

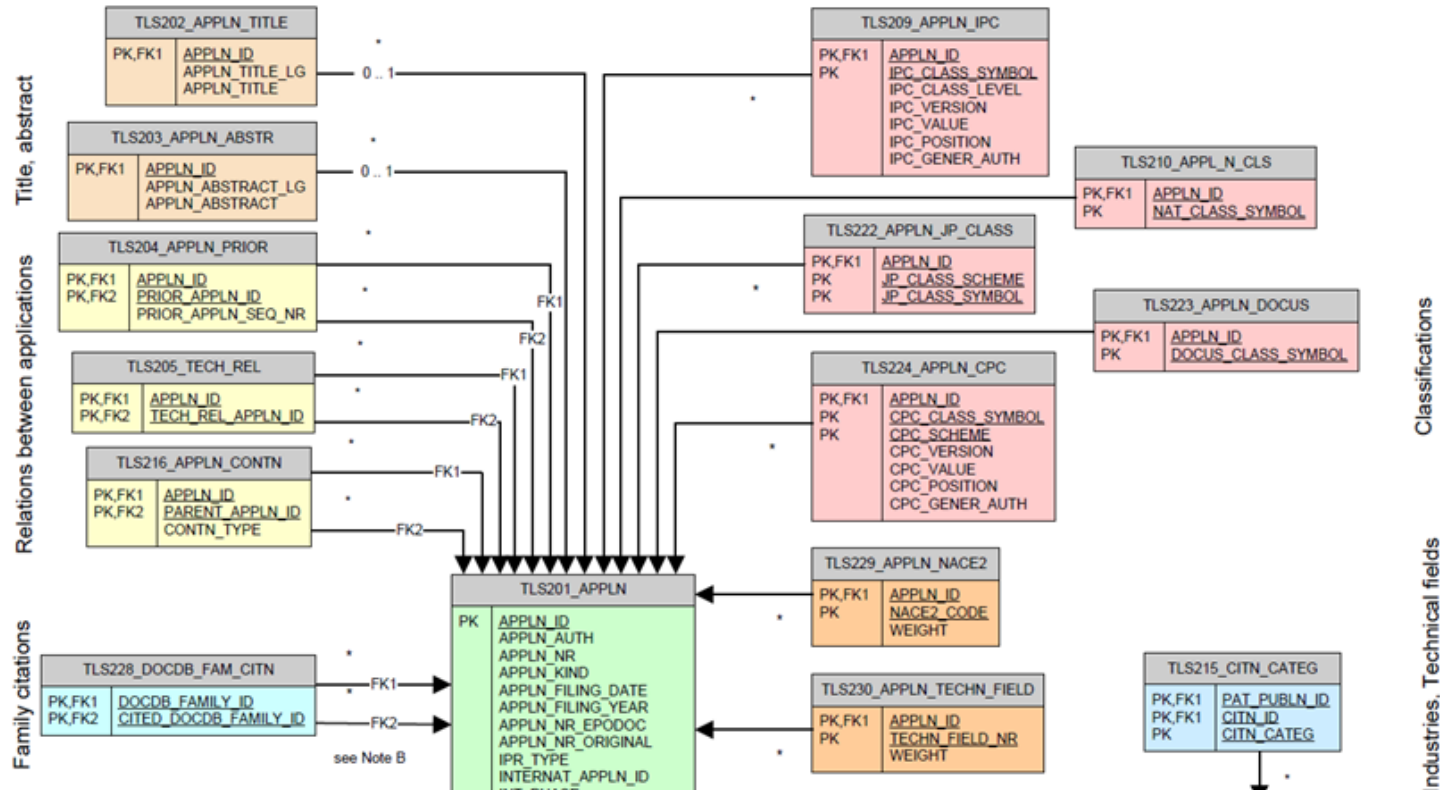
SQL BASICS

ELEMENTS OF SQL FOR PATSTAT



RELATIONAL DATABASES

3.2 Logical model diagram



SQL works on
relational DB

PATSTAT logical
model is a
relational
database

SQL: NOT ONLY FOR DBMS

```
IMPORT SQLITE3
IMPORT PANDAS AS PD
DB = SQLITE3.CONNECT('PATSTAT.DB')
DEF RUN_QUERY(QUERY):
    RETURN PD.READ_SQL_QUERY(QUERY,DB)
```

