CZECH TELECOM

Expanding a Datawarehouse in step with Oracle advancements

Paris, 20.09.2003

Birmingham, 10.12. 2003

Jan Medek – jan.medek@ct.cz

Paul Skaife – paul@yorviq.co.uk

Agenda

§ Introduction

§ Project History

§ Project Structure

§ Phased Approach

§ How it was done

§ DWH Now

§ Future

Czech Telecom

- the biggest fixed line telecom in the Czech Republic
- former monopolist
- approx. 4 mil. lines (10 mil. inhabitants in CR)
- since 7/2002 - carrier selection
- since 1/2003 - carrier preselection and number portability
- since 3/2003 – ADSL
- now buying mobile Eurotel

Project History (1)

Reasons for DW:

- comming liberalisation of Czech telecommunication market
- decreasing revenue
- ne
- his
- se
- information integration
- speed of getting information
- analyses

Unified view on a customer was needed

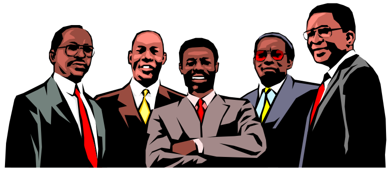
Project History (2)

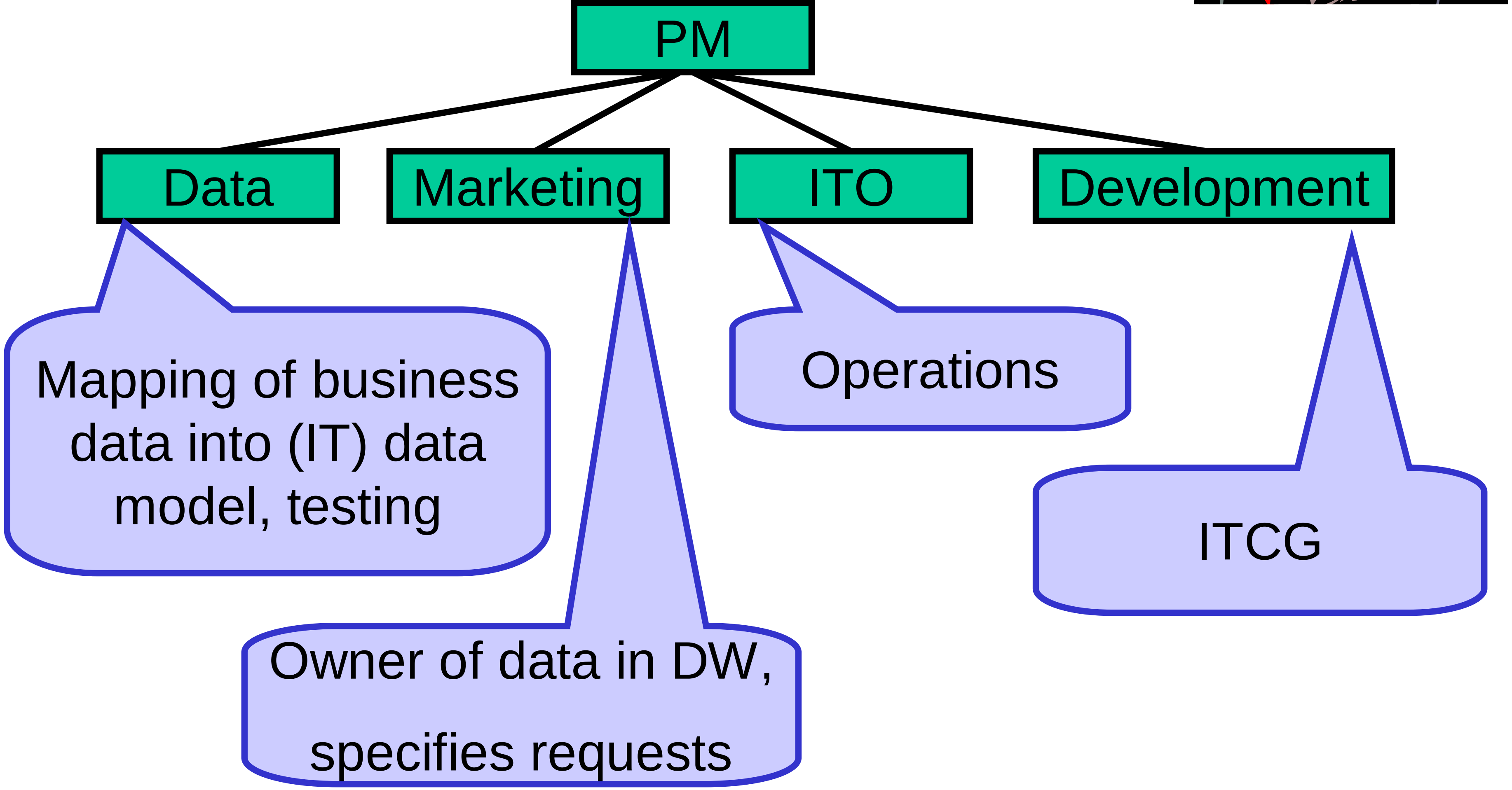
September 1999 - begin:

- OSS programme
- KADO
  - customer datawarehouse
  - bought from KPN
    - Ø similar market (CZ – NL)
    - Ø similar company (KPN – CTc)
    - Ø short implementation time

Project Structure (1)

Teams





PM

Data

Marketing

ITO

Development

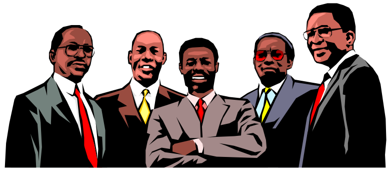
Mapping of business data into (IT) data model, testing

Owner of data in DW, specifies requests

Operations

ITCG

Project Structure (2)



Team structure

- at the beginning
  - Data team – 2 consultants (GB, USA), 1 CTc employee
    later – 6 consultants (GB, NL, SK, USA)
  - Marketing, ITO – CTc employees
  - Development – external supplier
- now
  - Data, Marketing, ITO – CTc employees
  - Development – external supplier

Phased Approach (1)

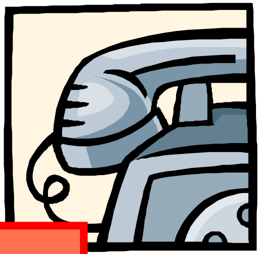
Development phases:

1. *Voice – customer, product, revenue, traffic*
2. *Data – customer, product, revenue, traffic*
3. *CRM – contact history*

4. Enhancements

Phased Approach (2)

Voice (since 9/1999)



- *Customer - identification, address, industry classification, no of employees etc.)*

Benefits:

