



POZNAN UNIVERSITY OF TECHNOLOGY

Introduction to Data Warehouse Benchmarking

Robert Wrembel
Poznan University of Technology
Institute of Computing Science
Robert.Wrembel@cs.put.poznan.pl
www.cs.put.poznan.pl/rwrembel



Outline

- ⇒ Introduction to benchmarks
- ⇒ Transaction Processing Council (TPC)
 - TPC-DS
 - TPC-H
- ⇒ Star schema benchmark
- ⇒ OLAP Council
 - APB-1 OLAP Benchmark



Benchmark

- ⇒ **DB schema definition**
- ⇒ **Scalable DB size**
- ⇒ **Workload (queries)**
- ⇒ **Performance metrics**

- ⇒ **DB benchmarks**
 - **Transaction Processing Council (TPC)**
 - www.tpc.org
- ⇒ **ETL benchmarks**
- ⇒ **Cloud benchmarks**
 - **Yahoo! Cloud Serving Benchmark**
 - **Development of Cloud & Big Data Verification & Validation Methodology**

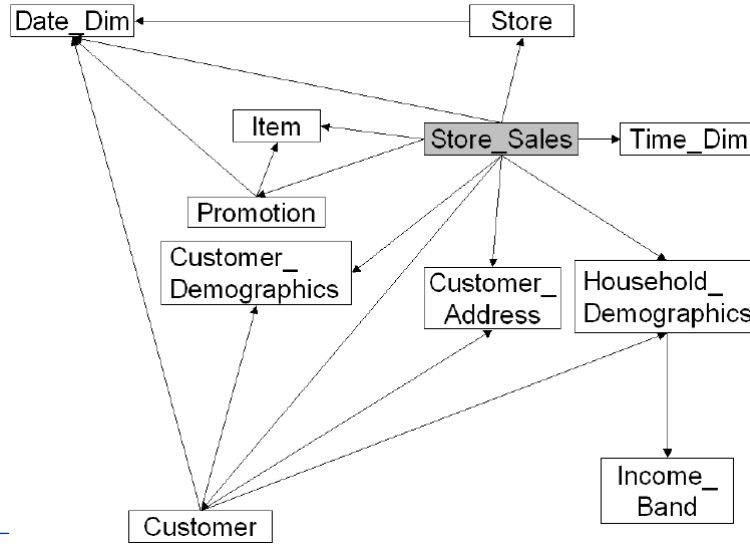


TPC-DS

- ⇒ **New DSS benchmark (Apr 2012)**
- ⇒ **Designed to be broadly representative of modern decision support systems**
- ⇒ **99 queries**
 - **various operational requirements and complexities (e.g., ad-hoc, reporting, iterative OLAP, data mining)**
 - **high CPU and IO load**



TPC-DS



R.Wrembel - Poznan University of Technology

5



TPC-DS

Table	Avr Row Size in bytes	Sample Row Counts. Number of rows are within 1/100 th Percent of these numbers							
		1GB	100GB	300GB	1TB	3TB	10TB	30TB	100TB
call_center	305	8	30	36	42	48	54	60	60
catalog_page	139	6	30	36	42	48	54	60	60
catalog_returns	166	11,718	20,400	26,000	30,000	36,000	40,000	46,000	50,000
catalog_sales	226	144,067	14,404,374	43,193,472	143,996,756	432,018,033	1,440,033,112	4,319,925,093	14,400,509,482
customer	132	1,441,546	143,997,065	431,969,836	1,439,980,416	4,320,078,880	14,399,964,710	43,200,404,822	144,001,282,896
customer_addr ess	110	100,000	2,000,000	5,000,000	12,000,000	30,000,000	65,000,000	80,000,000	100,000,000
customer_ demographics	42	50,000	1,000,000	2,500,000	6,000,000	15,000,000	32,500,000	40,000,000	50,000,000
date_dim	141	1,920,800	1,920,800	1,920,800	1,920,800	1,920,800	1,920,800	1,920,800	1,920,800
household_ demographics	21	73,049	73,049	73,049	73,049	73,049	73,049	73,049	73,049
income_band	16	7,200	7,200	7,200	7,200	7,200	7,200	7,200	7,200
inventory	16	20	20	20	20	20	20	20	20
item	281	11,745,000	399,330,000	585,684,000	783,000,000	1,033,560,000	1,311,525,000	1,627,857,000	1,965,337,830
promotions	124	18,000	204,000	264,000	300,000	360,000	402,000	462,000	502,000
reason	38	300	1,000	1,300	1,500	1,800	2,000	2,300	2,500
ship_mode	56	35	55	60	65	67	70	72	75
store	263	20	20	20	20	20	20	20	20
store_returns	134	12	402	804	1,002	1,350	1,500	1,704	1,902
store_sales	164	287,514	28,795,080	86,393,244	287,999,764	863,989,652	2,879,970,104	8,639,952,111	28,799,941,488
time_dim	59	2,880,404	287,997,024	864,001,869	2,879,987,999	8,639,936,081	28,799,983,563	86,399,341,874	287,998,696,432
warehouse	117	86,400	86,400	86,400	86,400	86,400	86,400	86,400	86,400
web_page	96	5	15	17	20	22	25	27	30
web_returns	162	60	2,040	2,604	3,000	3,600	4,002	4,602	5,004
web_sales	226	71,763	7,197,670	21,599,377	71,997,522	216,003,761	720,020,485	2,160,007,345	7,200,085,924
web_site	292	719,384	72,001,237	216,009,853	720,000,376	2,158,968,881	7,199,963,324	21,600,036,511	71,999,537,298

