Edgar Frank “Ted” Codd
Remembering Jim Gray

The Convergence of Social and Technological Networks
Real-World Concurrency
The Polaris Tableau System
Patent Exhaustion
Search Engine Advertising
@String{pub-AW = "Addison-Wesley"}
@String{pub-AW:adr = "Reading, MA, USA"}
@Book{Graham:1994:CM, 
author = "Ronald L. Graham and Donald E. Knuth and Oren Patashnik",
title = "Concrete Mathematics",
publisher = pub-AW, 
address = pub-AW:adr, 
edition = "Second",
pages = "xiii + 657",
year = "1994",
ISBN = "0-201-55802-5",
LCCN = "QA39.2 .G733 1994",
bibdate = "Wed Jul 6 14:39:36 1994",}
Relational databases

Reflect \texttt{\LaTeX} entry across its diagonal:

<table>
<thead>
<tr>
<th>key</th>
<th>author</th>
<th>title</th>
<th>year</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graham:1994:CM</td>
<td>Ronald L. Graham and Donald E. Knuth and Oren Patashnik</td>
<td>Concrete Mathematics</td>
<td>1994</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Relational databases: split into key/value tables

<table>
<thead>
<tr>
<th>key</th>
<th>author</th>
<th>title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graham:1994:CM</td>
<td>Ronald L. Graham and Donald E. Knuth and Oren Patashnik</td>
<td>Concrete Mathematics</td>
</tr>
<tr>
<td>Lamport:1994:LDP</td>
<td>Leslie Lamport</td>
<td>\LaTeX{} — A Document Preparation System</td>
</tr>
<tr>
<td>Knuth:1986:TB</td>
<td>Donald E. Knuth</td>
<td>The \TeX{}book</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
A single database can contain multiple tables, and tables can be indexed for rapid access. Tables may be physical data, or logical views created from subsets of table data.

For bibsql, we have three tables:

- `strtab`: BIBTEX @String{...} abbreviations
- `namtab`: Author/editor names
- `bibtab`: BIBTEX fields (author, title, year, ...) and complete entry (entry)
S is for **Structured**, not **Standard**.
Several supported statements, but we often need only select:

```sql
select fieldlist from table
  where      field1 like 'pattern'
    and field2 = 'value2'
    and field3 > 'value3'
  order by field3 desc
limit n;
```
select * from bibtab;

1||9|article|acmturingawards.bib|Perlis:1967:SAS|
Alan J. Perlis|||The Synthesis of Algorithmic Systems||
j-JACM|14||1|||19||jan|1|1967|JACOAH|
http://doi.acm.org/10.1145/321371.321372|||00045411
OR 00045411|
|||This is the 1966 ACM Turing Award Lecture, and the first award.|||
@Article{Perlis:1967:SAS,
  author = "Alan J. Perlis",
  title = "The Synthesis of Algorithmic Systems",
  \ldots{}
Sample SQL queries...

```
select year, author, title from bibtab
where author like '%Perlis%' and year = '1967';
1967|Alan J. Perlis|The Synthesis of Algorithmic Systems
1967|B. A. Galler and A. J. Perlis|A proposal for definitions

select year, author, title from bibtab
where author = 'Alan J. Perlis'
order by year;
1958|Alan J. Perlis|Announcement
1963|Alan J. Perlis|Computation’s development critical to our
1967|Alan J. Perlis|The Synthesis of Algorithmic Systems
...
How many variants are there of Guy Steele’s name?

```
select count, name from namtab
    where name like '%Steele%'
    order by 1 desc;
```

15|Guy L. Steele Jr.
3|Guy L. Steele
2|Guy L. Steele, Jr.
1|G. L. Steele, Jr.
1|G. Steele
Sample SQL queries...

Find five Knuth articles published between 1956 and 1969:

```sql
select distinct year, author, title from bibtab
    where author like '%D%Knuth'
    and '1955' < year
    and year < '1970'
    order by year desc
limit 5;
```

1969|Donald E. Knuth|Seminumerical Algorithms
1968|Donald E. Knuth|Very magic squares
1967|Donald E. Knuth|The Remaining Trouble Spots in ALGOL 60
1966|Donald E. Knuth|Errata: ‘Additional comments on a problem in concurrent ...’
1966|Donald E. Knuth|Letter to the Editor: Additional comments...
Sample SQL queries...

What is the percentage of journal articles that have each of one to five authors?

```sql
select round(100 * count(authorcount) / (select count(*) from bibtab
    where authorcount > 0 and bibtype = 'article')) || '%',
    authorcount
from bibtab
where authorcount > 0 and bibtype = 'article'
group by authorcount
order by count(authorcount) desc
limit 5;
```

47.0% | 1  
29.0% | 2  
14.0% | 3  
5.0%  | 4  
1.0%  | 5  
Database implementations

- MySQL
- PostgreSQL
- SQLite3
- IBM DB2
- Ingres
- Microsoft SQL Express
- Oracle
- Sybase

All but SQLite3 are **client/server databases**, and relatively complex to set up and manage. Some are licensed commercial systems ($$$). SQLite3 requires only **one** platform independent file, and its software is highly portable and in the **public domain**.
SQLite3 schemas

sqlite> .schema
CREATE TABLE bibtab (  
    authorcount INTEGER,  
    editorcount INTEGER,  
    pagecount INTEGER,  
    bibtype TEXT,  
    filename TEXT,  
    label TEXT,  
    author TEXT,  
    ...  
    ZMnumber TEXT,  
    entry TEXT NOT NULL UNIQUE  
);
CREATE TABLE namtab (  
    name        TEXT NOT NULL UNIQUE,  
    count       INTEGER  
);  
CREATE TABLE strtab (  
    key         TEXT,  
    value       TEXT,  
    entry       TEXT NOT NULL UNIQUE  
);  
CREATE INDEX bibididx on bibtab (author, title, label);  
CREATE INDEX bibtimestampidx on bibtab(bibtimestamp);  
CREATE INDEX isbn13idx on bibtab (isbn13);  
...
bibtosql: convert Bib\TeX entries to database input

% bibtosql --help
Usage: /usr/local/bin/bibtosql
[ --author ]
[ --create ]
[ --database dbname ]
[ --help ]
[ --version ]
[ --server ( MySQL | psql | PostgreSQL | SQLite ) ]
[ -- ] BibTeX files or <infile
>outfile

% bibtosql --create *.bib | sqlite3 bibtex.db
% bibsql --help
Usage: /usr/local/bin/bibsql
[  --author ]
[  --command ' command1; command2; ... ' ]
[  --database dbname ]
[  --help ]
[  --options ' ... server options ...' ]
[  --server ( MySQL | psql | PostgreSQL | SQLite ) ]
[  --user dbuser ]
[  --version ]

% bibsql -s psql
psql> ... user input here ...
Interfaces to SQL databases are available in common programming and scripting languages. Sample C code for interfacing to MySQL, PostgreSQL, and SQLite3 is included in the bibsql distribution:

http://www.math.utah.edu/pub/bibsql/