



CONSULTANCY

Copyright © 1991 - 2012 R20/Consultancy B.V.,  
The Hague, The Netherlands. All rights  
reserved. No part of this material may be  
reproduced, stored in a retrieval system, or  
transmitted in any form or by any means,  
electronic, mechanical, photographic, or  
otherwise, without the explicit written permission  
of the copyright owners.

---

## Module 2: Best Practices for Developing a Reusable and Flexible Data Model

---



by  
**Rick F. van der Lans**  
**R20/Consultancy BV**  
Twitter @rick\_vanderlans  
[www.r20.nl](http://www.r20.nl)

---

for **INFORMATICA**<sup>®</sup>  
The Data Integration Company™

---

## The World of BI is Changing

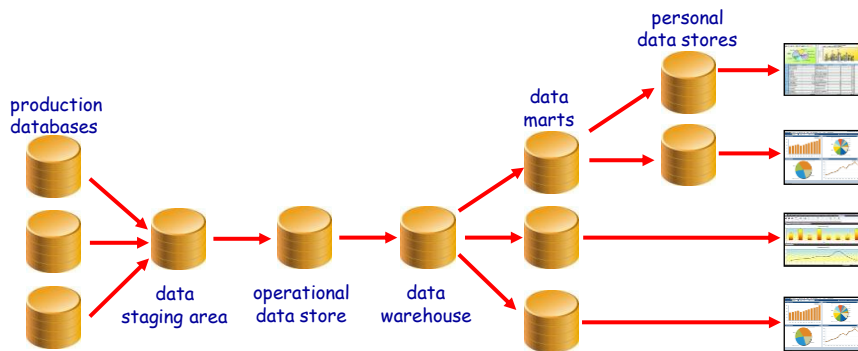
---



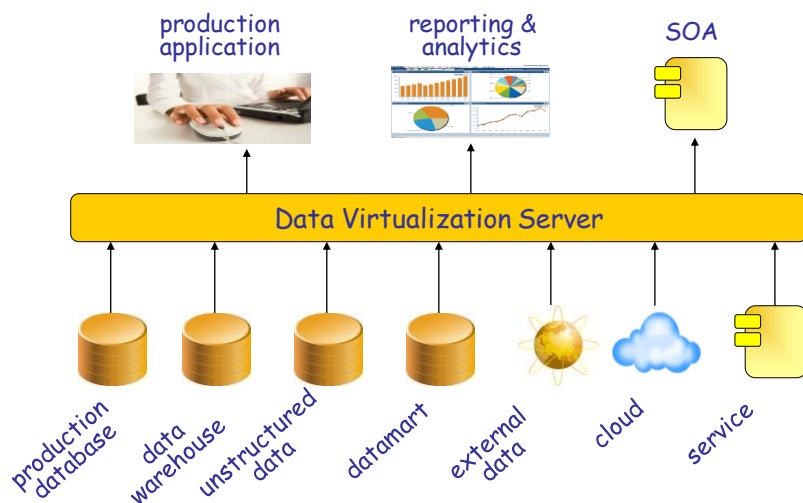
- New data sources
- New user groups
- New forms of reporting and analytics
- Operational BI
- More agility required
  - Time-to-market
- TCO has to go down



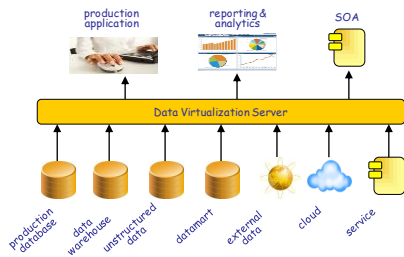
## Network of Databases (Labyrinth)



