Data warehousing
it doesn't get any easier!

Bart Heinsius
October 3, 2007

Content

1. Technology vs Business
2. Defining Definitions
3. Choosing Facts and Dimensions
4. Conclusions
5. Questions

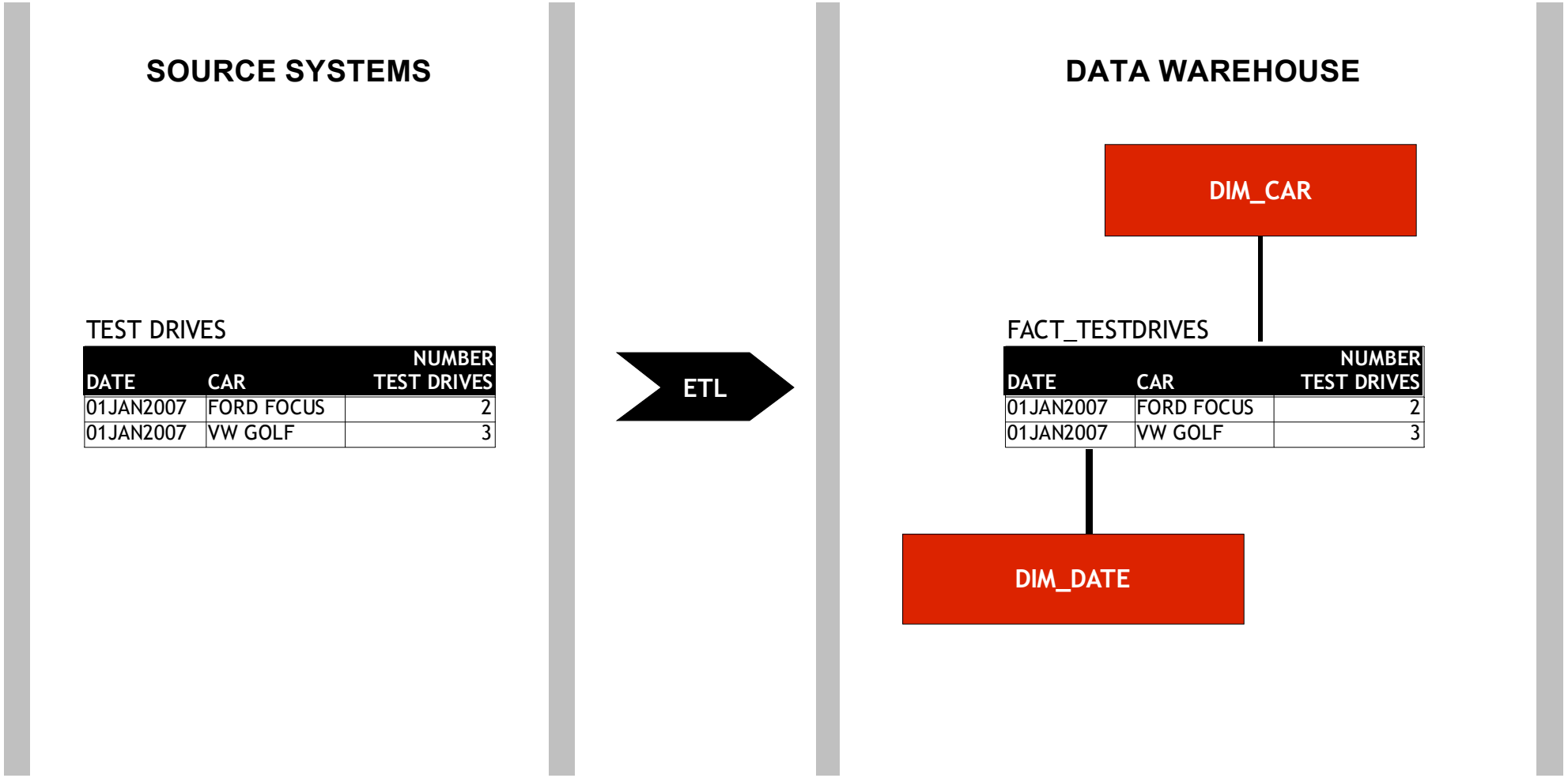
Technology vs Business

- Data warehouse: why?
  - ➔ information
- Project
  - list need for information
  - technical implementation

