

Stories from the trenches

Partitioning as a design pattern

ITU 2012

Kennie Nybo Pontoppidan

Who am I?

Kennie Nybo Pontoppidan

knp@rehfeld.dk



Program

- About the XXXX project
- Data warehouse heroes and architecture
- Partitioning in many disguises

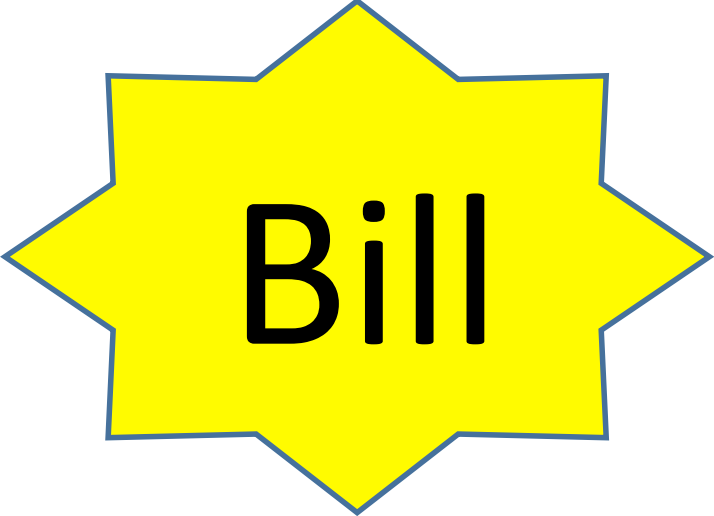
The XXXX project - Mission

- Jointly defined KPI's
 - View your own KPI's
 - Benchmark with "twins"
- Access to your own data
 - Common datamodel
 - Raw data

The XXXX project – DW Challenges

- Data from 30 different it-systems
- 7 subject areas
 - Citizens
 - Employees
 - Salaries
 - Absence
 - ERP
 - Budgets
 - Postings
 - Schools
 - ...
- Conformity
 - Within area
 - Between areas

Data warehouse heros



Bill



Ralph

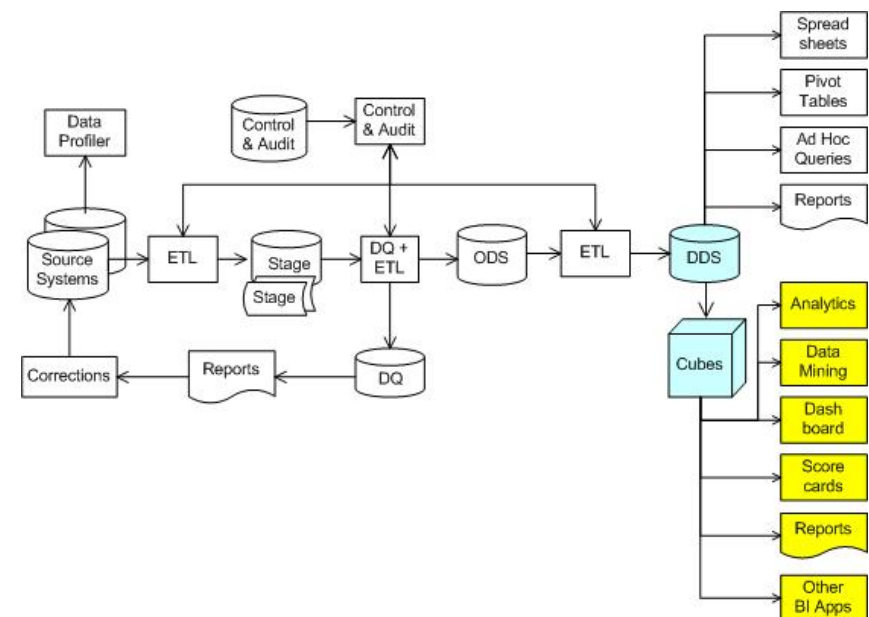


Dan

Bill Inmon

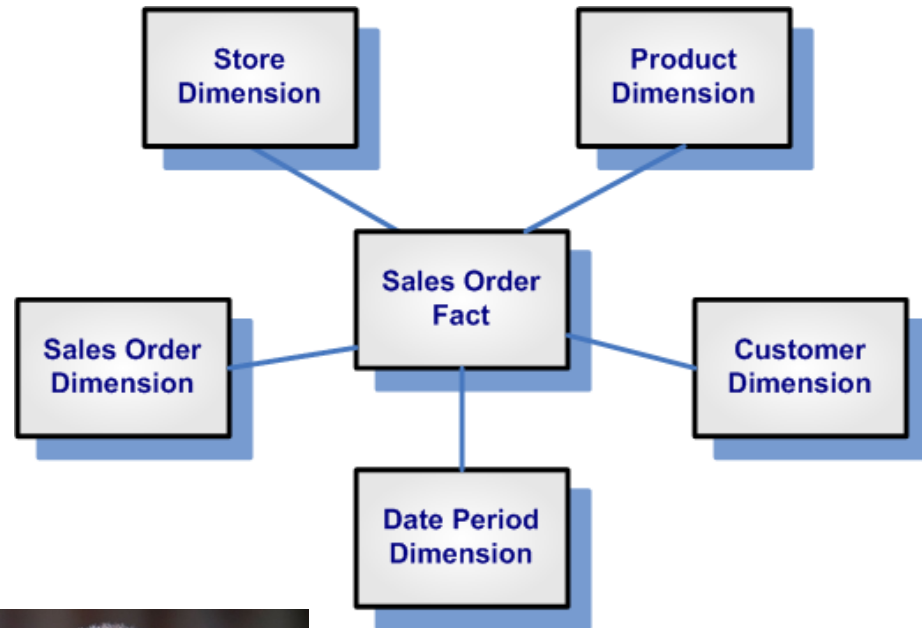


- Inventor of the word
 - Data warehouse
- Oracle reference architecture
- Top down approach



