Basic SQL for Institutional Research

Phil Rhodes

TAIR 2012

February 22, 2012

Topics

- Why SQL?
- What is SQL?
- Basic Queries
- Joining Two Tables
- Simple Summary Reports

Why SQL?

- Institutional Research is all about data
- Data usually 'locked up' in relational databases (Oracle, DB2, SQL Server)
- IR offices can:
  - Depend on IT to give timely, correct extracts
  - Extract data themselves
- There are other tools, but some operations are easier in SQL

What is SQL?

- "S-Q-L" vs "sequel"
- Developed at IBM in early 1970's
- Based on a 1970 paper by Edgar F. Codd
- Standardized first in 1986
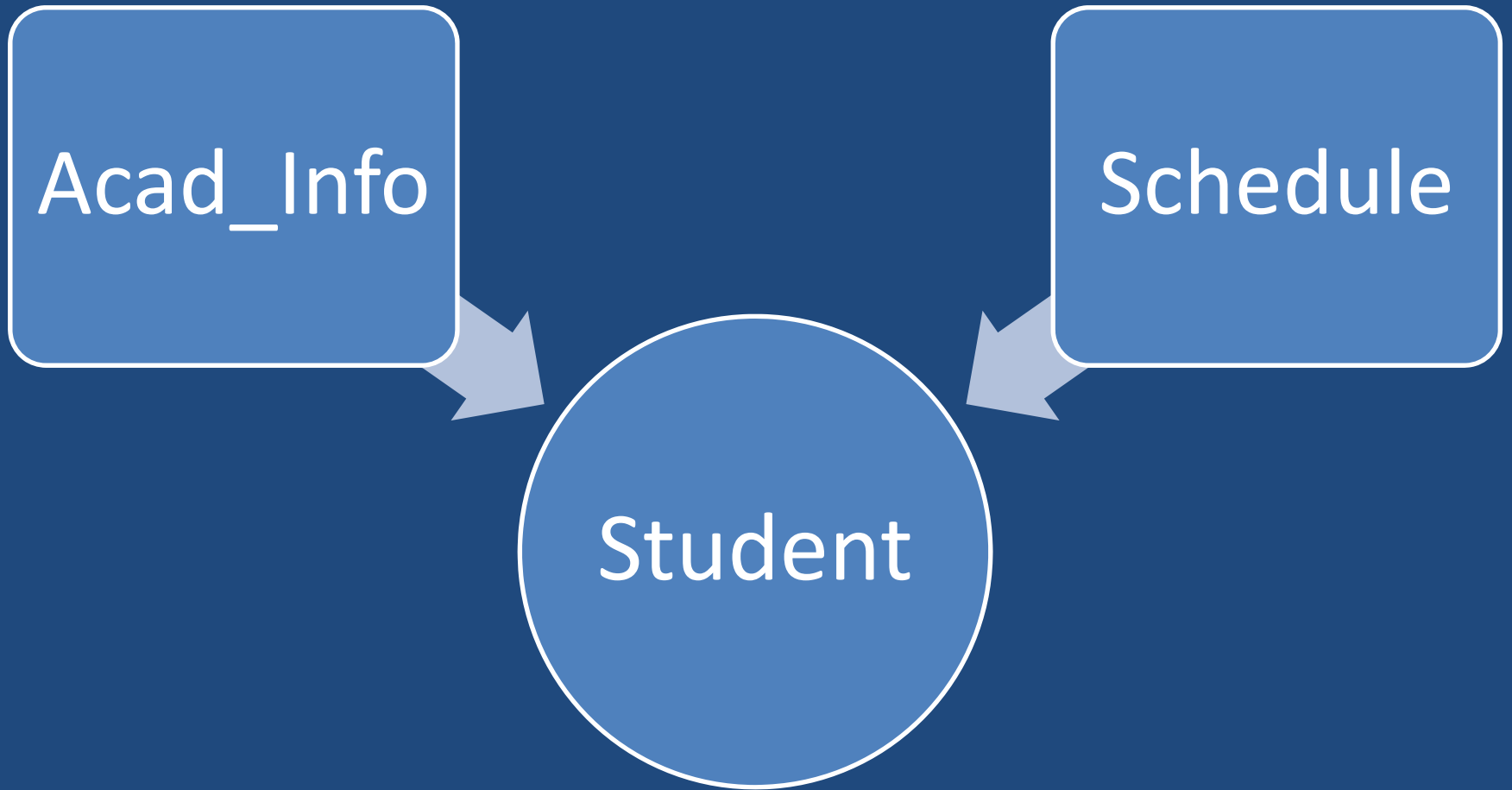- Many variants
  - "English"

What is SQL?