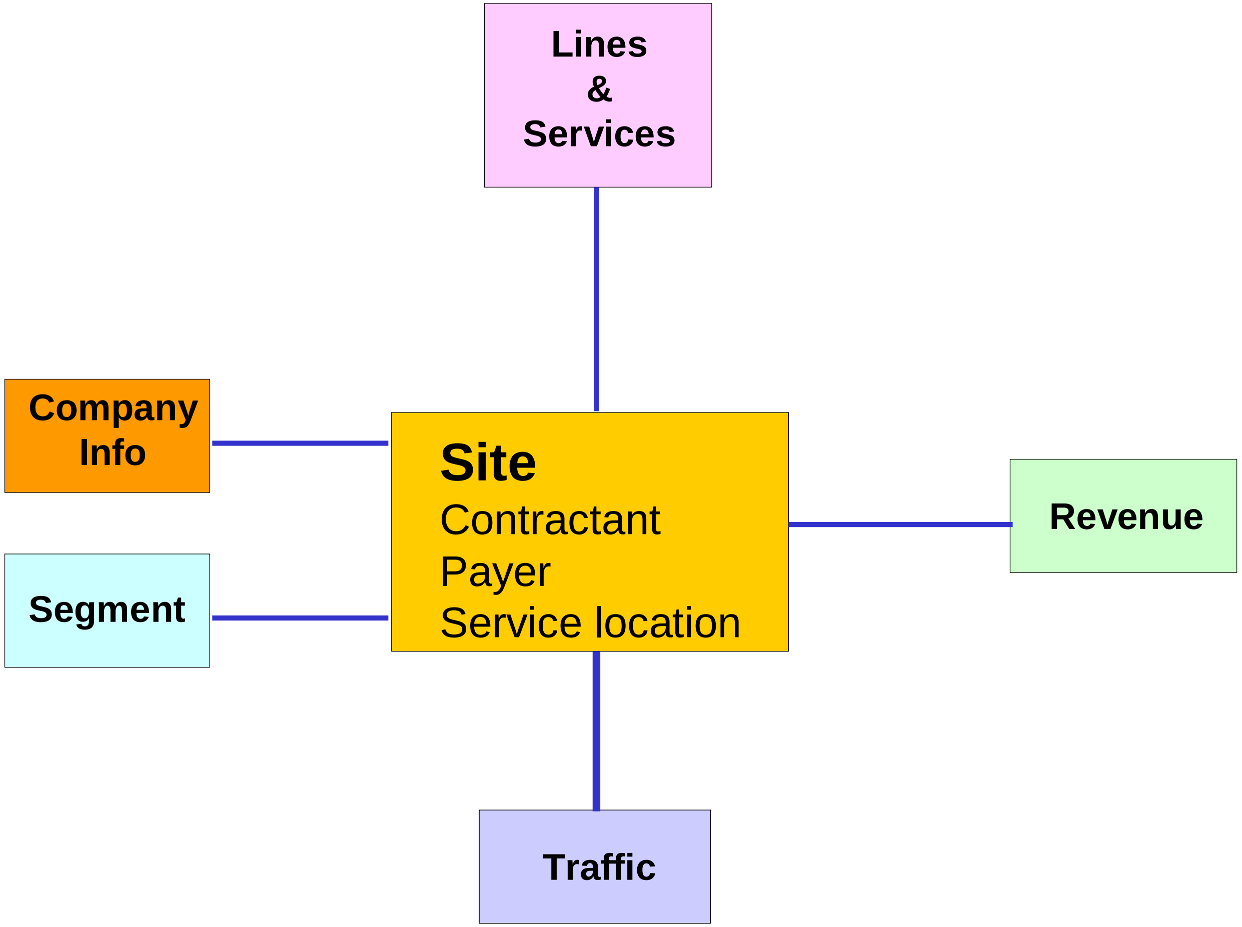
Unified view on a customer

Customer segmentation

Source systems

- *Ordering*
- *Billing (new billing system – migration)*
- *Company info*

Phased Approach (3)



Lines & Services

Company Info

Segment

Site
Contractant
Payer
Service location

Revenue

Traffic

Phased Approach (4)

Data (since 10/2000)

- *Customer - identification, address, industry classification, no of employees etc.)*
- *Products – leased lines and data services*
- *Revenue – total billed amount per customer*
- *Traffic – no of units, usage charges (if any)*

Source systems

- *Ordering for leased lines and data services (2⇨1)*
- *Billing for data services*

CZECH TELECOM

Phased Approach (5)

CRM (since 9/2001)

- *Customer contact person(s) - name, address, position, phone etc.*
- *Contact history – contacts (who, when, what etc.)*
- *Marketing campaigns (DW⇨CRM, CRM⇨DW)*

Source

- *CRM system*

Phased Approach (6)

Enhancements (since 2002)

Basic types:

- *new products/services*
- *more details*
- *changes within the source systems*
- *changes of the source systems*
- *speed up (earlier data availability)*
- *others*

Phased Approach (7)

Enhancements (since 2002)

- *Extended history to 36 months (customers, revenue, traffic)*
- *Revenue per service*
- *Revenue per line*
- *New products/services (price packages, multiple billing periods, ADSL etc.)*
- *Runtime scenarios*

IT Consultancy **G**roup bv
Friesestraatweg 215
9743 AD Groningen

I T C G

Paul Skaife
e-mail p.skaife@itcg.nl

How it was done

Contents

Implementation

Tools

Issues Encountered

- and solutions !!!

Testing

Oracle effects

Implementation

Architecture

Tools

Specifications

Issues Encountered

- QA
- Changes

Testing

Data Architecture

Data Architecture Concept

Data Architecture in Practise

Data Architecture

“raw”

