

HARVARD BUSINESS ANALYTICS PROGRAM

Data at Scale

Database

Relational Database

Structured Query Language (SQL)

NoSQL

Data Warehouse

Data Mart

Data Lake

Raw Data

<https://hbap.io/pds/favorites>

Flat-File Database

CSV

Relational Database

SQL

SQLite

C
R
U
D

CREATE

READ

UPDATE

DELETE

CREATE, INSERT

SELECT

UPDATE

DELETE, DROP

```
SELECT columns FROM table;
```

AVG
COUNT
DISTINCT
LOWER
MAX
MIN
UPPER
...

WHERE
LIKE
ORDER BY
LIMIT
GROUP BY
...

```
INSERT INTO table (column, ...) VALUES(value, ...);
```

```
UPDATE table SET column = value WHERE condition;
```



```
DELETE FROM table WHERE condition;
```

```
CREATE TABLE table (column type, ...);
```

BLOB

INTEGER

NUMERIC

REAL

TEXT

NOT NULL
UNIQUE

PRIMARY KEY
FOREIGN KEY

IMDb

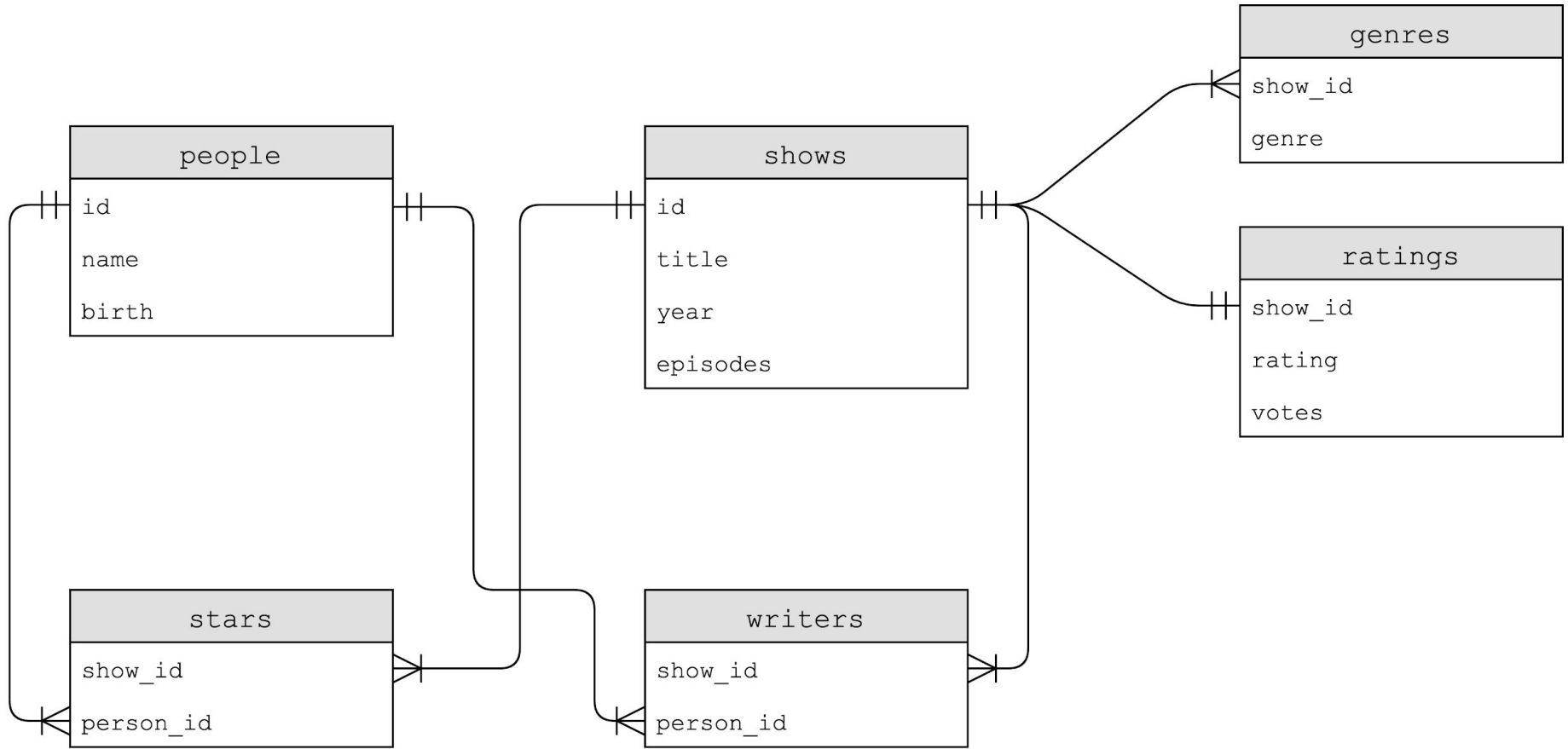
IMDb

IMDb

IMDb

IMDb

IMDb



JOIN

...