Technology vs Business

- Problem:
  - User is unaware of what he/ she wants
- Reason:
  - Used to query operational systems
- Consequence:
  - DWH grafted onto operational systems
  - DWH provides operational and technical details
  - DWH doesn't provide *information*

Technology vs Business



How many cars have been taken for a test drive? 2

Technology vs Business

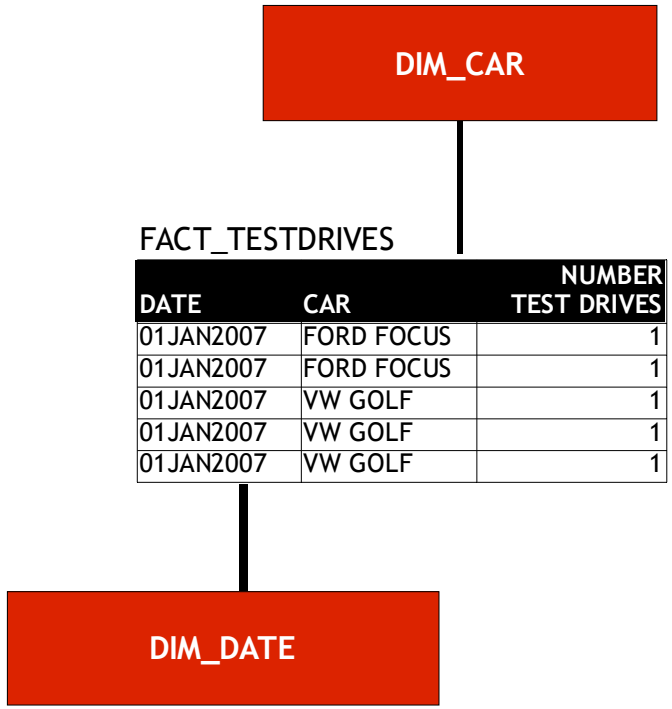
SOURCE SYSTEMS

TEST DRIVES

DATE	CAR	NUMBER TEST DRIVES
01JAN2007	FORD FOCUS	1
01JAN2007	FORD FOCUS	1
01JAN2007	VW GOLF	1
01JAN2007	VW GOLF	1
01JAN2007	VW GOLF	1