PYTHON

```
LIBRARY(SQLDF)
```

```
SQLDF('SELECT * FROM
TLS206_PERSONS WHERE
PERSON_CTRY_CODE = "DE"')
```

R

KEYS

- A KEY IS AN ATTRIBUTE (OR A SET OF ATTRIBUTES) WHICH CAN UNIQUELY IDENTIFY A SINGLE ROW OF ITS OWN TABLE
- EXAMPLES
 - TLS201_APPLN.APPLN_ID
(SHORT FOR: COLUMN APPLN_ID IN TLS201_APPLN)
 - IN TABLE TLS211_PAT_PUBLN:
COLUMN PAT_PUBLN_ID
- ONE KEY (THE SHORTEST ONE) IS THE PRIMARY KEY; OTHER KEYS MAY EXIST: ALTERNATIVE KEYS
- BY CONVENTION, KEYS ARE USUALLY THE LEFTMOST COLUMNS OF A TABLE

SELECT

- SELECT WILL BE THE FOUNDATION OF ALMOST EVERY QUERY YOU WRITE
- IT TELLS THE DATABASE WHICH COLUMNS YOU WANT TO SEE.
- YOU CAN EITHER SPECIFY COLUMNS BY NAME (SEPARATED BY COMMAS) OR USE THE WILDCARD * TO RETURN EVERY COLUMN.
- IN ADDITION TO THE COLUMNS, YOU ALSO HAVE TO TELL THE TABLE (USE THE KEYWORD FROM)

SELECT (EXAMPLES)

- `SELECT A.* FROM PATSTAT.TLS201_APPLN A;`
- `SELECT APPLN_ID, APPLN_FILING_YEAR, EARLIEST_FILING_YEAR
FROM PATSTAT.TLS201_APPLN;`
- FIRST QUERY CONTAINS ALIAS FOR TABLE 201;
- TABLE NAME AND ATTRIBUTE NAME ARE CONNECTED WITH A DOT:
<TABLE NAME>.<ATTRIBUTE NAME>

SELECT * FROM PATSTAT.TLS201_APPLN

#	APPLN_ID	APPLN_AUTH	APPLN_NR	APPLN_KIND	APPLN_FILING_DATE	APPLN_FILING_YEAR	APPLN_NR_EPODOC
	0	XX		D	9999-12-31	9999	
	1	EP	103094	A	2000-02-15	2000	EP20000103094
	2	EP	107845	A	1992-12-02	1992	EP20000107845
	3	EP	202556	A	2000-07-17	2000	EP20000202556
	4	EP	300208	A	2000-01-13	2000	EP20000300208

Appln_id

primary key: unique appln_id + appn_auth + appln_kind

WHERE: FILTERING RESULTS

- ```
SELECT APPLN_ID
FROM TLS201_APPLN
WHERE APPLN_AUTH = 'IT'
```

RETRIEVES ALL APPLICATION IDS WHICH WERE FILED IN ITALY.

- ```
SELECT <ATTRIBUTE NAME>
```

 → DEFINE LIST OF FIELDS (ATTRIBUTES) TO SHOW
- ```
FROM <TABLE NAME>
```

 → DEFINE TABLES USED
- ```
WHERE <CONDITION>
```

 → OPTIONAL; CONDITION (WHICH ROWS?)
- NOTE: SQL LANGUAGE IS NOT CASE-SENSITIVE

WHERE: FILTERING RESULTS (2)

- THE WHERE CLAUSE CONTAINS ONE OR MORE CONDITIONS.

