Tables:

article (article\_key, title, year, month, journal)

inproceedings (inproc\_key, crossref, year, booktitle, title)

proceedings (proc\_key, crossref, year, booktitle, title)

mastersthesis (masThesis\_key, year, title)

phdthesis (phdThesis\_key, year, title)

book (book\_key, year, title)

data (data\_key, year, title)

incollection (incol\_key, year, title)

www (www\_key, year, title)

author ()

article\_author (id, article\_key, title, author)

inproceedings\_author (id, inproc\_key, author)

proceedings\_author (id, proc\_key, author)

mastersthesis\_author (id, masThesis\_key, author)

phdsthesis\_author (id, phdThesis\_key, author)

incollection\_author (id, incol\_key, author)

book\_author (id, book\_key, author)

data\_author (id, data\_key, author)

www\_author (id, www\_key, author)

…

For any publication:

Pub (key, title, author, year, month, journal, booktitle, crossref)

Get journal or booktitle if there is any

导入流程：

下载 备份

删除author 全局删除重复项

删除key重复项

导入后设置primary key

Queries:

Q1

For each type of publication: year

CREATE TABLE Result1 (

publication\_type VARCHAR(50) PRIMARY KEY,

count INT

);

SET @article\_count :=

(SELECT COUNT(\*) FROM articles WHERE year BETWEEN 2010 AND 2019);

INSERT INTO Result1 (publication\_type, count)

VALUES (“Article”, @article\_count);

Q2

Conferences: num of papers (proceedings & inproceedings) in each year

SELECT booktitle

FROM (SELECT \* FROM Proceedings UNION SELECT \* FROM Inproceedings)

WHERE key LIKE ‘conf/%’

GROUP BY booktitle, year

HAVING COUNT(\*)>500;

Integrating Heterogeneous, Autonomous, Distributed Applications Using the DOM Prototype.

Q3

For each 10-year period, count num of conference publications

SET @conf\_pub\_count\_1970 := (

SELECT SUM(count),

FROM

(SELECT year, COUNT(\*)

FROM (SELECT \* FROM Proceedings UNION SELECT \* FROM Inproceedings)

WHERE year BETWEEN 1970 AND 2019 AND key LIKE ‘conf/%’

GROUP BY year)

WHERE year BETWEEN 1970 AND 1979

);

Q4

Author name, conference or journal name (from articles, proceedings & inproceedings), number of collaborators

## Find all authors

SELECT DISTINCT author

FROM Article\_author, Article

WHERE journal LIKE ’% data’ OR journal LIKE ’data %’ OR journal LIKE ’% data %’

OR journal LIKE ’% Data’ OR journal LIKE ’Data %’ OR journal LIKE ’% Data %’

OR journal LIKE ’% DATA’ OR journal LIKE ’DATA %’ OR journal LIKE ’% DATA %’

## Find num of collaborators

SELECT author, COUNT(collabrator)

FROM (

(SELECT DISTINCT author, collabrator

FROM Article\_author, Article\_author

WHERE author <> author)

UNION () UNION ()

)

GROUP BY author

## ORDER BY and find all the top authors

WITH TempV AS

(SELECT author, RANK() OVER (ORDER BY collab\_count DESC) AS R

FROM author\_collab\_count)

SELECT \*

FROM TempV

WHERE R = 1

Q5

top 10 authors, conference or journal, publication titile

SELECT author

FROM pubs

WHERE title LIKE ’% Data’ OR title LIKE ’Data %’ OR title LIKE ’% Data %’

AND publication\_year BETWEEN YEAR(2023 - 4 AND 2023)

ORDER BY pub\_count DESC

LIMIT 10;

Q6

Conference name, month, count number of proceedings

SELECT conference\_name

FROM Proceedings

WHERE month = “June”

GROUP BY key

HAVING COUNT(\*) > 100

Q7

1. Author, paper, year, family name

Inproceedings ()

Mastersthesis ()

Phdthesis ()

Proceedings ()

SELECT author

FROM papers

WHERE publication\_year BETWEEN YEAR(2023 - 29 AND 2023)

GROUP BY author

HAVING COUNT(DISTINCT publication\_year) = 30

1. year, author, count number of publications

## Find earliest pub

WITH TempV AS

(SELECT key, RANK() OVER (ORDER BY year ASC) AS R

FROM pub)

SELECT key

FROM TempV

WHERE R = 1;

## Find author with pub count

SELECT author, COUNT(key)

FROM pub-author

WHERE key IN earliest\_pub

GROUP BY author

Q8