- Queries
  - Retrieve data
- Data Manipulation Language (DML)
  - Add, update, and delete data
- Data Definition Language (DDL)
  - Manage structure of tables and indices
  - Character, numeric, date/date-time values
- Data Control Language (DCL)
  - Access control

What is SQL?

- Focus on Queries

- Examples from SAS and MS Access

Sample Data

Sample Data

- Student table
  - Keys: ID
  - Bio-Demo data
- Acad_info table
  - Keys: ID, TERM
  - College, Degree, Major, Classification
- Schedule table
  - Keys: ID, TERM, CRN
  - Course, Credit Hours

Basic SQL Queries

- Most basic query:

  select `[column-list]`

  from `[table]`

- Returns values of the columns in [column-list] for all rows of the table

- An asterisk '*' can be used as an alias for all columns

  - Useful shorthand, risky in production code

- Example 1 - SAS

Basic SQL Queries

- Restricting the rows retrieved

    select [column-list]

    from [table]

    where [logical-expression]

- Standard boolean operators AND, OR, NOT

    - Also IN, LIKE, BETWEEN

    - Three-valued logic – True, False, Null/Unknown

        - IS NULL

- Example 2 – SAS & Access

Joining Two Tables

- Often require data from more than one table
- Keys
  - Required to match rows from one table to another
  - Primary and secondary keys
- Two types of joins
  - Inner Joins
  - Outer Joins

Joining Two Tables

- Inner Joins

```
select [column-list]
from [table1] as a
        inner join [table2] as b
where a.key = b.key
```

- Returns data from rows in both tables where keys match

Joining Two Tables – Inner Join

Table 1

ID	Gender
1	M
2	M
3	F
4	F

Table 2

ID	Classification
1	Freshman
3	Sophomore
5	Junior
7	Senior

```
select a.id, a.gender,
       b.classification
from table1 as a
     inner join table2 as b
where a.id = b.id
```

Joining Two Tables – Inner Join

Result

ID	Gender	Classification
1	M	Freshman
3	F	Sophomore

```
select a.id, a.gender,
        b.classification
from table1 as a
        inner join table2 as b
where a.id = b.id
```

Joining Two Tables – Inner Join

- Example 3

Joining Two Tables

- Outer Joins

```
select [column-list]
from [table1] as a
     left join [table2] as b
on a.key = b.key
```

- Returns all rows from table1 plus matching rows in table2

Joining Two Tables – Outer Join

Table 1

ID	Gender
1	M
2	M
3	F
4	F

Table 2

ID	Classification
1	Freshman
3	Sophomore
5	Junior
7	Senior

```
select a.id, a.gender,
       b.classification
from table1 as a
     left join table2 as b
on a.id = b.id
```

Joining Two Tables – Outer Join

Result

ID	Gender	Classification
1	M	Freshman
2	M	
3	F	Sophomore
4	F	

```
select a.id, a.gender,
       b.classification
from table1 as a
     left join table2 as b
on a.id = b.id
```

Joining Two Tables – Outer Join

- Example 4

Joining Two Tables – Issues

- Bad or Missing WHERE statements
  - Cartesian products
- Multiplying Observations
  - Duplicated data if more than one row in a table matches criterion
  - Sometimes good, sometimes sign of an error

Basic Reporting

- GROUP BY
  - Combine rows with common values
  - Often used with SQL aggregation functions like SUM()
    - SUM, MIN, MAX, AVG, COUNT, VAR, STDEV
  - WHERE is applied <u>before</u> GROUP BY
- HAVING
  - Similar to WHERE, but applied <u>after</u> GROUP BY
  - Can use aggregation functions

Basic Reporting

- ORDER BY
  - Order the result data by the values of certain columns
  - ASCENDING or DESCENDING

Basic Reporting

- Example 5: Create a report of the number of students by home state.

- Example 6: Create a report by classification of students taking more than 18 hours

Resources

- SAS Online Documentation
  - support.sas.com/onlinedoc/913/docMainpage.jsp
    - Base SAS → Base SAS Procedures Guide → SAS SQL Procedure User's Guide
- MS Access SQL Documentation
  - office.microsoft.com/en-us/access-help/CH010072899.aspx
- SQL Tutorial
  - www.sqltutorial.org

Resources

- O'Reilly Books (shop.oreilly.com)
    - Learning SQL, 2nd Edition
    - SQL In a Nutshell, 3rd Edition
    - SQL Cookbook, 1st Edition

Questions?

Phil Rhodes

Houston Baptist University

prhodes@hbu.edu

281-649-3417