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
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.
# 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

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

<table>
<thead>
<tr>
<th>Rows</th>
<th>kB</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>993</td>
<td>40 dist</td>
<td></td>
</tr>
<tr>
<td>949</td>
<td>200 dr3tile</td>
<td></td>
</tr>
<tr>
<td>92,082,16,640</td>
<td>DWSZ_R17_primary</td>
<td></td>
</tr>
<tr>
<td>92,082,13,376</td>
<td>DWSZ_R25</td>
<td></td>
</tr>
<tr>
<td>14,400</td>
<td>904 galexfield1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>16 galexfield2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16 halfspace</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16 MyTable</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16 MyTable_0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16 MyTable_1</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>16 MyTable_2</td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td>40 radec</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16 roomba</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>16 roomba2</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>16 stat</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>24 test1</td>
<td></td>
</tr>
</tbody>
</table>
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?
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
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...
SQL by Examples
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

```sql
1 select X, Y
2 from Data
3 where X>0.9
```
SQL by Examples

- Interactive session with exercises
  - See handouts...
A Universal GUI

3.3 Highlights:

- Enhancements
  - Graph plugin now features a graphical query builder
  - New plugin "WIKI table configurations": save query results in WIKI 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
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?