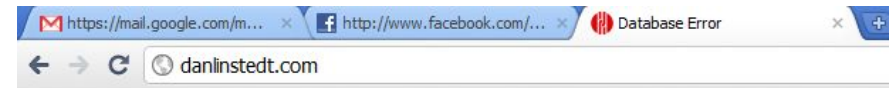
Ralph Kimball

- “Invented” dimension modelling
- Microsoft “reference architecture”
- Bottom up



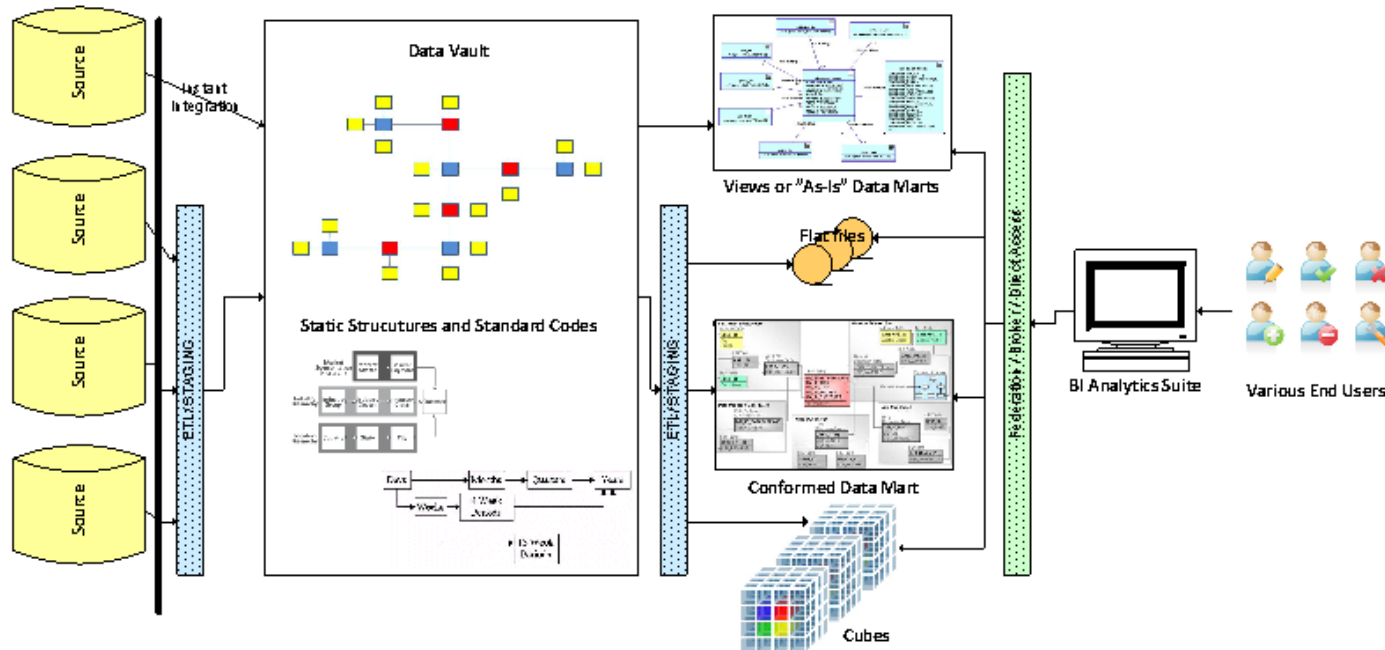
Dan Lindstedt

- Hybrid model

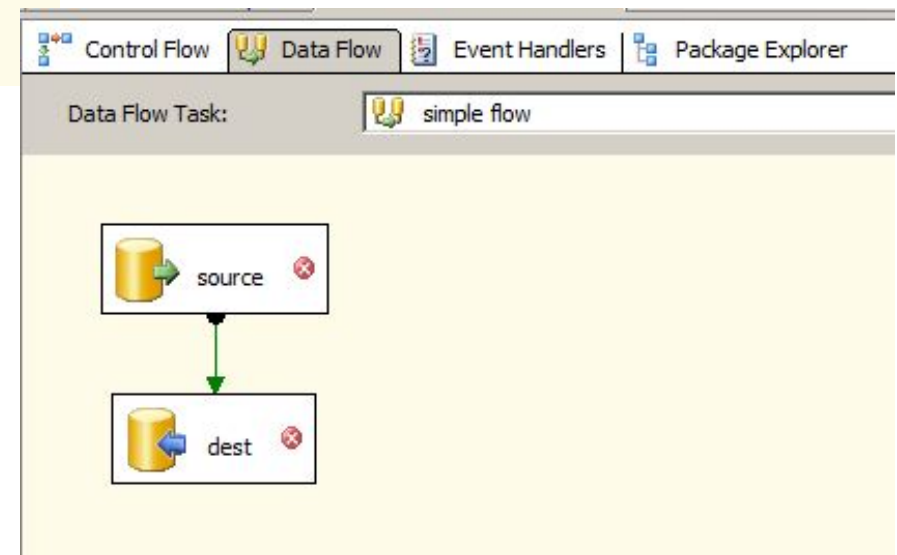
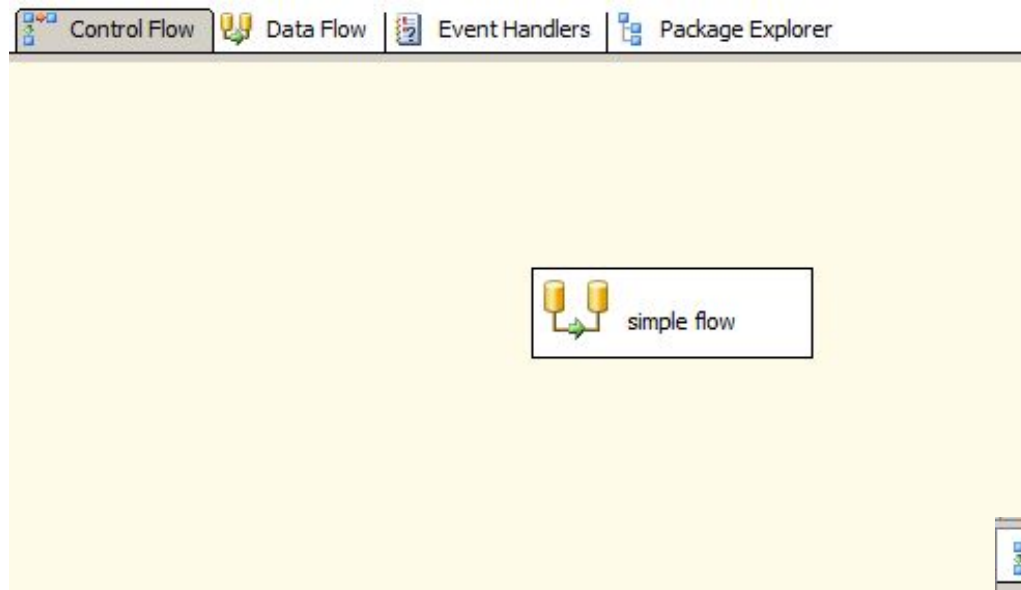