modelling

quality analysis

matching

integrity

transformations

tools

technology

knowledge

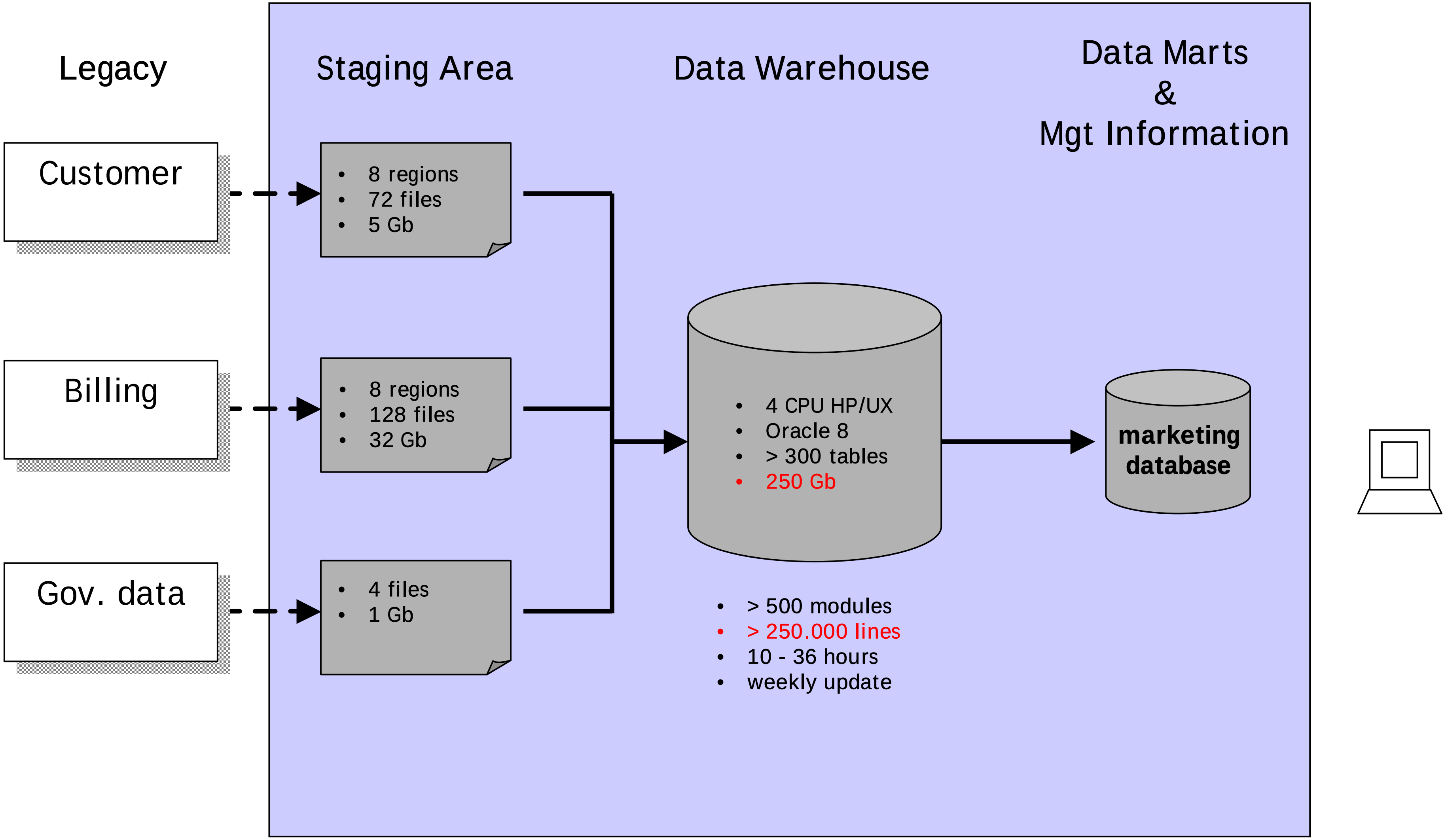
skills

integrated

uniform

consistent

Data Architecture in Practise



- 8 regions
- 72 files
- 5 Gb

- 8 regions
- 128 files
- 32 Gb

- 4 files
- 1 Gb

- 4 CPU HP/UX
- Oracle 8
- > 300 tables
- 250 Gb

- > 500 modules
- > 250.000 lines
- 10 - 36 hours
- weekly update

Contents

Implementation

Tools

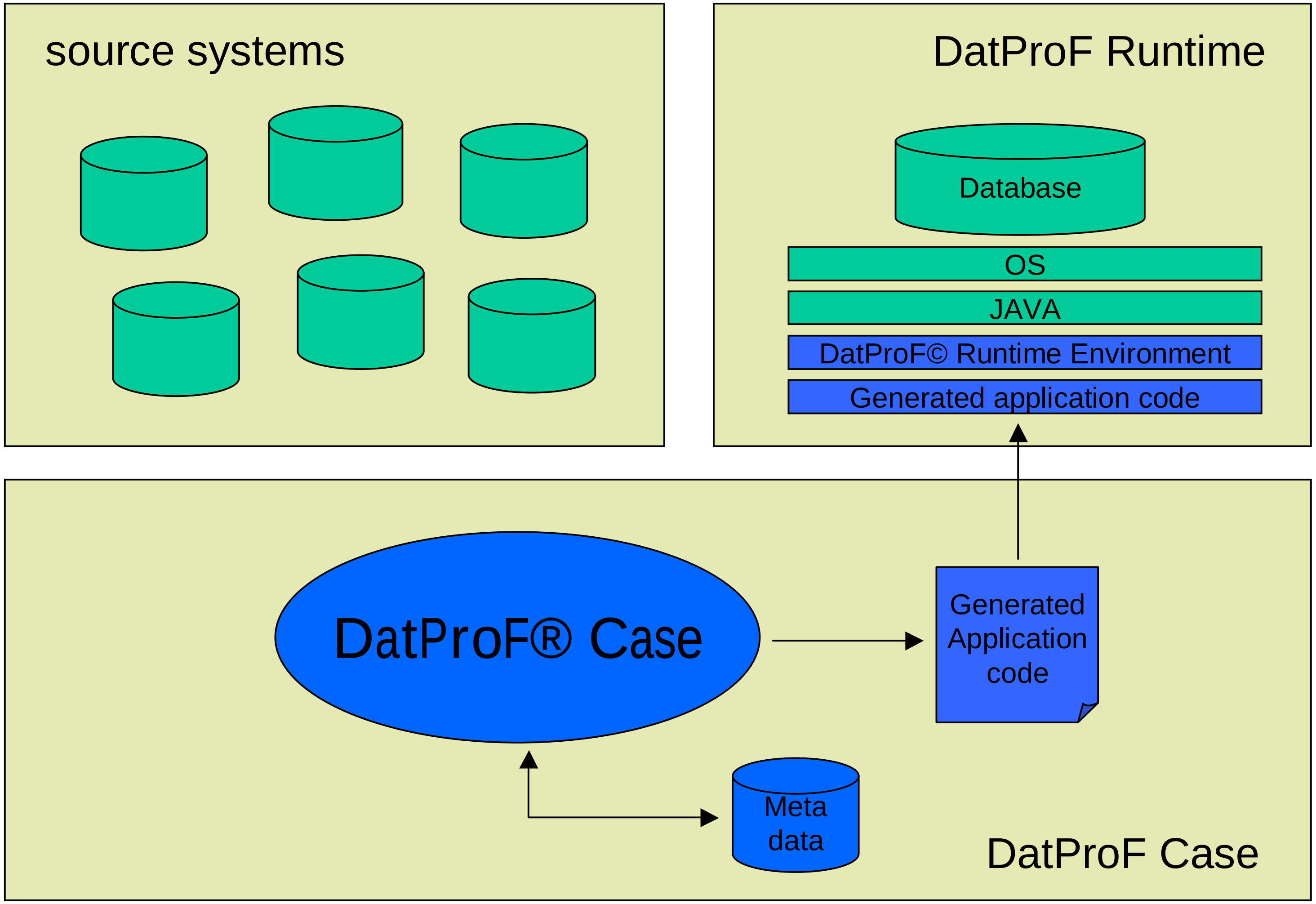
Issues Encountered

- and solutions !!!

Testing

Oracle effects

DatProF Case & Runtime



source systems
DatProF Runtime
Database
OS
JAVA
DatProF© Runtime Environment
Generated application code
DatProF® Case
Generated Application code
Meta data
DatProF Case

DatProF Case

Static models design

- Entities (tables/views)
- Columns
- Constraints
- References

Dynamic models design

- Data transformations
- Process dependencies
- Scenarios

DatProF Case features

Functional building blocks (design patterns)

Testdata integration

Impact analysis

HTML documentation generation

Visualization (data dictionary & processes)

Design wizards

- Data Quality Analysis
- Testdata Generator

100% code generation / Oracle integration

...

DatProF Runtime features

Installer

Scheduler

- Load balancing
- Parallel processing

Monitoring

