



Temporal Information Systems

(MA-INF 3302)

SS 2015

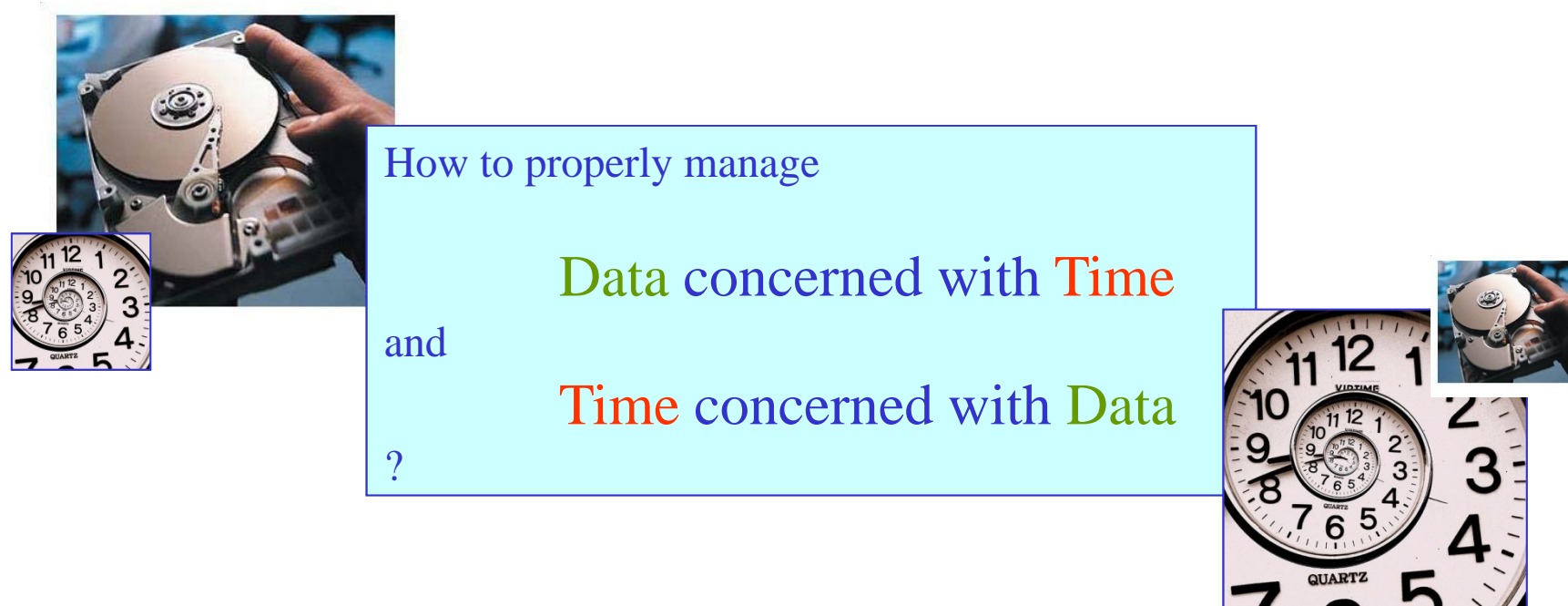
Prof. Dr. Rainer Manthey
Institut für Informatik III
Universität Bonn



TIS: Topic in a Nutshell

„Temporal“: „time-related“ (from lat. „tempus“: „**time**“)

In just one sentence, the **topic** of this lecture could be summarized as follows:



How to properly manage

Data concerned with **Time**

and

Time concerned with **Data**

?

Whereas the 1st phrase („Data concerned with Time“) may appear **reasonable**, the 2nd phrase („Time concerned with Data“) may be **irritating** – however, it is **meant exactly as stated!**