## Data Virtualization

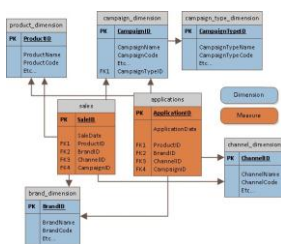


## Summary of Previous Module



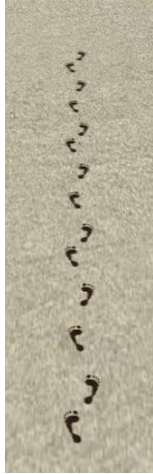
- Data virtualization leads to simplification of BI systems
- Data virtualization increases the agility of BI systems
- Data virtualization increases productivity, and improves maintenance

## The Data Model



- Classic components of a data model
  - Structural elements - entity, relation, attribute, integrity rule
  - Definitions
  - Examples
- Data integration components of a data model
  - Transformation specifications
  - Cleansing specifications
  - Integration specifications
  - Aggregation specifications

## Data Modeling Yesterday and Today



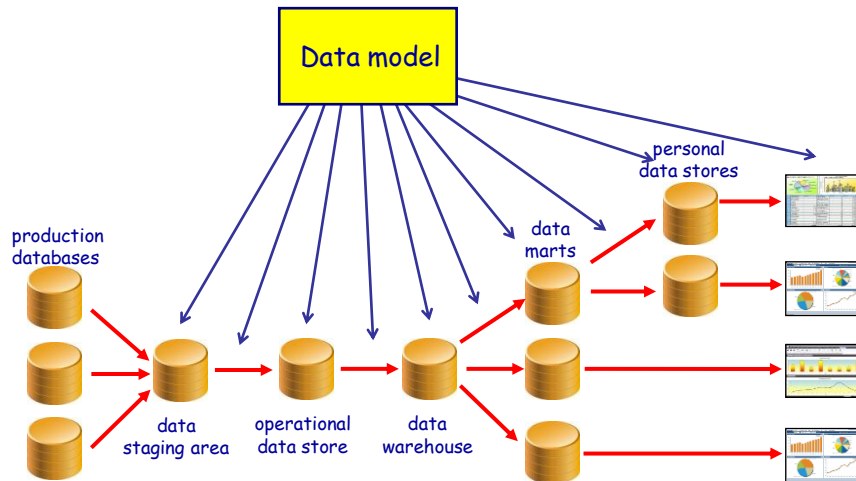
1. IT talks to user to understand information needs
2. IT creates and adapts data model (in isolation)
3. IT shows and explains data model to user
4. User gives feedback
5. IT goes back to drawing board and adapts data model
6. If not ok, go back to step 2
7. IT extends data model with integration specifications
8. Specialist implements and changes integration specifications
9. IT checks results (by looking at data)
10. If results not ok, go back to step 8
11. IT shows and explains data examples to user
12. If results not ok, go back to step 8
13. Done!

## Problems and Challenges



- IT not always an expert of the business area
- IT and users speak different languages
- Data model too abstract for users
- Time lost explaining to IT what information needs are
- Data model ends up at the wall
- Plans go out the window
  - "No battle plan survives contact with the enemy"
- Much time lost waiting
- Data model specifications end up everywhere