Error establishing a database connection

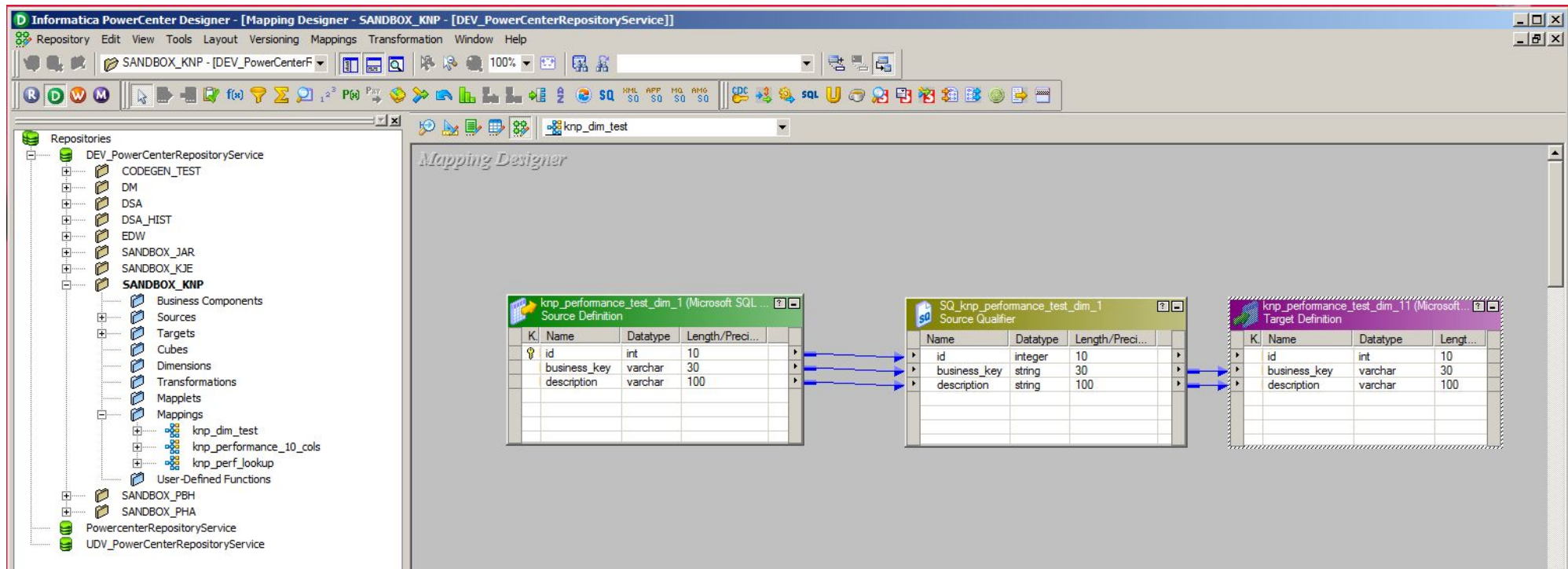
- Hubs, satellites and Links



SSIS package XML



Informatica PowerCenter XML



The screenshot displays the Informatica PowerCenter Designer interface. The left-hand pane shows the 'Repositories' tree, with 'Sandbox_KNP' expanded to show 'Mappings'. The main workspace, titled 'Mapping Designer', shows a mapping from 'knp_performance_test_dim_1' to 'knp_performance_test_dim_11'. The mapping is defined by a source qualifier 'SQ_knp_performance_test_dim_1'. The source and target tables are defined as follows:

K	Name	Datatype	Length/Preci...
	id	int	10
	business_key	varchar	30
	description	varchar	100

Source Definition: knp_performance_test_dim_1 (Microsoft SQL ...)

Name	Datatype	Length/Preci...
id	integer	10
business_key	string	30
description	string	100

Source Qualifier: SQ_knp_performance_test_dim_1

K	Name	Datatype	Lengt...
	id	int	10
	business_key	varchar	30
	description	varchar	100

Target Definition: knp_performance_test_dim_11 (Microsoft ...)

par·ti·tion (pär-tshn)*n.*

- 1.a. The act or process of *dividing something into parts*.
- b. The *state of being so divided*.
- 2.a. Something that divides or separates,
as a wall dividing one room or cubicle from another.
- b. A wall, septum, or other separating membrane in an organism.
- 3. A part or section into which something has been divided.
- ...

- MCM videos
 - Sqlskills.com



Partitioning as a design pattern

Dividing something into parts

- Files
- Database
- Table

State of being so divided

- Informatica PowerCenter (ETL vs. Customer)
- Developers vs (evil?) DBA's
- Backups (what is a full backup anyway)

Dividing something into parts – files

- Xml, csv, xls, fixed format
- 1 file
 - 1 or more tables
 - Data from 1 or more municipality
- 30 different file naming schemes

The collage illustrates various data formats and processing steps:

- Excel Icon:** A green icon with a white 'X' and a bar chart, representing spreadsheet data (xls).
- XML Document:** A document icon showing XML code for a quiz question and answer, with an 'XML' label in an orange box.
- ASCII Input Format Dialog:** A screenshot of a dialog box titled 'ASCII Input Format' with fields for 'Date Value Format' and 'Numeric Missing Value', and an 'Ok' button.
- Table:** A large table with columns A through L, containing student data including names, sex, locale, profile URLs, affiliations, and counts.

XML Code:

```
<?xml version="1.0"?>
<quiz>
  <question>
    Who was the forty-second
    president of the U.S.A.?
  </question>
  <answer>
    William Jefferson Clinton
  </answer>
  <!-- Note: We need to add
  more questions later.-->
</quiz>
```

Table Data (Columns A-L):

