INTRODUCTION TO DATABASES AND SQL

Tamás Budavári / The Johns Hopkins University

Databases

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

Databases

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

SQL: Structured Query Language

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

Sloan Digital Sky Survey / SkyServer

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...

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

SDSS is supported by

Powered by Microsoft

MyDB | Local Only

Views
Tables
Functions
Procedures

Sort by... | All selected...

	Rows	kB	Name
☐	993	40	dist
☐	949	200	dr3tile
☐	92,082	16,640	DWSZ_R17_primary
☐	92,082	13,376	DWSZ_R25
☐	14,400	904	galexfield1
☐	3	16	galexfield2
☐	1	16	halfspace
☐	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
- Mangage 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

$Name: v3_5_16 $,$Revision: 1.64 $, Last modified: Tuesday, January 27, 2009 at 3:19:32 PM

Transactions

ACID Properties

The Elevator Problem

- 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

- 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?

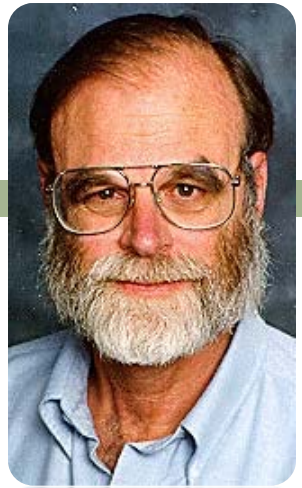
Image from Wikipedia

Database Transactions

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

ACID Properties

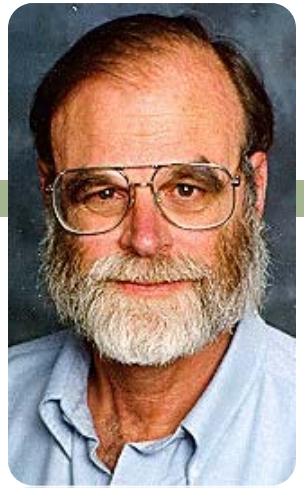
- Requirements
  - Atomicity
  - Consistency
  - Isolation
  - Durability



Defined by Jim Gray

Atomicity

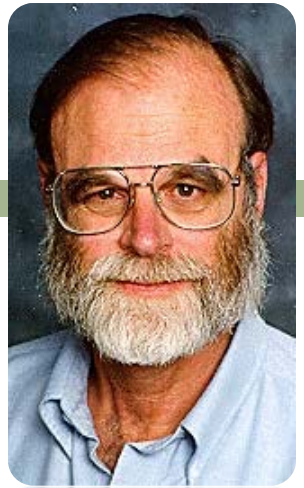
- All parts of a transaction succeed
- Or rollback to previous state



Defined by Jim Gray

Consistency

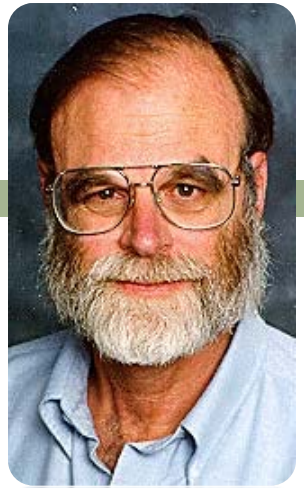
- Data always meets validation rules
  - Any type of constraints



Defined by Jim Gray

Isolation

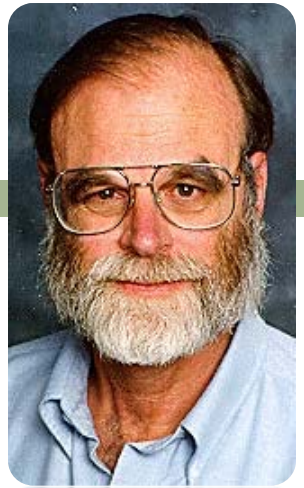
- No inference across transactions
  - Even if concurrent



