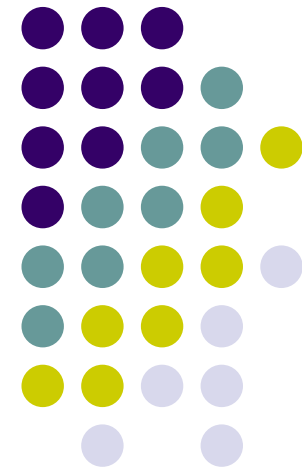


Automatic Data Fusion with HumMer

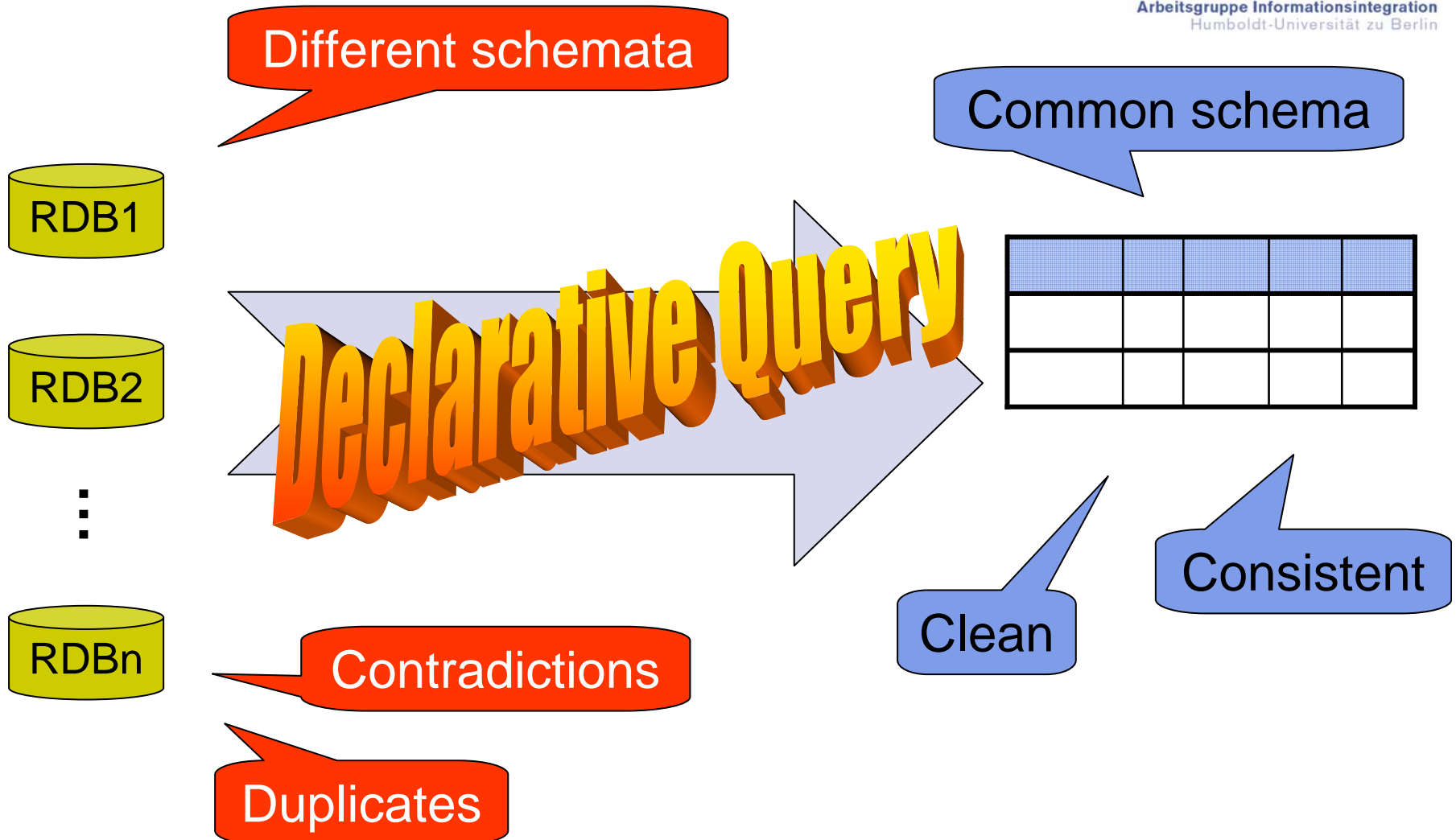


Trondheim, 30.8.2005
Alexander Bilke, Jens
Bleiholder, Christoph Böhm,
Karsten Draba, Felix
Naumann, Melanie Weis
Humboldt-Universität zu
Berlin
University of Technology
Berlin