## Implementation of Data Model

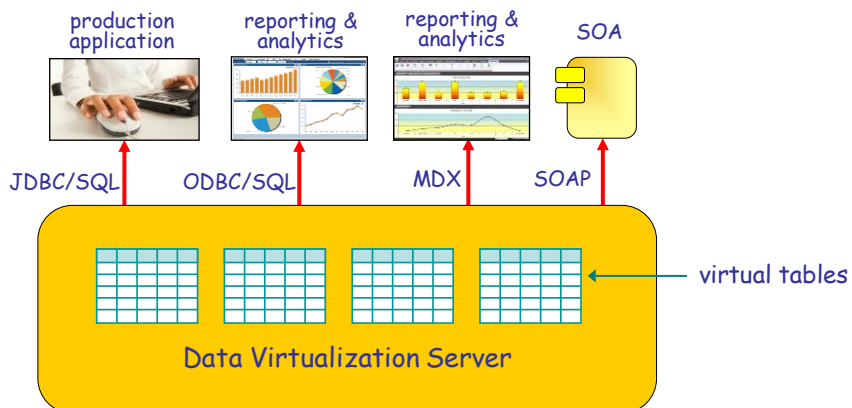


Copyright © 1991 - 2012 R20/Consultancy B.V., The Hague, The Netherlands



9

## The View From the Applications

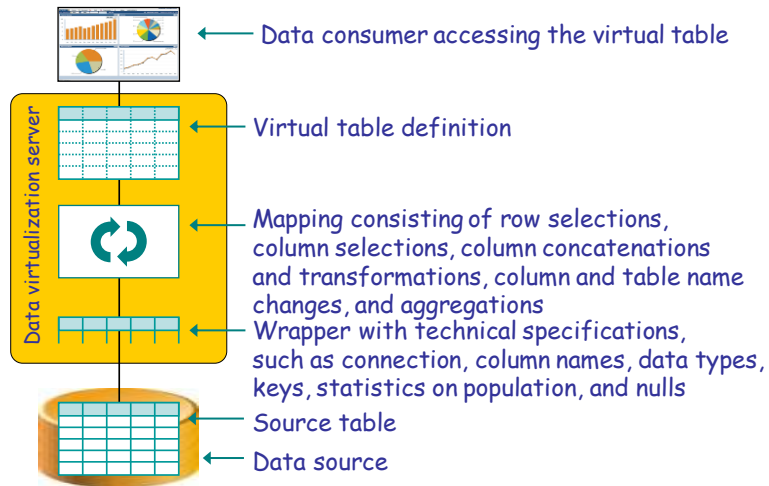


Copyright © 1991 - 2012 R20/Consultancy B.V., The Hague, The Netherlands

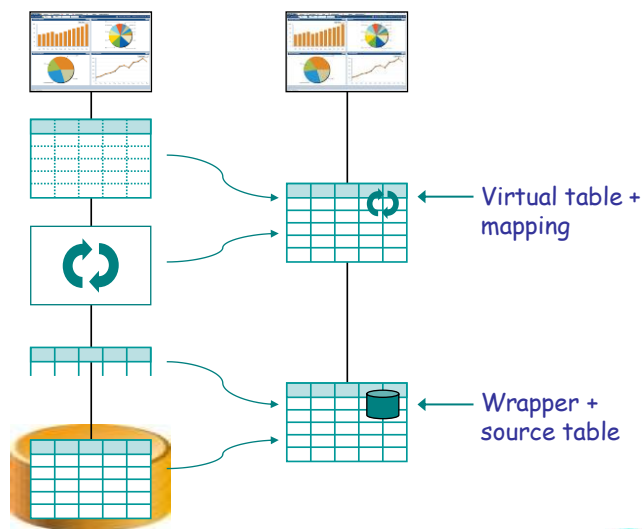


10

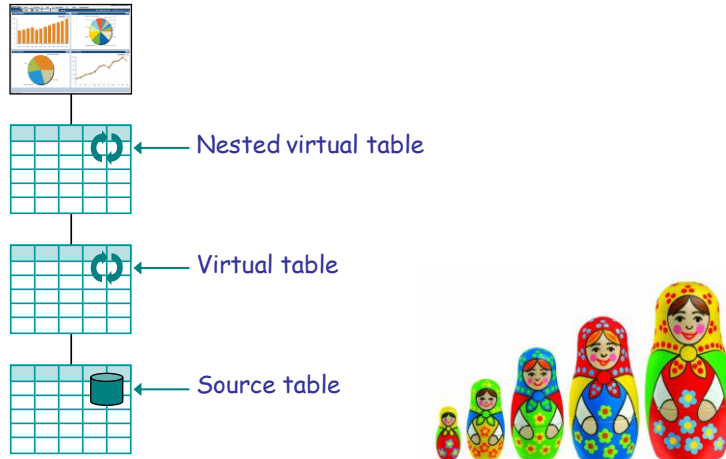
## The Virtual Table



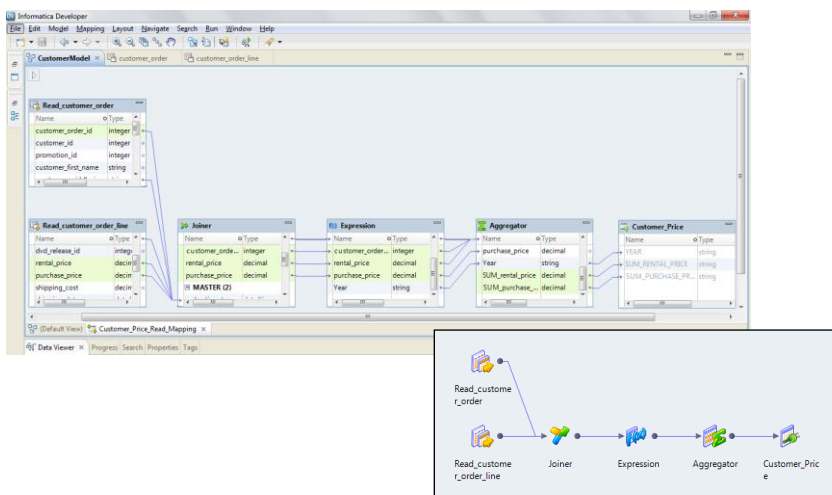
## Virtual Tables and Physical Tables



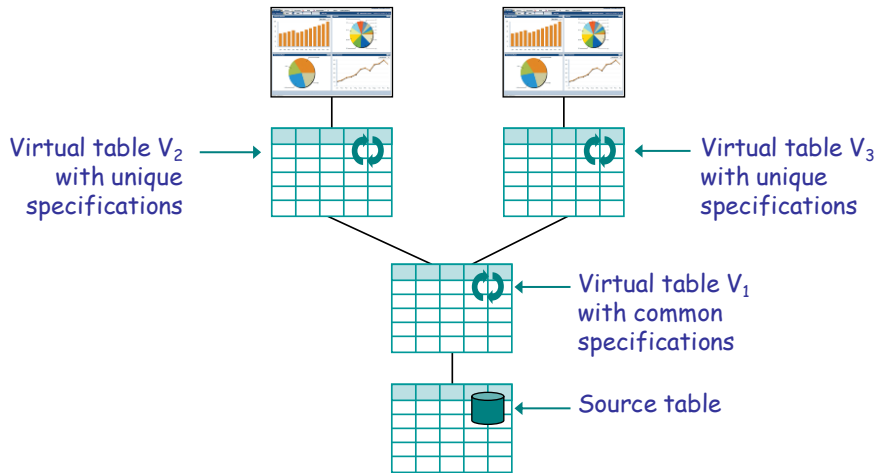
# Nesting Virtual Tables



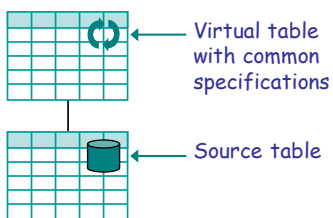
# The Mapping of a Virtual Table



## Sharing Common Specifications

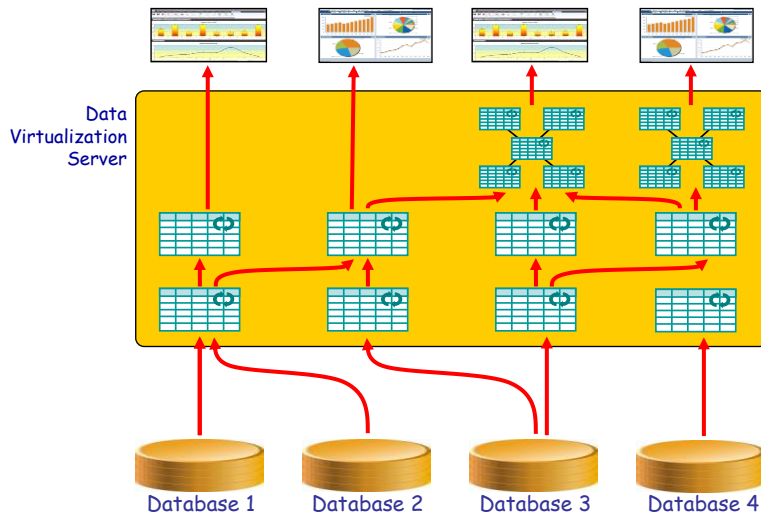


## Examples of Common Specifications



- Data integration specifications
  - Homogeneous and heterogeneous
- Transformations
  - Standardization of structure and values
- Cleansing operations
  - From simple to complex
- Aggregations

## Layers of Virtual Tables

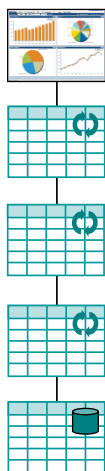


