NANYANG TECHNOLOGICAL UNIVERSITY

School of Computer Engineering

Laboratory Report

Subject: CZ 4031

Querying Databases Efficiently

Group ID: 26
Guo Jiachun
Xu Mengxing
Zhou Xinzi

19 October 2015

0. Introduction

In this laboratory project, PostgreSQL is used, because it is very fast, advanced and flexible and we want to learn about it.
Our testing database is running on a virtual machine in cloud cluster. The virtual machine runs Ubuntu 14.04 and PostgreSQL 9.3.9. It occupies one core of a 6 Core processor (Intel(R) Xeon(R) CPU E5-2630L v2 @ 2.40GHz) with 1 GB memory. The cloud cluster uses SSD hard disk, so the hard disk access time is fast.

1. Schema Design and Data Acquisition

1.1 Schema Design

We created 7 tables for this project. They are article, author, book, incollection, inproceedings, pub_author, publication.

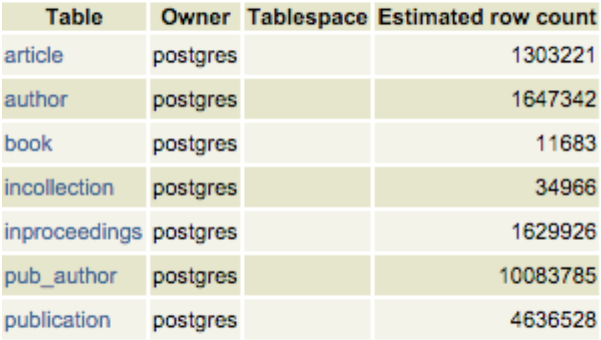


Table	Owner	Tablespace	Estimated row count
article	postgres		1303221
author	postgres		1647342
book	postgres		11683
incollection	postgres		34966
inproceedings	postgres		1629926
pub_author	postgres		10083785
publication	postgres		4636528

Figure 1 Table information in database

The ER diagram of the schema is provided below.

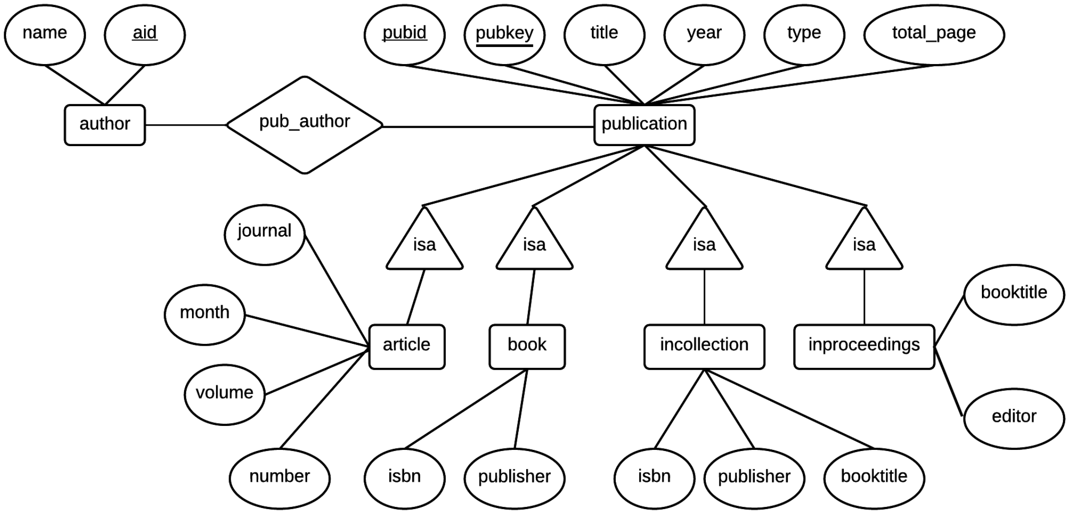


Figure 2 ER diagram

Please refer to our source codes for the detailed table design.

1.2 Data Acquisition

We used python and lxml to read the data from DBLP's xml file. We generated INSERT statements in SQL to files. And we executed the SQL files in PostgreSQL to import the data.

1.3 Assumptions

The DBLP does not provide the page number count for each publication directly. Base of our research, we create our own decode algorithm for calculating page numbers in data processing. Please refer to our Python code for more detail.

Articles and inproceedings may be published in either journal or conferences. We extracted this information from publication's attribute "pub_key" and added a new column "type" in relation publication. The value is "journals" for articles or inproceedings published in journal, and "conf" for those published in conferences. We used this attribute in Query 3, 4 and 9 For incollection and book, this attribute is not relevant.

2. Queries

2.1 Introduction to our queries

We used a python program to run the queries in the server and recorded the execution time automatically.

In Query 3A, we set X (author) to "Yan Zhang" and in Query 3B and Y (year) to 2012, we set X (author) to "Wei Wang", Y (year) to 2009 and Z (conference) to CSCWD.

Queries 8 and 9 are our custom queries, where Query 8 selects the authors who have written more than 4000 pages of publication and Query 9 selects the top ten prolistic authors in all the conferences.

Please refer to our appendix for the SQL queries and results.

2.2 Analysis

	Full (s)	Half (s)	Quarter (s)
Analyze	6.4986	4.7777	3.8524
Q1	0.3426	0.2291	0.1746
Q2A	22.6443	10.4221	5.1452
Q2B	32.6716	21.5816	8.5167
Q3A	4.1945	2.3362	0.8583
Q3B	0.7304	0.3856	0.2787
Q4A	0.4514	0.2315	0.1261
Q4B	0.4279	0.2179	0.1182
Q5	10.0300	3.4353	1.1560
Q6	101.9929	48.3984	18.0066
Q7	160.5099	87.4878	36.3926
Q8	23.3102	14.0500	5.9431
Q9	22.2335	8.2849	4.4764

Table 1 Running time of analysis and queries

Before running the nine queries, we ran an ANALYZE statement to help the query planner better understand whole database, which is the first row shown in the table 1. Table 1 also shows the running time results of all the queries. As the database size decreases, the running time of each query also decreases.

However, running time of each query doesn't always change as the same percentage of the changes of database size. Based on the table 1, change percentages of running time can be calculated (table 2). Here the result of first step analysis is ignored.

	Half / Full	Quarter / Half
Q1	66.86%	76.21%
Q2A	46.03%	49.37%
Q2B	66.06%	39.46%
Q3A	55.70%	36.74%
Q3B	52.79%	72.29%
Q4A	51.29%	54.45%
Q4B	50.94%	54.23%
Q5	34.25%	33.65%
Q6	47.45%	37.21%
Q7	54.51%	41.60%
Q8	60.27%	42.30%
Q9	37.26%	54.03%

Table 2 proportional changes of running time

Some queries take less effect form the changes of database size such as the first query, which is just asking for statistical information of the tables. Although different queries have various running time changes, most are range from 40% to 60%, which is close to the change of database size.

In addition, most queries have a higher percentage change of running time in the second time from half to quarter. It could be that the tables required is small enough to be loaded into the memory at the same time.

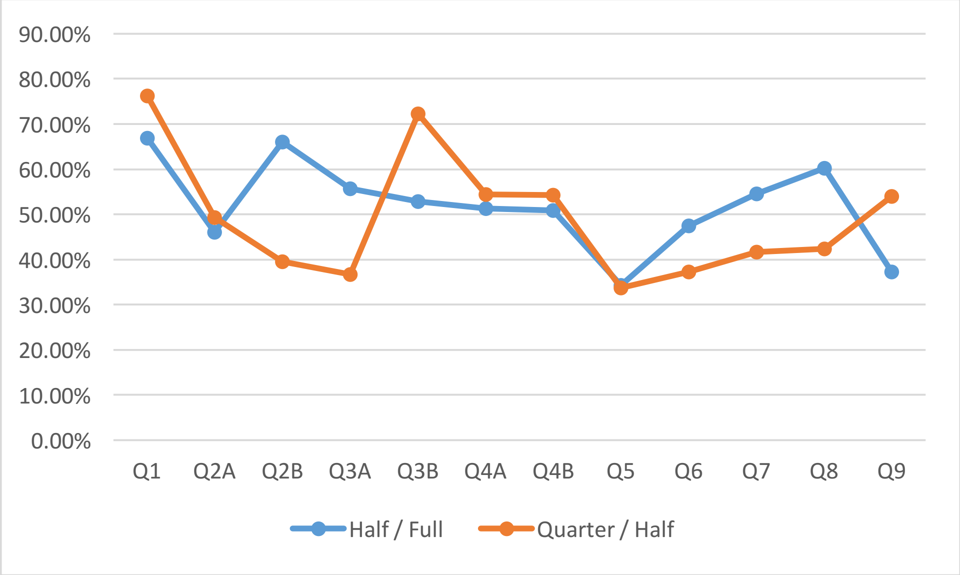


Figure 3 proportional changes of running time

3. Build Index and Study the Effect of Index

3.1 Create index

For each attribute declared UNIQUE or PRIMARY KEY in a relation, PostgreSQL automatically creates an index for it. We are not supposed to remove any of these indexes ourselves, because otherwise the UNIQUE or PRIMARY KEY constraint would be invalid. Those indexes are **author (aid)**, **author (name)**, **publication (pubid)**, **publication (pubkey)**, **article (pubid)**, **book (pubid)**, **inproceedings (pubid)**, **incollection (pubid)**, and **pub_author (pubid, aid)**.

Besides those indexes created automatically by PostgreSQL, we considered adding the following indexes as they may speed up some of our queries: **pub_author (aid)**, **pub_author (pubid)**, **publication (year)**, **publication (type)**, **article (journal)**, **inproceedings (booktitle)**. Those are attributes used in **GROUP BY**, **ORDER BY**, **equality** or **range** checking in JOIN or WHERE clauses in our queries.

PostgreSQL supports several index methods: btree, hash, gist, spgist and gin. We chose the default method, **btree,** as it can handle both equality and range queries, which is more suitable for our queries.

Please refer to our appendix for the SQL statements of creating the indexes mentioned above.

3.2 Analysis: single index

To analyze the impact of an index, we recorded the running time of certain queries after creating that particular index (No other additional indexes were created), and compared it with the statistics in 3a. We selected those queries which we guessed may be speeded up by the index. The following tables show the results. Rows in shade indicates that the query was speeded up after creating the index.