ETL

DATA WAREHOUSE



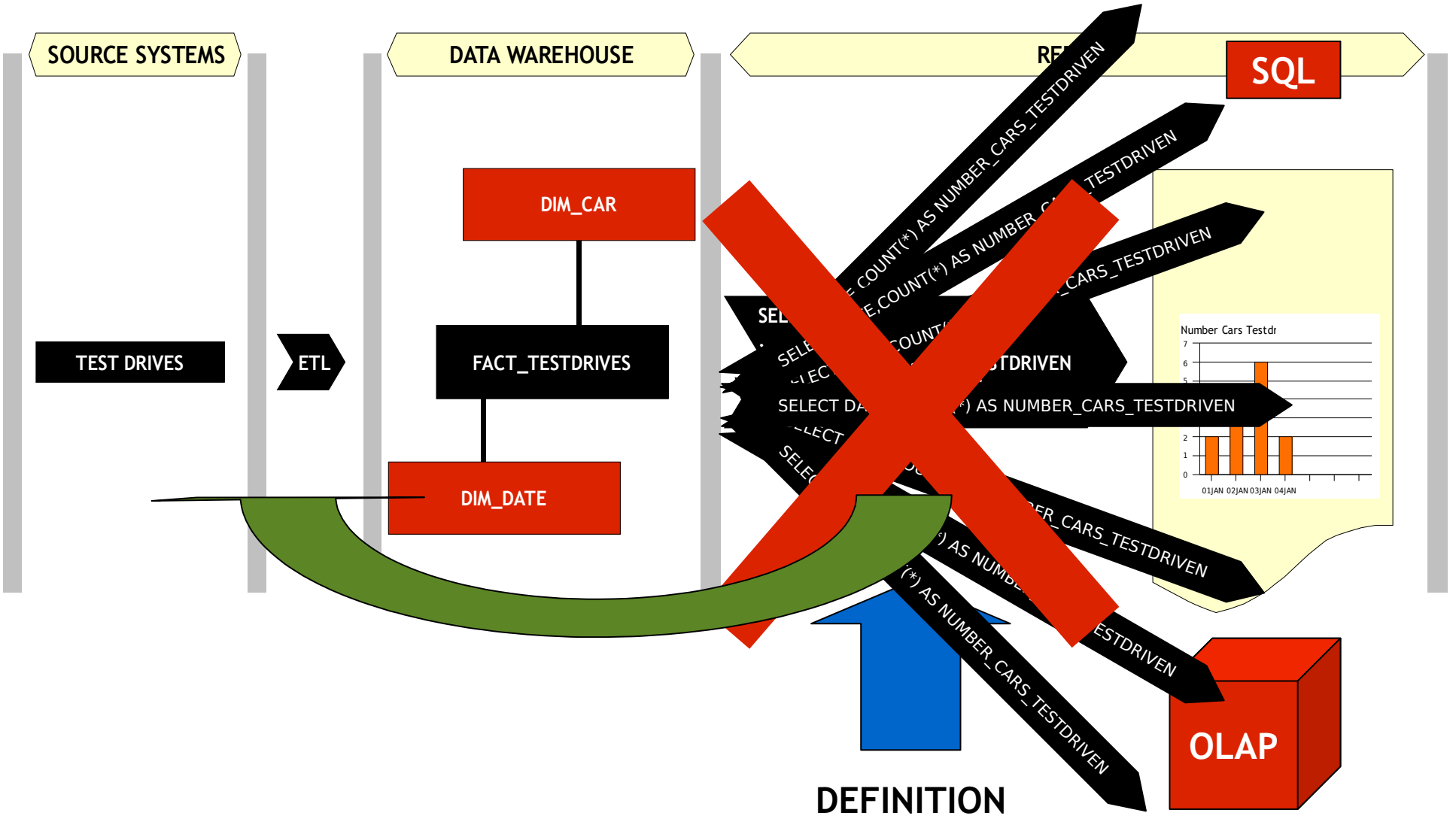
DATE	CAR	NUMBER TEST DRIVES
01JAN2007	FORD FOCUS	1
01JAN2007	FORD FOCUS	1
01JAN2007	VW GOLF	1
01JAN2007	VW GOLF	1
01JAN2007	VW GOLF	1