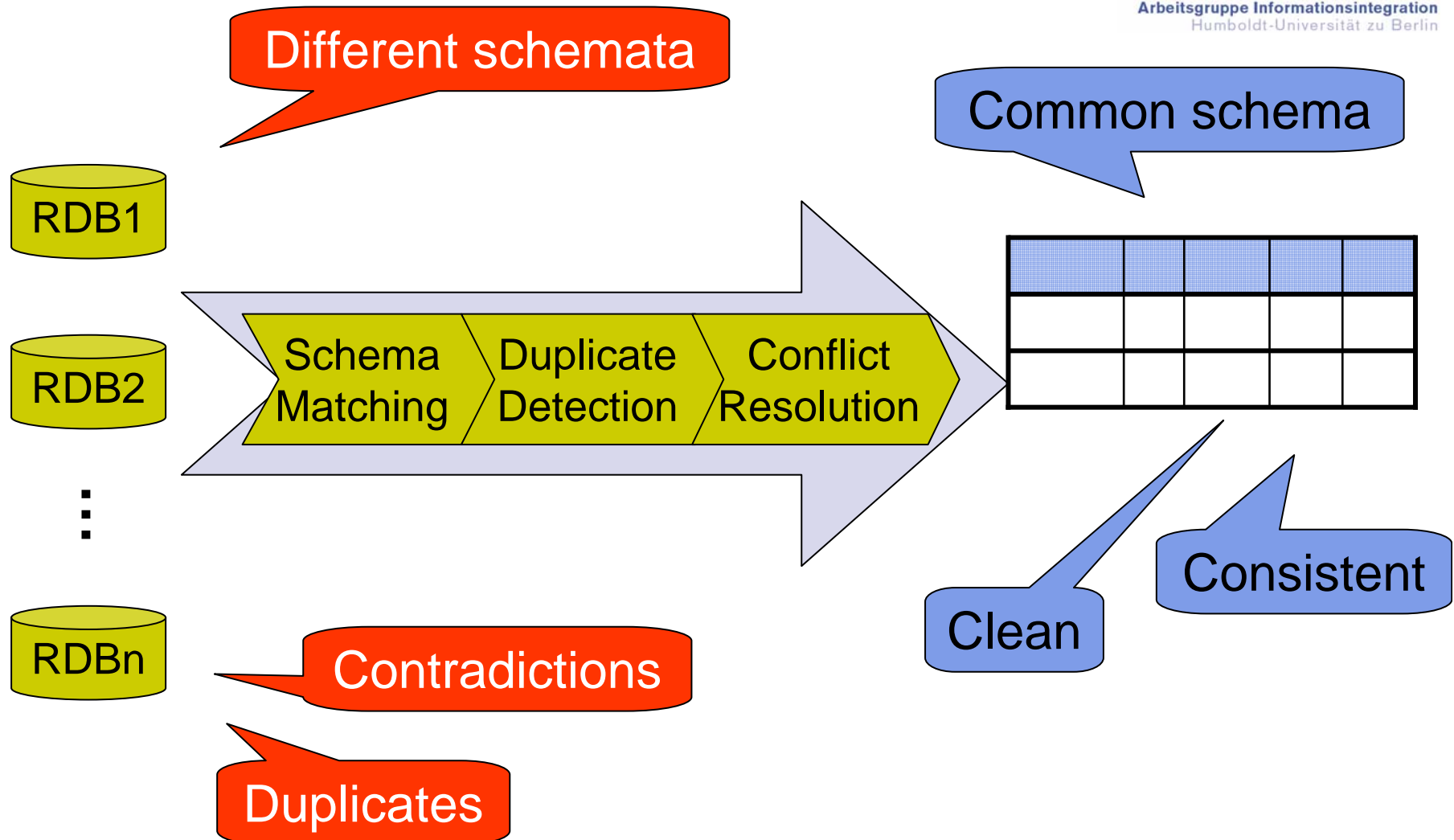


The Problem





The Solution: HumMer





Fully Automated Data Fusion

HumMer
Arbeitsgruppe Informationsintegration
Humboldt-Universität zu Berlin

Schema
Matching

Duplicate
Detection

Conflict
Resolution

Neat trick:
1. Tuple-
based
schema
matching
[ICDE 2005]

Neat trick:
1. Compare
data values
2. Compare
values of
„children“
[SIGMOD 2005]

Neat trick:
1. Extend SQL
by **FUSE BY**
clause
2. Optimize like
grouping
[ADBIS 2005]