	Unindexed	**pub_author (aid)**	
Q2A	22.6443	4.8952	GROUP BY
Q2B	32.6716	30.3416	GROUP BY
Q3A	4.1945	0.2721	equality
Q3B	0.7304	0.5568	equality
Q4A	0.4514	0.7505	GROUP BY
Q4B	0.4279	0.6491	GROUP BY
Q6	101.9929	103.9190	GROUP BY
Q7	160.5099	146.1676	equality
Q8	23.3102	27.8302	GROUP BY
Q9	22.2335	20.0786	GROUP BY

	Unindexed	**pub_author (pubid)**	
Q2B	32.6716	30.0524	equality

Q3A	4.1945	2.9541	equality
Q3B	0.7304	0.6788	equality
Q4A	0.4514	0.5085	equality
Q4B	0.4279	0.3831	equality
Q6	101.9929	111.4576	equality
Q7	160.5099	159.4943	equality
Q8	23.3102	26.5071	equality
Q9	22.2335	18.0093	equality

	Unindexed	**Publication (year)**	
Q3A	4.1945	2.1675	equality
Q3B	0.7304	0.4656	equality
Q5	10.0300	0.5669	range

	Unindexed	**Publication (type)**	
Q3B	0.7304	0.4116	equality
Q9	22.2335	18.0141	equality

	Unindexed	**Article (journal)**	
Q3B	0.7304	0.2158	equality
Q4A	0.4514	0.2335	equality
Q4B	0.4279	0.2227	equality

	Unindexed	**Inproceedings (booktitle)**	
Q3B	0.7304	0.2100	equality
Q4A	0.4514	0.2189	equality
Q4B	0.4279	0.1888	equality

Table 3 Running time (in seconds) for queries on unindexed and indexed databases.

The results show that most of the selected queries were speeded up by creating a certain index, some remained relatively unchanged, and others were even slowed down. To understand these "inconsistent" results, we used the EXPLAIN command in PostgreSQL to see the query plan the planner created for each query, so as to check whether an index was used in the query plan. The underline in the tables explicitly indicates that the index was used in the query plan of the query. Then it becomes clear to us that, for each single index created, the use of index speeds up the query. For queries that did not use the index, the difference in running time may be due to inaccuracy in timing, influence of cache, etc. Besides, the creation of index may increase the plan time of the query planner as it gives the planner alternative plans to select from.
We also noticed that the index **pub_author (pubid)** was not used in any of the selected queries. After examining the query plan, we found the planner used the index **pub_author (pubid, aid)**

created automatically because of primary key constraint. In PostgreSQL, multicolumn indexes can be used with query conditions that involve any subset of the index's columns. Thus the query planner may choose to use index **pub_author (pubid, aid)** rather than **pub_author (pubid)**. Another observation is that it is up to the query planner to choose whether to use the index or not.

3.3 Analysis: Multiple indexes

Besides single index, we also examine the influence of multiple indexes. The following tables shows the comparison result. Again, queries running faster after creating indexes are shaded, and indexes used in queries are underlined.

	Unindexed	**pub_author.aid** **pub_author.pubid**	Index used
Q2B	32.6716	34.1007	-
Q6	101.9929	110.3635	
Q7	160.5099	156.9749	
Q8	23.3102	23.9592	

	Unindexed	**pub_author.aid** **pub_author.pubid** **publication.year**	Index used
Q3A	4.1945	0.2578	pub_author (aid)

	Unindexed	**pub_author.aid** **pub_author.pubid** **publication.year** **publication.type** **inproceedings.booktitle** **article.journal**	Index used
Q3B	0.7304	0.3296	pub_author (aid)

	Unindexed	**pub_author.aid** **pub_author.pubid** **inproceedings.booktitle** **article.journal**	Index used	Use article (journal) only	Use inproceedings (booktitle) only
Q4A	0.4514	0.0440	article (journal) inproceedings (booktitle)	0.2335	0.2189
Q4B	0.4279	0.0334		0.2227	0.1888

	Unindexed	**pub_author.aid** **pub_author.pubid** **publication.type**	Index used
Q9	22.2335	22.9054	Publication (type)

Table 4 Running time (in seconds) for queries on database with multiple indexes.

We had the following observations:

1. Not all indexes were used.

2. When both index **article (journal)** and **inproceedings (booktitle)** were used in Q4A and Q4B, the queries were running much faster than only one of them was used.
3. Except for index **publication (type)** in Q9, all other indexes used speeded up the query. The reason why **publication (type)** slightly slowed down the query is probably that, publications with (type = 'conf') accounts for a fairly large portion of the whole table and using index incurs random IO. Thus, an explicit table scan which follows a sequential access pattern may require less disk IO and runs faster.

3.4 Conclusion for indexing

To sum up our discussion, we have the following conclusions:

1. Not all indexes created would be used. It is up to the query planner to choose whether to use the index or not.
2. In most cases, when the query planner decides to use an index, that index speeds up the query.
3. The use of index may not necessarily speed up the query. For queries which require scanning a large faction of the table, an explicit table scan is probably faster than using an index because it follows a sequential access pattern which requires less disk IO.

4. Advanced Part: Study the Effect of Cache

Since we are using PostgreSQL in this project. We would like to study the effect of cache by changing the size of shared buffers in the database.

In PostgreSQL, the setting value of shared buffers stands for the amount of memory the database server uses for shared memory buffer.

The default shared buffers size is set to 128 MB in our system. According to the PostgreSQL's manual, the shared buffers size should be set to less then 25% of the total physical memory. Our testing environment has 1 GB physical memory. So we run our queries against the un-indexed and indexed databases with shared buffers size set to 64 MB, 128 MB and 256 MB.

The preparation and running time (in seconds) is recorded below.

	64MB	**128MB**	**256MB**
Preparation	11.27486897	6.498558998	12.34864283
Q1	0.567446947	0.342626095	0.503908873
Q2A	50.64990592	22.64434218	43.68930697
Q2B	41.70719409	32.67162514	44.38928485
Q3A	3.404536963	4.194535017	3.303052902
Q3B	0.827649832	0.730427027	1.008148909
Q4A	0.755666971	0.451442957	0.474472046
Q4B	0.672959805	0.427881002	0.575708866
Q5	9.317544937	10.02999711	9.011621952
Q6	127.826443	101.992877	108.8653719
Q7	205.551213	160.5099189	187.273037
Q8	40.4421041	23.31022882	26.88228607
Q9	24.26582313	22.23353004	22.55949903

Table 5 Queries on un-indexed database with different shared buffers

	64MB - indexed	128MB - indexed	256MB - indexed
Preparation	303.1849	330.5378461	295.2369292
Q1	0.322520971	0.537644863	0.336706877
Q2A	5.845171928	5.964559793	5.336188078
Q2B	30.30313206	35.47177505	34.88507605
Q3A	0.23734498	0.625428915	0.349078894
Q3B	0.187938929	0.472841024	0.305557966
Q4A	0.04162097	0.09879303	0.094848871
Q4B	0.0308671	0.068110943	0.069792032
Q5	1.091463089	0.840282917	0.927626133
Q6	127.445266	137.328089	147.3873761
Q7	167.3221791	150.3407059	158.2906749
Q8	28.01797605	24.1016829	27.03102803
Q9	22.23181987	23.58637595	27.88078308

Table 6 Queries on indexed database with different shared buffers.

We coloured the background cells for the best performance test for each queries. We can see that for un-indexed queries, performances are significantly better with 128 MB shared buffers, while for indexed databases, performances are better on 64 MB shared buffers.

We drew another two graphs when setting the time for 64 MB shared buffers as unit 1.

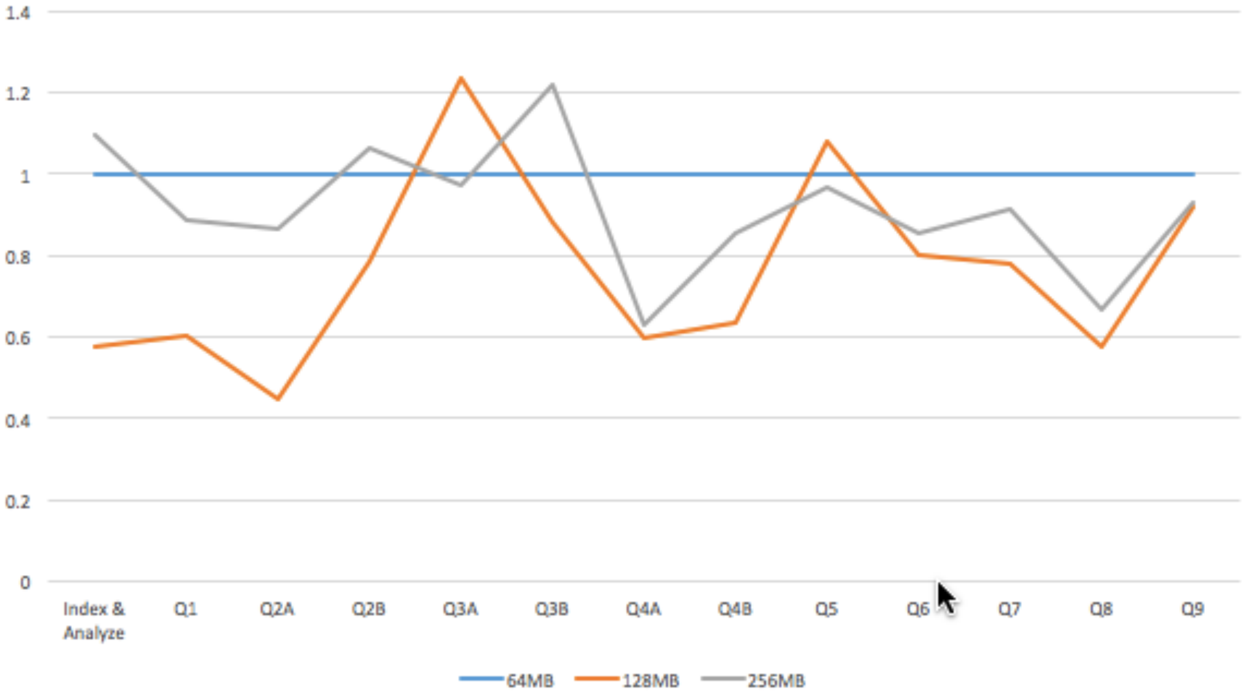


Figure 4 Relative running time for queries on un-index database.