- Text based monitor
- SMTP Notifier using Subscription Model

Logging

DatProF Specifications

Easy to write

- No specific tools needed
- Written by ITCG and/or customer

Easy to read

- Communication between customer and implementer
- In terms of customer (“natural language”)
- In terms of implementer (“building blocks”)

Easy to implement

- unambiguous
- highly structured

Contents

Implementation

Tools

Issues Encountered

- and solutions !!!

Testing

Oracle effects

Issues Encountered

Data Quality

Expanding Source Systems

Changes in Source Systems

Quality Analysis

Never Trust a Stranger

Data Quality

Duplicates

Incorrect Information

Relational Dependencies

Non Domain Values

Formats

etc.

Duplicate & Incorrect

Tony Blair
10 Downing street
W1
London
UK

Mr T. Blair
Downing street
London
UK

Anthony Blair
Downing street
London

Relational Dependencies

Customer

ID	Name	Revenue
07532	Mr Anthony Blair	?????

Party

ID	Name	Expired
025673	Tony Blair	
765000	Mr. T. Blair	
004832	Anthony Blair	01/03/2002

Revenue

ID	Customer_ID	Revenue
0001	025673	100
0002	765000	450
0003	765000	400
0004	765000	250
0005	765000	450
0006	004832	5000
0007	004832	5000

What is the revenue of Mr Anthony Blair   100?   1700?   11800?

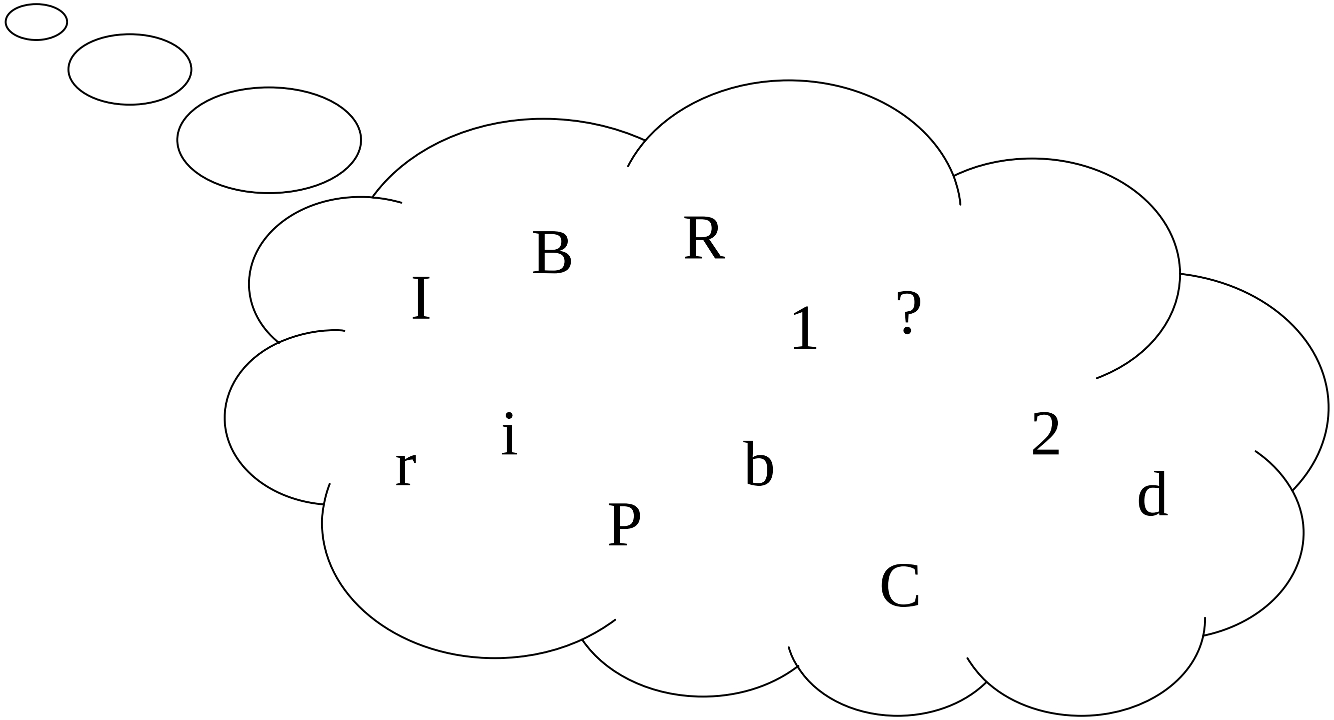
Erroneous Data

CUSTOMER_TYPE in (I,B,R)

