

DATABASES: FROM A TO Y

8/3/2016

Tamás Budavári / The Johns Hopkins University

Databases

Tamás Budavári

- Store your bytes
 - ▣ And return them
- Efficient filtering
 - ▣ Thousands of man-years
 - ▣ Optimally pick from many strategies



Databases

Tamás Budavári

- SQL basics
 - ▣ filtering, aggregation, joins
- SQL programming
 - ▣ variables, functions, procedures
 - ▣ data management, transactions
- Tools



SQL: Structured Query Language

Tamás Budavári

- Standard declarative language
- Filter the data
- Powerful analysis tool
- Possible to extend





Sloan Digital Sky Survey / SkyServer

SDSS



[Home](#) [Tools](#) [Schema](#) [Projects](#) [Astronomy](#) [SDSS](#) [Contact Us](#) [Download](#) [Site Search](#) [Help](#)

Due to system maintenance this site will be unavailable Thursday March 17th from 7:00AM central until 7:30AM central. We apologize for the inconvenience.

Welcome to the DR7 site!!!

This website presents data from the Sloan Digital Sky Survey, a project to make a map of a large part of the universe. We would like to show you the beauty of the universe, and share with you our excitement as we build the largest map in the history of the world.

News

The site hosts data from Data Release 7 (DR7). What's new in DR7, what's new on this site, and known problems. [More...](#)

For Astronomers

A separate branch of this website for professional astronomers (English). [More...](#)

SDSS is supported by



Powered by

Microsoft

SkyServer Tools

[Famous places](#)

[Get images](#)

[Visual Tools](#)

[Explore](#)

[Search](#)

[Object Cross-ID](#)

[CasJobs](#)

Science Projects

[Basic](#)

[Advanced](#)

[Challenges](#)

[For Kids](#)

[Games and Contests](#)

[Teachers](#)

[Links to other projects](#)

Info Links

[About Astronomy](#)

[About the SDSS](#)

[About the SkyServer](#)

[SDSS Data Release 7](#)

[SDSS Project Website](#)

[Open SkyQuery](#)

[Images of RC3 Galaxies](#)

Help

[Getting Started](#)

[FAQ](#)

[How To](#)

[Glossary](#)

[Schema Browser](#)

[Sample SQL Queries](#)

[Details of SDSS Data](#)



MyDB Local Only

Views

Tables

Functions

Procedures

Sort by... All selected...

Rows	kB	Name
993	40	dist
949	200	dr3tile
92,082	16,640	DW5Z_R17_primary
92,082	13,376	DW5Z_R25
14,400	904	galaxfield1
3	16	galaxfield2
1	16	natiguera
1	16	MyTable
1	16	MyTable_0
1	16	MyTable_1
23	16	MyTable_2
1,000	40	radec
1	16	roomba
20	16	roomba2
1	16	stat
11	24	test1

Tamas Budavari 's MyDB

20,992 kB of 100,000 kB used

From this page you can get various information about the contents of both your MyDB and shared tables within your groups. Click the left table links to get information about a specific table, such as rows, columns or size. From the table pages you can also perform various table-specific tasks, such as:

- Download a table
- Manage your group tables
- Rename a table
- Drop a table

Sizes are approximations only.

Row counts are approximations only. For exact value run a count.

There's always some overhead, even empty MyDB's take up space.

Group tables do not count towards your MyDB size limit.

Contact

SNome: v3_5_16 8.3Revision: 1.64 L Last modified: Tuesday, January 27, 2009 at 3:19:32 PM

Sandbox

The screenshot shows a SQL query execution interface. At the top, there are buttons for 'Refresh', 'Syntax', 'Execute', 'Plan', and a dropdown menu '(select sample query)'. Below these buttons, the status 'Ready' is displayed. On the left side, there is a tree view showing the database structure: 'Tables' (expanded) with sub-items 'dbo.Data', 'dbo.Instruments', 'dbo.Runs', and 'dbo.Users'; 'Views'; and 'Stored procedures'. The main area on the right displays the SQL query:

```
1 select X, Y
2 from Data
3 where X>0.9
```

Below the query, the results are shown in a table with two columns: 'X' (float) and 'Y' (float). The table contains five rows of data.