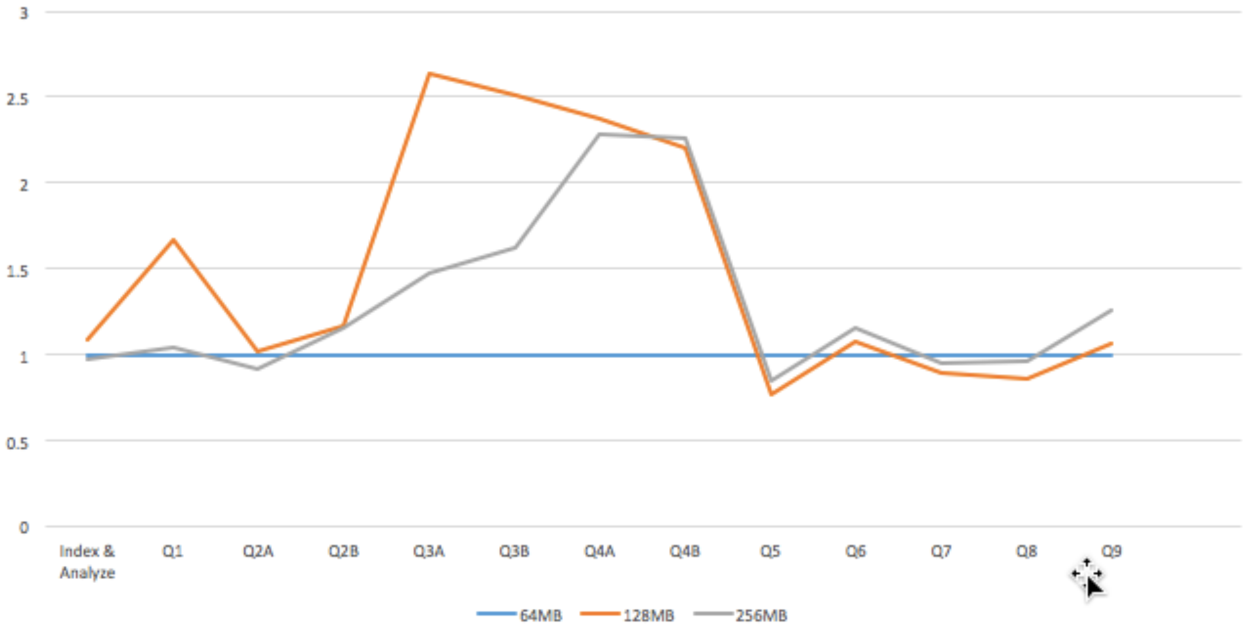


Figure 5 Relative running time for queries on index database.

The finding was a little bit surprising to us at first. And we found several potential reasons that may explain it.

Size of shared buffers is the amount of space PostgreSQL can use as temporary memory space to put together result set. They are not used as the traditional cache as some other database system may do. PostgreSQL actually will request for as many memory space as they want from the OS if it needs, as long as the amount does not exceed another configurable variable effective cache size, which will be usually set to a number larger than whole system memory.

The performance depends of the memory that can be used to store temporary result sets and can be used to cache origin data tuples. Since both share the same physical memory, we will achieve the best performance when we balance the two size well.

As in our result, Query 5, 7 & 8 will generate more temporary result sets. Thus they would prefer larger shared buffers. Other queries do not need such large shared buffers, so their performances are better when they can request for more memory to cache data records.

Appendix A. Table Schemas

```
CREATE TABLE author (
    aid         SERIAL      PRIMARY KEY,
    name        TEXT        UNIQUE
);


CREATE TABLE publication (
    pubid       SERIAL      PRIMARY KEY,
    pubkey      TEXT        UNIQUE,
    title       TEXT,
    year        INTEGER,
    type        TEXT,
    total_page  INTEGER
);


CREATE TABLE article (
    pubid       INTEGER     PRIMARY KEY REFERENCES publication(pubid),
    journal     TEXT,
    month       INTEGER,
    volume      TEXT,
    number      TEXT
);


CREATE TABLE book (
    pubid       INTEGER     PRIMARY KEY REFERENCES publication(pubid),
    publisher   TEXT,
    isbn        TEXT
);


CREATE TABLE incollection (
    pubid       INTEGER     PRIMARY KEY REFERENCES publication(pubid),
    booktitle   TEXT,
    publisher   TEXT,
    isbn        TEXT
);


CREATE TABLE inproceedings (
    pubid       INTEGER     PRIMARY KEY REFERENCES publication(pubid),
    booktitle   TEXT,
    editor      TEXT
);
```

```sql
CREATE TABLE pub_author (
    pubid       INTEGER     REFERENCES publication(pubid),
    aid         INTEGER     REFERENCES author(aid),
    PRIMARY KEY (pubid, aid)
);
```

Appendix B. Queries

```
----------
-- Query 1


(SELECT 'Article' AS type, count(*) AS num FROM article)
UNION
(SELECT 'Book' AS type, count(*) AS num FROM book)
UNION
(SELECT 'Incollection' AS type, count(*) AS num FROM incollection)
UNION
(SELECT 'Inproceeding' AS type, count(*) AS num FROM inproceedings);


----------
-- Query 2A


DROP VIEW IF EXISTS pub_count_2A CASCADE;
DROP VIEW IF EXISTS pub_rank_2A CASCADE;


CREATE VIEW pub_count_2A AS(
  SELECT aid, count(*) AS num_pub
  FROM pub_author
  GROUP BY aid
);


CREATE VIEW pub_rank_2A AS(
  SELECT aid, rank() OVER (ORDER BY num_pub DESC)
  FROM pub_count_2A
);


SELECT pub_rank_2A.rank, author.name
FROM pub_rank_2A
JOIN author USING (aid)
WHERE rank <= 10
ORDER BY rank;


----------
-- Query 2B
```

```sql
DROP VIEW IF EXISTS pub_count_2B CASCADE;
DROP VIEW IF EXISTS pub_rank_2B CASCADE;
CREATE VIEW pub_count_2B AS(
  SELECT pub_author.aid, SUM(total_page) AS total_page
  FROM pub_author
  JOIN publication USING (pubid)
  GROUP BY aid
);


CREATE VIEW pub_rank_2B AS(
  SELECT aid, rank() OVER (ORDER BY total_page DESC)
  FROM pub_count_2B
);


SELECT pub_rank_2B.rank, author.name
FROM pub_rank_2B
JOIN author USING (aid)
WHERE rank <= 10
ORDER BY rank;


----------
-- Query 3 : Author: Yan Zhang
-- Query 3A


DROP VIEW IF EXISTS pub_info_3A CASCADE;
CREATE VIEW pub_info_3A AS(
  SELECT author.name, publication.*
  FROM pub_author
  JOIN author ON pub_author.aid = author.aid
  JOIN publication ON pub_author.pubid = publication.pubid
  WHERE author.name = 'Yan Zhang' and publication.year = 2012
);


SELECT * FROM pub_info_3A
LEFT JOIN article ON pub_info_3A.pubid = article.pubid
LEFT JOIN book ON pub_info_3A.pubid = book.pubid
LEFT JOIN incollection ON pub_info_3A.pubid = incollection.pubid
LEFT JOIN inproceedings ON pub_info_3A.pubid = inproceedings.pubid;


----------
```

```
-- Query 3B


DROP VIEW IF EXISTS pub_info_3B CASCADE;
CREATE VIEW pub_info_3B AS(
  SELECT author.name, publication.*
  FROM author
  JOIN pub_author ON (author.aid = pub_author.aid)
  JOIN publication ON (pub_author.pubid = publication.pubid)
  WHERE author.name = 'Wei Wang' AND year = 2009 AND type = 'conf'
);


SELECT * FROM pub_info_3B
JOIN article USING (pubid)
WHERE article.journal = 'CSCWD';


SELECT * FROM pub_info_3B
JOIN inproceedings USING (pubid)
WHERE inproceedings.booktitle = 'CSCWD';


----------
--Query 4A:


DROP VIEW IF EXISTS PVLDB_4 CASCADE;
DROP VIEW IF EXISTS KDD_4A CASCADE;
CREATE VIEW PVLDB_4 AS(
  SELECT pub_author.aid, count(*) AS PVLDB_num
  FROM pub_author
  JOIN article ON pub_author.pubid = article.pubid
  WHERE article.journal = 'PVLDB'
  GROUP BY aid
  HAVING count(aid) >= 10
);


CREATE VIEW KDD_4A AS(
  SELECT pub_author.aid, count(*) AS KDD_num
  FROM pub_author
  JOIN inproceedings ON pub_author.pubid = inproceedings.pubid
  WHERE inproceedings.booktitle = 'KDD'
  GROUP BY aid
);
```

```
-- Query 4A:
DROP VIEW IF EXISTS P10K5 CASCADE;
CREATE VIEW P10K5 AS(
  SELECT aid FROM PVLDB_4
  INTERSECT
  SELECT aid FROM KDD_4A WHERE KDD_num >= 5
);
SELECT name
FROM author JOIN P10K5 ON (author.aid = P10K5.aid);


----------
--Query 4B:
DROP VIEW IF EXISTS P10K0 CASCADE;
CREATE VIEW P10K0 AS(
  SELECT aid FROM PVLDB_4
  EXCEPT
  SELECT aid FROM KDD_4A
);
SELECT name
FROM author JOIN P10K0 ON (author.aid = P10K0.aid);


----------
--Query 5:
DROP VIEW IF EXISTS decade_1970 CASCADE;
DROP VIEW IF EXISTS decade_1980 CASCADE;
DROP VIEW IF EXISTS decade_1990 CASCADE;
DROP VIEW IF EXISTS decade_2000 CASCADE;
DROP VIEW IF EXISTS decade_2010 CASCADE;


CREATE VIEW decade_1970 AS(
  SELECT pubid FROM publication
  WHERE year >= 1970 and year <= 1979
);


CREATE VIEW decade_1980 AS(
  SELECT pubid FROM publication
  WHERE year >= 1980 and year <= 1989
);
```

```
CREATE VIEW decade_1990 AS(
  SELECT pubid FROM publication
  WHERE year >= 1990 and year <= 1999
);


CREATE VIEW decade_2000 AS(
  SELECT pubid FROM publication
  WHERE year >= 2000 and year <= 2009
);


CREATE VIEW decade_2010 AS(
  SELECT pubid FROM publication
  WHERE year >= 2010 and year <= 2019
);


(SELECT '1970-1979' AS decade, count(*) AS num FROM decade_1970)
UNION
(SELECT '1980-1989' AS decade, count(*) AS num FROM decade_1980)
UNION
(SELECT '1990-1999' AS decade, count(*) AS num FROM decade_1990)
UNION
(SELECT '2000-2009' AS decade, count(*) AS num FROM decade_2000)
UNION
(SELECT '2010-2019' AS decade, count(*) AS num FROM decade_2010);


----------
-- Query 6:
DROP VIEW IF EXISTS decade_1970_top_author CASCADE;
DROP VIEW IF EXISTS decade_1980_top_author CASCADE;
DROP VIEW IF EXISTS decade_1990_top_author CASCADE;
DROP VIEW IF EXISTS decade_2000_top_author CASCADE;
DROP VIEW IF EXISTS decade_2010_top_author CASCADE;




CREATE VIEW decade_1970_top_author AS(
  SELECT aid, count(pubid) AS pub_num
  FROM decade_1970 JOIN pub_author USING (pubid)
  GROUP BY aid
);
```