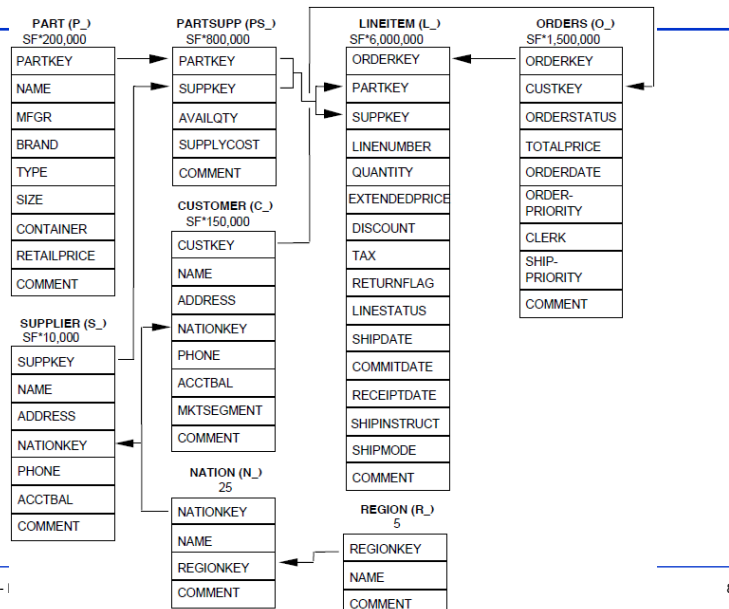


TPC-H

- ⇒ A suite of business oriented ad-hoc queries and concurrent data modifications
- ⇒ 22 queries
- ⇒ 2 refresh functions
 - inserts
 - deletes
- ⇒ Scaling
 - 1GB, 10GB, 30GB, 100GB, 300GB, 1TB, 3TB, 10TB, 30TB, 100TB



TPC-H



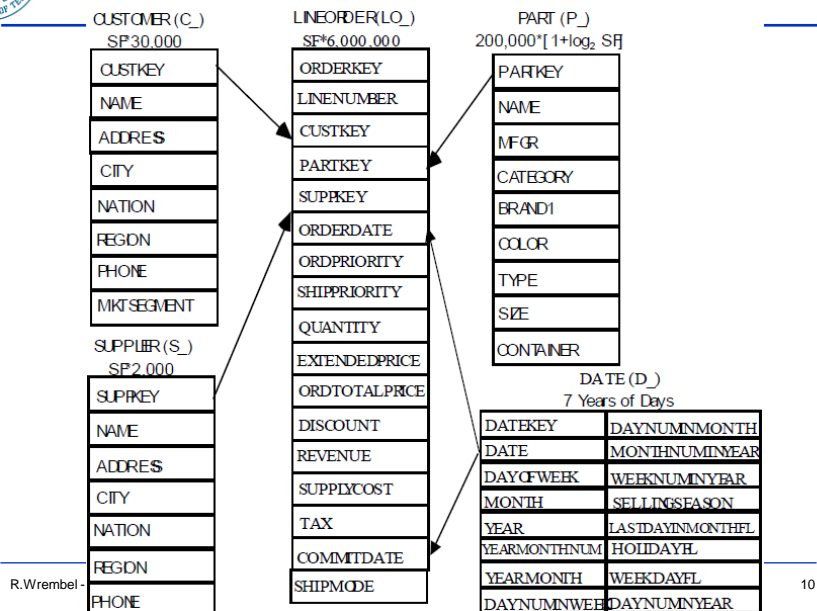


Star Schema Benchmark

- Modified TPC-H
 - typical star schema
- 13 queries organized in 4 categories



SSB





APB-1

➤ 10 Queries

- 10% Query 1 - Channel Sales Analysis
- 10% Query 2 - Customer Margin Analysis
- 15% Query 3 - Product Inventory Analysis
- 3% Query 4 - Time Series Analysis
- 5% Query 5 - Customer Budget
- 5% Query 6 - Product Budget
- 15% Query 7 - Forecast Analysis
- 20% Query 8 - Budget Performance
- 15% Query 9 - Forecast Performance
- 2% Query 10 - Ad Hoc



APB-1