X	Y
0.981793399711043	2.97527245762281
0.994198126637278	2.48971566147368
0.918508748491438	1.57256024952415
0.974056254131808	2.53658738117638

SQL by Examples

Tamás Budavári

- Interactive session with exercises
 - ▣ See handouts...

Database Systems

Tamás Budavári

- SQLite
 - ▣ To start with the smallest
 - ▣ Command line executable
 - Single library for coding
 - ▣ Also a pure C# implementation
- Hands-on...

Database Systems

Tamás Budavári

- ❑ MySQL – open source
- ❑ PostgreSQL and Greenplum
- ❑ Microsoft SQL Server (Express Edition)
 - ▣ Same as the one under our LabDB
- ❑ IBM DB2 (Express-C)
- ❑ Oracle Database (XE-Express Edition)

Database Systems

Tamás Budavári

- MonetDB
 - ▣ Column store

- SciDB
 - ▣ Array database under development

Database Systems

Tamás Budavári

- Many others

Server – Client

Tamás Budavári

- Multiple clients
- Different client apps
 - ▣ Graphical UI
 - ▣ Command line
 - ▣ Your custom analysis

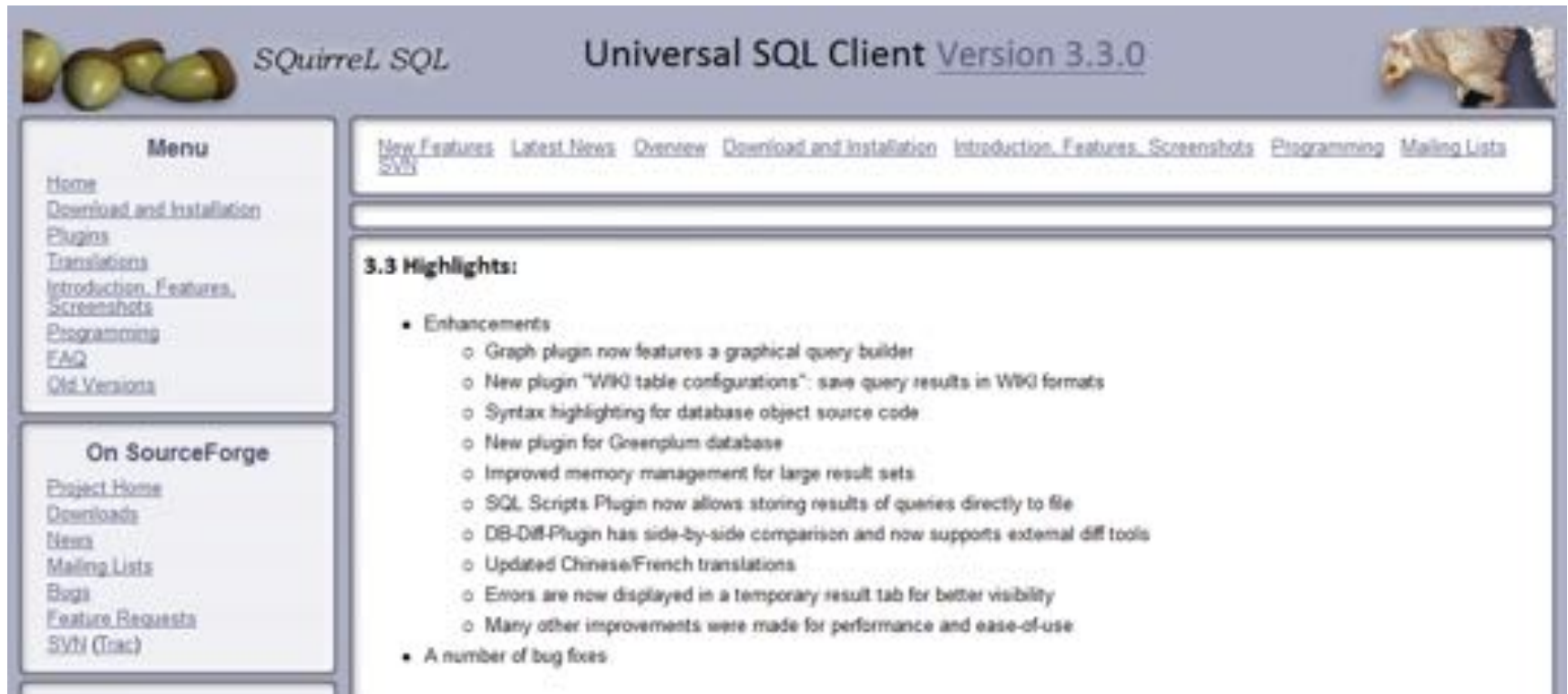
Programmatic Interface

Tamás Budavári

- Send SQL commands
- Read out the results
- Standard ODBC
 - ▣ Open Database Connectivity
 - ▣ C interface
- JDBC for Java

A Universal GUI

Tamás Budavári



The screenshot shows the homepage of the Squirrel SQL Universal SQL Client. The header features the Squirrel SQL logo (a squirrel) on the left and the text "Universal SQL Client Version 3.3.0" on the right, accompanied by a small image of a squirrel. Below the header, there is a navigation bar with links: [New Features](#), [Latest News](#), [Overview](#), [Download and Installation](#), [Introduction, Features, Screenshots](#), [Programming](#), and [Mailing Lists](#). The main content area is divided into two columns. The left column contains a "Menu" section with links: [Home](#), [Download and Installation](#), [Plugins](#), [Translations](#), [Introduction, Features, Screenshots](#), [Programming](#), [FAQ](#), and [Old Versions](#). Below the menu is an "On SourceForge" section with links: [Project Home](#), [Downloads](#), [News](#), [Mailing Lists](#), [Bugs](#), [Feature Requests](#), and [SVN \(Trac\)](#). The right column features a "3.3 Highlights:" section with a bulleted list of enhancements and bug fixes.