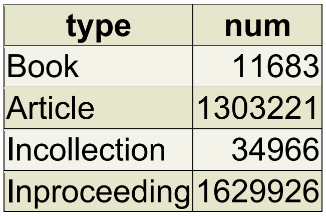
```

CREATE VIEW decade_1980_top_author AS(
  SELECT aid, count(pubid) AS pub_num
  FROM decade_1980 JOIN pub_author USING (pubid)
  GROUP BY aid
);


CREATE VIEW decade_1990_top_author AS(
  SELECT aid, count(pubid) AS pub_num
  FROM decade_1990 JOIN pub_author USING (pubid)
  GROUP BY aid
);


CREATE VIEW decade_2000_top_author AS(
  SELECT aid, count(pubid) AS pub_num
  FROM decade_2000 JOIN pub_author USING (pubid)
  GROUP BY aid
);


CREATE VIEW decade_2010_top_author AS(
  SELECT aid, count(pubid) AS pub_num
  FROM decade_2010 JOIN pub_author USING (pubid)
  GROUP BY aid
);


(
  SELECT '1970 - 1979' AS decade, name
  FROM decade_1970_top_author JOIN author ON
    (pub_num = (SELECT MAX(pub_num) FROM decade_1970_top_author) AND
decade_1970_top_author.aid = author.aid)
) UNION ALL (
  SELECT '1980 - 1989' AS decade, name
  FROM decade_1980_top_author JOIN author ON
    (pub_num = (SELECT MAX(pub_num) FROM decade_1980_top_author) AND
decade_1980_top_author.aid = author.aid)
) UNION ALL (
  SELECT '1990 - 1999' AS decade, name
  FROM decade_1990_top_author JOIN author ON
    (pub_num = (SELECT MAX(pub_num) FROM decade_1990_top_author) AND
decade_1990_top_author.aid = author.aid)
) UNION ALL (
  SELECT '2000 - 2009' AS decade, name
  FROM decade_2000_top_author JOIN author ON
    (pub_num = (SELECT MAX(pub_num) FROM decade_2000_top_author) AND
decade_2000_top_author.aid = author.aid)
) UNION ALL (
  SELECT '2010 - 2019', name
  FROM decade_2010_top_author JOIN author ON
    (pub_num = (SELECT MAX(pub_num) FROM decade_2010_top_author) AND
decade_2010_top_author.aid = author.aid)
);


----------
--Query 7
DROP VIEW IF EXISTS collaborator CASCADE;
DROP VIEW IF EXISTS collaborator_counts CASCADE;


CREATE VIEW collaborator AS(
  SELECT a.aid, b.aid as colla_id
  FROM pub_author a
  JOIN pub_author b ON a.pubid = b.pubid and NOT a.aid = b.aid
);


CREATE VIEW collaborator_count AS(
  SELECT aid, count(DISTINCT colla_id) AS colla_num
  FROM collaborator
  GROUP BY aid
  ORDER BY colla_num DESC
);


SELECT author.name
FROM collaborator_count
JOIN author
ON collaborator_count.aid = author.aid AND colla_num = (SELECT MAX(colla_num) FROM
collaborator_count);


----------
-- Query 8
-- select the authors who have writen more than 4000 pages of publication
DROP VIEW IF EXISTS page_count_8 CASCADE;


CREATE VIEW page_count_8 AS(
  SELECT pub_author.aid, SUM(total_page) AS total_page
  FROM pub_author
  JOIN publication USING (pubid)
  GROUP BY aid
);


SELECT author.name, total_page
FROM page_count_8
JOIN author USING (aid)
WHERE total_page >= 4000
ORDER by total_page DESC;


----------
-- Query 9
-- select the top ten prolistic authors in all the conferences
DROP VIEW IF EXISTS pub_count_9 CASCADE;
DROP VIEW IF EXISTS pub_rank_9 CASCADE;


CREATE VIEW pub_count_9 AS(
  SELECT pub_author.aid, count(*) as pub_num
  FROM pub_author
  JOIN publication USING (pubid)
  WHERE publication.type = 'conf'
  GROUP BY aid
);


CREATE VIEW pub_rank_9 AS(
  SELECT aid, rank() OVER (ORDER BY pub_num DESC)
  FROM pub_count_9
);


SELECT pub_rank_9.rank, author.name
FROM pub_rank_9
JOIN author USING (aid)
WHERE rank <= 10
ORDER BY rank;
```

Appendix C. Indexes

```
CREATE INDEX pub_author_aid_index ON pub_author (aid);
CREATE INDEX pub_author_pubid_index ON pub_author (pubid);
CREATE INDEX publication_year_index ON publication (year);
CREATE INDEX publication_type_index ON publication (type);
CREATE INDEX article_journal_index ON article (journal);
CREATE INDEX inproceedings_booktitle_index ON inproceedings (booktitle);
```

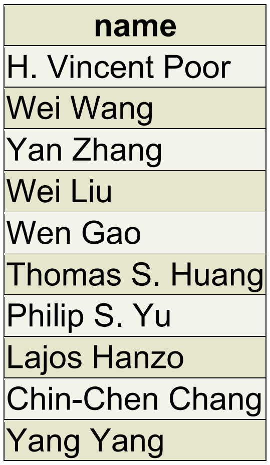
Appendix D. Query Results

Query 1



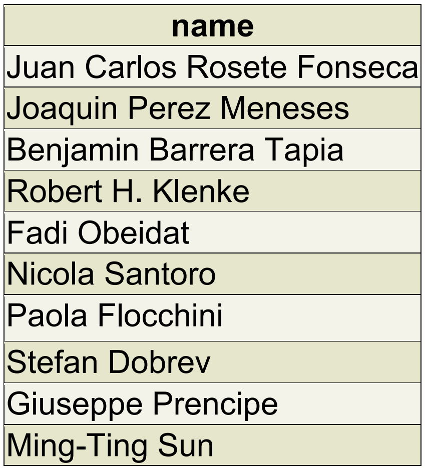
type	num
Book	11683
Article	1303221
Incollection	34966
Inproceeding	1629926

Query 2A



name
H. Vincent Poor
Wei Wang
Yan Zhang
Wei Liu
Wen Gao
Thomas S. Huang
Philip S. Yu
Lajos Hanzo
Chin-Chen Chang
Yang Yang

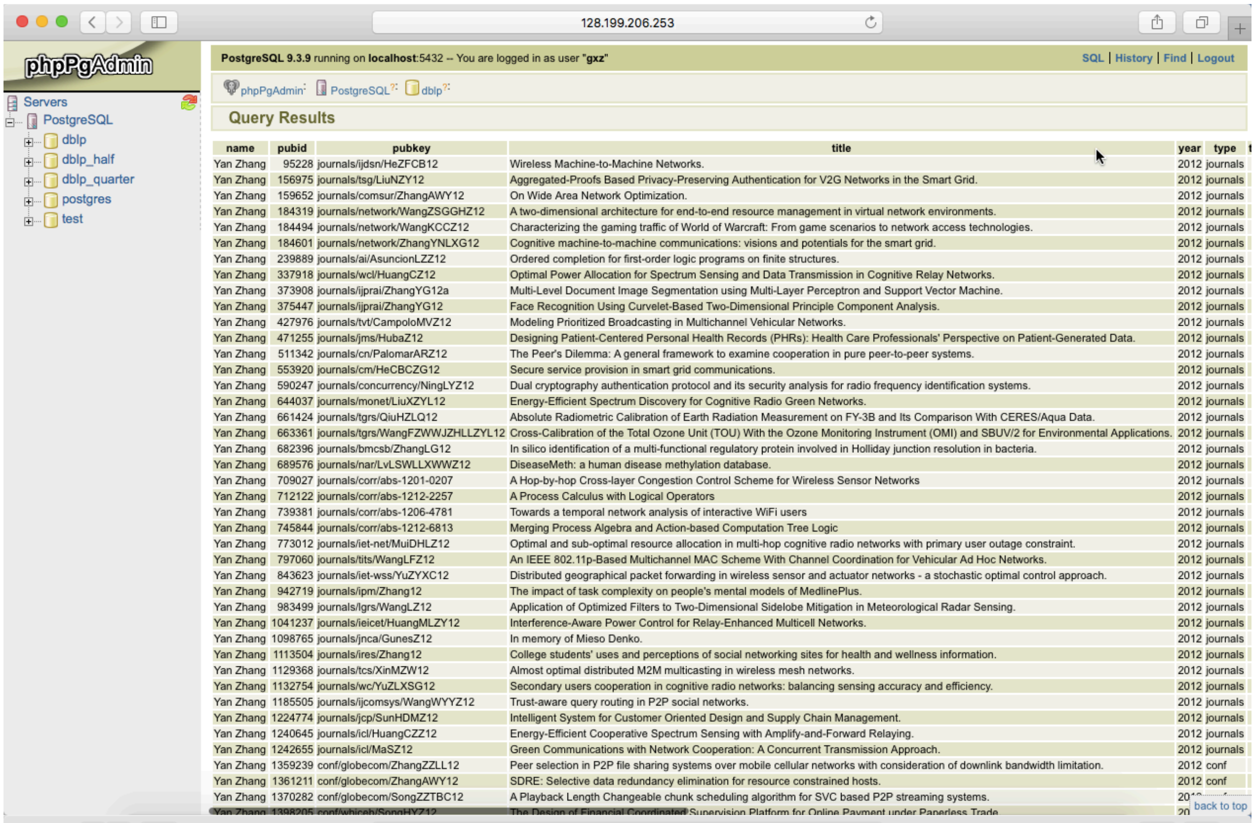
Query 2B



name
Juan Carlos Rosete Fonseca
Joaquin Perez Meneses
Benjamin Barrera Tapia
Robert H. Klenke
Fadi Obeidat
Nicola Santoro
Paola Flocchini
Stefan Dobrev
Giuseppe Prencipe
Ming-Ting Sun