Temporal Information Systems

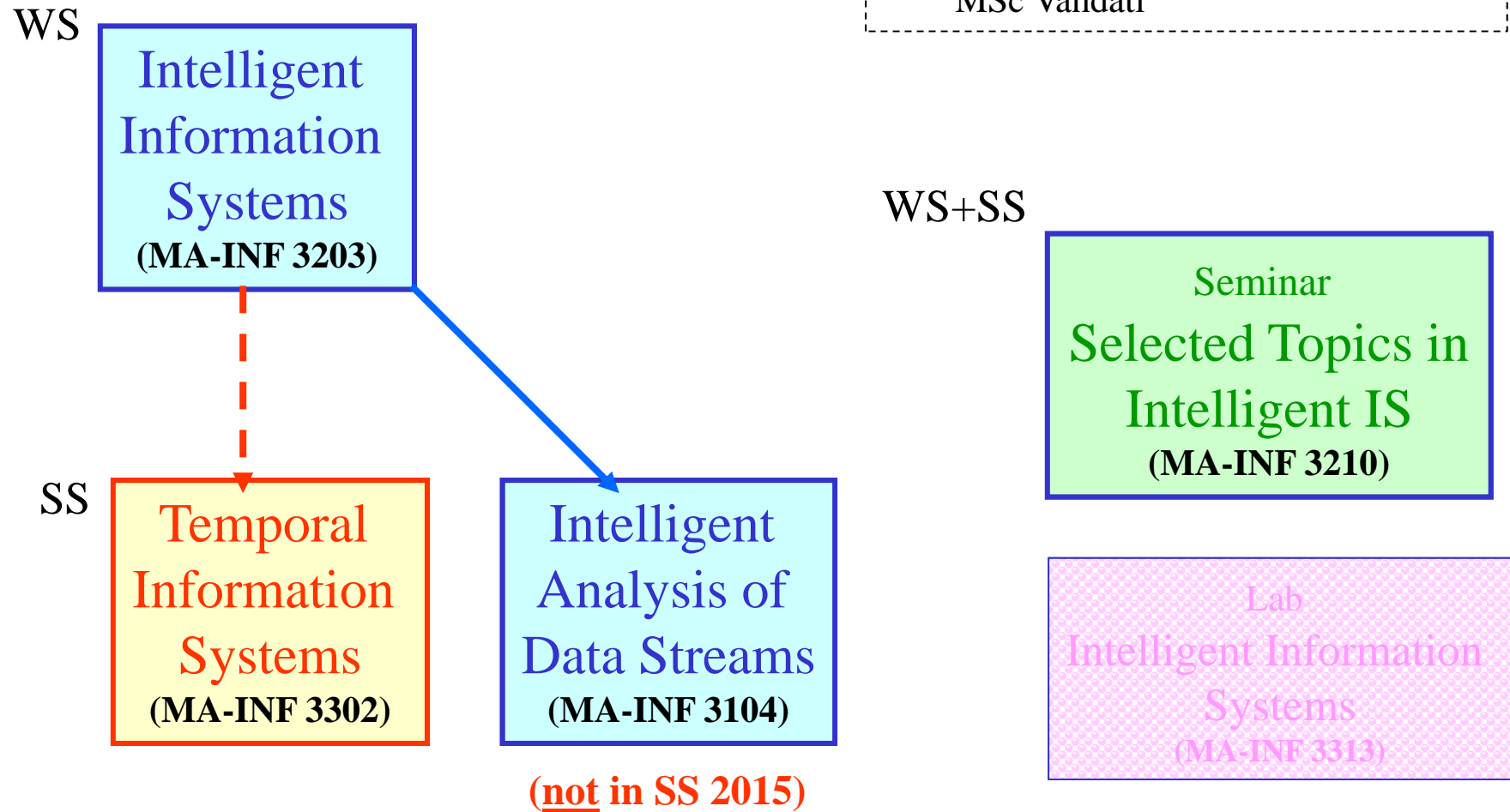
SS 2015

Organisation



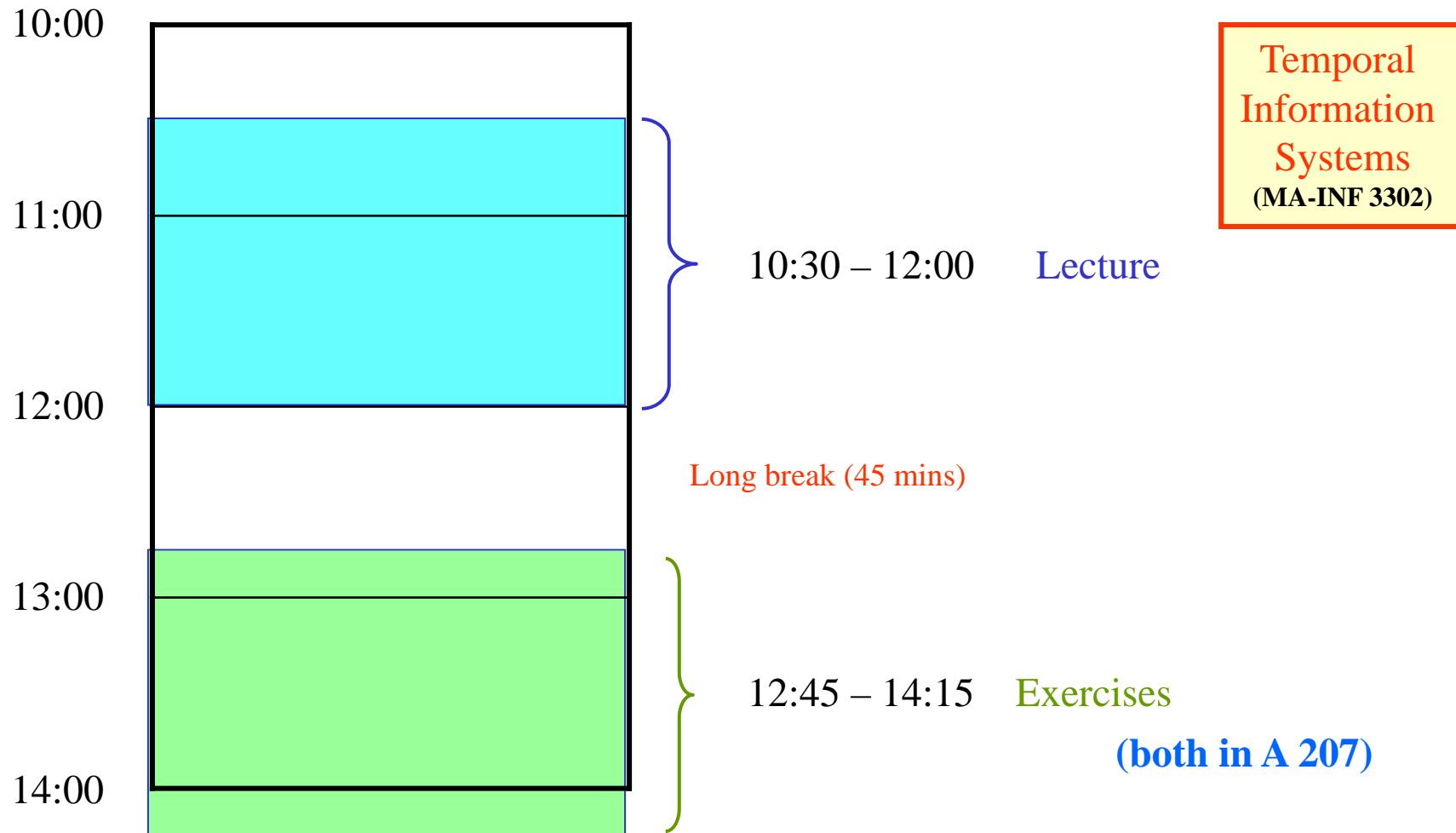
Modules Offered by the IDB Group

IDB (Intelligent Databases) Group:
Prof. Dr. Manthey, PD Dr. Behrend,
MSc Vahdati



TIS Weekly Schedule

(nearly) every **Wednesday** during this semester



Calendar of this Semester

	8.4.		L
	15.4.		L+E
	22.4.		L+E
	29.4.		L+E
	6.5.		L+E
	13.5.		L+E
	20.5.	Dies academicus	
	27.5.	Whitsun week	
Relevant for exam →	3.6.		L+E
	10.6.		
	17.6.		L+E
	24.6.		L+E
	1.7.		L+E
	8.7.		L+E
	15.7.		L (+E)
			<hr/>
			13 L+12 E

Exam (written): End of July

Temporal Information Systems - SS 11 - Mozilla Firefox

http://www.iai.uni-bonn.de/III/lehre/vorlesungen/TemporalIS/SS15/

Master Programme in Computer Science - Diplomstudiengang Informatik

Lecture "Temporal Information Systems" (SS 2014)

News

- **Welcome** to the homepage of the new module "Temporal Information Systems" (TIS), created on March 24, 2011! TIS is a **new module** within the curriculum of the area "Information and Communication Management" of our Master Programme, therefore no module description is in existence yet. A **brief summary of the topics** to be covered during the semester has been prepared on the follow-up page entitled "About" (see side bar).
- Both, lecture and exercises will take place **each Monday afternoon consecutively**: The lecture will begin at 12:45 (s.t., i.e., sharp) and last 90 minutes till 14:15. After 15 minutes break, we will continue with the exercises from 14.30 till 16:00. Both, lecture and

Fertig

Slides for download

Exercises and Admission to Exams

Exercises will be offered as a regular opportunity to improve your understanding of the topics covered and to train for the final exam „online“ – during the semester.