HumMer-Demo File Extra Help

0. Sources Data Sources

1. Matching

2. Duplicate Definition

3. Duplicate Detection

4. Conflict

5. Result

HumMer-Demo File Extra Help

0. Sources

1. Matching

2. Duplicate Definition

3. Duplicate Detection

4. Conflict

5. Result

HumMer-Demo File Extra Help

0. Sources

1. Matching

2. Duplicate Definition

3. Duplicate Detection

4. Conflict

5. Result

HumMer-Demo File Extra Help

0. Sources

1. Matching

2. Duplicate Definition

3. Duplicate Detection

4. Conflict

5. Result

Start Over

Start Over

Start Over

Start Over

Start Over

Start Over

Done

Back Next

ationsintegration
versität zu Berlin

Result

Choose the fusion implementation to use default Execute

back next

	CLU...	TITLE	VERSI...	COUN...	YEAR	ORIGI...	GENRE	DIREC.
	VOTE	COALESCE	COALES...	MAX	COALES...	LAST	COALES.	
13	87	HOPE FLOATS	engl...	USA	1998	Hop...	Unterhaltu...	Fore...
14	84	GOOD WILL H...	engl...	USA	1998	Goo...	Drama	Gus...
15	83	GODZILLA	engl...	USA	1998	God...	Fantasy, S...	Rola...
16	80	Gadjo Dilo GADJO DILO	franz...	F/Rum	1998	Gadj...	Unterhaltu...	Ton...
17	77	Deconstructin...	engl...	USA	1998	Dec...	Komödie/...	Woo...
18	74	City Of Angels	engl...	USA	1998	City ...	Drama	Brad...
19	69	BOOGIE NIGH...	engl...	USA	1998	Boo...	biografisc...	Paul ...
20	65	Antz	engl...	USA	1998	Antz	Animation,...	Darn...
21	57	SPIDER			2002		Drama	
22	51	SECRETARY			2002		Komödie	
23	49	S.F.W.			1994		Komödie	
24	31	Intolerable Cr...			2003		Komödie	
25	25	GANGSTER N...			2000		Gangsterfi...	
26	24	From Hell			2001			
27	17	DEATHWATCH			2002		Kriegsfilm	
28	15	CHARLOTTE ...			2001		Melodram	
29	11	Big Fish			2003		Drama	

Rows: 0:99

Duplicate Contradiction Uncertainty Unique



What can I click on?

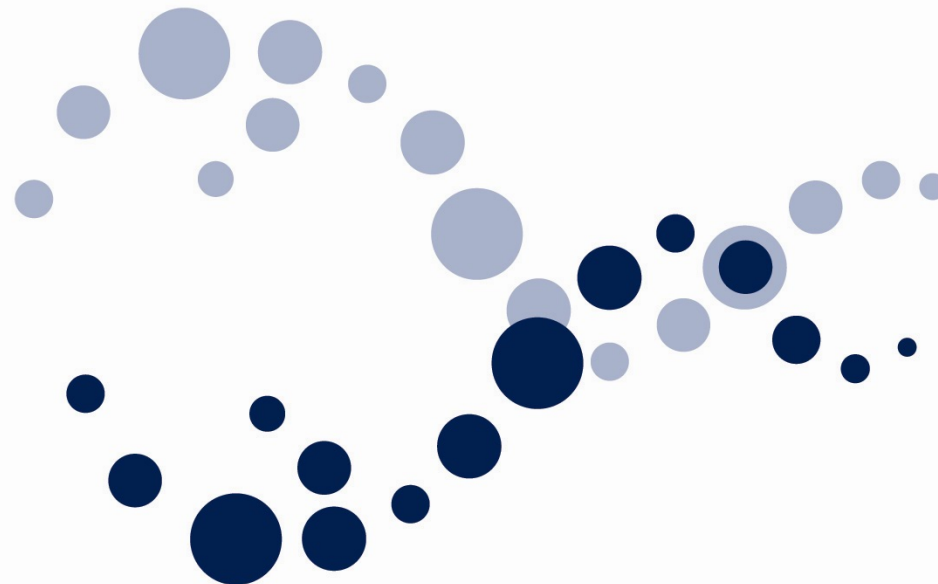
- Specify data sources
- View and adjust **Matching**
 - optional
- View and adjust **Duplicates**
 - optional
- Pick **Conflict Resolution Function**
 - optional
- See **Query Result and its Lineage**

Visit us!

- Today 2pm
- Thursday 11am
- Room: right here



HumMer
Arbeitsgruppe Informationsintegration
Humboldt-Universität zu Berlin



HumMer
Arbeitsgruppe Informationsintegration
Humboldt-Universität zu Berlin