Squirrel SQL Universal SQL Client Version 3.3.0

[New Features](#) [Latest News](#) [Overview](#) [Download and Installation](#) [Introduction, Features, Screenshots](#) [Programming](#) [Mailing Lists](#)

Menu

- [Home](#)
- [Download and Installation](#)
- [Plugins](#)
- [Translations](#)
- [Introduction, Features, Screenshots](#)
- [Programming](#)
- [FAQ](#)
- [Old Versions](#)

On SourceForge

- [Project Home](#)
- [Downloads](#)
- [News](#)
- [Mailing Lists](#)
- [Bugs](#)
- [Feature Requests](#)
- [SVN \(Trac\)](#)

3.3 Highlights:

- Enhancements
 - Graph plugin now features a graphical query builder
 - New plugin "WFO table configurations": save query results in WFO formats
 - Syntax highlighting for database object source code
 - New plugin for Greenplum database
 - Improved memory management for large result sets
 - SQL Scripts Plugin now allows storing results of queries directly to file
 - DB-Diff-Plugin has side-by-side comparison and now supports external diff tools
 - Updated Chinese/French translations
 - Errors are now displayed in a temporary result tab for better visibility
 - Many other improvements were made for performance and ease-of-use
- A number of bug fixes

Research Them!

Tamás Budavári

- Which one?
- Why?
- Things to consider
 - ▣ How much data? Scale to my problem?
 - ▣ Extensibility for scientific analysis?
 - ▣ Hardware requirements? What OS?



Concurrency

Parallel actions

The Elevator Problem

Tamás Budavári

- People on multiple levels
 - ▣ Press the button...

Mutual Exclusion

- Multiple processes or threads
 - ▣ Access shared resources in critical sections
 - E.g., call the elevator when it's time to go
- Locking
 - ▣ Elevators, etc...

Dining Philosophers

Tamás Budavári

- Five silent philosophers sit at the table
 - ▣ Alternate between eating and thinking
 - ▣ Need both forks left & right to eat
 - Must be picked up one by one!
 - ▣ Infinite food in front of them
- How can they all think & eat forever?