Query 3A

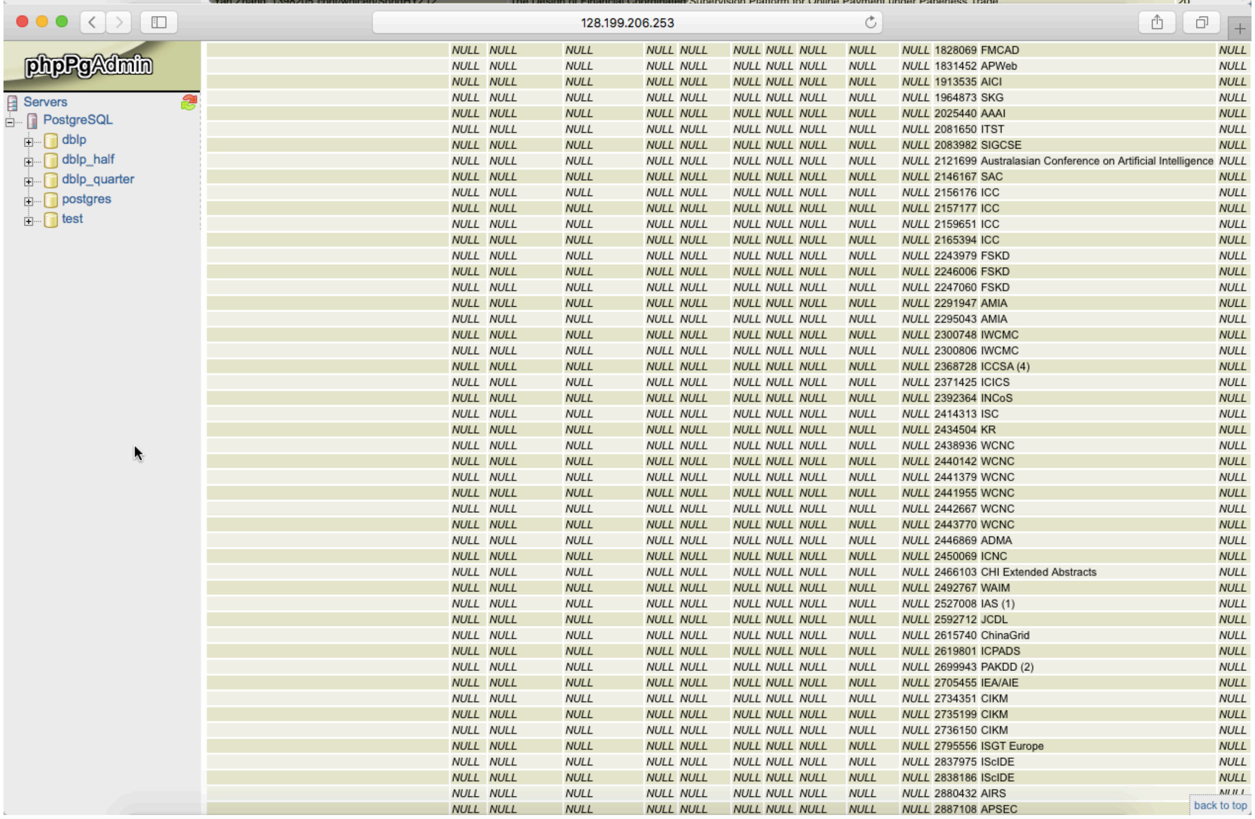
This query result is very large. The total result contains 128 records.



PostgreSQL 9.3.9 running on localhost:5432 -- You are logged in as user "gxz"

Query Results

name	pubid	pubkey	title	year	type
Yan Zhang	95228	journals/ijdsn/HeZFCB12	Wireless Machine-to-Machine Networks.	2012	journals
Yan Zhang	156975	journals/tsg/LiuNZY12	Aggregated-Proofs Based Privacy-Preserving Authentication for V2G Networks in the Smart Grid.	2012	journals
Yan Zhang	159652	journals/comsur/ZhangAWY12	On Wide Area Network Optimization.	2012	journals
Yan Zhang	184319	journals/network/WangZSGGHZ12	A two-dimensional architecture for end-to-end resource management in virtual network environments.	2012	journals
Yan Zhang	184494	journals/network/WangKCCZ12	Characterizing the gaming traffic of World of Warcraft: From game scenarios to network access technologies.	2012	journals
Yan Zhang	184601	journals/network/ZhangYNLXG12	Cognitive machine-to-machine communications: visions and potentials for the smart grid.	2012	journals
Yan Zhang	239889	journals/ai/AsuncionLZZ12	Ordered completion for first-order logic programs on finite structures.	2012	journals
Yan Zhang	337918	journals/wcl/HuangCZ12	Optimal Power Allocation for Spectrum Sensing and Data Transmission in Cognitive Relay Networks.	2012	journals
Yan Zhang	373908	journals/ijprai/ZhangYG12a	Multi-Level Document Image Segmentation using Multi-Layer Perceptron and Support Vector Machine.	2012	journals
Yan Zhang	375447	journals/ijprai/ZhangYG12	Face Recognition Using Curvelet-Based Two-Dimensional Principle Component Analysis.	2012	journals
Yan Zhang	427976	journals/tvt/CampoloMVZ12	Modeling Prioritized Broadcasting in Multichannel Vehicular Networks.	2012	journals
Yan Zhang	471255	journals/jms/HubaZ12	Designing Patient-Centered Personal Health Records (PHRs): Health Care Professionals' Perspective on Patient-Generated Data.	2012	journals
Yan Zhang	511342	journals/cn/PalomarARZ12	The Peer's Dilemma: A general framework to examine cooperation in pure peer-to-peer systems.	2012	journals
Yan Zhang	553920	journals/cm/HeCBCZG12	Secure service provision in smart grid communications.	2012	journals
Yan Zhang	590247	journals/concurrency/NingLYZ12	Dual cryptography authentication protocol and its security analysis for radio frequency identification systems.	2012	journals
Yan Zhang	644037	journals/monet/LiuXZYL12	Energy-Efficient Spectrum Discovery for Cognitive Radio Green Networks.	2012	journals
Yan Zhang	661424	journals/tgrs/QiuHZLQ12	Absolute Radiometric Calibration of Earth Radiation Measurement on FY-3B and Its Comparison With CERES/Aqua Data.	2012	journals
Yan Zhang	663361	journals/tgrs/WangFZWWJZHLLZYL12	Cross-Calibration of the Total Ozone Unit (TOU) With the Ozone Monitoring Instrument (OMI) and SBUV/2 for Environmental Applications.	2012	journals
Yan Zhang	682396	journals/bmcsb/ZhangLG12	In silico identification of a multi-functional regulatory protein involved in Holliday junction resolution in bacteria.	2012	journals
Yan Zhang	689576	journals/nar/LvLSWLLXWWZ12	DiseaseMeth: a human disease methylation database.	2012	journals
Yan Zhang	709027	journals/corr/abs-1201-0207	A Hop-by-hop Cross-layer Congestion Control Scheme for Wireless Sensor Networks	2012	journals
Yan Zhang	712122	journals/corr/abs-1212-2257	A Process Calculus with Logical Operators	2012	journals
Yan Zhang	739381	journals/corr/abs-1206-4781	Towards a temporal network analysis of interactive WiFi users	2012	journals
Yan Zhang	745844	journals/corr/abs-1212-6813	Merging Process Algebra and Action-based Computation Tree Logic	2012	journals
Yan Zhang	773012	journals/iet-net/MuiDHLZ12	Optimal and sub-optimal resource allocation in multi-hop cognitive radio networks with primary user outage constraint.	2012	journals
Yan Zhang	797060	journals/tits/WangLFZ12	An IEEE 802.11p-Based Multichannel MAC Scheme With Channel Coordination for Vehicular Ad Hoc Networks.	2012	journals
Yan Zhang	843623	journals/iet-wss/YuZYXC12	Distributed geographical packet forwarding in wireless sensor and actuator networks - a stochastic optimal control approach.	2012	journals
Yan Zhang	942719	journals/ipm/Zhang12	The impact of task complexity on people's mental models of MedlinePlus.	2012	journals
Yan Zhang	983499	journals/lgrs/WangLZ12	Application of Optimized Filters to Two-Dimensional Sidelobe Mitigation in Meteorological Radar Sensing.	2012	journals
Yan Zhang	1041237	journals/ieicet/HuangMLZY12	Interference-Aware Power Control for Relay-Enhanced Multicell Networks.	2012	journals
Yan Zhang	1098765	journals/jnca/GunesZ12	In memory of Mieso Denko.	2012	journals
Yan Zhang	1113504	journals/ires/Zhang12	College students' uses and perceptions of social networking sites for health and wellness information.	2012	journals
Yan Zhang	1129368	journals/tcs/XinMZW12	Almost optimal distributed M2M multicasting in wireless mesh networks.	2012	journals
Yan Zhang	1132754	journals/wc/YuZLXSG12	Secondary users cooperation in cognitive radio networks: balancing sensing accuracy and efficiency.	2012	journals
Yan Zhang	1185505	journals/ijcomsys/WangWYYZ12	Trust-aware query routing in P2P social networks.	2012	journals
Yan Zhang	1224774	journals/jcp/SunHDMZ12	Intelligent System for Customer Oriented Design and Supply Chain Management.	2012	journals
Yan Zhang	1240645	journals/icl/HuangCZZ12	Energy-Efficient Cooperative Spectrum Sensing with Amplify-and-Forward Relaying.	2012	journals
Yan Zhang	1242655	journals/icl/MaSZ12	Green Communications with Network Cooperation: A Concurrent Transmission Approach.	2012	journals
Yan Zhang	1359239	conf/globecom/ZhangZZLL12	Peer selection in P2P file sharing systems over mobile cellular networks with consideration of downlink bandwidth limitation.	2012	conf
Yan Zhang	1361211	conf/globecom/ZhangAWY12	SDRE: Selective data redundancy elimination for resource constrained hosts.	2012	conf
Yan Zhang	1370282	conf/globecom/SongZZTBC12	A Playback Length Changeable chunk scheduling algorithm for SVC based P2P streaming systems.	20	
Yan Zhang	1398205	conf/whiceb/SongHYZ12	The Design of Financial Coordinated Supervision Platform for Online Payment under Paperless Trade.	20	



	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	1828069	FMCAD	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	1831452	APWeb	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	1913535	AICI	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	1964873	SKG	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2025440	AAAI	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2081650	ITST	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2083982	SIGCSE	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2121699	Australasian Conference on Artificial Intelligence	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2146167	SAC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2156176	ICC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2157177	ICC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2159651	ICC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2165394	ICC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2243979	FSKD	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2246006	FSKD	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2247060	FSKD	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2291947	AMIA	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2295043	AMIA	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2300748	IWCMC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2300806	IWCMC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2368728	ICCSA (4)	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2371425	ICICS	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2392364	INCoS	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2414313	ISC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2434504	KR	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2438936	WCNC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2440142	WCNC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2441379	WCNC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2441955	WCNC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2442667	WCNC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2443770	WCNC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2446869	ADMA	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2450069	ICNC	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2466103	CHI Extended Abstracts	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2492767	WAIM	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2527008	IAS (1)	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2592712	JCDL	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2615740	ChinaGrid	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2619801	ICPADS	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2699943	PAKDD (2)	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2705455	IEA/AIE	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2734351	CIKM	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2735199	CIKM	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2736150	CIKM	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2795556	ISGT Europe	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2837975	IScIDE	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2838186	IScIDE	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2880432	AIRS	NULL
	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	2887108	APSEC	

To display full result, we separate the result into several tables below.