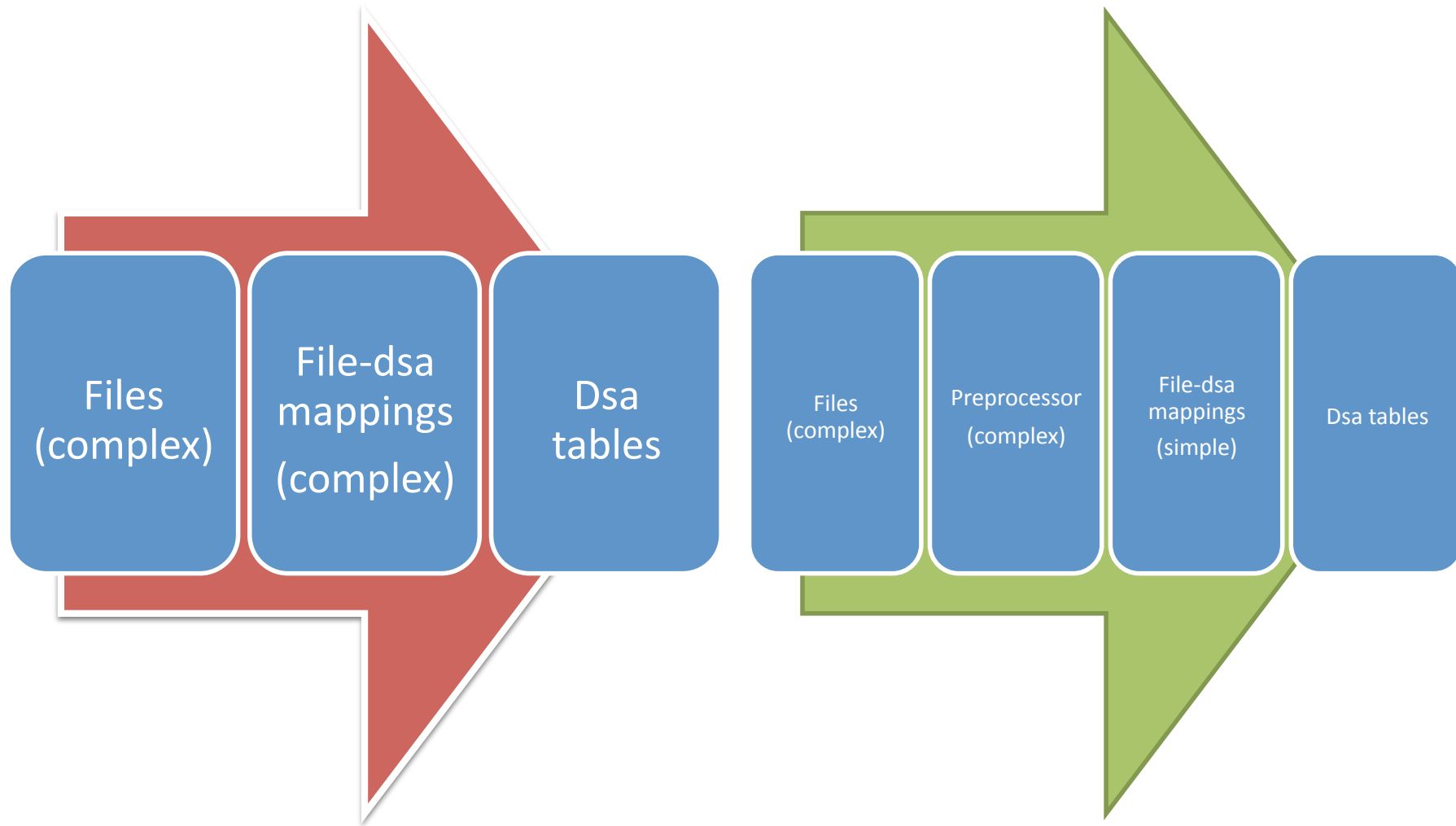
A	B	C	D	E	F	G	H	I	J	K	L	
1	first_name	last_name	sex	locale	profile_url	pic_big	affiliation	affiliation	affiliation	profile_url	affiliation	count
2	Tyrone	Roussel	male	en_US	http://www	http://pr	Yale Univ	college	Alumnus	2011-01-22	2006	595
3	Ted	Fronk	male	en_US	http://www	http://pr	New Haver	college	Undergra	2011-02-09	2011	492
4	Roslyn	Heitmann	female	en_US	http://www	http://pr	Princeton	Universit		2011-01-27		1576
5	Clare	Gostomski	female	en_US	http://www	http://pr	Princeton	college	Alumnus	2011-01-30	0	686
6	Jessie	Mahnke	female	en_US	http://www	http://pr	Stanford	college	Alumnus	2011-02-15	2008	1926
7	Lonnie	Pinales	male	en_US	http://www	http://pr	Stanford	college	Undergra	2011-01-18	0	
8	Carlene	Voorhies	female	en_US	http://www	http://pr	Columbia	Universit		2011-01-06		1656
9	Kurt	Metayer	male	en_US	http://www	http://pr	New York	college	Undergra	2010-10-04	2009	317
10	Maricela	Seel	female	en_US	http://www	http://pr	Massachu	college	Undergra	2010-09-29	2010	159
11	Erik	Gunnels	female	en_US	http://www	http://pr	Cambridg	college	Alumnus	2011-02-05	2009	2194
12	Darren	Thrall	female	en_US	http://www	http://pr	Brown Un	college	Alumnus	2011-02-10	2007	2477
13	Earnestin	Pruden	female	en_US	http://www	http://pr	Providen	college	Alumnus	2011-02-04	2007	1562
14	Dona	Petros	female	en_GB	http://www	http://pr	Dartmouth	College		2011-02-09		2396