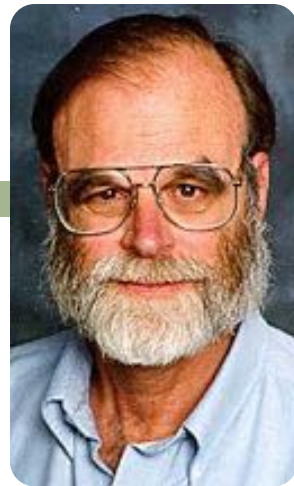
Database Transactions

Tamás Budavári

- Every time somebody swipes a card
- Commands in SQL
 - ▣ BEGIN TRANSACTION [name]
 - ▣ COMMIT TRANSACTION [...]
 - ▣ ROLLBACK TRANSACTION [...]
- Nested transactions

ACID Properties

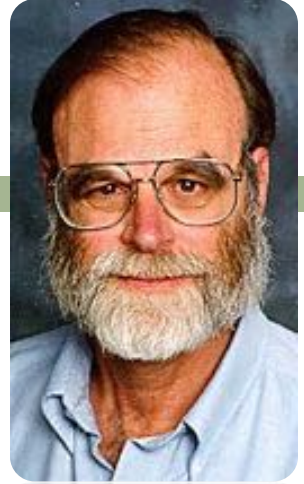
- Requirements
 - ▣ Atomicity
 - ▣ Consistency
 - ▣ Isolation
 - ▣ Durability



Defined by Jim Gray

Atomicity

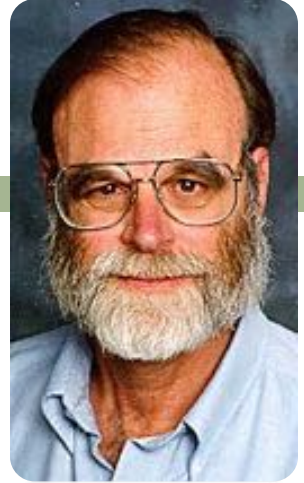
- All parts of a transaction succeed
- Or rollback to previous state



Defined by Jim Gray

Consistency

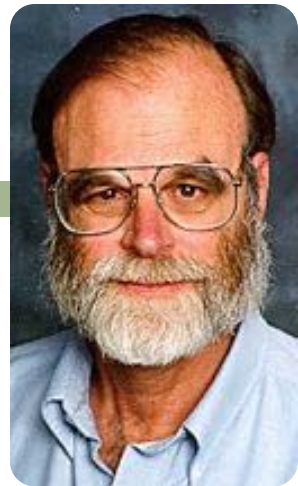
- Data always meets validation rules
 - ▣ Any type of constraints



Defined by Jim Gray

Isolation

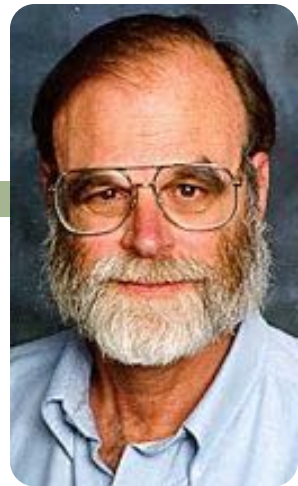
- No inference across transactions
 - ▣ Even if concurrent



Defined by Jim Gray

Durability

- Committed transaction will remain so
 - ▣ Even in the event of power failure, error, ...
 - Caching in harddrives, etc...



Defined by Jim Gray

Locks

- Resources

- ▣ Database, Tables, Pages, Ranges, Rows, ...

- Different modes

- ▣ Shared for reading, Update, Exclusive, etc...

Examples in LabDB

Tamás Budavári

- Transactions
 - ▣ BEGIN / COMMIT / ROLLBACK
- Exceptions
 - ▣ TRY ... CATCH

In Practice

```
1 BEGIN TRANSACTION T1
2 BEGIN TRY
3     INSERT Instruments VALUES ('New Instr')
4     DECLARE @i int = 1/0 -- Generate divide-by-zero error
5     COMMIT TRAN T1
6 END TRY
7 BEGIN CATCH
8     -- Deal with the error
9     SELECT @@ERROR as ErrorCode -- Print
10    ROLLBACK TRAN T1
11 END CATCH
```