How many cars have been taken for a test drive? 5

Technology vs Business

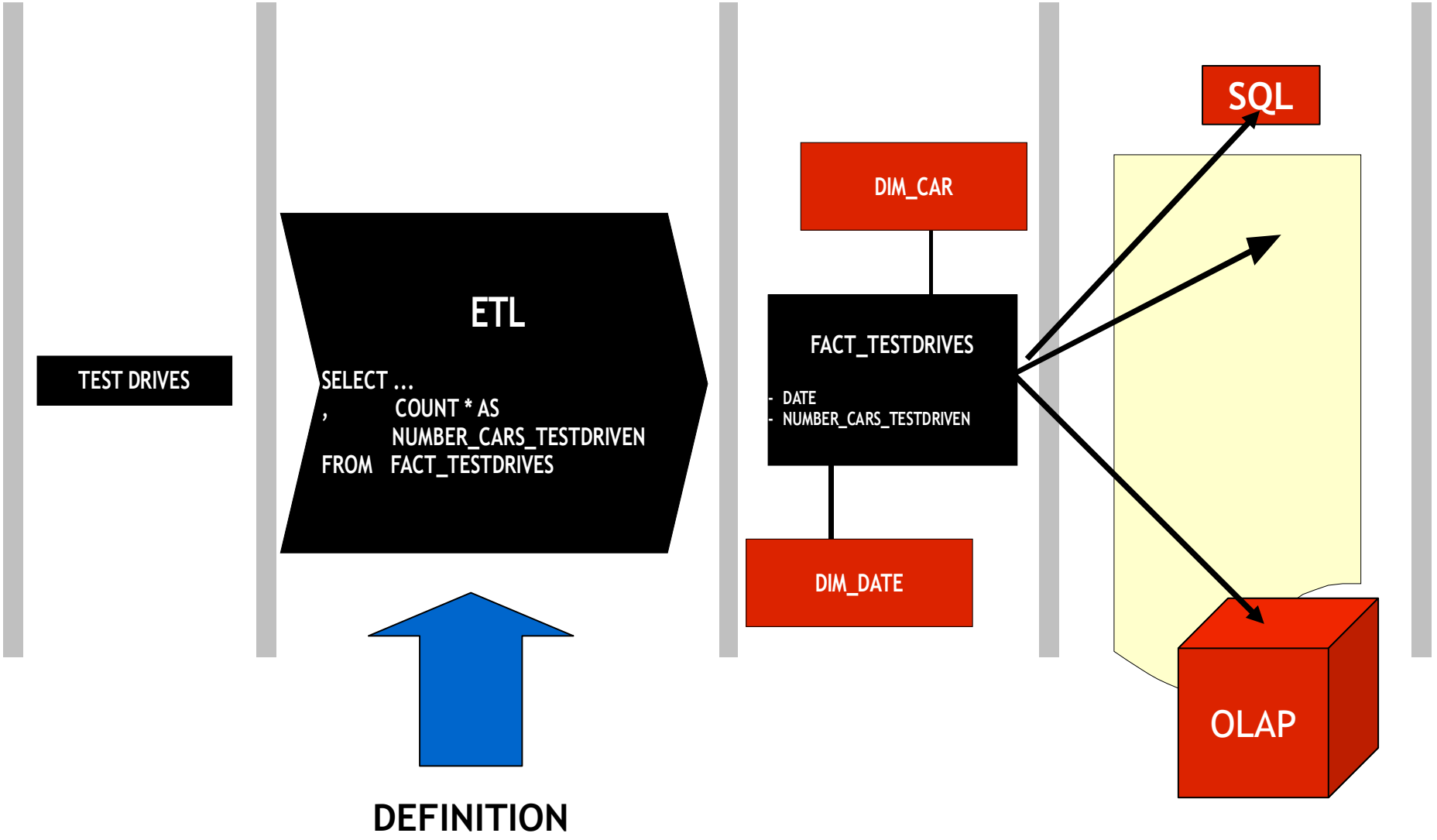
- Don't report technology!
  - User requested #cars a day
  - Make that information available in the data warehouse

Defining Definitions



NEW: NUMBER_CARS_TESTDRIVEN = COUNT(*) WHERE DAY <> SUNDAY !