- THE GENERAL FORM OF A CONDITION IS:

<ATTRIBUTE OR CONSTANT> <COMPARISON OPERATOR> <ATTRIBUTE OR CONSTANT>

- THE MOST COMMON COMPARISON OPERATORS AND THEIR MEANINGS ARE

= EQUAL

<> NOT EQUAL

>, >= LARGER, LARGER OR EQUAL

<, <= SMALLER, SMALLER OR EQUAL

YOU CAN ALSO COMPARE 2 ATTRIBUTES: APPLN_FILING_YEAR = EARLIEST_PUBLN_YEAR

WHERE: FILTERING RESULTS (3)

- OTHER OPERATORS:

- LIKE: PERSON_NAME LIKE 'APPLE%'
- IN: APPLN_AUTH IN ('EP', 'JP', 'US')

- BOOLEAN LOGIC APPLIES:

```
WHERE    (APPLN_AUTH = 'EP'
          AND APPLN_FILING_YEAR = 2008)
OR        (APPLN_AUTH <> 'EP'
          AND APPLN_FILING_YEAR = 2010)
```

LIMITING & SORTING RESULTS

- ORDER BY <ATTRIBUTE> [ASC/DESC]:

```
SELECT * FROM TLS211_PAT_PUBLN ORDER BY PUBLN_DATE
```

- * LIMIT / TOP: RETRIVES ONLY X RECORDS [MYSQL VS SQL SERVER]

```
SELECT TOP 5 * FROM TLS211_PAT_PUBLN
```

AGGREGATIONS, SUMMARIES ETC

- GROUP BY SEPARATES THE ROWS INTO GROUPS BASED ON THE CONTENTS OF A PARTICULAR COLUMN AND ALLOWS US TO PERFORM AGGREGATE FUNCTIONS ON EACH GROUP
- OTHER COLUMNS CAN BE ARGUMENT OF SUMMARIES FUNCTIONS AS:
SUM, COUNT, AVG, MAX, MIN
KEYWORD DISTINCT IS FOR UNIQUE VALUES

AGGREGATIONS, SUMMARIES (2)

- SELECT APPLN_AUTH, COUNT(APPLN_ID) AS C FROM PATSTAT.TLS201_APPLN GROUP BY APPLN_AUTH
- SELECT APPLN_AUTH, MAX(FILING_DATE) AS MAXD, MIN(FILING_DATE) AS MIND FROM PATSTAT.TLS201_APPLN GROUP BY APPLN_AUTH
- SELECT MAX(FILING_DATE) AS MAXD, MIN(FILING_DATE) AS MIND FROM PATSTAT.TLS201_APPLN → NO GROUP BY

AGGREGATIONS, SUMMARIES (3)

- HAVING CLAUSE ALLOWS FILTERING ON GROUP BY RESULTS

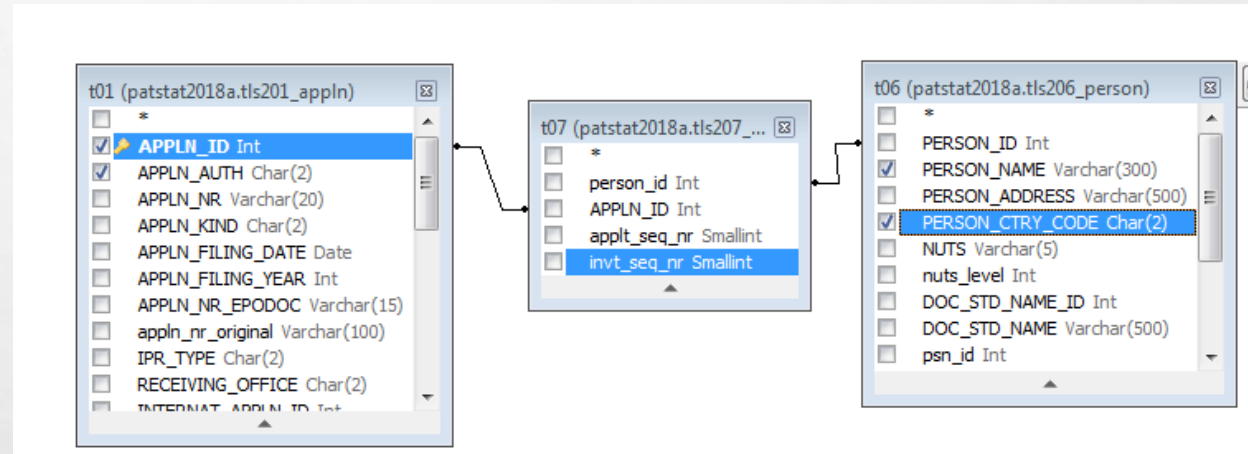
```
SELECT
  T01.APPLN_AUTH, COUNT(DISTINCT T01.APPLN_ID) AS COUNT_APPLN_ID
FROM
  PATSTAT.TLS201_APPLN T01
GROUP BY
  T01.APPLN_AUTH
HAVING
  COUNT(DISTINCT T01.APPLN_ID) > 1000
```