15	Alejandra	Ram	female	en_US	http://www	http://pr	Hanover	college	Undergra	2011-01-27	2009	2678
16	Karina	Foord	female	en_US	http://www	http://pr	Universit	college	Grad Stud	2011-02-15	2009	3079
17	Kathrine	Berquist	female	en_US	http://www	http://pr	Philadelphia	PA		2011-01-14		130
18	Ericka	Nitta	male	en_US	http://www	http://pr	California	college	Undergra	2011-01-03	2009	1110
19	Jamie	Neeley	female	en_US	http://www	http://pr	Pasadena	college	Undergra	2011-02-15	0	1763
20	Hugh	Wakemar	female	en_US	http://www	http://pr	Washingt	college	Undergra	2011-02-17	2009	861
21	Clayton	Steuck	male	en_US	http://www	http://pr	Saint Loui	college	Undergra	2011-02-01	2009	338
22	Loraine	Girton	female	en_US	http://www	http://pr	Georgeto	college	Undergra	2011-02-17	0	2634
23	Althea	Seager	female	en_US	http://www	http://pr	Washingt	college	Undergra	2011-02-12	2013	1011
24	Tanisha	Hannum	female	en_US	http://www	http://pr	Cornell U	college	Grad Stud	2011-02-15	0	2285
25	Rae	Boos	female	en_US	http://www	http://pr	Ithaca, NY	college	Alumnus	2011-02-05	0	564
26	Odessa	Gerardo	male	en_US	http://www	http://pr	Duke University			2011-02-08		3357
27	Nelson	Balls	en_US	en_US	http://www	http://pr	Durham, NC			2011-01-23		126
28	Ted	Allsup	female	en_US	http://www	http://pr	University of Califo			2011-01-01		858
29	Kurt	Creceilius	female	en_US	http://www	http://pr	Berkeley, work		Undergra	2011-02-14	0	803