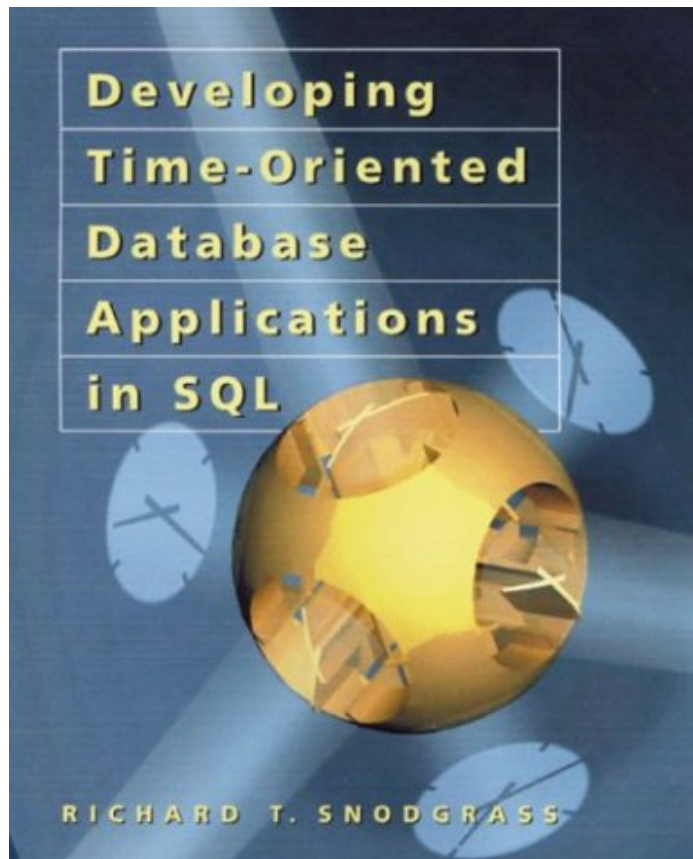
I don't control attendance – it is up to you if you make use of this offer or not.



This semester, there is **no** further achievement required for being **admitted** to the written exam!

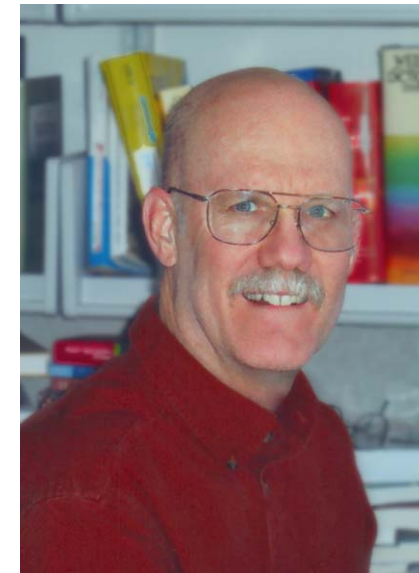
A **midterm exam** is **not** planned this time, due to lack of time (and other reasons).

Main Source for Additional Reading: „The Snodgrass Book“



Probably the most active and prominent researcher in temporal databases:

Richard T. Snodgrass
(Univ. of Arizona)



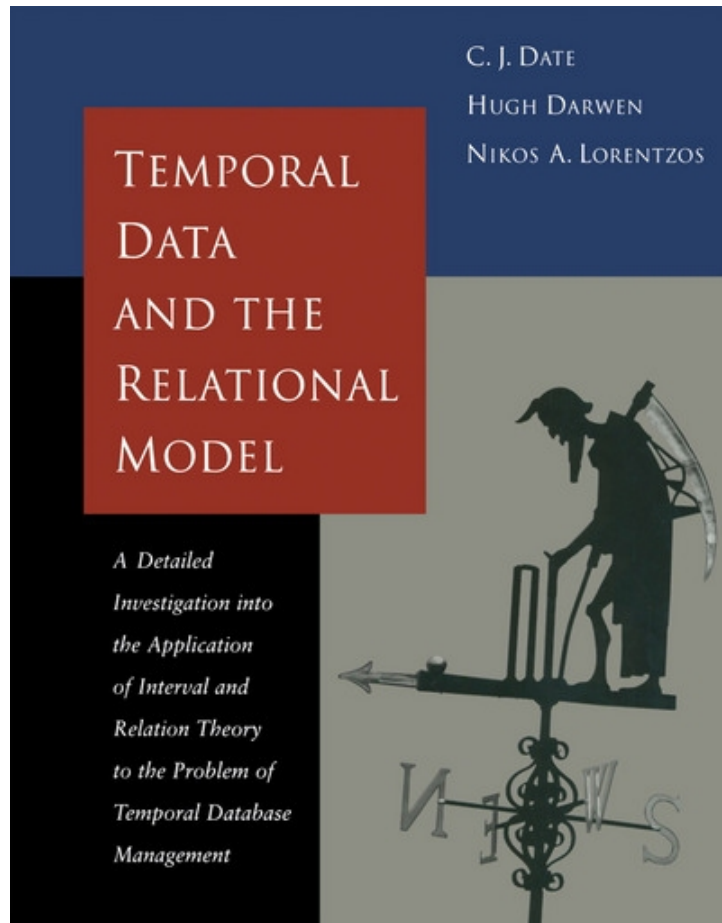
Freely downloadable as PDF file from:

<http://www.cs.arizona.edu/people/rts/tddbbook.pdf>

originally published by:

Morgan Kaufman, 1999,
522 pp.,
ISBN 978-1558604360

A Different Approach – Written by a Relational DB „Guru“

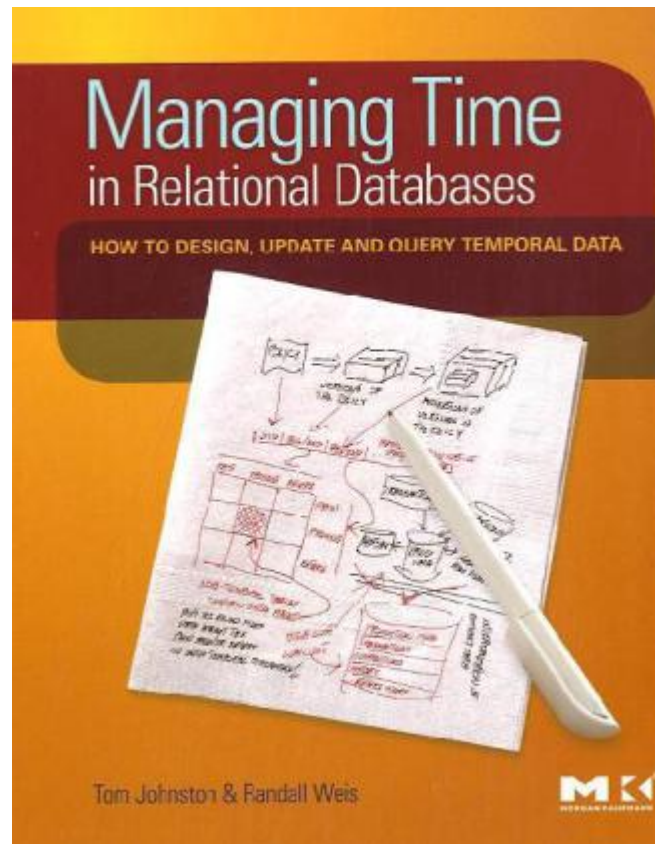


Chris J. Date, Hugh Darwen,
Nikos A. Lorentzos:

“Temporal Data and the Relational Model”

Morgan Kaufman Publ., 2002, 422 pp.
ISBN 978-1558608559
(~ 60 €)

The Most Recent Textbook on TIS: More on the Line of Snodgrass



Tom Johnston, Randall Weis:

“Managing Time in Relational Databases:
How to Design, Update and Query Temporal Data”

Morgan Kaufman Publ., 2010, 478 pp.
ISBN 0-12-375041-5
(~ 50 €)

"The authors present an original and comprehensive conceptual approach called Asserted Versioning, which includes support for bi-temporality and is a significant advance in the theory and practice of managing time-varying data."

Richard Snodgrass' comment on this book

Prerequisites

- Attending TIS doesn't require much of a specific background; thus, it is even **well-suited for beginners** in our master programme.
- The only **prerequisite** expected is that you have a good basic understanding of **relational databases** and some solid basic skills in **SQL** programming.
- In SQL, you should master
 - the most common DDL commands (CREATE TABLE, CREATE VIEW) including various forms of integrity constraints (e.g., KEY, CHECK, REFERENCES)
 - the main concepts of the **query** sublanguage of SQL (SELECT-FROM-WHERE incl. AS, JOIN/UNION/MINUS, AND/OR/NOT, EXISTS, GROUP BY/HAVING, aggregate functions),
 - the basic data **modification** commands incl. transactions and integrity checking (INSERT, DELETE, UPDATE; COMMIT).
- Some of the **more specific aspects of SQL** will be briefly revisited on the first exercise date on April 16th (constraints incl. assertions, expressing forall conditions in terms of NOT EXISTS etc.), but it is up to you to **make sure** that you are sufficiently **„fit“** in SQL by the end of April.