Copyright © 1991 - 2012 R20/Consultancy B.V., The Hague, The Netherlands



17

## (1) Bottom-Up Modeling



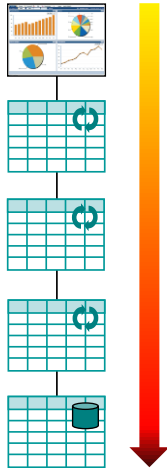
- From stored tables to virtual tables
- Design based on data available in existing data sources (less focus on the requirements of reports)
- Fits a more classic approach
- Much focus on sharing common specifications

Copyright © 1991 - 2012 R20/Consultancy B.V., The Hague, The Netherlands



18

## (2) Top-Down Modeling



- From virtual tables to stored tables
- Design based on requirements of reports
- Fits self-service BI and an agile approach
- However, what about sharing common specifications?

Copyright © 1991 - 2012 R20/Consultancy B.V., The Hague, The Netherlands



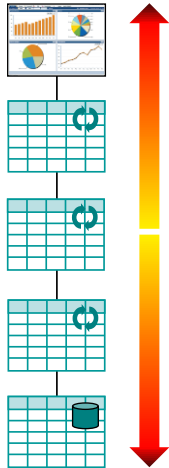
19

## Modeling Virtual Tables

Unbound virtual table

Column	DataType	Length	Nullable
FirstName	varchar	32	0
LastName	varchar	32	0
SSN	varchar	32	0
Age	int	10	0
Gender	varchar	32	0
DateOfBirth	date	10	0

### (3) Inside-Out Modeling



- From virtual tables with common specifications to stored tables and to reports