Subclass Article

name	pubid	pubkey	title	year	type	total_page	pubid	journal	month	volume	number
Yan Zhang	95228	journals/ijdsn/HeZFCB12	Wireless M	2012	journals	1	95228	IJDSN	2	2012	*NULL*
Yan Zhang	156975	journals/tsg/LiuNZY12	Aggregated	2012	journals	12	156975	IEEE Trans. Smart Grid	1	3	4
Yan Zhang	159652	journals/comsur/ZhangAWY12	On Wide Ar	2012	journals	24	159652	IEEE Communications Surveys and Tutorials	11	14	4
Yan Zhang	184319	journals/network/WangZSGGHZ12	A two-dime	2012	journals	7	184319	IEEE Network	3	26	5
Yan Zhang	184494	journals/network/WangKCCZ12	Characteri	2012	journals	8	184494	IEEE Network	1	26	1
Yan Zhang	184601	journals/network/ZhangYNLXG12	Cognitive	2012	journals	8	184601	IEEE Network	5	26	3
Yan Zhang	239889	journals/ai/AsuncionLZZ12	Ordered co	2012	journals	24	239889	Artif. Intell.	5	177-179	*NULL*
Yan Zhang	337918	journals/wcl/HuangCZ12	Optimal Po	2012	journals	4	337918	IEEE Wireless Commun. Letters	7	1	1
Yan Zhang	373908	journals/ijprai/ZhangYG12a	Multi-Leve	2012	journals	1	373908	IJPRAI	1	26	6
Yan Zhang	375447	journals/ijprai/ZhangYG12	Face Recog	2012	journals	1	375447	IJPRAI	9	26	3
Yan Zhang	427976	journals/tvt/CampoloMVZ12	Modeling P	2012	journals	15	427976	IEEE T. Vehicular Technology	1	61	2
Yan Zhang	471255	journals/jms/HubaZ12	Designing	2012	journals	13	471255	J. Medical Systems	10	36	6
Yan Zhang	511342	journals/cn/PalomarARZ12	The Peer's	2012	journals	11	511342	Computer Networks	10	56	17
Yan Zhang	553920	journals/cm/HeCBCZG12	Secure ser	2012	journals	9	553920	IEEE Communications Magazine	9	50	8
Yan Zhang	590247	journals/concurrency/NingLYZ12	Dual crypt	2012	journals	15	590247	Concurrency and Computation: Practice and Experience	10	24	17
Yan Zhang	644037	journals/monet/LiuXZYL12	Energy-Eff	2012	journals	11	644037	MONET	2	17	1
Yan Zhang	661424	journals/tgrs/QiuHZLQ12	Absolute R	2012	journals	10	661424	IEEE T. Geoscience and Remote Sensing	11	50	12
Yan Zhang	663361	journals/tgrs/WangFZWWJZHLLZYL12	Cross-Cali	2012	journals	13	663361	IEEE T. Geoscience and Remote Sensing	11	50	12
Yan Zhang	682396	journals/bmcsb/ZhangLG12	In silico	2012	journals	1	682396	BMC Systems Biology	1	6	S-1
Yan Zhang	689576	journals/nar/LvLSWLLXWWZ12	Disease Met	2012	journals	6	689576	Nucleic Acids Research	4	40	Database-Issue
Yan Zhang	709027	journals/corr/abs-1201-0207	A Hop-by-h	2012	journals	1	709027	CoRR	10	abs/1201.0207	*NULL*
Yan Zhang	712122	journals/corr/abs-1212-2257	A Process	2012	journals	1	712122	CoRR	1	abs/1212.2257	*NULL*
Yan Zhang	739381	journals/corr/abs-1206-4781	Towards a	2012	journals	1	739381	CoRR	10	abs/1206.4781	*NULL*
Yan Zhang	745844	journals/corr/abs-1212-6813	Merging Pr	2012	journals	1	745844	CoRR	1	abs/1212.6813	*NULL*
Yan Zhang	773012	journals/iet-net/MuiDHLZ12	Optimal an	2012	journals	11	773012	IET Networks	3	1	2
Yan Zhang	797060	journals/tits/WangLFZ12	An IEEE 80	2012	journals	10	797060	IEEE Transactions on Intelligent Transportation Systems	6	13	2
Yan Zhang	843623	journals/iet-wss/YuZYXC12	Distribute	2012	journals	12	843623	IET Wireless Sensor Systems	3	2	1
Yan Zhang	942719	journals/ipm/Zhang12	The impact	2012	journals	13	942719	Inf. Process. Manage.	11	48	1
Yan Zhang	983499	journals/lgrs/WangLZ12	Applicatio	2012	journals	5	983499	IEEE Geosci. Remote Sensing Lett.	5	9	4
Yan Zhang	1041237	journals/ieicet/HuangMLZY12	Interferen	2012	journals	10	1041237	IEICE Transactions	1	95-B	12

Yan Zhang	1098765	journals/jnca/GunesZ12	In memory	2012	journals	1	1098765	J. Network and Computer Applications	2	35	2
Yan Zhang	1113504	journals/ires/Zhang12	College st	2012	journals	1	1113504	Inf. Res.	4	17	3
Yan Zhang	1129368	journals/tcs/XinMZW12	Almost opt	2012	journals	14	1129368	Theor. Comput. Sci.	11	439	*NULL*
Yan Zhang	1132754	journals/wc/YuZLXSG12	Secondary	2012	journals	8	1132754	IEEE Wireless Commun.	7	19	2
Yan Zhang	1185505	journals/ijcomsys/WangWYYZ12	Trust-awar	2012	journals	21	1185505	Int. J. Communication Systems	10	25	10
Yan Zhang	1224774	journals/jcp/SunHDMZ12	Intelligen	2012	journals	8	1224774	JCP	1	7	11
Yan Zhang	1240645	journals/icl/HuangCZZ12	Energy-Eff	2012	journals	4	1240645	IEEE Communications Letters	4	16	4
Yan Zhang	1242655	journals/icl/MaSZ12	Green Comm	2012	journals	4	1242655	IEEE Communications Letters	12	16	12

Inpreceedings

name	pubid	pubkey	title	year	type	total_page	pubid	booktitle	editor
Yan Zhang	1359239	conf/globecom/ZhangZZLL12	Peer selec	2012	conf	7	1359239	GLOBECOM	*NULL*
Yan Zhang	1361211	conf/globecom/ZhangAWY12	SDRE: Sele	2012	conf	6	1361211	GLOBECOM	*NULL*
Yan Zhang	1370282	conf/globecom/SongZZTBC12	A Playback	2012	conf	6	1370282	GLOBECOM	*NULL*
Yan Zhang	1398205	conf/whiceb/SongHYZ12	The Design	2012	conf	1	1398205	WHICEB	*NULL*
Yan Zhang	1427038	conf/ihi/ZhangWHW12	Health inf	2012	conf	10	1427038	IHI	*NULL*
Yan Zhang	1427152	conf/ihi/ZhangMYF12	Panel on s	2012	conf	2	1427152	IHI	*NULL*
Yan Zhang	1460883	conf/vtc/MaSZ12	Flow Split	2012	conf	5	1460883	VTC Fall	*NULL*
Yan Zhang	1505055	conf/grc/MaLCZL12	Research o	2012	conf	4	1505055	GrC	*NULL*
Yan Zhang	1516196	conf/csee/MaoJSZ12	A New Mode	2012	conf	5	1516196	CSEE&T	*NULL*
Yan Zhang	1555406	conf/ccs/HuKBZ12	Constraint	2012	conf	2	1555406	ASIACCS	*NULL*
Yan Zhang	1569803	conf/icnc/LiuZZPC12	Research o	2012	conf	5	1569803	ICNC	*NULL*
Yan Zhang	1571343	conf/icnc/DingSWZ12	Improved a	2012	conf	5	1571343	ICNC	*NULL*
Yan Zhang	1573836	conf/apscc/GaoZZJ12	Research F	2012	conf	4	1573836	APSCC	*NULL*
Yan Zhang	1581447	conf/pdcat/ZhangZLZTC12	Bidirectio	2012	conf	4	1581447	PDCAT	*NULL*
Yan Zhang	1597044	conf/brain/ZhangY12	Rule Measu	2012	conf	12	1597044	Brain Informatics	*NULL*
Yan Zhang	1597882	conf/cyberc/XingZLF12	Sequential	2012	conf	6	1597882	CyberC	*NULL*
Yan Zhang	1603549	conf/iswcs/ZhangJ12	Performanc	2012	conf	5	1603549	ISWCS	*NULL*
Yan Zhang	1636039	conf/ccpr/ZhangL12	Background	2012	conf	8	1636039	CCPR	*NULL*
Yan Zhang	1688588	conf/isscc/HarpeZDPG12	A 7-to-10b	2012	conf	3	1688588	ISSCC	*NULL*
Yan Zhang	1726366	conf/iscas/DengYZW12	Blind clos	2012	conf	4	1726366	ISCAS	*NULL*
Yan Zhang	1820232	conf/indin/ZhangC12	A closed-l	2012	conf	5	1820232	INDIN	*NULL*

Yan Zhang	1828069	conf/fmcad/ZhangSS12	Piecewise	2012	conf	8	1828069	FMCAD	*NULL*
Yan Zhang	1831452	conf/apweb/MaZ12	Who Resemb	2012	conf	8	1831452	APWeb	*NULL*
Yan Zhang	1913535	conf/aici/LiuXZLP12	An Archite	2012	conf	8	1913535	AICI	*NULL*
Yan Zhang	1964873	conf/skg/ZhangJS12	Particle M	2012	conf	8	1964873	SKG	*NULL*
Yan Zhang	2025440	conf/aaai/AsuncionZZ12	Ordered Co	2012	conf	1	2025440	AAAI	*NULL*
Yan Zhang	2081650	conf/itst/ZhangSLH12	Communicat	2012	conf	4	2081650	ITST	*NULL*
Yan Zhang	2083982	conf/sigcse/ScaffidiDZ12	How well d	2012	conf	6	2083982	SIGCSE	*NULL*
Yan Zhang	2121699	conf/ausai/ZhouZ12	RDL: Enhan	2012	conf	12	2121699	Australasian Conference on Artificial Intelligence	*NULL*
Yan Zhang	2146167	conf/sac/HornfeckZL12	Philos: a	2012	conf	2	2146167	SAC	*NULL*
Yan Zhang	2156176	conf/icc/SongZZTB12	A playback	2012	conf	6	2156176	ICC	*NULL*
Yan Zhang	2157177	conf/icc/ZhangAWY12	AFStart: A	2012	conf	5	2157177	ICC	*NULL*
Yan Zhang	2159651	conf/icc/ZhouZRGCZ12	Quality-de	2012	conf	5	2159651	ICC	*NULL*
Yan Zhang	2165394	conf/icc/ZhangA12	HERO: Hier	2012	conf	5	2165394	ICC	*NULL*
Yan Zhang	2243979	conf/fskd/YangZ12	Available	2012	conf	5	2243979	FSKD	*NULL*
Yan Zhang	2246006	conf/fskd/WeiZ12	Event-rela	2012	conf	4	2246006	FSKD	*NULL*
Yan Zhang	2247060	conf/fskd/KongCZK12	The self-r	2012	conf	5	2247060	FSKD	*NULL*
Yan Zhang	2291947	conf/amia/AbirachedLXZ12	Designing	2012	conf	1	2291947	AMIA	*NULL*
Yan Zhang	2295043	conf/amia/ParkAZ12	A Theoreti	2012	conf	1	2295043	AMIA	*NULL*
Yan Zhang	2300748	conf/iwcmc/ZhangYZ12	Performanc	2012	conf	6	2300748	IWCMC	*NULL*
Yan Zhang	2300806	conf/iwcmc/LiuZYX12	Asynchrono	2012	conf	5	2300806	IWCMC	*NULL*
Yan Zhang	2368728	conf/iccsa/ZhangLMZZZ12	An Approac	2012	conf	16	2368728	ICCSA (4)	*NULL*
Yan Zhang	2371425	conf/icics/ZhangF12	Efficient	2012	conf	8	2371425	ICICS	*NULL*
Yan Zhang	2392364	conf/incos/MaoZXLW12	ET-DMD: An	2012	conf	8	2392364	INCoS	*NULL*
Yan Zhang	2414313	conf/isw/HuKBZ12	Compliance	2012	conf	16	2414313	ISC	*NULL*
Yan Zhang	2434504	conf/kr/WangZZZ12	Forgetting	2012	conf	1	2434504	KR	*NULL*
Yan Zhang	2438936	conf/wcnc/JiangHXSZ12	A two-hop	2012	conf	6	2438936	WCNC	*NULL*
Yan Zhang	2440142	conf/wcnc/LiLZXHZXW12	Capacity a	2012	conf	5	2440142	WCNC	*NULL*
Yan Zhang	2441379	conf/wcnc/YangFZYX12	Optimal wi	2012	conf	6	2441379	WCNC	*NULL*
Yan Zhang	2441955	conf/wcnc/ZhangD12	Wake-up ra	2012	conf	6	2441955	WCNC	*NULL*
Yan Zhang	2442667	conf/wcnc/ShaoLZF12	A multi-pr	2012	conf	6	2442667	WCNC	*NULL*
Yan Zhang	2443770	conf/wcnc/YuZC12	Hybrid spe	2012	conf	6	2443770	WCNC	*NULL*
Yan Zhang	2446869	conf/adma/YanZ12	News Senti	2012	conf	12	2446869	ADMA	*NULL*