I = Internal

B= Business

R= Residential



Pollution Model

Pollution is everywhere

Data Analysis - What

Analysis

- Statistics
- Characteristics
- Constraint controls
- Duplicates

Solution

- Correction in the source
- Correction in the target

Data Quality – How

Correct use of :

- Knowledge
- Experience
- Tools

Domains

```
====================================
column_name        : TYP
minimum            : ,
maximum            : M-L
average            : N/A
# distinct         : 27
# nulls            : 0
frequent val 1     : R [ 7,579,759]
frequent val 2     : B [ 3,146,005]
frequent val 3     : I  [150,675]
frequent val 4     : A [44]
frequent val 5     : C [19]
infrequent val 1   : - [1]
infrequent val 2   : 3 [1]
infrequent val 3   : 4 [1]
infrequent val 4   : 5 [1]
infrequent val 5   : K [1]
```

- 27 Values found
- 3 Values in domain

Min – Max Values

```
====================================
column_name       :  CYCLE_RUN_YEAR
minimum           :       1993
maximum           :       2000
average           :  1999.98504
# nulls           :  0
frequent val 1    :  2000 [1599782]
frequent val 2    :  1999 [21363]
frequent val 3    :  1998 [546]
frequent val 4    :  1997 [293]
frequent val 5    :  1996 [104]
infrequent val 1  :  1993 [2]
infrequent val 2  :  1994 [42]
infrequent val 3  :  1995 [49]
infrequent val 4  :  1996 [104]
infrequent val 5  :  1997 [293]
```

- Operational for 2 years
- Test data present in production

Amounts

```
====================================
column_name         : TOTAL_DUE_AMT
minimum             :  -260,665.4
maximum             : 12,693,212.9
average             : 6192.81162
# nulls             : 1509
frequent val 1      : 181.30 [42845]
frequent val 2      : 175.00 [12998]
frequent val 3      : 190.00 [12900]
frequent val 4      : 223.30 [6363]
frequent val 5      : 362.60 [3657]
infrequent val 1    : 0.10 [1]
infrequent val 2    : 0.30 [1]
infrequent val 3    : 0.80 [1]
infrequent val 4    : 1.70 [1]
infrequent val 5    : 2.90 [1]
```

- Very large bills and refunds
- Null bills
- Frequent values show standard charges

Erroneous Data

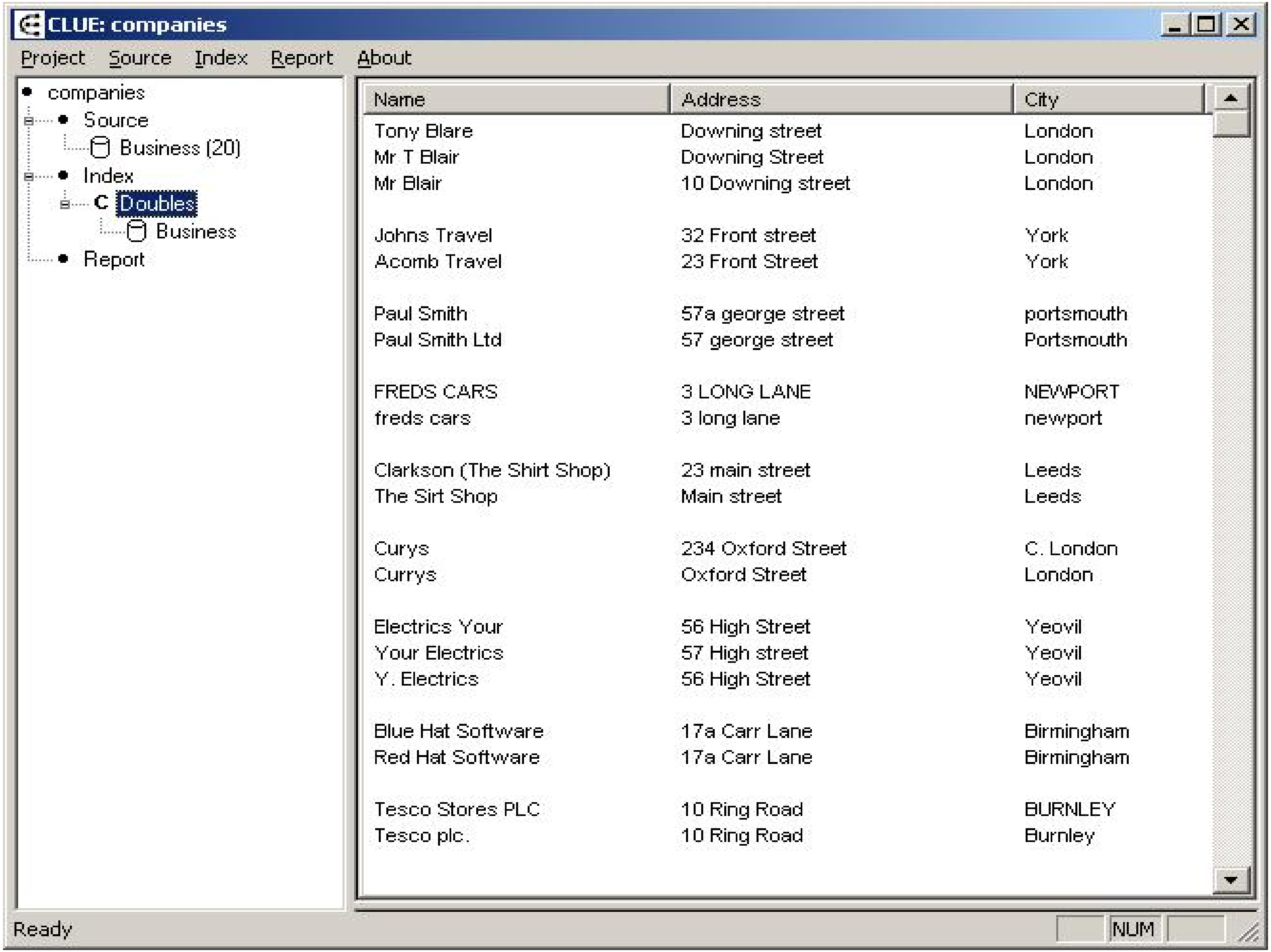
```
====================================

column_name       : DATE_BILL

minimum           : 19180298

maximum           : 20011130

average           : 19949140.4

shortest          : N/A [3 pos]

longest           : N/A [3 pos]

# distinct        : 3495

# nulls           : 0

  frequent val 1  : 19960701 [10622]

  frequent val 2  : 20000701 [9731]

  frequent val 3  : 19970701 [9288]

  frequent val 4  : 19990701 [8951]

  frequent val 5  : 19980701 [8793]

infrequent val 1  : 19841203 [1]
```

19180298
ddmmyy

Constraint Checks

```
Constraint failure:


BILL            BILL_AMOUNT

-------         -------------

55397                2.432,34

55396                1.562,54

55395               10.321,56

55394                  435,99

55393               -2.432,32

55392                233,65

61672                6.443,32
```

Bills where customers do not exist



Name	Address	City
Tony Blare	Downing street	London
Mr T Blair	Downing Street	London
Mr Blair	10 Downing street	London
Johns Travel	32 Front street	York
Acomb Travel	23 Front Street	York
Paul Smith	57a george street	portsmouth
Paul Smith Ltd	57 george street	Portsmouth
FREDS CARS	3 LONG LANE	NEWPORT
freds cars	3 long lane	newport
Clarkson (The Shirt Shop)	23 main street	Leeds
The Sirt Shop	Main street	Leeds
Curys	234 Oxford Street	C. London
Currys	Oxford Street	London
Electrics Your	56 High Street	Yeovil
Your Electrics	57 High street	Yeovil
Y. Electrics	56 High Street	Yeovil
Blue Hat Software	17a Carr Lane	Birmingham
Red Hat Software	17a Carr Lane	Birmingham
Tesco Stores PLC	10 Ring Road	BURNLEY
Tesco plc.	10 Ring Road	Burnley