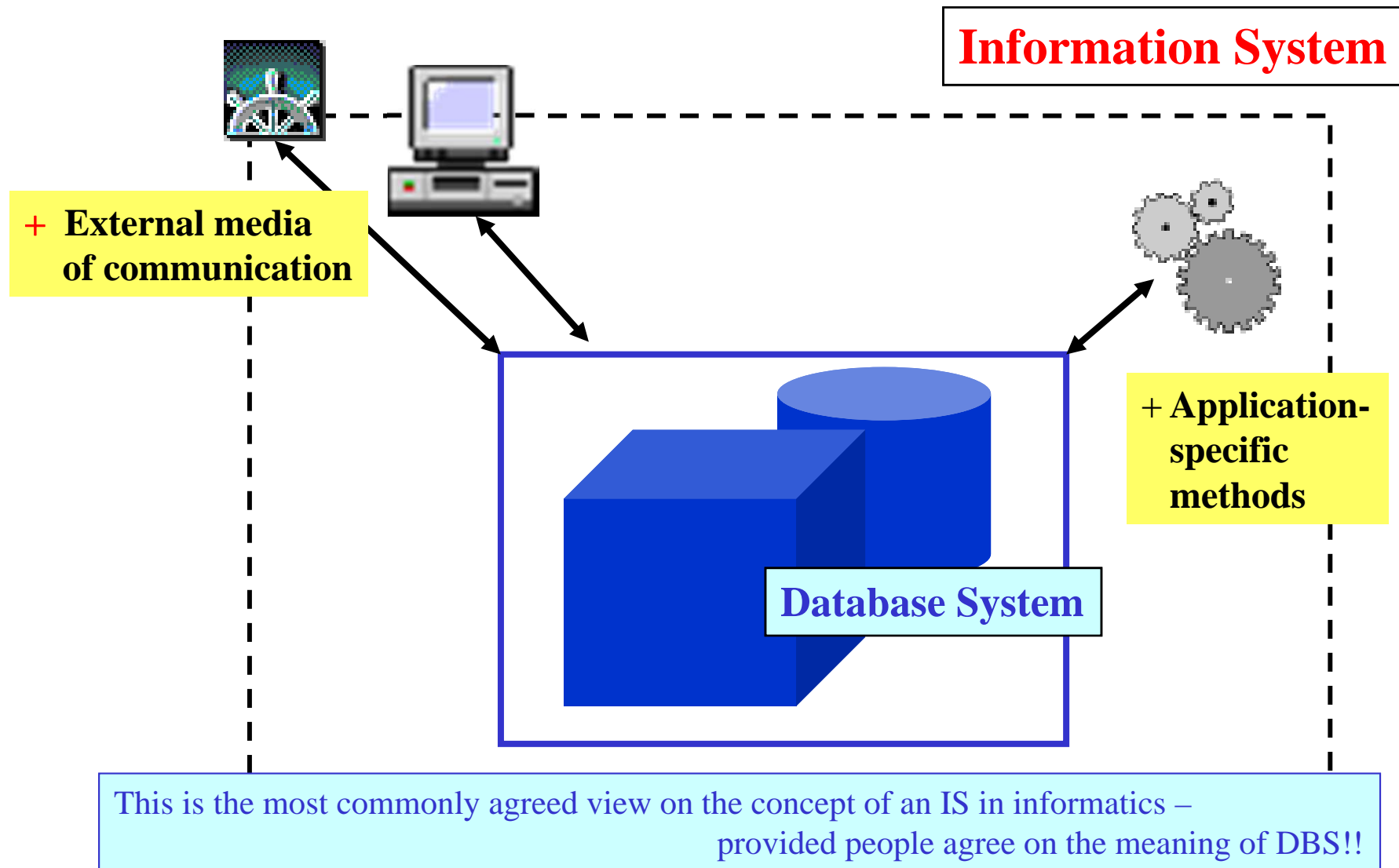
Temporal Information Systems

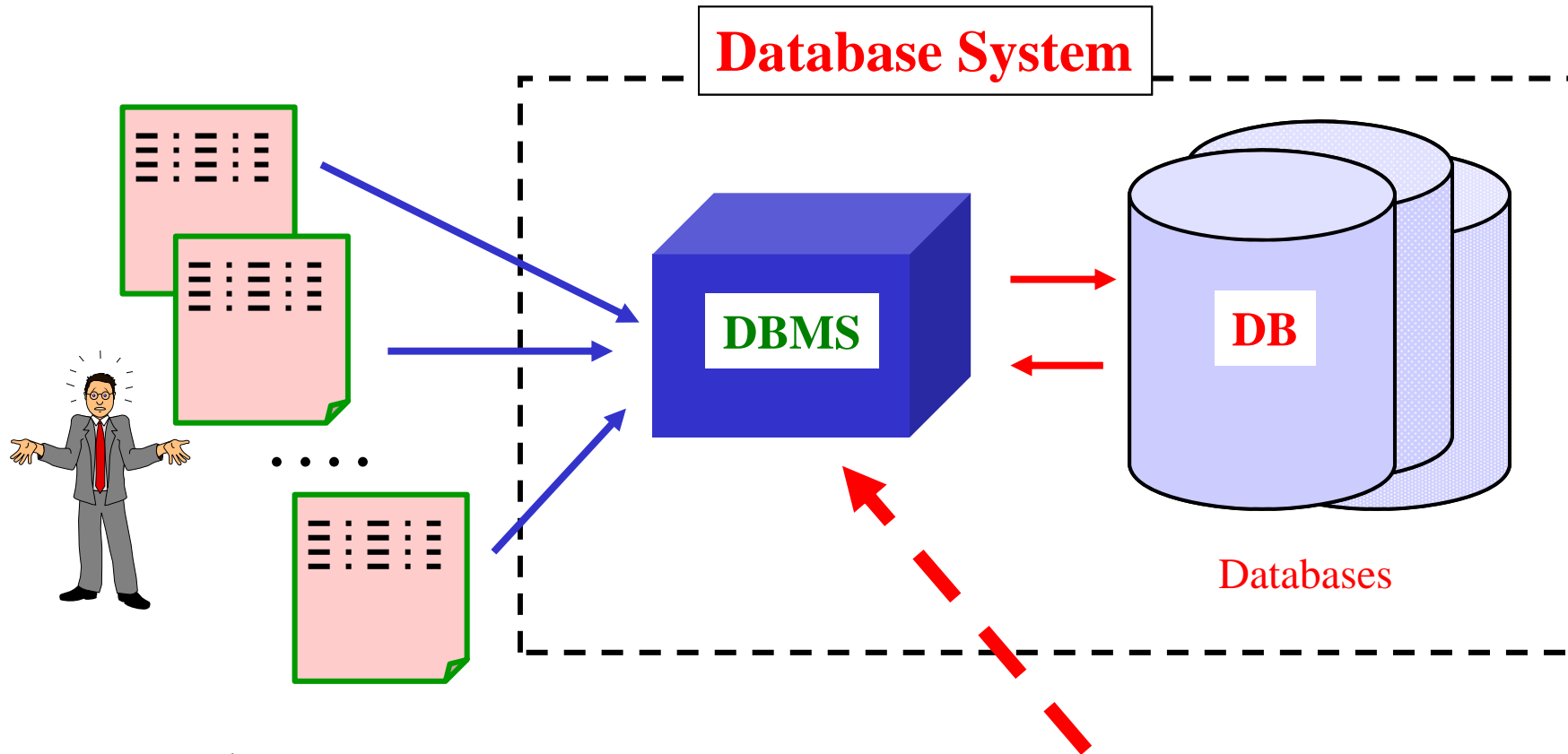
SS 2015



Motivation

Information system: The DB-centered View





Users and
application programmes

DBMS: **Data Base Management System**

(Many powerful **application-independent** services: schema mgt, query optimization, storage mgt, transaction mgt, etc.)

Temporal IS vs. Temporal DB

„The modifier *temporal* is used to indicate that the modified concept concerns some aspect of time.“

(A Consensus Glossary of Temporal Database Concepts, 1993)

- The lecture is called „Temporal Information Systems“, because
 - ... it thus refers to the more general concept (IS) and therefore can be used in a wider sense, and ...
 - ... this is in line with other namings in our MSc programme, e.g., Spatial IS.
- The term nearly exclusively used in scientific literature (in CS) is the related notion „Temporal Databases“ – even though literally speaking this is too limited!
- Certainly managing temporal data requires involving the DBMS as well, therefore the term „Temporal Database Systems“ would be more appropriate than just TDB.

In this lecture, we will focus on **temporal database** research, which actually means research on **temporal database systems**!

Temporal Database: Attempts of Definition (1)

„Databases which store all the past history as is best known at every state of the database are called **temporal databases**.“

(Sripada, 1988)

„A database that maintains past, present, and future data is called a **temporal database**.“

(Tansel, 1993)

„A **temporal database** is a repository of temporal information.“

(Chomicki, 1995)

„A **temporal database** is one that supports some aspect of time.“

(Snodgrass, 1995)

„A **temporal database** can be thought of, very loosely, as a database that contains historical data instead of or in addition to current data.“

(Date/Darwen/Lorentzos, 2002)

Temporal Database: Attempts of Definition (2)

Even now, ten years later, there is **no agreement** on an exact definition of the term among scientists – the spectrum ranges from very „liberal“ to rather „strict“ claims:

wide

„A **temporal database** is one that supports some aspect of time.“

(Snodgrass, 1995)



„Databases which store all the past history as is best known at every state of the database are called **temporal databases**.“

narrow

(Sripada, 1988)

In our lecture, we will **start from the most „liberal definition“** and first try to **understand** what the more restrictive approaches are really talking about.

Why Temporal Databases? (1)

„Most applications of database technology are temporal in nature.

Examples include financial applications such as portfolio management, accounting, and banking; record-keeping applications such as personnel, medical record, and inventory management; scheduling applications such as airline, train, and hotel reservations and project management; and scientific applications such as weather monitoring.