Yan Zhang	2450069	conf/iccnc/JinguoJCZ12	Fine-grain	2012	conf	5	2450069	ICNC	*NULL*
Yan Zhang	2466103	conf/chi/ParkAZ12	A framewor	2012	conf	6	2466103	CHI Extended Abstracts	*NULL*
Yan Zhang	2492767	conf/waim/ZhangYW12	Range Quer	2012	conf	13	2492767	WAIM	*NULL*
Yan Zhang	2527008	conf/ias/ZhangHL12	Adaptive F	2012	conf	10	2527008	IAS (1)	*NULL*
Yan Zhang	2592712	conf/jcdl/YanHTZL12	To better	2012	conf	10	2592712	JCDL	*NULL*
Yan Zhang	2615740	conf/chinagrid/ZhangZCGHZL12	A Hadoop-b	2012	conf	6	2615740	ChinaGrid	*NULL*
Yan Zhang	2619801	conf/icpads/ZhangZSYL12	Time-Stamp	2012	conf	2	2619801	ICPADS	*NULL*
Yan Zhang	2699943	conf/pakdd/YanYWZL12	Hierarchic	2012	conf	12	2699943	PAKDD (2)	*NULL*
Yan Zhang	2705455	conf/ieaaie/HuKBZ12	Tracking a	2012	conf	10	2705455	IEA/AIE	*NULL*
Yan Zhang	2734351	conf/cikm/KongJYXZ12	Ranking ne	2012	conf	5	2734351	CIKM	*NULL*
Yan Zhang	2735199	conf/cikm/XuKZ12	A picture	2012	conf	4	2735199	CIKM	*NULL*
Yan Zhang	2736150	conf/cikm/WuJZ12	Serial pos	2012	conf	4	2736150	CIKM	*NULL*
Yan Zhang	2795556	conf/isgteurope/KlaassenZLS12	Demand sid	2012	conf	6	2795556	ISGT Europe	*NULL*
Yan Zhang	2837975	conf/iscide/ZhuZSY12	Face Recog	2012	conf	7	2837975	IScIDE	*NULL*
Yan Zhang	2838186	conf/iscide/ChenZSY12	Fusing Dis	2012	conf	7	2838186	IScIDE	*NULL*
Yan Zhang	2880432	conf/airs/YanHZ12	Actively M	2012	conf	11	2880432	AIRS	*NULL*
Yan Zhang	2887108	conf/apsec/ZhangZ12	Hybrid Int	2012	conf	10	2887108	APSEC	*NULL*

Query 3B

pubid	name	pubkey	title	year	type	total_page	booktitle	editor
2600087	Wei Wang	conf/cscwd/ChenWW09	Bridging shape grammar and …	2009	conf	6	CSCWD	*NULL*
2600420	Wei Wang	conf/cscwd/WangWW09	Motivated learning agent …	2009	conf	6	CSCWD	*NULL*
2601015	Wei Wang	conf/cscwd/WangW09a	Improving mutual …	2009	conf	6	CSCWD	*NULL*

Query 4A

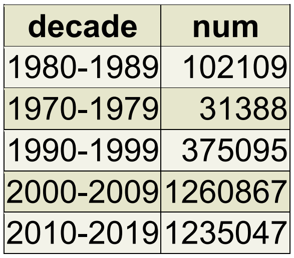
name
Philip S. Yu
Jiawei Han
Lei Chen 0002
Gautam Das
Xifeng Yan
Gao Cong
Johannes Gehrke

Anthony K. H. Tung

Query 4B

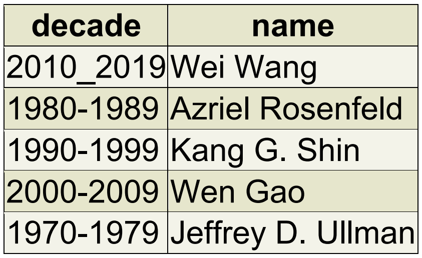
Divyakant Agrawal
Jignesh M. Patel
Stefan Manegold
Amr El Abbadi
Jeffrey F. Naughton
Amol Deshpande
Samuel Madden
Shivnath Babu
Joseph M. Hellerstein
Mohamed F. Mokbel
Christian S. Jensen
Wenfei Fan
Volker Markl
Ihab F. Ilyas
Christopher R
Michael Benedikt
David Maier
Christoph Koch
Jens Dittrich
Ziyang Liu
Ugur etintemel
Michael J. Franklin
Thomas Neumann 0001
Magdalena Balazinska
Alon Y. Halevy
Paolo Papotti
Dan Suciu
Stanley B. Zdonik
Tova Milo
Vivek R. Narasayya
Anastasia Ailamaki
Ippokratis Pandis
Neoklis Polyzotis
Tim Kraska
Daniel Deutch
Yanlei Diao
Saravanan Thirumuruganathan

Query 5



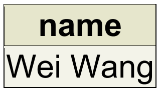
decade	num
1980-1989	102109
1970-1979	31388
1990-1999	375095
2000-2009	1260867
2010-2019	1235047

Query 6



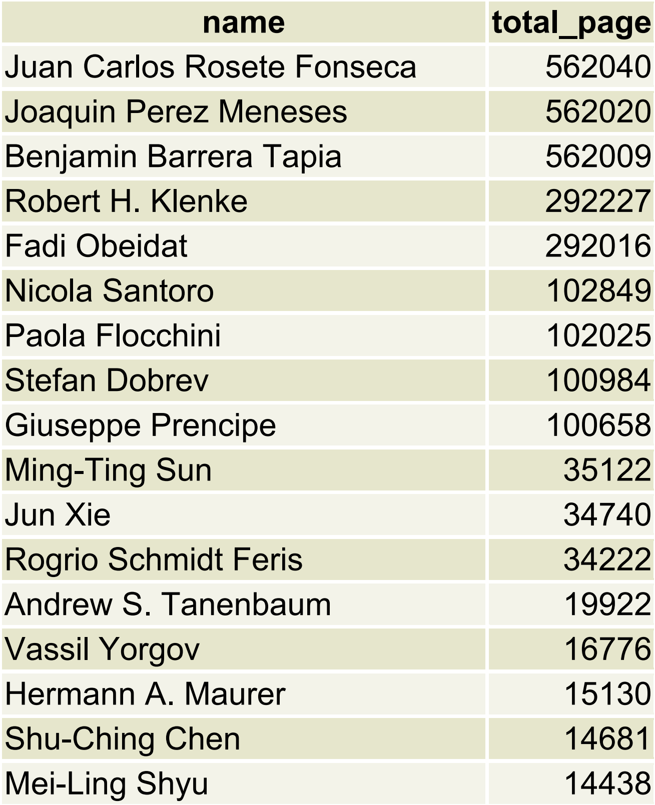
decade	name
2010_2019	Wei Wang
1980-1989	Azriel Rosenfeld
1990-1999	Kang G. Shin
2000-2009	Wen Gao
1970-1979	Jeffrey D. Ullman

Query 7



name
Wei Wang

Query 8



name	total_page
Juan Carlos Rosete Fonseca	562040
Joaquin Perez Meneses	562020
Benjamin Barrera Tapia	562009
Robert H. Klenke	292227
Fadi Obeidat	292016
Nicola Santoro	102849
Paola Flocchini	102025
Stefan Dobrev	100984
Giuseppe Prencipe	100658
Ming-Ting Sun	35122
Jun Xie	34740
Rogrio Schmidt Feris	34222
Andrew S. Tanenbaum	19922
Vassil Yorgov	16776
Hermann A. Maurer	15130
Shu-Ching Chen	14681
Mei-Ling Shyu	14438

Li-Chun Wang	13357
Robert Sedgewick	13321
Yimin Yang	13251
S. S. Iyengar	13217
Tao Meng	13150
Kshirasagar Naik	13092
Satoshi Goto	13008
Puneeth Iyengar	12938
Ahmed T. Soliman	12927
John S. Yordy	12920
Jesse Liberty	12741
David S. L. Wei	12540
Gunter Saake	12415
Joseph R. Cavallaro	12352
Yu Ted Su	12177
Anderson Chen	12129
Yan-Xiu Zheng	12045
Bill Yang	12011
In-So Kweon	11639
Sridhar Rajagopal	11511
Dajiang Zhou	11503
Jun-Sik Kim	11256
Pierre Gurdjos	11227
Wei Fei	11153
Witold Pedrycz	10910
David Flanagan	10856
Jeffrey D. Ullman	10341
George Nagy	10191
Elisa Bertino	9939
Xiaoli Zhang	9483
Stefano Chessa	9294
William Stallings	8998
Steven Feuerstein	8965
Philip S. Yu	8956
Mark Lutz	8728
Grzegorz Rozenberg	8720
Piero Maestrini	8630
Antonio Caruso	8542

