dimensions (cont'd)

School of Applied Technology

School of Applied Technology ILLINOIS INSTITUTE OF TECHNOLOGY

Mini-Dimension to the Rescue Dealing with Monster Dimensions:

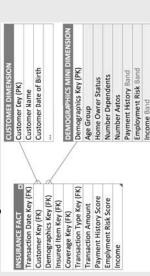
attributes into their

Break off the hot

own separate mini

dimension

➤ The solution is to break off the hot attributes into their own separate mini dimension



- The mini dimension contains one row for each possible combination of the attributes A
- Value bands are used in the mini-dimension to reduce the number of rows overall

2

School of Applied Technology ILLINOIS INSTITUTE OF TECHNOLOGY

Dealing with Rapidly Changing Monster Dimensions:

Monster Dimensions (cont'd

Customer dimension sample row:

Date of Birth Demographics mini-dimension sample row: 1984-02-10 Customer Name Demographics Key Age Group John Smith Customer Key 123456

Fact table sample row:

\$50,000 - \$59,999 \$50,000 - \$59,999 666,69\$ - 000,09\$

25 - 29

Income Band

		40331 123456 40430 123456
--	--	------------------------------

15

DEMOGRAPHICS MINI DIMENSION Employment RiskScore Payment History Score Number Dependents Home Owner Status Customer Key (PK) Customer Address CUSTOMER DIMENSION Customer Key (PK) Demographics Key (PK) Customer Name Customer Date of Birth Payment History Band Employment Risk Band Number Autos Number Dependents Home Owner Status Age Group Customer Name Number Autos Transaction Type Key (FK) Transaction Date Key (FK) Transaction Type Key (FK) ransaction Date Key (FK) Insured Item Key (FK) Demographics Key (FK) Payment History Score **Transaction Amount Emplyment Risk Score** INSURANCE FACT Insured Item Key (FK) Customer Key (FK) Coverage Key (FK) Fransaction Amount INSURANCE FACT Customer Key (FK) Coverage Key (FK)

combination of the

attributes

each possible

It has one row for

Value bands are used to reduce the number

4

of rows overall

ILLINOIS INSTITUTE OF TECHNOLOGY

School of Applied Technology

Dimension Value Band Example

Age Group 71 to 80 11 to 20 21 to 30 31 to 40 41 to 50 51 to 60 61 to 70 0 to 10 Daily Calories Burned 1001-1500 1501-2000 2001-2500 2501-3000 3001-3500 3501-4000 501-1000 0-500 Sugar Level 101-115 116-130 131-145 146-160 81-95 96-100 59-65 08-99

ILLINOIS INSTITUTE OF TECHNOLOGY

Resolving Multivalued Relationships

using Bridge Tables: Examples

Resolving Multivalued Relationships using Bridge Tables

- value consistent with the fact table's grain dimension attached to a fact has a single ➤ In a classic dimensional schema, each
- ➤ But there are a number of situation in which a dimension is legitimately multivalued

	ĺ	
Sales Fact Table	D	Sales Reason Dime
Other dimension keys	7	Sales Reason Key (P
Sales Reason (FK)	2	Sales Reason Name
Sales Qty		Sales Reason Descr
Sales Amt	7	Sales Reason Type

nsion

ILLINOIS INSTITUTE OF TECHNOLOGY

School of Applied Technology

Other attributes

102 Product Quality; Promotion 103 Product Quality; Convenient Location Sales Reason Group Name Sales Reason Group Key Sales Reason Key Weighting Factor 101 1 1 1 102 2 103 3 3 Sales Reason Group
Sales Reason Group Key (PK) Sales Reason Group Name Sample rows from Sales Reason Bridge Sample rows from Sales Reason Group Product Quality Sales Fact Table
Other dimension keys...
Sales Reason Group Key (FK) Sales Reason Group Ker Sales Qty Sales Amt

Sales Reason Ker Sales Reason Name Tother attributes Sample rows from Sales Reason Dimension 3 Convenient Location . Product Quality

Resolving Multivalued Relationships using Bridge Tables Multivalued Sales Reasons Bridge Sales Reason Key (PK)

Sales Reason Name Sales Reason Descr Sales Reason Type Other attributes Sales Reason Group Key (FK) Weighting Factor

The Sales Reason Group table may be required by relationships. It provides no useful information at query time and is often your modeling tool to resolve FK/PK omitted.

➤ Many sales reasons on a single transaction ➤ Many customers in a bank account

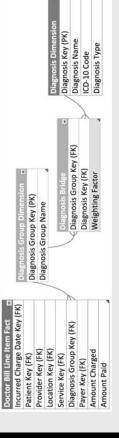
> Many diagnosis at a time of treatment ➤ Many witnesses to an accident

Many options on a car

LLINOIS INSTITUTE OF TECHNOLOGY

School of Applied Technology

Resolving Multivalued Relationships using Bridge Tables **Multivalued Diagnosis Bridge**



➤ The weighing factor is an explicit allocation

groups, especially for out patient treatments case it seems reasonable to re-use diagnosis ➤ Records in the Diagnosis Group Dimension can be made for each patient, but in this where many groups would be repeated

Resolving Multivalued Relationships using Bridge Tables **Multivalued Bank Account Customer Bridge**

Week 11 Topic: Dimensional Modeling: More

Dimension Patterns and Considerations

School of Applied Technology

ILLINOIS INSTITUTE OF TECHNOLOGY

Customer Dimension Customer Key (PK)
Customer Name Customer Date of Birth Customer Address Account Key (FK)
Customer Key (FK) Weighting Factor Begin Date End Date Account Dimension F Account Key (PK) Primary Holder Primary Address Account Type Monthly Account Snapshot Fact Month End Date Key (FK)
Account Key (FK) Number of Transactions Interest Paid Fees Charged Month Ending Balance Average Daily Balance Household Key (FK) Branch Key (FK)

Associate customers to accounts where these have a many-to-many relationship

Query account balances by individual customer or groups of customers A

(prorated) by individual customers to avoid Show account balances correctly weighted double counting A

7

Questions?