Changes

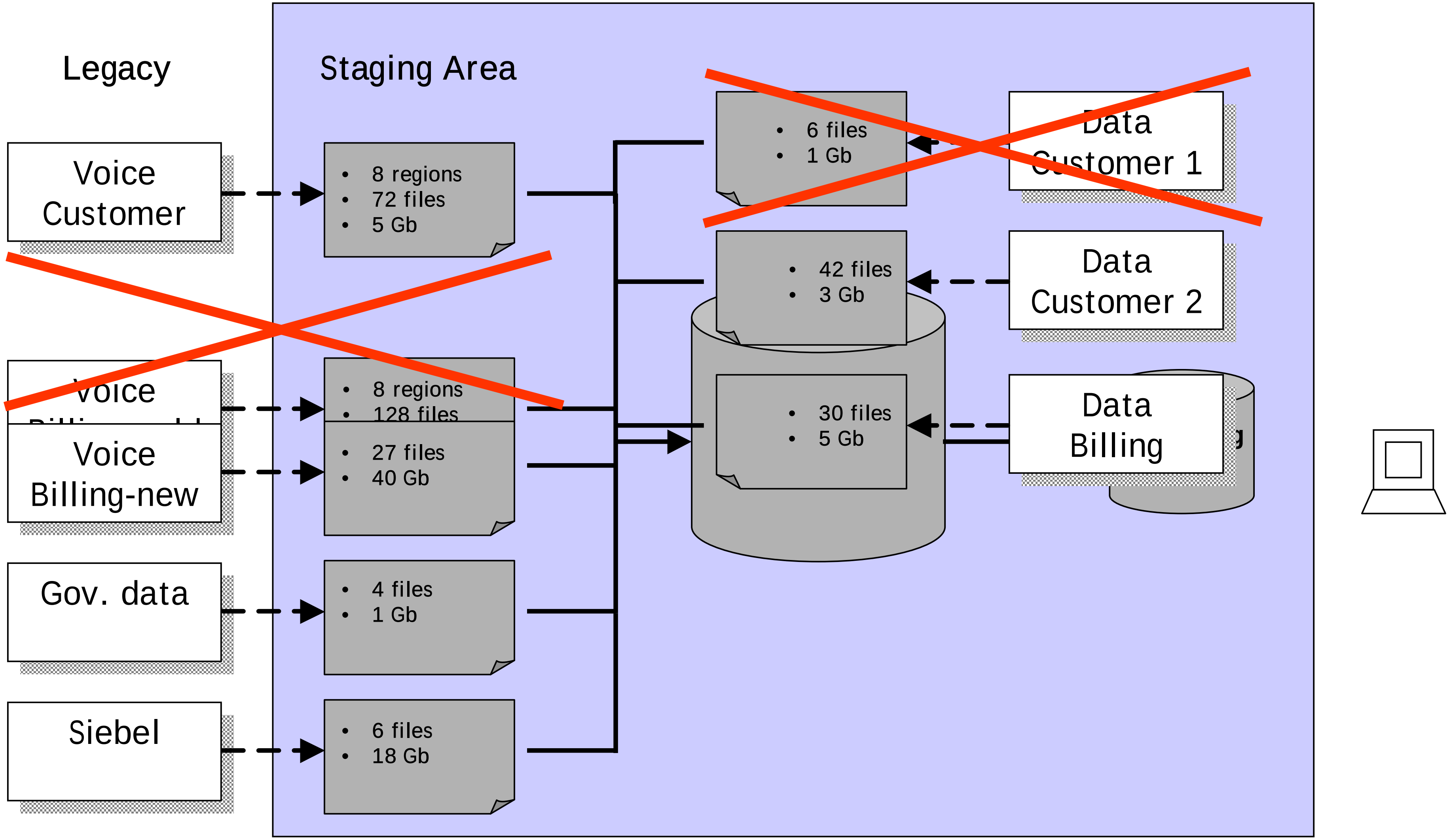
Additional Source Systems

Consolidation of Source Systems

Changes within a source

User requirements

Additional Source Systems



Legacy

Staging Area

Voice Customer

- 8 regions
- 72 files
- 5 Gb

~~Voice Billi~~

- ~~8 regions~~
- ~~128 files~~

Voice Billing-new

- 27 files
- 40 Gb

Gov. data

- 4 files
- 1 Gb

Siebel

- 6 files
- 18 Gb

- ~~6 files~~
- ~~1 Gb~~

~~Data Customer 1~~

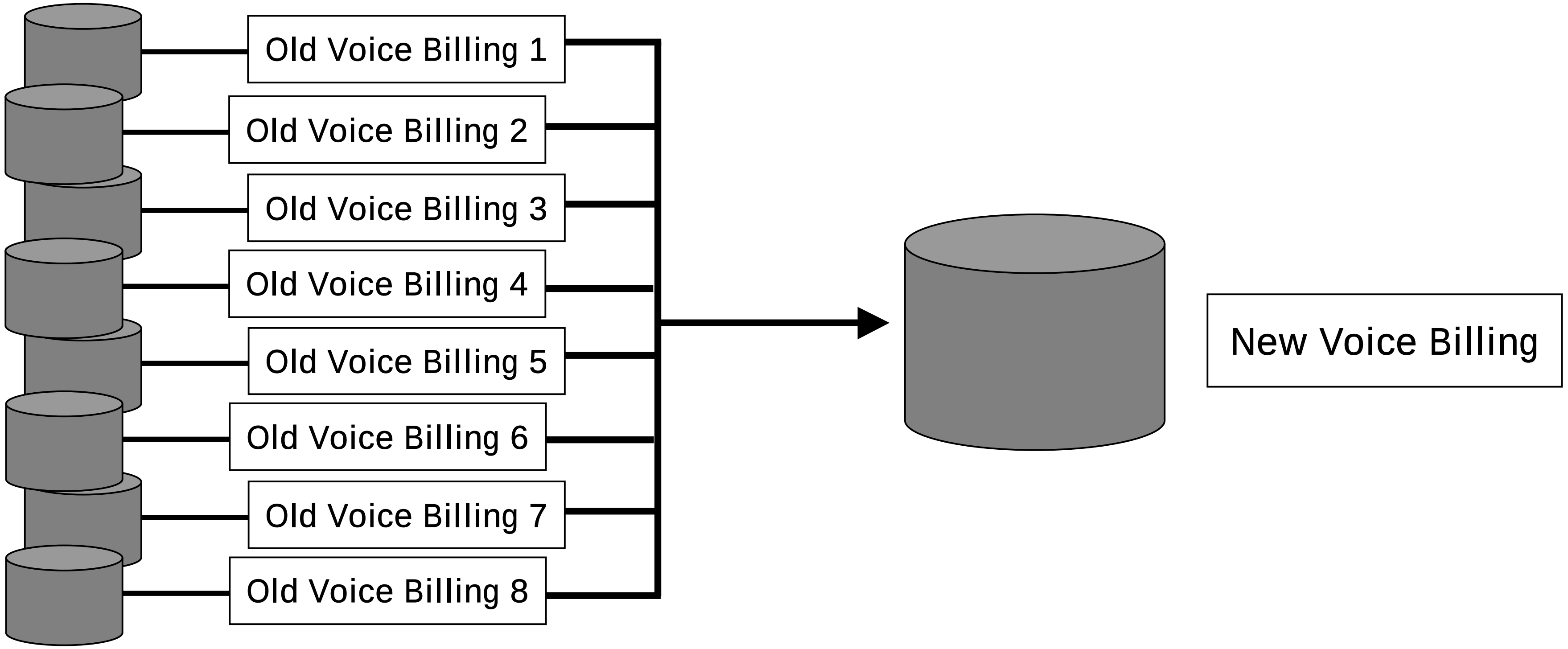
- 42 files
- 3 Gb

Data Customer 2

- 30 files
- 5 Gb

Data Billing

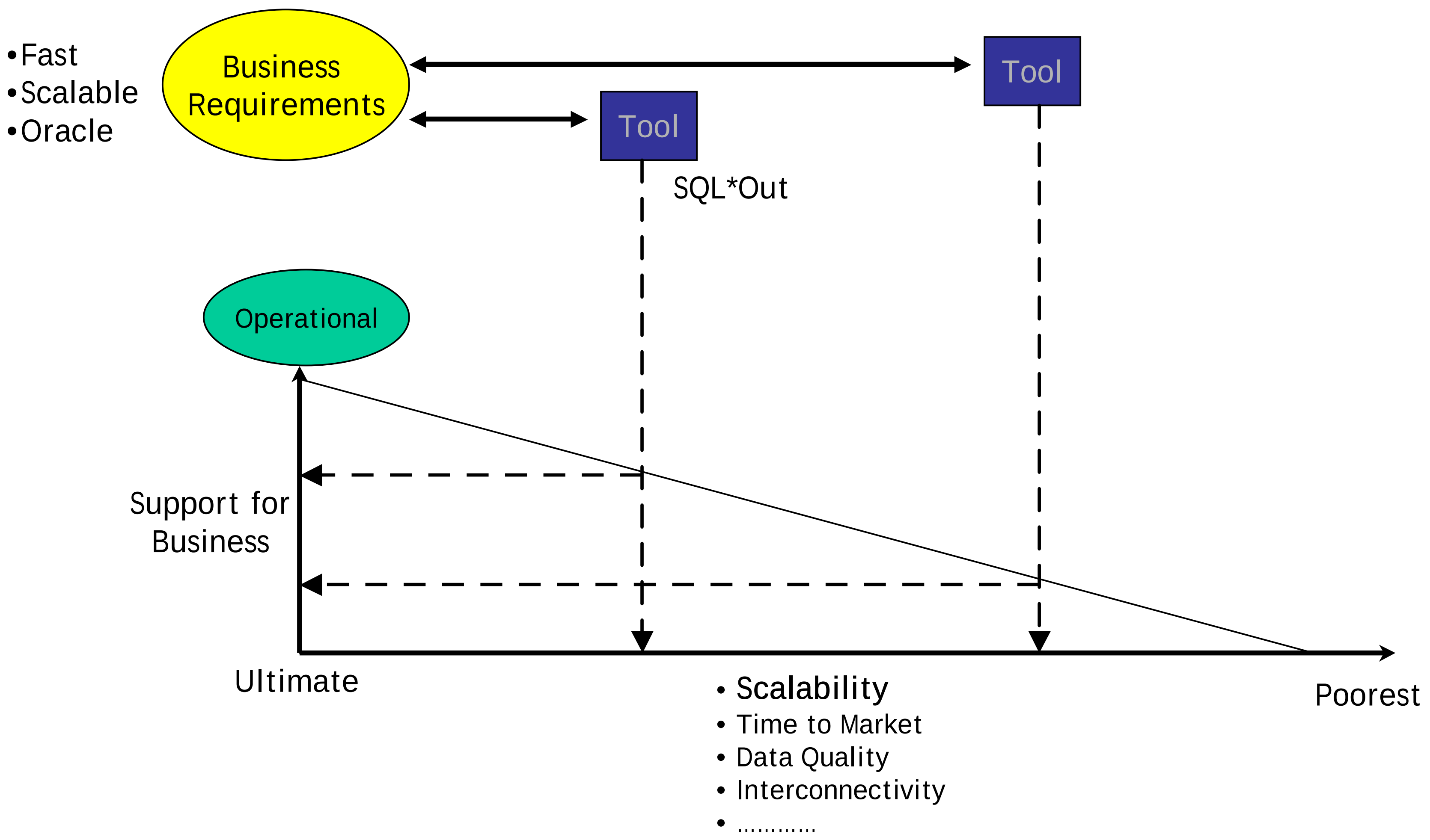
Consolidation of Source Systems



Consolidation

- Longer to extract (source)
  - ~8 * Data
- Longer to load ?
- Longer to process ?

Tool Selection Matrix - Extract



Changes within a source

- traffic_a, traffic_b, traffic_c

  Source System does job of Source System
- 8 traffic types on my bill ?
- Nobody told the DW

ČESKÝ TELECOM

Contents

Implementation

Tools

Issues Encountered

- and solutions !!!