JOINS (MULTITABLE QUERIES)

- IN RELATIONAL DBS DATA ARE BROKEN IN MULTIPLE TABLES
- JOIN CLAUSE CONNECTS 2 TABLES
- *SELECT * FROM TABLE1 A INNER JOIN TABLE2 B ON A.FIELD1=B.FIELD2*
- SPECIFIES THE TABLE TO BE JOINED WITH THE OTHER TABLES IN THE QUERY THE JOIN CONDITION (ON) SPECIFIES THE ATTRIBUTES WHICH SHOULD BE JOINED

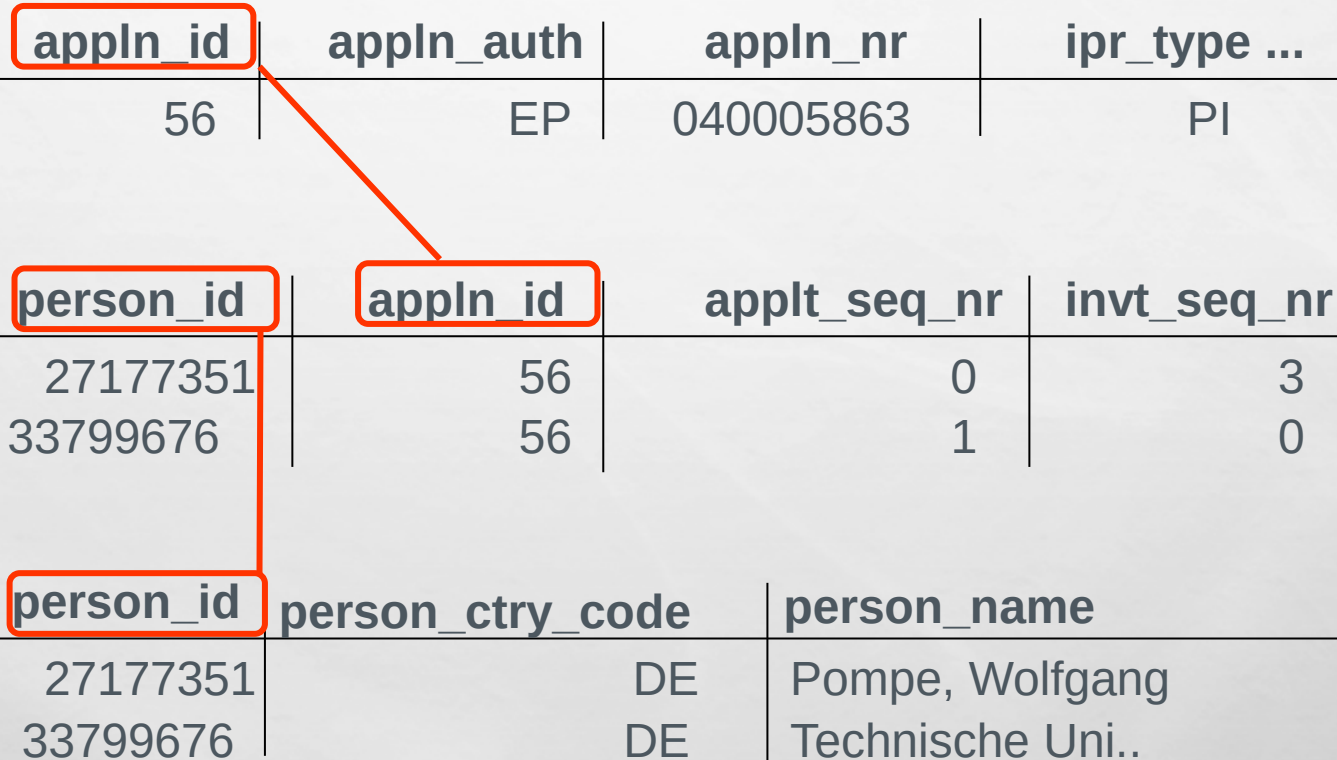
JOINS (FIND EP APPLICANTS)