Defining Defenitions



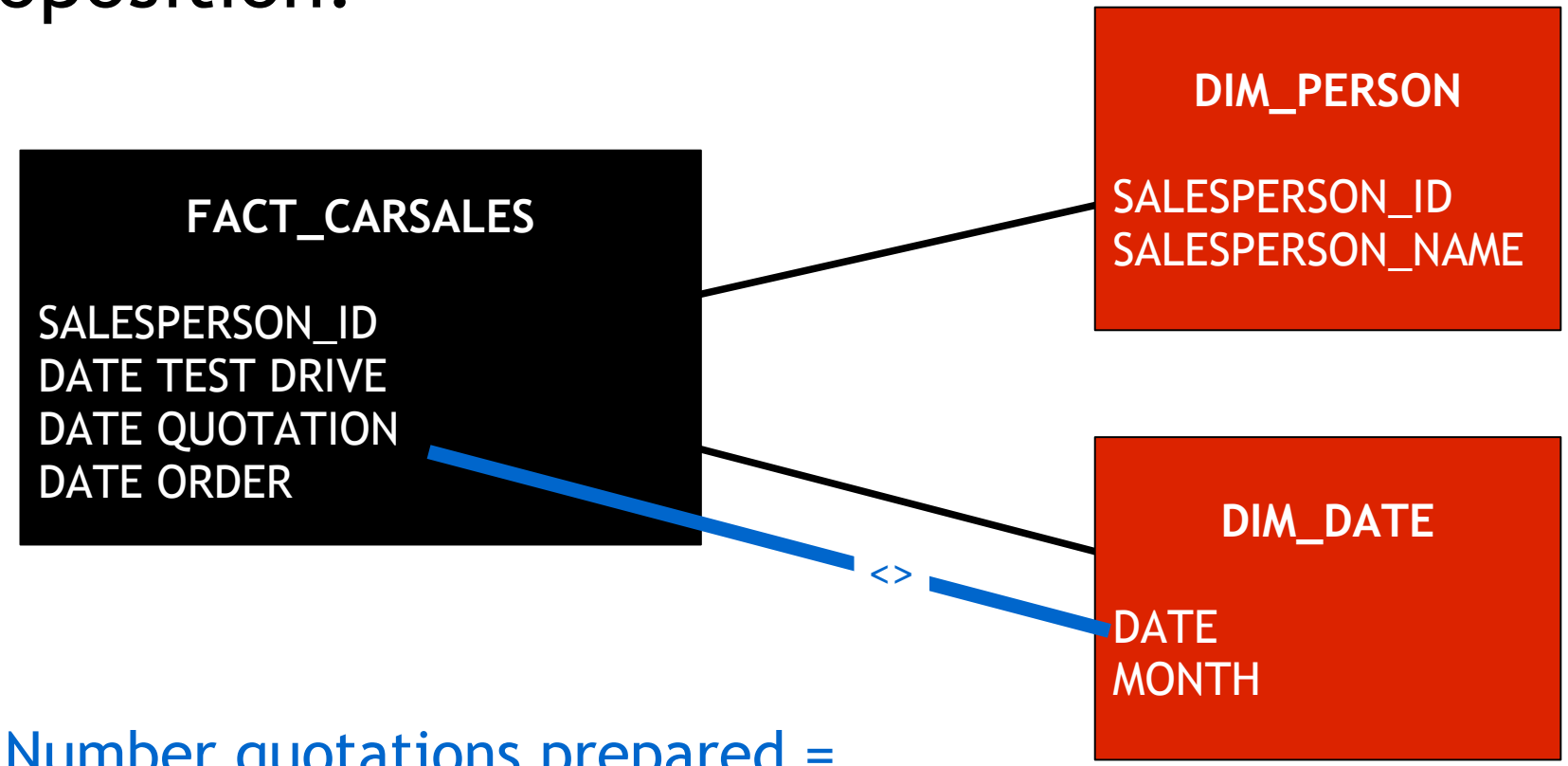
NEW: NUMBER_CARS_TESTDRIVEN = COUNT(*) WHERE DAY <> SUNDAY !

Choosing Facts and Dimensions

- Case study: car sales
  - Need for information:
    - # orders
    - # quotations
    - # test drives
    - flow test drive -> quotation -> order
    - per sales person
    - per month

Choosing Facts and Dimensions

- Proposition:



Number quotations prepared =
COUNT(*) WHERE QUOTATION_DATE = .

-> Definition!

Choosing Facts and Dimensions

- loading FACT_CARSALES:
  - Test drive -> new record

SALESPERSON_ID	TESTDRIVE_DATE	QUOTATION_DATE	ORDER_DATE
1	01JAN2007		

  - Quotation prepared -> update record

SALESPERSON_ID	TESTDRIVE_DATE	QUOTATION_DATE	ORDER_DATE
1	01JAN2007	08JAN2007	

  - Order placed -> update record

SALESPERSON_ID	TESTDRIVE_DATE	QUOTATION_DATE	ORDER_DATE
1	01JAN2007	08JAN2007	15JAN2007

- #Quotations prepared:
  COUNT(*) WHERE QUOTATION_DATE = .