JOIN

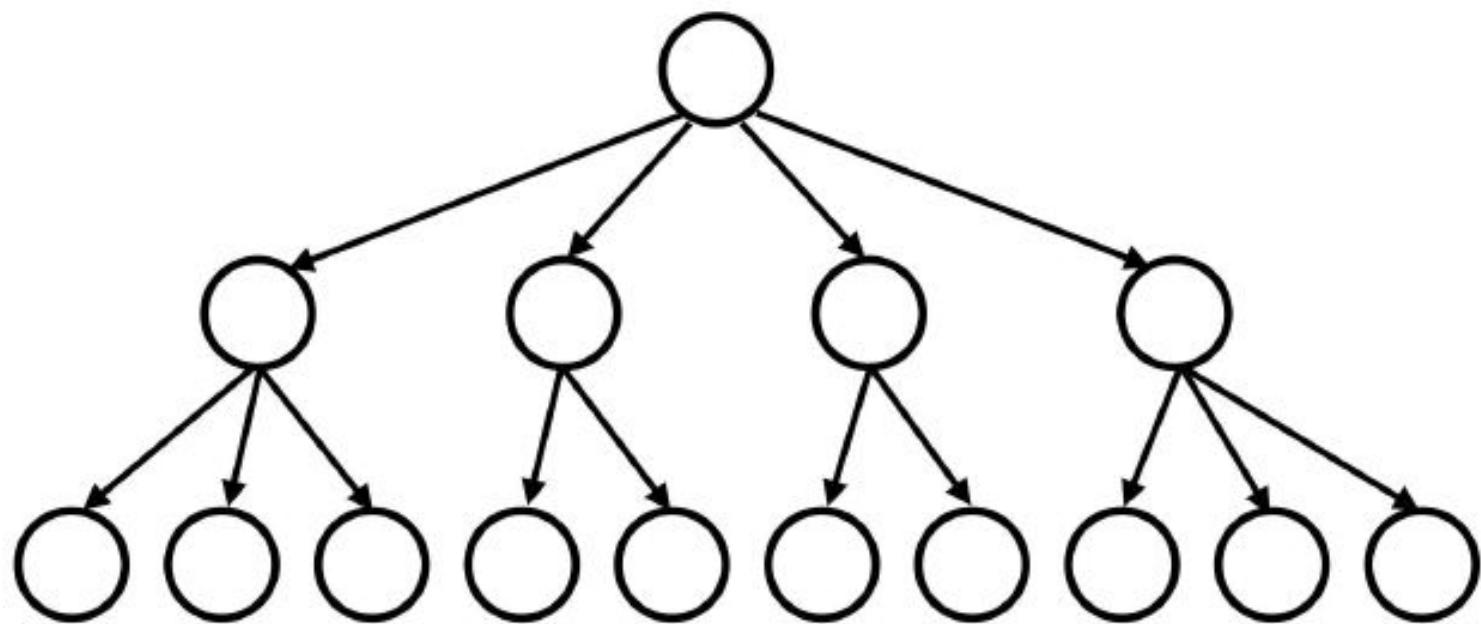
...

JOIN

...

Index


```
CREATE INDEX name ON table (column, ...);
```



View

```
CREATE VIEW name AS SELECT ...;
```

SQL Injection Attacks

```
SELECT * FROM users
WHERE username = '{username}'
AND password = '{password}'
```

```
SELECT * FROM users
WHERE username = 'malan@harvard.edu'--'
AND password = '{password}'
```

```
SELECT * FROM users
WHERE username = 'malan@harvard.edu'--'
AND password = '{password}'
```



```
SELECT * FROM users
WHERE username = 'malan@harvard.edu\!--'
AND password = '{password}'
```

Race Conditions



```
SELECT likes FROM posts ...  
UPDATE posts SET likes = ...
```

BEGIN TRANSACTION
COMMIT
ROLLBACK

Data at Scale