Dividing something into parts - files

- Csv
- 1 file
 - 1 tables
 - Data from just 1 municipality
- 1 file naming scheme



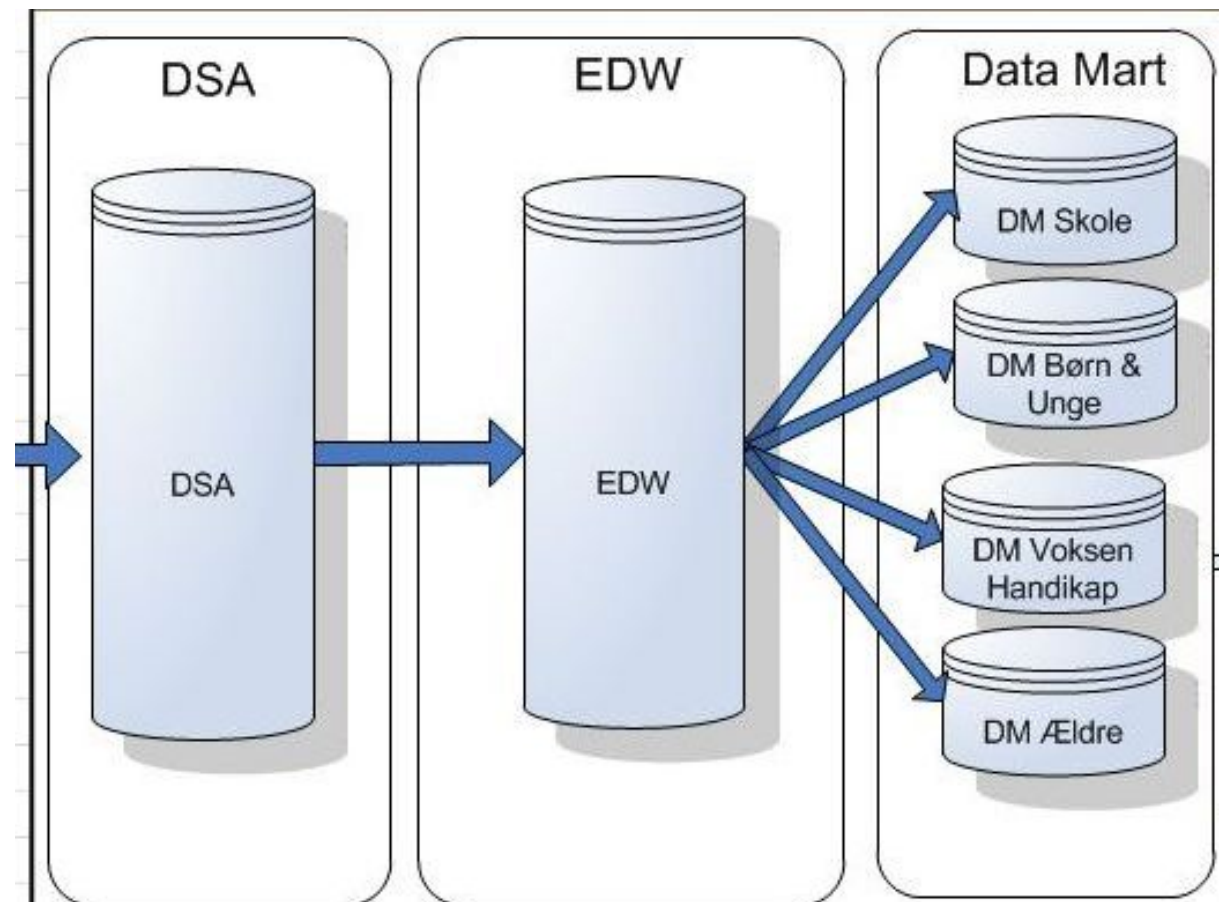
Dividing something into parts - files



Dividing something into parts - Database

1 data warehouse layer = 1 database

- Scaling
- IO-pattern
for a data flow



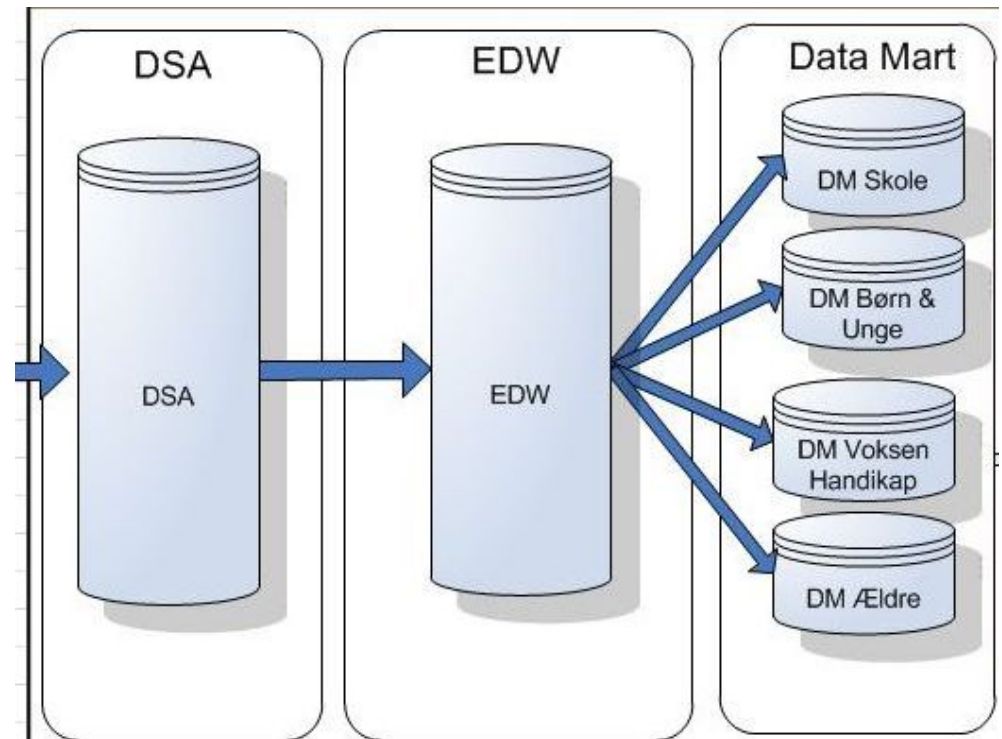