Choosing Facts and Dimensions

- loading F_CARSALES:
  - Test drive -> new record

SALESPERSON_	RIVE_DATE	QUOT	E	ORDER_DATE
1	07			

  - Quotation prep -> upd ecord

SALESPERSON_ID	TEST		TION_DATE	ORDER_DATE
1	01JAN20		N2007	

  - Placed order -> upd rd

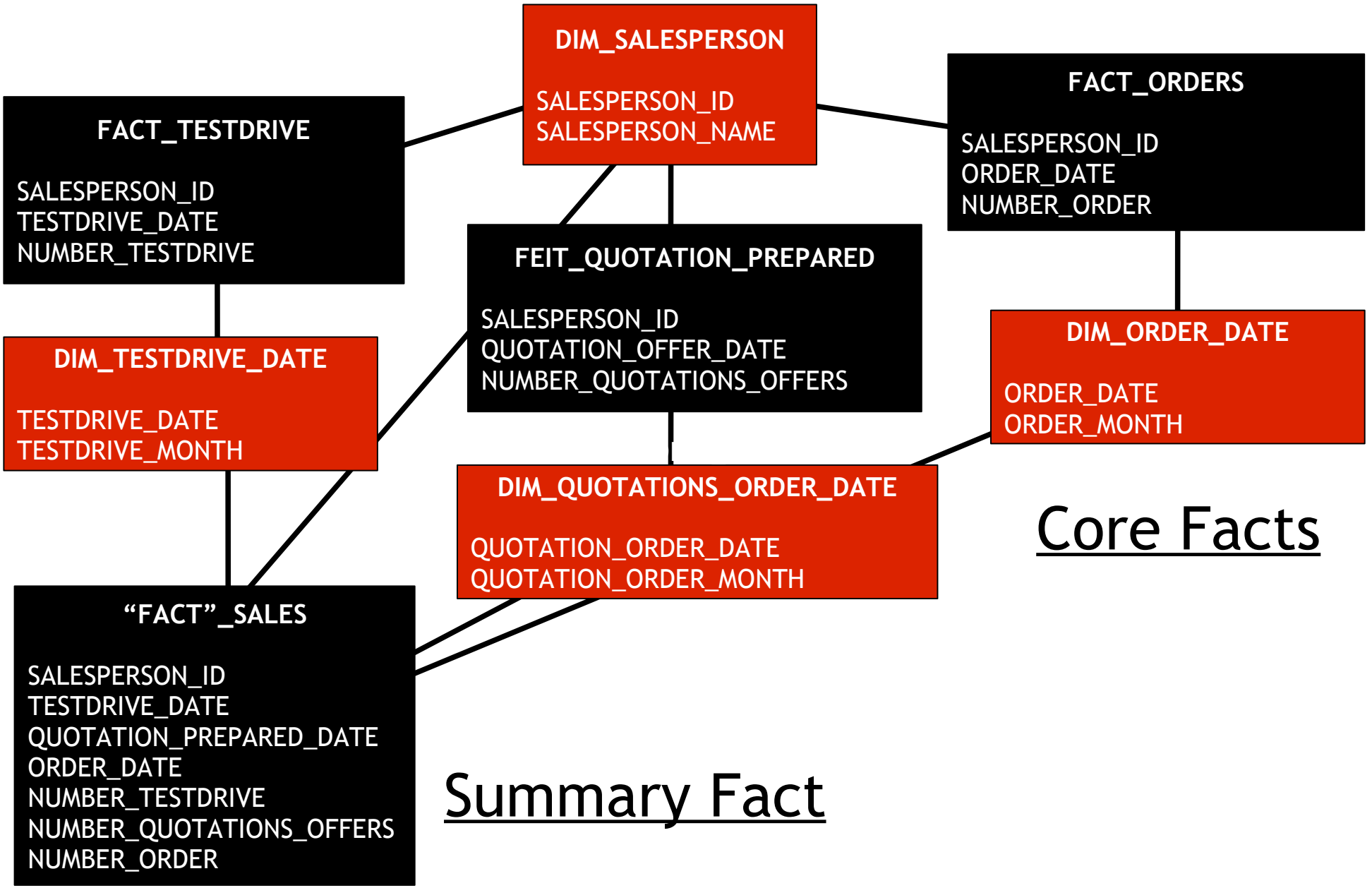
SALESPERSON_ID	TEST		TION_DATE	ORDER_DATE
1	01J		7	15JAN2007

- Slowly Chan Fact!
- Updates ading takes a long e!

Choosing Facts and Dimensions

- What is a FACT?
  - Longman: “a piece of information that is known to be true”
  - > FACT_SALES doesn't contain facts!
- What then are the facts?
  - Test drives
  - Quotations
  - Orders

Choosing Facts and Dimensions



Conclusions

- Make end user forget all about technology
- Implement definitions centrally and NOT in the actual report
- Choose real facts

QUESTIONS