Applications such as these rely on *temporal databases*, which record time-referenced data.“

(Christian Jensen in „Introduction to Temporal Database Research“)

Why Temporal Databases? (2)

„Two decades of research into temporal databases have unequivocally shown that a **time-varying table**, containing certain kinds of DATE columns, is a **completely different animal** than its cousin, the table without such columns.“

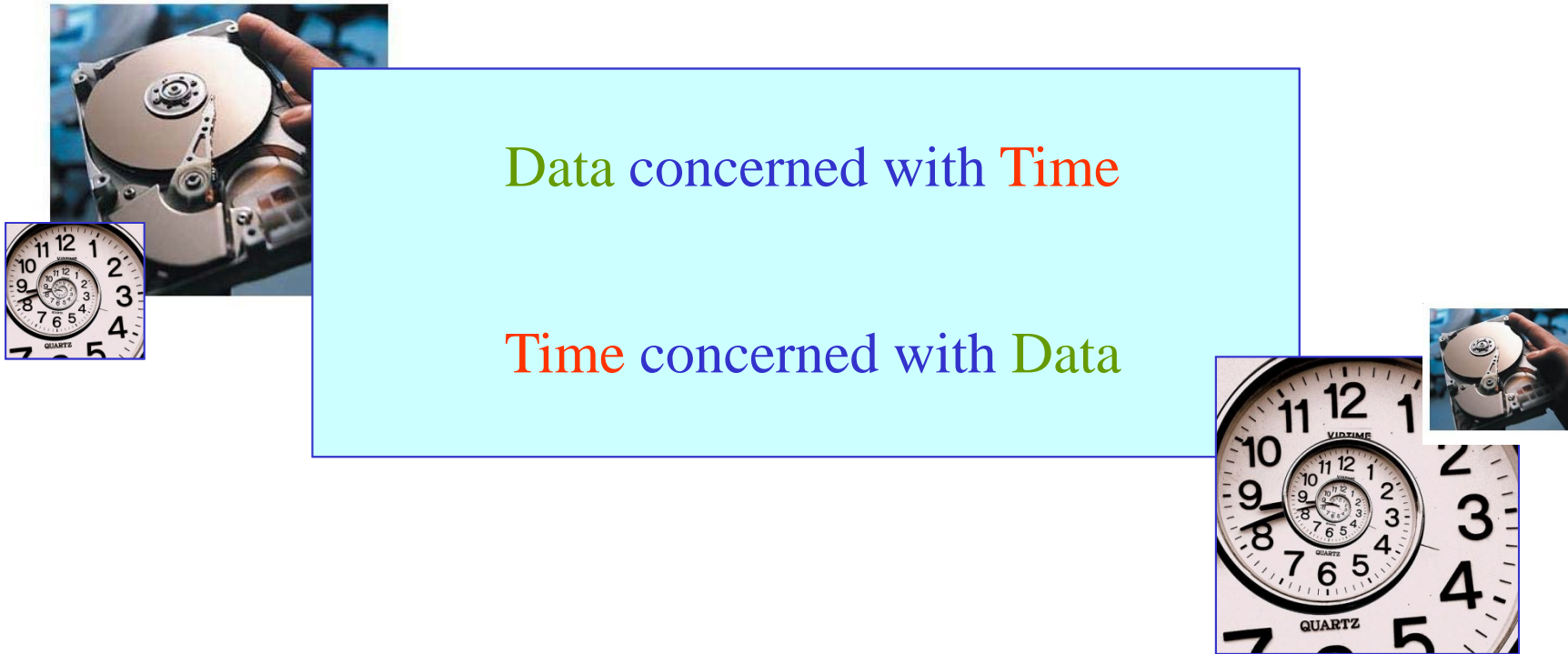
(R. Snodgrass in his book)



„Despite the near universality of time and the time-varying nature of the enterprise being modeled – a static and unmalleable configuration is rare and uninteresting – **SQL quite frankly does a lousy job** in capturing those aspects that are changing in time, or in providing constructs to effectively model, query, or modify such information.“

(R. Snodgrass again)

Time about Data vs. Data about Time: An Intuitive Introduction



1st Example Database: History of American Presidents

Presidents of the USA

Presidency	President	Birthday	From	To	Term
1	George Washington	22.2.1732	30.4.1789	4.3.1793	1
...					
22	Grover Cleveland	18.3.1837	4.3.1885	4.3.1889	25
23	Benjamin Harrison	20.8.1833	4.3.1889	4.3.1893	26
24	Grover Cleveland	18.3.1837	4.3.1893	4.3.1897	27
...					
43	George W. Bush	6.7.1946	20.1.2001	20.1.2005	54
43	George W. Bush	6.7.1946	20.1.2005	20.1.2009	55
44	Barack Obama	4.8.1961	20.1.2009	20.1.2013	56
44	Barack Obama	4.8.1961	20.1.2013		57

What do these time-valued attributes mean?

What do they refer to?

Do all three time columns play the same role?

1st Example Database: Meaning of Temporal Values (1)

Presidents of the USA

Presidency	President	Birthday	From	To	Term
1	George Washington	22.2.1732	30.4.1789	4.3.1793	1
...					
22	Grover Cleveland	18.3.1837	4.3.1885	4.3.1889	25
23	Benjamin Harrison	20.8.1833	4.3.1889	4.3.1893	26
24	Grover Cleveland	18.3.1837	4.3.1893	4.3.1897	27
...					
43	George W. Bush	6.7.1946	20.1.2001	20.1.2005	54
43	George W. Bush	6.7.1946	20.1.2005	20.1.2009	55
44	Barack Obama	4.8.1961	20.1.2009	20.1.2013	56
44	Barack Obama	4.8.1961	20.1.2013		57

1st president of the US was inaugurated!

Barack Obama was born!

Bush jr. leaves office!

Each of the dates appearing in a field of this table refers a particular (past) **event** that has **happened** in reality.

1st Example Database: Meaning of Temporal Values (2)

Presidents of the USA

Presidency	President	Birthday	From	To	Term
1	George Washington	22.2.1732	30.4.1789	4.3.1793	1
...					
22	Grover Cleveland	18.3.1837	4.3.1885	4.3.1889	25
23	Benjamin Harrison	20.8.1833	4.3.1889	4.3.1893	26
24	Grover Cleveland	18.3.1837	4.3.1893	4.3.1897	27
...					
43	George W. Bush	6.7.1946	20.1.2001	20.1.2005	54
43	George W. Bush	6.7.1946	20.1.2005	20.1.2009	55
44	Barack Obama	4.8.1961	20.1.2009	20.1.2013	56
44	Barack Obama	4.8.1961	20.1.2013		57

Cleveland serves as 22nd president of the US!

Entries in columns *From* and *To* together in addition represent time **periods** during which a particular fact was true in reality!

(How does the *Birthday* column fit with this view?)

1st Example DB: How to Represent Now and Future?

Presidency	President	Birthday	From	To	Term
1	George Washington	22.2.1732	30.4.1789	4.3.1793	1
...					
22	Grover Cleveland	18.3.1837	4.3.1885	4.3.1889	25
23	Benjamin Harrison	20.8.1833	4.3.1889	4.3.1893	26
24	Grover Cleveland	18.3.1837	4.3.1893	4.3.1897	27
...					
43	George W. Bush	6.7.1946	20.1.2001	20.1.2005	54
43	George W. Bush	6.7.1946	20.1.2005	20.1.2009	55
44	Barack Obama	4.8.1961	20.1.2009	20.1.2013	56
44	Barack Obama	4.8.1961	20.1.2013	?	57

What does the **empty** To-field for Obama mean?