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

SQL by Examples

Database Systems

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

Database Systems

- 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)

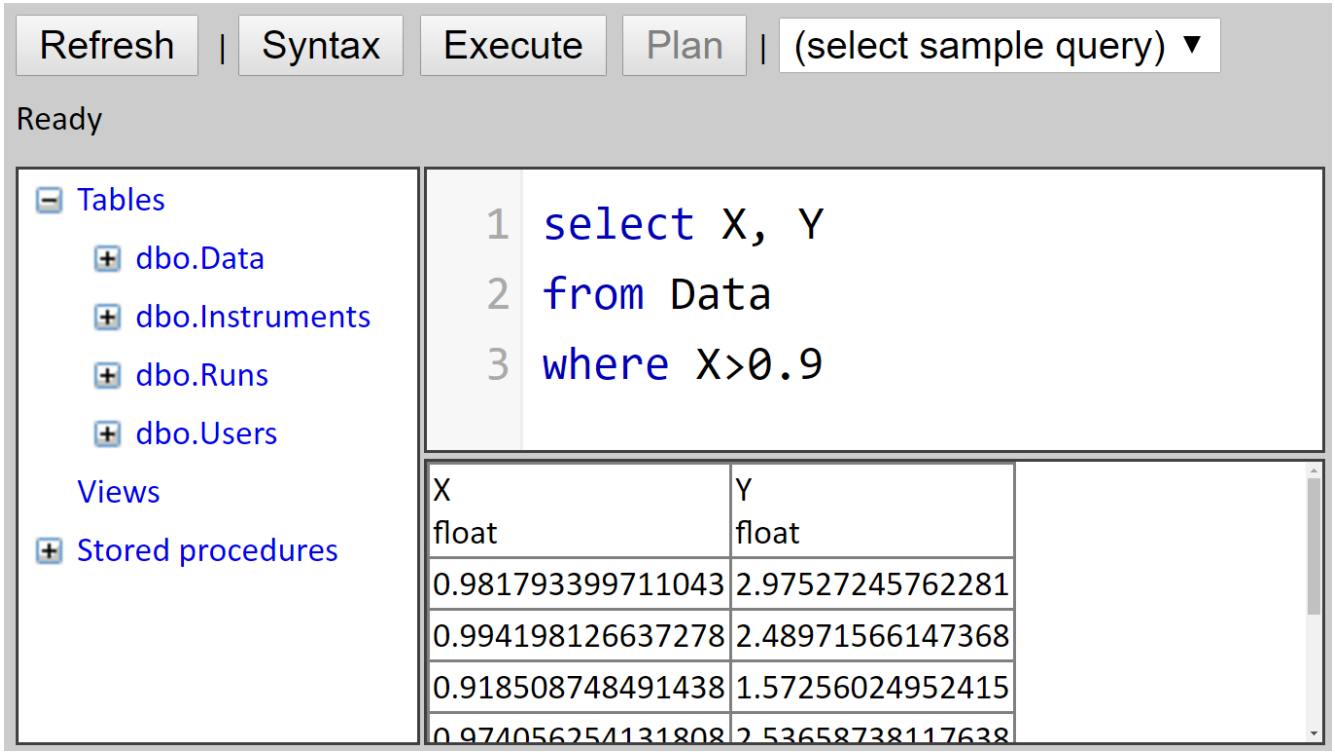
Database Systems

- MonetDB
  - Column store
- SciDB
  - Array database under development

Server – Client

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

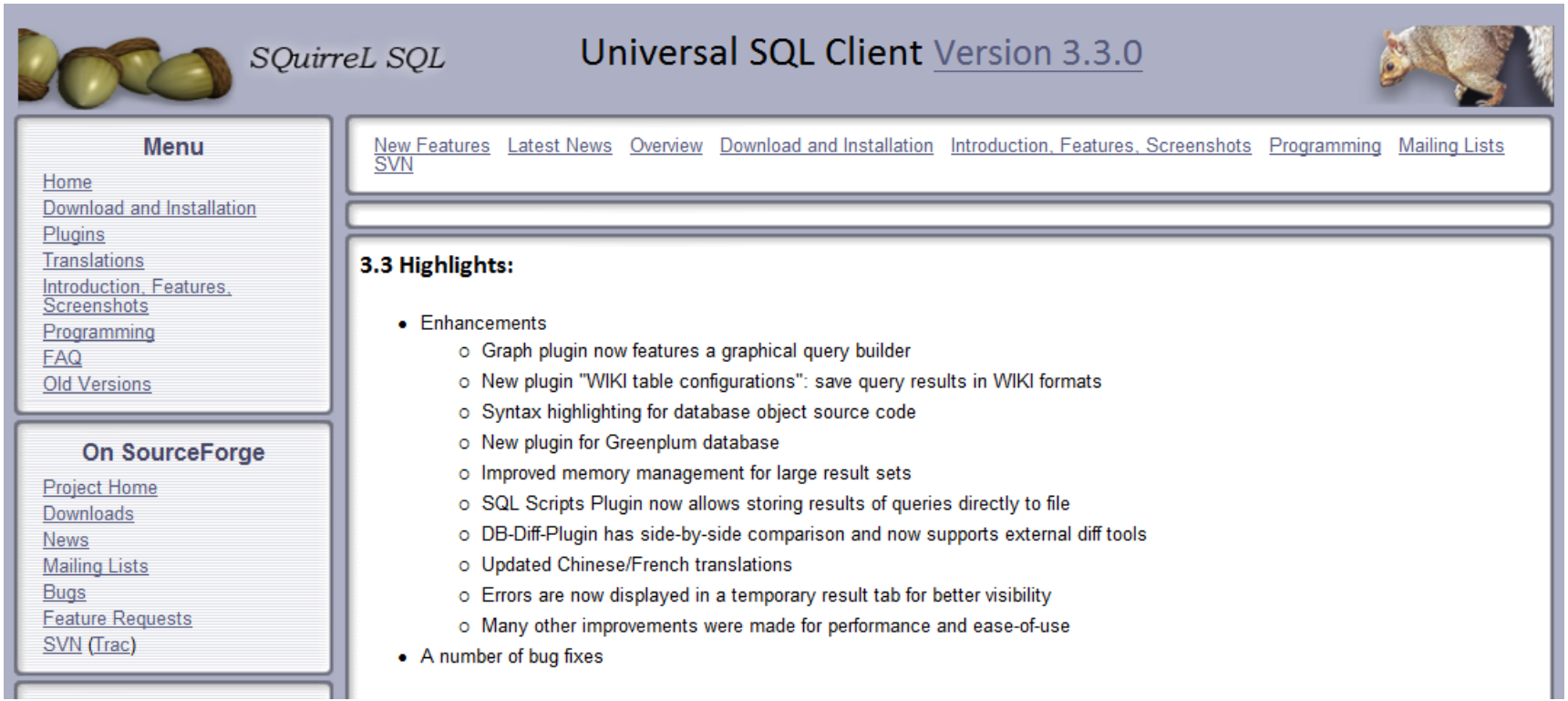
Sandbox


Refresh | Syntax Execute Plan | (select sample query)
Ready
Tables
dbo.Data
dbo.Instruments
dbo.Runs
dbo.Users
Views
Stored procedures
1 select X, Y
2 from Data
3 where X>0.9
X
float
Y
float
0.981793399711043
2.97527245762281
0.994198126637278
2.48971566147368
0.918508748491438
1.57256024952415
0.974056254131808
2.53658738117638

SQL by Examples

- Interactive session with exercises
  - See handouts...

A Universal GUI



Programmatic Interface

- Send SQL commands
- Read out the results

- Standard ODBC
  - Open Database Connectivity
  - C interface
- JDBC for Java

Research Them!

- Which one?
- Why?

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