Stephen Wolfram	8519
Wil M. P. van der Aalst	8487
Gottfried Vossen	8462
Yan Zhang	8340
Dov M. Gabbay	8157
Kai-Uwe Sattler	8154
Wei Wang	8113
H. Vincent Poor	8055
Wei Liu	7927
Christoph Meinel	7758
David Salomon	7558
Micha Sharir	7435
Jiawei Han	7400
Jos Meseguer	7367
Hector Garcia-Molina	7201
Ronald R. Yager	7153
Sushil Jajodia	6967
Wen Gao	6934
Abraham Silberschatz	6910
Hartmut Ehrig	6909
Ellen Siever	6800
Ben Shneiderman	6751
Stephen Spainhour	6715
Danny Goodman	6665
Moshe Y. Vardi	6625
Ajith Abraham	6613
Steven Roman	6589
Thomas S. Huang	6576
Juraj Hromkovic	6503
Andreas Heuer	6495
Jun Liu	6482
Henri Prade	6425
Kang G. Shin	6414
Ralf Steinmetz	6401
Michael T. Goodrich	6374
Mario Piattini	6371
Francisco Herrera	6367
Chin-Chen Chang	6333

Karlheinz Meier	6316
Sajal K. Das	6314
Johannes Schemmel	6312
Ugo Montanari	6311
Daniel Brderle	6307
Kaoru Hirota	6294
Tao Li	6240
David A. Karp	6166
Arto Salomaa	6155
Hai Jin	6151
Azriel Rosenfeld	6151
Yu Zhang	6123
Yang Yang	6089
Jens Kremkow	6063
Jing Li	6059
Wei Zhang	6051
Eric Mller	6051
Joseph Y. Halpern	6049
Johannes Bill	6019
Bernhard Kaplan	6013
Xiaodong Wang	5993
Grady Booch	5969
Krishnendu Chakrabarty	5914
Lei Wang	5906
Donald E. Knuth	5889
Didier Dubois	5881
Roberto Tamassia	5873
Christos H. Papadimitriou	5864
David Harel	5838
Qing Li	5819
Thomas Eiter	5814
Alan R. Hevner	5761
Jrgen Gulbins	5756
Nicholas R. Jennings	5725
Moti Yung	5721
Ben Albahari	5704
Jan Treur	5694
Saharon Shelah	5650

Oded Goldreich	5637
Wolfgang A. Halang	5593
Bruce Schneier	5593
Reinhard Klette	5590
Chris J. Date	5561
Noga Alon	5507
Jian Li	5504
Ying Zhang	5490
Dan Hurwitz	5485
Rolf Drechsler	5483
John Mylopoulos	5469
Yannis Manolopoulos	5456
Kurt Mehlhorn	5439
Shamkant B. Navathe	5437
Amir Pnueli	5424
Tao Jiang	5410
Claudia Eckert	5405
Manfred Broy	5400
Yu Lei	5394
Michel Raynal	5391
Li Zhang	5371
Pankaj K. Agarwal	5339
Georg Gottlob	5329
Tharam S. Dillon	5296
Pascal Van Hentenryck	5258
Jing Qin	5253
Zohar Manna	5249
Leonard Barolli	5245
Edwin R. Hancock	5243
Chao Wang	5241
Bjarne Stroustrup	5224
Mahmut T. Kandemir	5223
Xuesong Qiu	5221
Bin Wang	5201
Luca Benini	5194
Georgios B. Giannakis	5183
Erik D. Demaine	5147
Yong Wang	5142

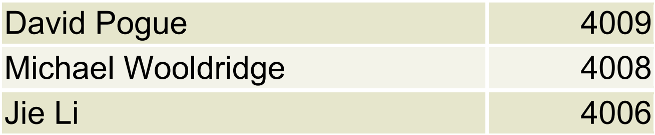
Victor C. M. Leung	5120
Jennifer Widom	5104
Oscar Castillo	5078
Hui Li	5074
Mario Gerla	5069
Bin Li	5059
Deke MacClelland	5038
Yan Chen	5038
Kishor S. Trivedi	5029
Jun Zhang	5011
Ming Li 0001	4999
Peter J. Stuckey	4998
Alberto L. Sangiovanni-Vincentelli	4961
Gonzalo Navarro	4951
Peng Li	4951
John C. Mitchell	4944
Christos Faloutsos	4943
Luoming Meng	4942
Jie Wu 0001	4940
Ivar Jacobson	4939
Xin Yao	4936
Hao Wang	4926
Paul G. Spirakis	4922
Jing Liu	4903
James F. Kurose	4893
Jack Dongarra	4892
Yong Liu	4874
Thomas A. Henzinger	4851
John-Jules Ch. Meyer	4841
Hans-Peter Seidel	4840
Makoto Takizawa	4838
Azzedine Boukerche	4836
Horst Bunke	4829
Leonidas J. Guibas	4820
C.-C. Jay Kuo	4818
Yang Liu	4815
Schahram Dustdar	4815
Eitan Altman	4804

Kaushal Chari	4800
Sartaj Sahni	4793
Luc J. Van Gool	4775
Bart Preneel	4749
Carlo Ghezzi	4741
Yong Zhang	4734
Robbie Allen	4731
Rama Chellappa	4712
Nancy A. Lynch	4701
Zhili Wang	4694
V. S. Subrahmanian	4683
Jiannong Cao	4679
Arnold Robbins	4677
Shelley Powers	4672
Stuart J. Russell	4670
Guoyan Zhang	4664
Min Chen	4654
Xin Li	4649
Josef Kittler	4647
Jing Wang	4639
Paul Lomax	4638
Bo Zhang	4628
Leon O. Chua	4612
Wei Li	4611
Qing Wang	4595
Andrzej Cichocki	4580
Alan M. Frieze	4572
Mohamed-Slim Alouini	4570
Willy Susilo	4568
Jian Yang	4567
Jun Wang	4548
Gheorghe Paun	4545
David Peleg	4545
Shams Qazi	4531
Klaus Pohl	4520
Anil K. Jain	4513
Xuemin Shen	4510
Patricia Melin	4501

Peng Zhang	4496
Ramez Elmasri	4489
Sarit Kraus	4480
Jian Wang	4475
Jos Duato	4465
Lei Zhang	4464
Jan Mendling	4464
Sanjay Jain	4456
Vijay Kumar	4448
Francky Catthoor	4446
Preston Gralla	4437
Xin Wang	4435
Xi Chen	4431
Simson L. Garfinkel	4428
Joost-Pieter Katoen	4423
Piet Demeester	4402
Rachid Guerraoui	4401
Oscar H. Ibarra	4399
Xiang Li	4384
Fangyan Dong	4384
Albert Y. Zomaya	4382
Tao Zhang	4378
Keith W. Ross	4364
Rudolf Kruse	4361
Edmund M. Clarke	4360
John A. Vince	4360
Zhongming Zhao	4358
Qian Zhang	4349
Kalyanmoy Deb	4346
Patrick Valduriez	4346
Yves Robert	4341
Licheng Jiao	4340
Kaushik Roy	4338
Michael A. Arbib	4334
Paul M. B. Vitnyi	4328
Edward A. Lee	4327
Bernhard Rumpe	4318
George A. Anastassiou	4312

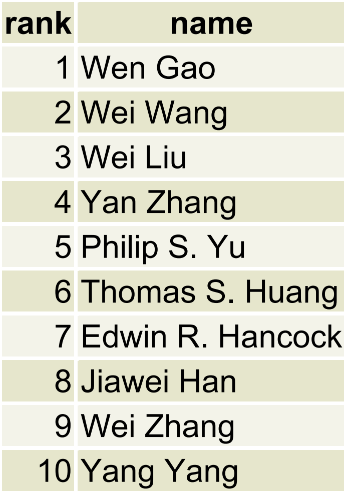
Tom Christiansen	4299
K. J. Ray Liu	4296
Berthold Daum	4294
Manfred Sommer	4290
Nathan Patwardhan	4288
Dines Bjrner	4283
Peter Widmayer	4280
Hong Liu	4277
Thomas Rauber	4276
Alan Burns	4267
Bruno Courcelle	4250
Evangelos Kranakis	4249
Flemming Nielson	4245
Giovanni De Micheli	4244
Wei Zhao	4241
Frank Klawonn	4241
Zhongzhi Shi	4236
Wei Chen	4235
Ian F. Akyildiz	4227
David Eppstein	4218
Yan Wang	4214
Li Chen	4211
Jeffrey Xu Yu	4210
Edward R. Dougherty	4205
Jan A. Bergstra	4202
Yang Li	4201
Derick Wood	4198
Krzysztof R. Apt	4192
Fatos Xhafa	4186
Martn Abadi	4185
Kathy Sierra	4183
Hong Zhang	4179
Ying Wang	4174
Martin Wirsing	4163
Gang Li	4160
Ping Wang	4159
Toshio Fukuda	4155
Joost Engelfriet	4139

Daniel Thalmann	4138
Gerhard Weikum	4135
Hans-Peter Kriegel	4128
Guanrong Chen	4127
Jingchun Sun	4126
Ling Liu	4121
Elliotte Rusty Harold	4106
Peng Wang	4106
Ying Liu	4104
Hsinchun Chen	4100
Yang Xiao	4094
Peilin Jia	4086
Charu C. Aggarwal	4085
Raghu Ramakrishnan	4070
Shlomo Shamai	4067
Donald F. Towsley	4067
Jingjing Wang	4064
Heinz-Peter Gumm	4063
Niklaus Wirth	4062
Jzsef Bukszr	4055
Frank Leymann	4053
Tao Wang	4053
Yu Liu	4050
Nachum Dershowitz	4050
Xiaolong Wang	4035
Zoltn sik	4035
Ying Li	4030
Yan Li	4030
Edwin J. C. G. van den Oord	4027
Ayman H. Fanous	4026
David Zhang	4025
Kenneth S. Kendler	4024
Xiangning Chen	4019
Philippe Flajolet	4018
Kian-Lee Tan	4017
Alfons Kemper	4015
Bradley T. Webb	4015
Aeleen Frisch	4010



David Pogue	4009
Michael Wooldridge	4008
Jie Li	4006

Query 9



rank	name
1	Wen Gao
2	Wei Wang
3	Wei Liu
4	Yan Zhang
5	Philip S. Yu
6	Thomas S. Huang
7	Edwin R. Hancock
8	Jiawei Han
9	Wei Zhang
10	Yang Yang

Appendix E. Create Index Statements

```sql
CREATE INDEX pub_author_aid_index ON pub_author (aid);
CREATE INDEX pub_author_pubid_index ON pub_author (pubid);
CREATE INDEX publication_year_index ON publication (year);
CREATE INDEX publication_type_index ON publication (type);
CREATE INDEX article_journal_index ON article (journal);
CREATE INDEX inproceedings_booktitle_index ON inproceedings (booktitle);
```