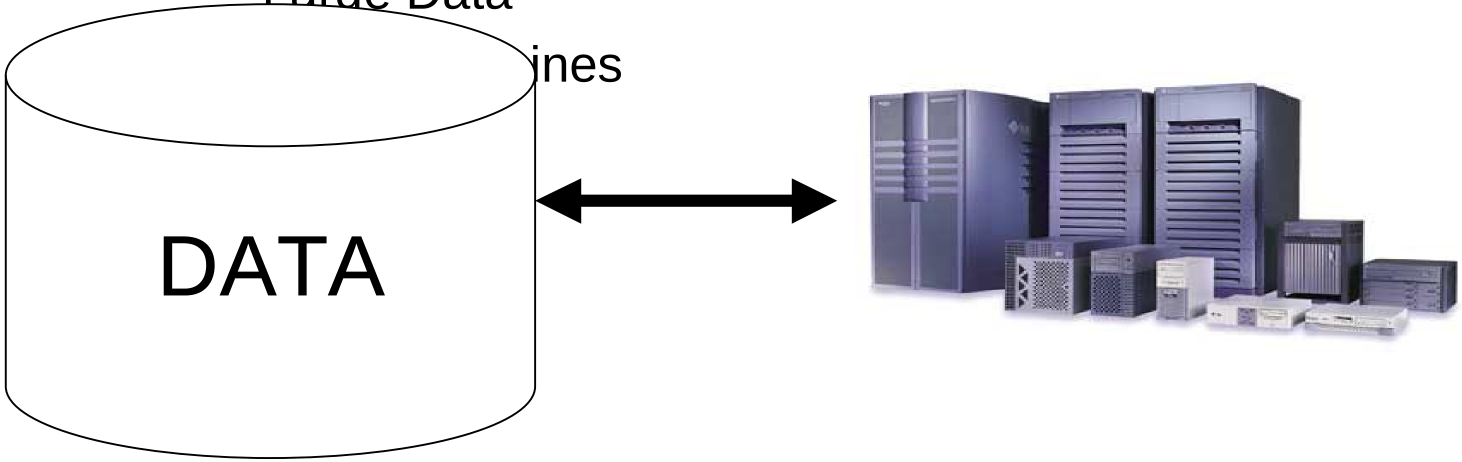
Testing

Oracle effects

Production Testing

Testing on Production Like

- Large Data
- ines



Production Testing

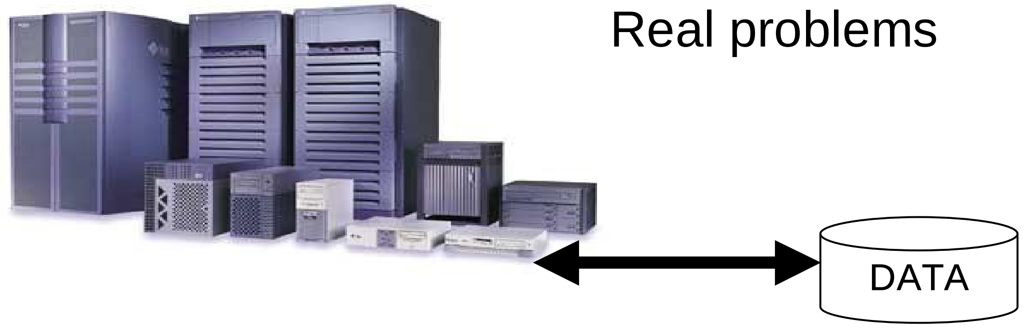
We do not have

- Time
  - Processing , Data, Personnel, Resources
- Money
  - Time, Resources

Idea

Create small consistent test sets

- Small
  - Less Cost (time , money, processing, resources)
- Consistent



Real problems

DATA

Problems

Complex Models

Complex Queries

Time

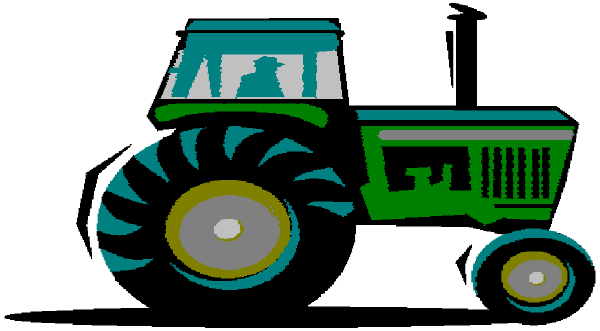
Cost

Resources

Problem Solutions

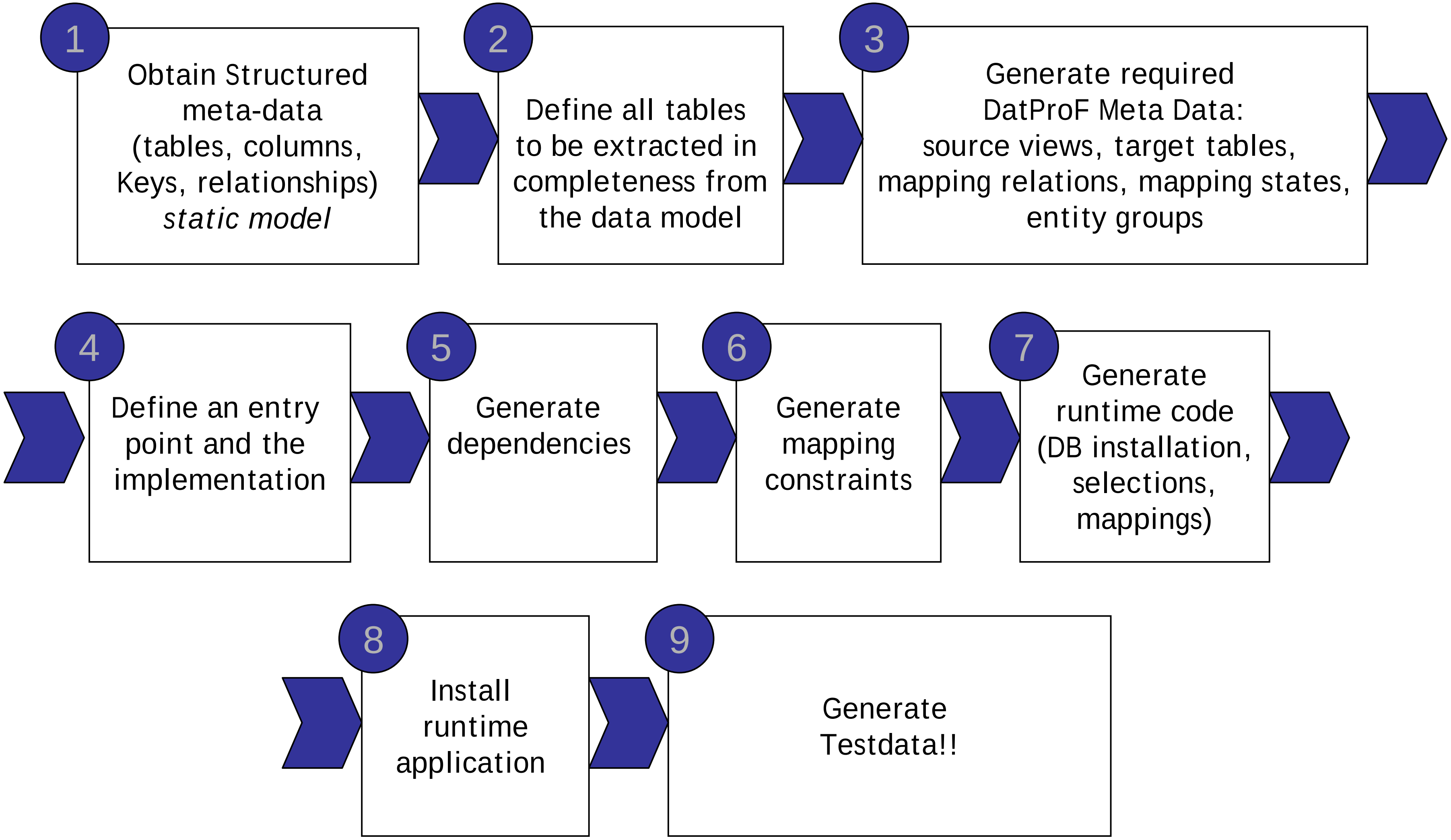
Methodology

Automation



ČESKÝ TELECOM

ITCG Approach



1 Obtain Structured meta-data (tables, columns, Keys, relationships) *static model*

2 Define all tables to be extracted in completeness from the data model

3 Generate required DatProF Meta Data: source views, target tables, mapping relations, mapping states, entity groups

4 Define an entry point and the implementation

5 Generate dependencies

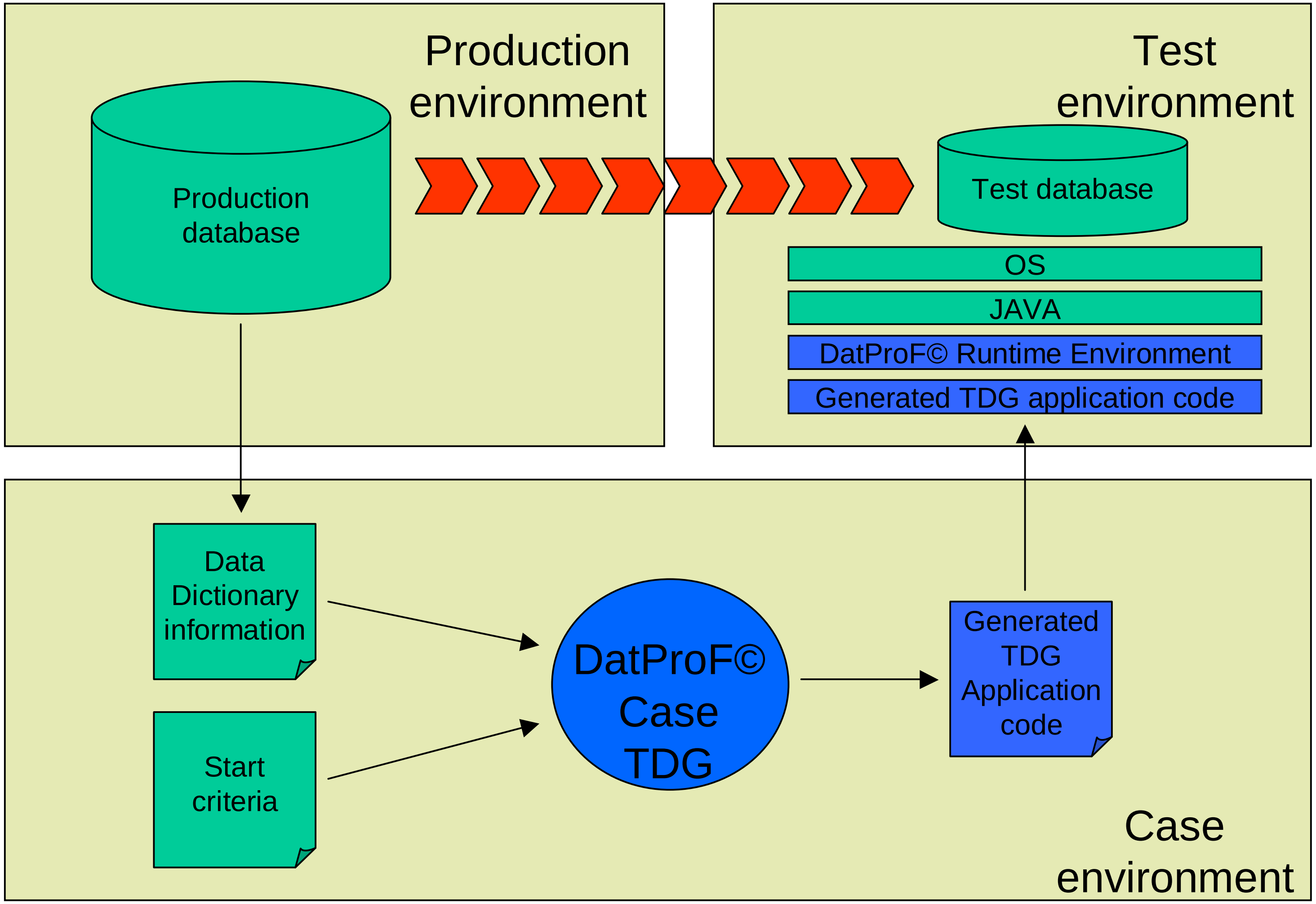
6 Generate mapping constraints

7 Generate runtime code (DB installation, selections, mappings)

8 Install runtime application

9 Generate Testdata!!

Subset Generation



Contents

Implementation

Tools

Issues Encountered

- and solutions !!!