Should there better be a **concrete entry** instead?

If so, which one: The **current date** (changing every midnight),
or the scheduled **end of term** (20.1.2017)?

1st Example DB: An Alternative Design – Any Better?

Presidency	President	Birthday	From Since		Term
1	George Washington	22.2.1732	30.4.1789		1
...					
22	Grover Cleveland	18.3.1837	4.3.1885		25
23	Benjamin Harrison	20.8.1833	4.3.1889		26
24	Grover Cleveland	18.3.1837	4.3.1893		27
...					
43	George W. Bush	6.7.1946	20.1.2001		54
43	George W. Bush	6.7.1946	20.1.2005		55
44	Barack Obama	4.8.1961	20.1.2009		56
44	Barack Obama	4.8.1961	20.1.2013		57

If we mean „until now“ in the Obama case, **dropping** „To“ altogether and **renaming** „From“ into „Since“ might be the better choice – if we can rely on „gap-free“ presidencies (which we can, up till now)!

1st Example DB: Queries Referring to Past, Present and Future

Presidency	President	Birthday	From	To	Term
1	George Washington	22.2.1732	30.4.1789	4.3.1793	1
...					
22	Grover Cleveland	18.3.1837	4.3.1885	4.3.1889	25
23	Benjamin Harrison	20.8.1833	4.3.1889	4.3.1893	26
24	Grover Cleveland	18.3.1837	4.3.1893	4.3.1897	27
...					
43	George W. Bush	6.7.1946	20.1.2001	20.1.2005	54
43	George W. Bush	6.7.1946	20.1.2005	20.1.2009	55
44	Barack Obama	4.8.1961	20.1.2009	20.1.2013	56
44	Barack Obama	4.8.1961	20.1.2013		57

How to express these queries in **SQL**?
What answers to expect?

Who is the current president?
Who was president on July 4, 1890?
Who will be president on July 4, 2015?

1st Example DB: Same Queries – Different Design

Presidency	President	Birthday	Since	Term
1	George Washington	22.2.1732	30.4.1789	1
...				
22	Grover Cleveland	18.3.1837	4.3.1885	25
23	Benjamin Harrison	20.8.1833	4.3.1889	26
24	Grover Cleveland	18.3.1837	4.3.1893	27
...				
43	George W. Bush	6.7.1946	20.1.2001	54
43	George W. Bush	6.7.1946	20.1.2005	55
44	Barack Obama	4.8.1961	20.1.2009	56
44	Barack Obama	4.8.1961	20.1.2013	57

Is a different color still appropriate?

How to express these queries for the other design?

Who is the current president?
 Who was president on July 4, 1890?
 Who will be president on July 4, 2015?

1st Example DB: Two Tables – Lifetime of Presidents Included

Presidency	President	From	To	Term
1	George Washington	30.4.1789	4.3.1793	1
...				
16	Abraham Lincoln	4.3.1861	15.4.1865	20
...				
35	John F. Kennedy	20.1.1961	22.11.1963	44
...				
37	Richard Nixon	20.1.1973	9.8.1974	47
...				
44	Barack Obama	20.1.2013		

Presidency

Lifetime

President	Born	Died
George Washington	22.2.1732	14.12.1799
...		
Abraham Lincoln	12.2.1809	15.4.1865
...		
John F. Kennedy	29.5.1917	22.11.1963
...		
Richard Nixon	9.1.1913	22.4.1994
...		
Barack Obama	4.8.1961	

Which presidents died in office?
 Which presidents resigned before the regular end of their term (alive)?
 Who was the oldest president at the begin of his presidency?
 How many former presidents are still alive today?

1st Example Database: More Non-Obvious Queries

Presidency	President	Birthday	From	To	Term
1	George Washington	22.2.1732	30.4.1789	4.3.1793	1
...					
22	Grover Cleveland	18.3.1837	4.3.1885	4.3.1889	25
23	Benjamin Harrison	20.8.1833	4.3.1889	4.3.1893	26
24	Grover Cleveland	18.3.1837	4.3.1893	4.3.1897	27
...					
43	George W. Bush	6.7.1946	20.1.2001	20.1.2005	54
43	George W. Bush	6.7.1946	20.1.2005	20.1.2009	55
...					

Which period is covered by the 43rd presidency?
 How long did the 43rd presidency last?
 For how many years did Cleveland serve as president?
 Which presidents served during different presidencies?

1st Example Database: A Purely Historical Table

Presidents of the USA

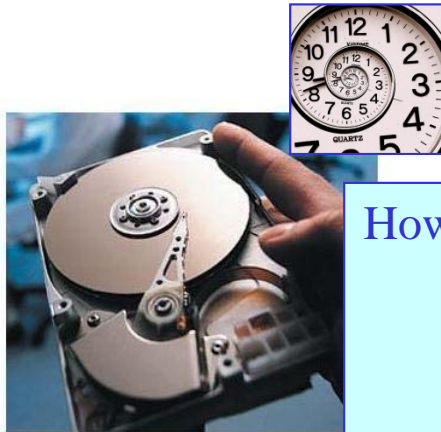
Presidency	President	Birthday	From	To	Term
1	George Washington	22.2.1732	30.4.1789	4.3.1793	1
...					
22	Grover Cleveland	18.3.1837	4.3.1885	4.3.1889	25
23	Benjamin Harrison	20.8.1833	4.3.1889	4.3.1893	26
24	Grover Cleveland	18.3.1837	4.3.1893	4.3.1897	27
...					
43	George W. Bush	6.7.1946	20.1.2001	20.1.2005	54
43	George W. Bush	6.7.1946	20.1.2005	20.1.2009	55
44	Barack Obama	4.8.1961	20.1.2009	20.1.2013	56
44	Barack Obama	4.8.1961	20.1.2013		57

fixed forever

The facts recorded in this table are (presumably) all settled, known to be **historically correct**, never to be reconsidered.

The only modifications to be expected are **appending** new rows and filling in missing To-dates.

The First Example: Managing Data about Time (History)



How to properly manage

Data concerned with Time

and

Time concerned with Data

?

2nd Example Database : Classes and Exams in a University

Exams

Student	Class	Signed_up	Dropped	Grade	Exam Date
John	1203	11.11.2010		1,3	13.2.2011
Jack	1203	19.11.2010	2.1.2011		
Tim	1203	21.11.2010		3,0	18.3.2011
Pete	1203	27.11.2010	3.2.2011	5,0	18.3.2011
John	2201	11.11.2010		1,7	19.2.2011
Jack	2201		2.1.2011		
Tim	3203	2.12.2010		3,7	1.4.2010

Assume, this is how the „Exams“ table looked like on Sunday, April 1, 2011.

At first glance, this looks like another „settled history“ table, but ...