- SELECT
- T01.APPLN_ID,
- T06.PERSON_NAME,
- T06.PERSON_CTRY_CODE
- FROM
- PATSTAT.TLS201_APPLN T01
- INNER JOIN PATSTAT.TLS207_PERS_APPLN T07 ON T07.APPLN_ID = T01.APPLN_ID
- INNER JOIN PATSTAT.TLS206_PERSON T06 ON T06.PERSON_ID = T07.PERSON_ID
- WHERE
- T07.APPLT_SEQ_NR > 0 AND T01.APPLN_AUTH = "EP"



JOINS (MULTITABLE QUERIES 2)

tls201_appln	appln_id	appln_auth	appln_nr	ipr_type ...
	56	EP	040005863	PI
tls207_pers_appln	person_id	appln_id	applt_seq_nr	invt_seq_nr
	27177351	56	0	3
	33799676	56	1	0
tls206_person	person_id	person_etry_code	person_name	
	27177351	DE	Pompe, Wolfgang	
	33799676	DE	Technische Uni..	



JOINS – DIFFERENT TYPES

- **(INNER) JOIN**: Returns records that have matching values in both tables
- **LEFT (OUTER) JOIN**: Return all records from the left table, and the matched records from the right table
- **RIGHT (OUTER) JOIN**: Return all records from the right table, and the matched records from the left table
- **FULL (OUTER) JOIN**: Return all records when there is a match in either left or right table



SUBQUERIES

- NESTING QUERIES WITHIN ANOTHER QUERY
- SUBQUERY IS WITHIN ()

```
SELECT * FROM TLS211_PAT_PUBLN  
WHERE YEAR(PUBLN_DATE) =  
      (SELECT MIN(YEAR(PUBLN_DATE))  
       FROM TLS211_PAT_PUBLN)
```

IN GENERAL

SELECT [DISTINCT] [TOP N] ATTRIBUTE(S) OR EXPRESSION(S)

FROM TABLE

JOIN TABLE ON JOIN CONDITION(S) -- JOIN CLAUSE MAY REPEAT

WHERE CONDITION(S)

GROUP BY ATTRIBUTE(S) OR EXPRESSION(S)

HAVING GROUP CONDITION(S)

ORDER BY ATTRIBUTE(S) OR EXPRESSION(S)

- [HTTPS://WWW.W3SCHOOLS.COM/SQL/](https://www.w3schools.com/sql/)
- [USING PATSTAT WITH SQL FOR BEGINNERS](https://www.epo.org/searching-for-patents/business/patstat.html)
([HTTPS://WWW.EPO.ORG/SEARCHING-FOR-PATENTS/BUSINESS/PATSTAT.HTML](https://www.epo.org/searching-for-patents/business/patstat.html))
- [HTTPS://WWW.DATAQUEST.IO/BLOG/SQL-BASICS/](https://www.dataquest.io/blog/sql-basics/)