Testing

Oracle effects

Oracle Version Upgrades

Generated Code

Simple Upgrade paths

8.0 – 9.2.0.4

8.16 -> 8.17

- Slowed down
- Bugs in Oracle
- Change the generator

9.2

- Bugs fixed
- Change generator

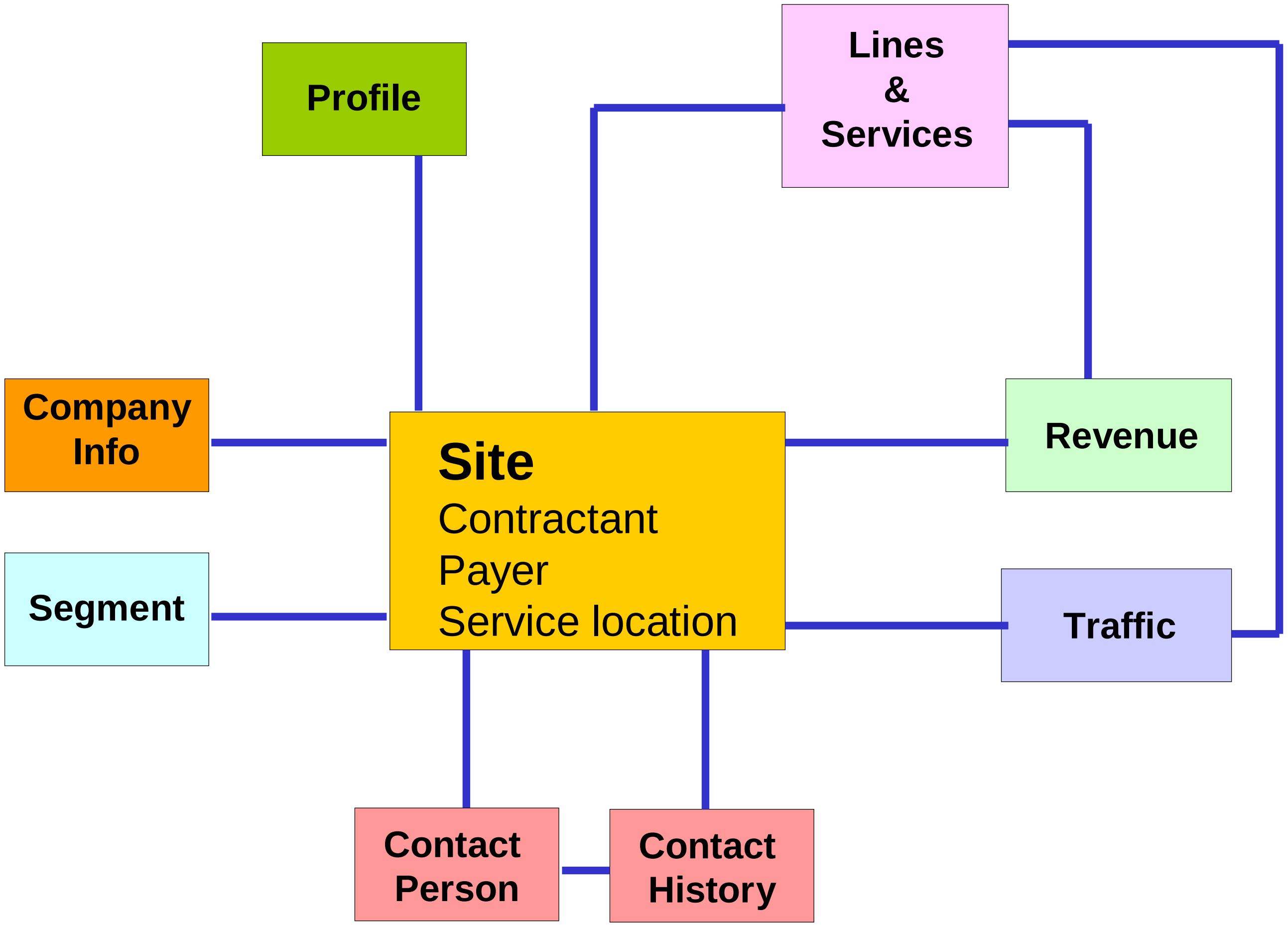
Added Functionality

Partitioning

Locally Managed Table spaces

OLAP functions

DWH Now (1)

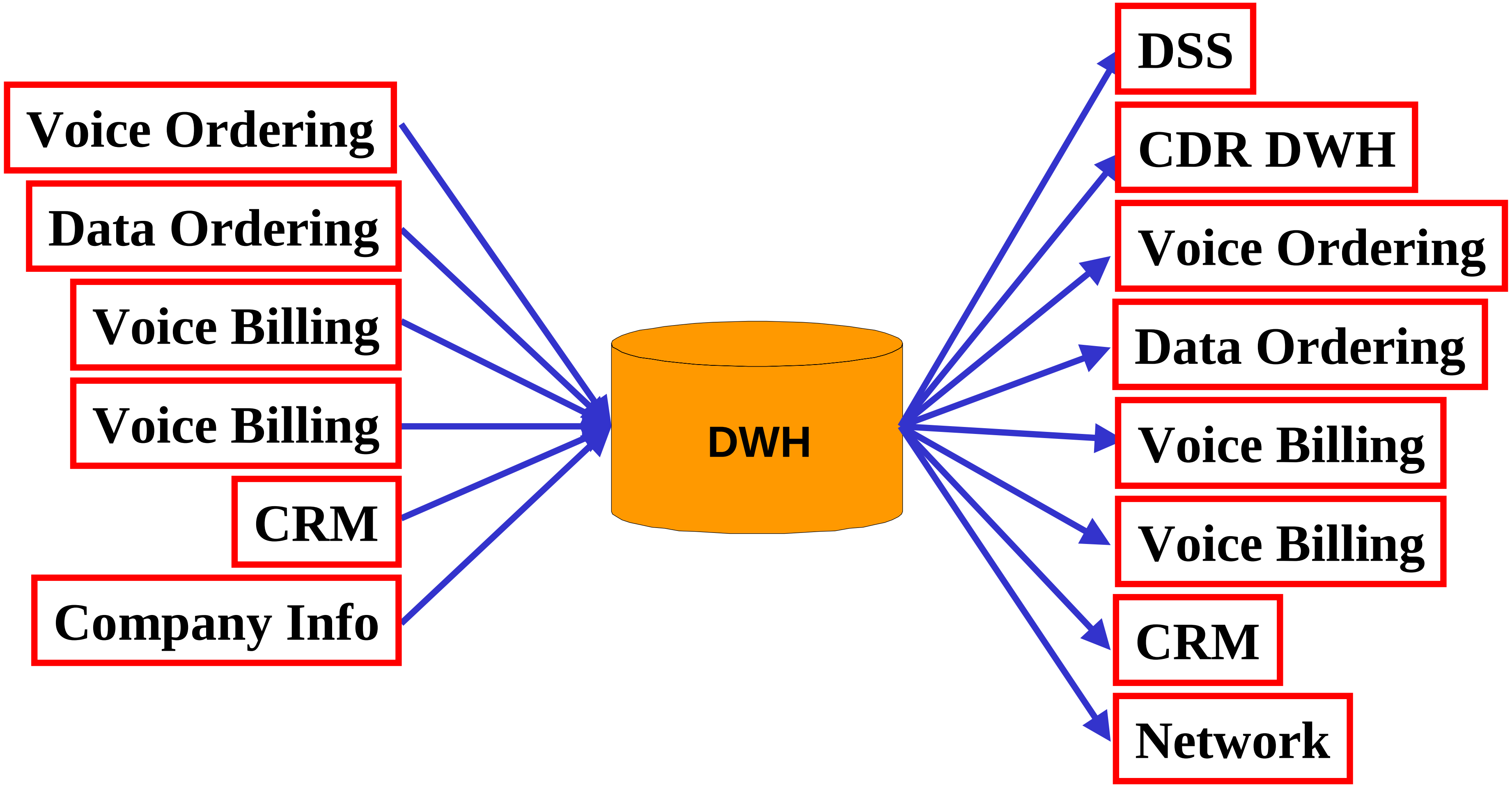

Profile
Lines & Services
Company Info
Site
Contractant
Payer
Service location
Revenue
Segment
Traffic
Contact Person
Contact History

DWH Now (2)

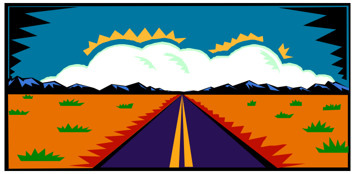
Use

- ***Customer segmentation (created in DW and distributed to all downstream systems)***
- ***Analyses (marketing, product development etc.)***
- ***Customer selections for marketing campaigns***
- ***Commission calculation (sales reps, dealers etc.)***
- ***Reporting (regular, ad-hoc)***
- ***Planning of network upgrade/tuning (based on traffic analyses)***

DWH Now (3)


Voice Ordering
Data Ordering
Voice Billing
Voice Billing
CRM
Company Info
DWH
DSS
CDR DWH
Voice Ordering
Data Ordering
Voice Billing
Voice Billing
CRM
Network

Future



Future

- *new ordering*

Tracking of:

Ø new products/services

Ø changes within the source systems

Ø internal customer requests

Thank you.

Q & A

Jan Medek - jan.medek@ct.cz

Paul Skaife - p.skaife@itcg.nl