- Central virtual tables form a *canonical data model*
- Much focus on sharing common specifications
- Fits an agile and a classic approach
- High level of independence between reports and data sources

### Testing by Example

YEAR	SUM_RENTAL_PRICE	SUM_PURCHASE_PRICE
1 2000	2380.00	6164.31
2 2001	19090.00	91337.11

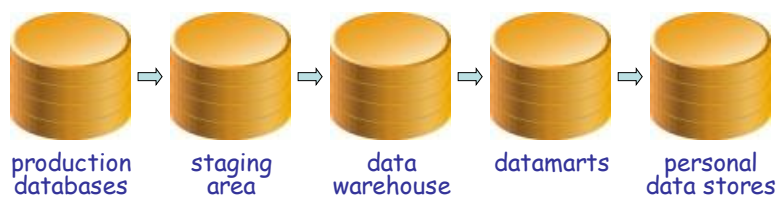
Virtual contents of virtual table

## Early and Iterative Participation by Analyst

The screenshot shows the Informatica Analyst interface. The main window displays a 'Column Profiling' table for the project 'Profile\_CUSTOMER\_PURCHASES'. The table has columns for Name, Unique Value, Unique %, NULL, NULL %, Datatype, Inferred %, Documented Datatype, Min Value, and Max Value. The 'ADDRESS\_STATE' column is highlighted in green, indicating a 'Fixed Length String' with 100% inferred and documented data. To the right, a 'Statistics' pane shows details for the selected column, including 'Maximum Length: 2', 'Minimum Length: 2', and 'Bottom (5)' values: CA, FL, IL, IN, KS, VA, OH.