2nd Example DB: Potential Mistakes in Exam DB

Student	Class	Signed_up	Dropped	Grade	Exam Date
John	1203	11.11.2010		1,3	13.2.2011
Jack	1203	19.11.2010	2.1.2011		
Tim	1203	21.11.2010		3,0	18.3.2011
Pete	1203	27.11.2010	3.2.2011	5,0	18.3.2011
John	2201	11.11.2010		1,7	19.2.2011
Jack	2201		2.1.2011		
Tim	3203	2.12.2010		3,7	1.4.2010

Failed or
dropped?

Sign-up date missing?

Mistyped?

There are several „suspicious“ entries in this table!

2nd Example DB: Typo Corrected – History Cleaned Up

On April 4th, 2011, the typo of Tim's exam date is discovered and corrected:

Student	Class	Signed_up	Dropped	Grade	Exam Date
John	1203	11.11.2010		1,3	13.2.2011
Jack	1203	19.11.2010	2.1.2011		
Tim	1203	21.11.2010		3,0	18.3.2011
Pete	1203	27.11.2010	3.2.2011	5,0	18.3.2011
John	2201	11.11.2010		1,7	19.2.2011
Jack	2201		2.1.2011		
Tim	3203	2.12.2010		3,7	1.4.2011

So this is the „version of history“ as known on April 4th, 2011.

2nd Example DB: Failed Exam Discovered as a Mistake

On April 6th, 2011, it turns out that Pete didn't appear for the exams in class 1203 because he had indeed dropped the course – but the examiner didn't know:

Student	Class	Signed_up	Dropped	Grade	Exam Date
John	1203	11.11.2010		1,3	13.2.2011
Jack	1203	19.11.2010	2.1.2011		
Tim	1203	21.11.2010		3,0	18.3.2011
Pete	1203	27.11.2010	3.2.2011	5,0	18.3.2011
John	2201	11.11.2010		1,7	19.2.2011
Jack	2201		2.1.2011		
Tim	3203	2.12.2010		3,7	1.4.2011

This is the next „version of history“ as known on April 6th, 2011.

2nd Example DB: Missing Sign-up Date Discovered and Added

On **April 7th, 2011**, Jack was asked when he actually signed-up for 2201. It turned out that Jack **never** signed up for this class, but erroneously entered a drop date for that course (and couldn't get rid of it anymore):

Student	Class	Signed_up	Dropped	Grade	Exam Date
John	1203	11.11.2010		1,3	13.2.2011
Jack	1203	19.11.2010	2.1.2011		
Tim	1203	21.11.2010		3,0	18.3.2011
Pete	1203	27.11.2010	3.2.2011		
John	2201	11.11.2010		1,7	19.2.2011
Jack	2201		2.1.2011		
Tim	3203	2.12.2010		3,7	1.4.2011

After deleting this error, we have a version as known on April 7th, 2011.

2nd Example DB: Missing Information About Improved Grade

Finally, on April 8th, 2011, Tim complains about the grade he found in the DB for 1203. It had been improved to 2,7 by the examiner after reconsidering the evaluation of one answer:

Student	Class	Signed_up	Dropped	Grade	Exam Date
John	1203	11.11.2010		1,3	13.2.2011
Jack	1203	19.11.2010	2.1.2011		
Tim	1203	21.11.2010		2,7	18.3.2011
Pete	1203	27.11.2010	3.2.2011		
John	2201	11.11.2010		1,7	19.2.2011
Tim	3203	2.12.2010		3,7	1.4.2011

Since April 8th, 2011, this is the version stored in the DB – it seems to be correctly reflecting what happened in reality.

2nd Example DB: A Solution Keeping Track of All Changes

Student	Class	Signed_up	Dropped	Grade	Exam Date	From	To
John	1203	11.11.2010				11.11.2010	14.2.2011
John	1203	11.11.2010		1,3	13.2.2011	14.2.2011	
Jack	1203	19.11.2010				19.11.2010	2.1.2011
Jack	1203	19.11.2010	2.1.2011			2.1.2011	
Tim	1203	21.11.2010				21.11.2010	20.3.2011
Tim	1203	21.11.2010		3,0	18.3.2011	20.3.2011	8.4.2011
Tim	1203	21.11.2010		2,7	18.3.2011	8.4.2011	
Pete	1203	27.11.2010				27.11.2010	3.2.2011
Pete	1203	27.11.2010	3.2.2011			3.2.2011	21.3.2011
Pete	1203	27.11.2010	3.2.2011	5,0	18.3.2011	21.3.2011	6.4.2011
Pete	1203	27.11.2010	3.2.2011			6.4.2011	
John	2201	11.11.2010				11.11.2010	21.2.2011
John	2201	11.11.2010		1,7	19.2.2011	21.2.2011	
Jack	2201		2.1.2011			2.11.2010	7.4.2011
Tim	3203	2.12.2010		3,7	1.4.2010	1.4.2011	4.4.2011
Tim	3203	2.12.2010		2,7	1.4.2011	4.4.2011	

Keeping Track of All Changes?

- There is an increasing number of **real-life applications** where keeping track of all changes ever performed to the „real“ database is useful, sometimes even required by law (**auditing**, liability, etc.).
- Transaction processing, in particular recovery, already requires to track changes made during a transaction at least **temporarily: transaction log**.
- Keeping track of changes „**forever**“ is a different affair, but it is a form of „logging“, too.
- However, it is not clear whether this kind of „history log“ should be part of the table to be logged itself:
 - Should the „history log“ be **changeable by users**, or by the system only?
 - Should users be able to „**query the log**“?
 - Which tables should be logged at all?
- Recently, big vendors of relational DBMS started to include system-controlled history keeping into their systems (e.g., „**Total Recall**“ of Oracle 11g).

Second Example: Managing Time About (the History of Changes of) Data

How to properly manage

Data concerned with Time

and

Time concerned with Data

?



Some of the Problems to be Addressed in this Lecture

What is the correct **meaning** of temporal data wrt real world?

How to **model** temporal aspects of change appropriately?

How to **express** temporal **queries** adequately?

How to design good **query languages** supporting time?

How to **evaluate** temporal **queries** efficiently?

How to **update** temporal data correctly and consistently?

How to **reason** about temporal data?

Outline of the Lecture

Last year, the lecture TIS covered the following chapters – it is likely to be structured like this again (though some „evolution“ is planned):

- 0 Organisation and Motivation
- 1 Some General Reflections About Time
- 2 Representing Time in SQL
- 3 Recording „Time About Data“ – Keeping a History of Change
- 4 Managing „Data About Time“ – In Combination with History Keeping
- 5 Temporal Perspectives for SQL

Next week !



Some General Reflections About Time

Chapter 1