Dividing something into parts - Tables

Table partitioning (2005 EE feature)

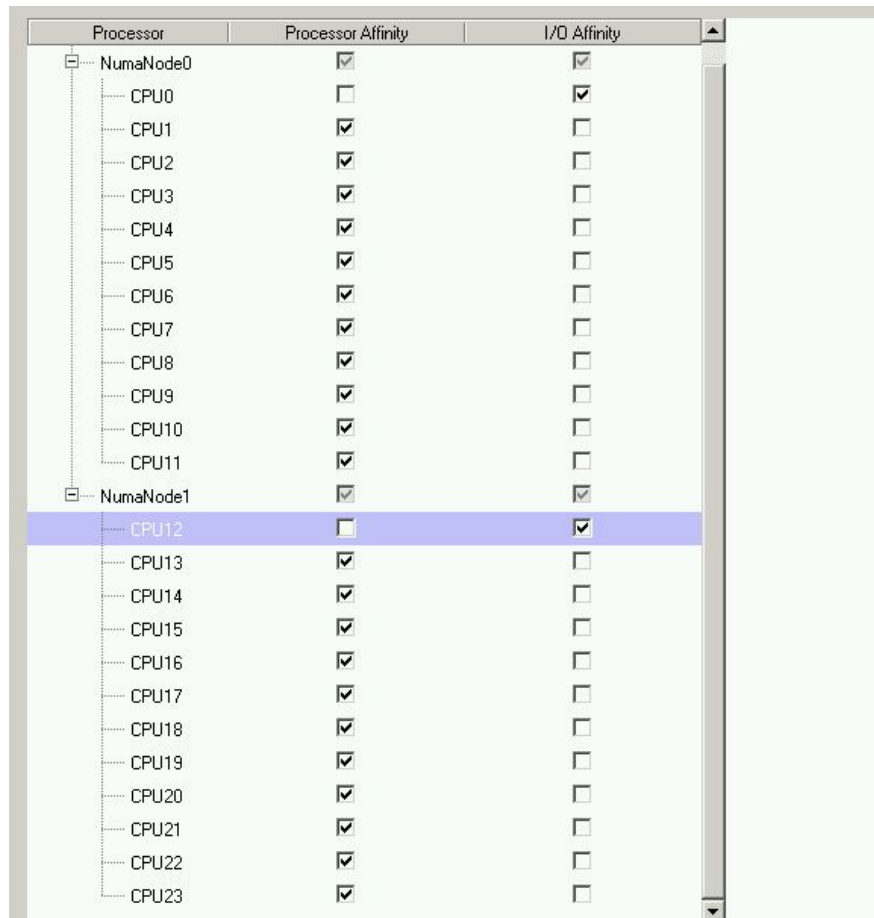
- Pruning
 - Divide the data warehouse into 100 parts 1/100 the size
- Switching
 - Separation of readers and writers
 - Fast

Dividing something into parts - Tables

- DSA
 - Partition by municipality_id, month_year
- EDW and data marts
 - Partition by municipality_id