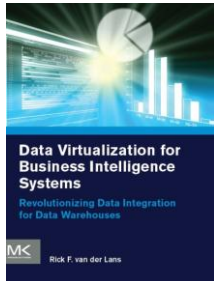
Name	Unique Value	Unique %	NULL	NULL %	Datatype	Inferred %	Documented Datatype	Min Value	Max Value
ADDRESS_LINE1	35	70.00	-	-	String(28)	100.00	varchar(50)	100 CARDI PO BOX 20	
ADDRESS_LINE2	1	2.00	50	100.00	-	-	varchar(50)		
ADDRESS_LINE3	1	2.00	50	100.00	-	-	varchar(50)		
ADDRESS_CITY	23	46.00	-	-	String(15)	100.00	varchar(50)	BALTIMORE WASHINGT	
ADDRESS_STATE	11	22.00	-	-	Fixed Length String	100.00	varchar(250)	CA VA	
ADDRESS_POSTAL	26	52.00	11	22.00	Integer(9)	100.00	varchar(50)	11372 95207	
PHONE	41	82.00	-	-	Fixed Length String	100.00	varchar(50)	(202) 785- (972) 583-	

## The Classic Approach



- Data model specifications are implemented independent of the data integration specifications
- Data model and integration specifications dispersed
- The business user not involved - minimal interaction
- Much time lost waiting
- Testing the specifications is time- and resource consuming

## The Data Virtualization Approach



- Data modeling becomes more collaborative and iterative
- More agile approach
- The data model becomes the reusable and flexible foundation for any BI project
- The data model is part of the running system (not a picture on a wall)
- IT can select different data modeling approaches
- Centralized implementation of common integration specifications - higher level of reuse
- No time lost waiting

## Five Modules



- **Module 1: How to Architect for Agility and Productivity in Your Projects**
- **Module 2: Best Practices for Developing a Reusable and Flexible Data Model**
- **Module 3: Best Practices for Real-Time Data Profiling and Data Quality**
- **Module 4: Best Practices for Ensuring Managed Self-Service BI**
- **Module 5: Best Practices for Scaling with Enterprise Needs**

## Rick F. van der Lans

---



**Rick F. van der Lans** is an independent consultant, lecturer, and author. He specializes in data warehousing, business intelligence, service oriented architectures, and database technology. He is managing director of R20/Consultancy B.V.. Rick has been involved in various projects in which SOA, data warehousing, and integration technology was applied.

Rick van der Lans is an internationally acclaimed lecturer. He has lectured professionally for the last twenty years in many of the European and Middle East countries, the USA, South America, and in Australia. He has been invited by several major software vendors to present keynote speeches.

He is the author of several books on computing, including his new *Data Virtualization for Business Intelligence Systems*. Some of these books are available in different languages. Books such as the popular *Introduction to SQL* is available in English, Dutch, Italian, Chinese, and German and is sold world wide. He also authored *The SQL Guide to Ingres* and *SQL for MySQL Developers*.

As author for BeyeNetwork.com, writer of whitepapers, as chairman for the annual European Data Warehouse and Business Intelligence Conference, and as columnist for a few IT magazines, he has close contacts with many vendors.

**R20/Consultancy B.V.** is located in The Hague, The Netherlands, [www.r20.nl](http://www.r20.nl). You can get in touch with Rick via:  
Email: [rick@r20.nl](mailto:rick@r20.nl)  
Twitter: [@Rick\\_vanderlans](https://twitter.com/Rick_vanderlans)  
LinkedIn: <http://www.linkedin.com/pub/rick-van-der-lans/9/207/223>

---

Copyright © 1991 - 2012 R20/Consultancy B.V., The Hague, The Netherlands



27