Dividing something into parts – CPU's

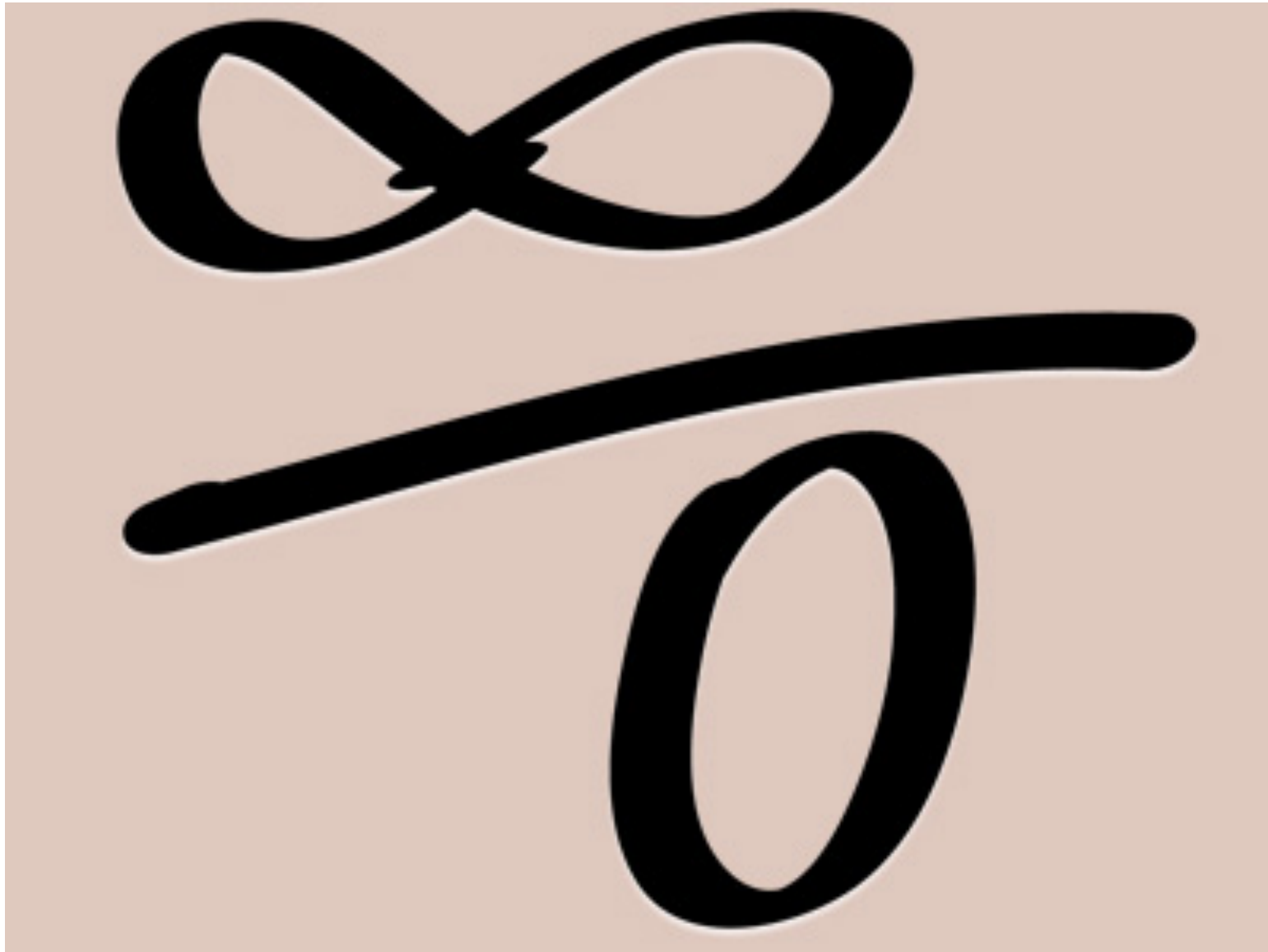


The screenshot shows a Windows Task Manager Performance tab with the 'CPU' section selected. It displays a list of processors organized into two NUMA nodes. The 'Processor Affinity' column shows checkboxes for each processor, and the 'I/O Affinity' column shows checkboxes for each processor. The row for CPU12 is highlighted in blue.

Processor	Processor Affinity	I/O Affinity
NumaNode0	☒	☒
CPU0	☐	☒
CPU1	☒	☐
CPU2	☒	☐
CPU3	☒	☐
CPU4	☒	☐
CPU5	☒	☐
CPU6	☒	☐
CPU7	☒	☐
CPU8	☒	☐
CPU9	☒	☐
CPU10	☒	☐
CPU11	☒	☐
NumaNode1	☒	☒
CPU12	☐	☒
CPU13	☒	☐
CPU14	☒	☐
CPU15	☒	☐
CPU16	☒	☐
CPU17	☒	☐
CPU18	☒	☐
CPU19	☒	☐
CPU20	☒	☐
CPU21	☒	☐
CPU22	☒	☐
CPU23	☒	☐

- Affinity masking
- Not really possible in Oracle – even on Windows

State of being so divided



State of being so divided

- Informatica PowerCenter (ETL vs. Customer)
- Meta data
 - File definitions
 - Table definitions in all layers
 - Simple transformations
- Autogenerate mapping code from meta data

State of being so divided

Developers vs. (evil) DBA's

- Meta data on table definitions
 - Script ddl from meta data
 - Hide partitioning in ddl
- Partition Scheme
 - Change File group design