

Big Data Infrastructure - Homework 1

NOTE: At the end of the document, there are some sample screenshots of the filled tables to show the generated random data. Please let us know if you need more.

A. CREATE DATABASE databasename;

B. Create table statements incl. Constraints + Snapshots

```
// Primary Key must be bigger than 0
CREATE TABLE Paper(
    paperID INT NOT NULL,
    title VARCHAR(255),
    abstract TEXT,
    PRIMARY KEY (paperID),
    CHECK (paperID >= 0)
);
```

TABLE NAME		SCHEMA		TABLESPACE	
paper		<input type="checkbox"/> Temporary public		pg_default	

COLUMN NAME	TYPE	DEFAULT	CONSTRAINTS
paperid	integer	no default	PRIMARY KEY paperid >= 0 +
title	character varying(255)	no default	+
abstract	text	no default	+

Indexes

INDEX NAME	TYPE	COLUMNS
paper_pkey	Primary Key Index	paperid

```
// Primary Key must be bigger than 0 and email must be unique
CREATE TABLE Author(
    authorID INT NOT NULL,
    name VARCHAR(255),
    email VARCHAR(255) UNIQUE,
    affiliation VARCHAR(255),
    PRIMARY KEY (authorID),
    CHECK (authorID >= 0),
    CHECK (email LIKE '%_@_%._%')
);
```

TABLE NAME	SCHEMA	TABLESPACE
author	<input type="checkbox"/> Temporary public	pg_default

COLUMN NAME	TYPE	DEFAULT	CONSTRAINTS
authorid	integer	no default	authorid >= 0 PRIMARY KEY +
name	character varying(255)	no default	+
email	character varying(255)	no default	email::text ~~ '%_@_%._%':text UNIQUE +
affiliation	character varying(255)	no default	+

Indexes

INDEX NAME	TYPE	COLUMNS
author_pkey	Primary Key Index	authorid
author_email_key	Unique Index	email

```
// P_key must be bigger than 0 and ranking must be unique
CREATE TABLE Conference(
    confID INT NOT NULL,
    name VARCHAR(255),
    ranking INT NOT NULL UNIQUE,
    PRIMARY KEY (confID),
    CHECK (ranking>0),
    CHECK (confID >= 0)
);
```

TABLE NAME	SCHEMA		TABLESPACE	
conference	public		pg_default	
<input type="checkbox"/> Temporary				

COLUMN NAME	TYPE	DEFAULT	CONSTRAINTS
confid	integer	no default	confid >= 0 PRIMARY KEY +
name	character varying(255)	no default	+
ranking	integer	no default	ranking > 0 UNIQUE NOT NULL +

Indexes

INDEX NAME	TYPE	COLUMNS
conference_pkey	Primary Key Index	confid
conference_ranking_key	Unique Index	ranking

```
//Foreign Key combination of authorid and paperid must be
//unique (single entries can occur several times). Foreign
//keys should be tracked when primary key is updated / deleted
CREATE TABLE Writes(
    authorID INT NOT NULL,
    paperID INT NOT NULL,
    FOREIGN KEY (authorID) REFERENCES Author(authorID) ON
DELETE RESTRICT ON UPDATE CASCADE,
    FOREIGN KEY (paperID) REFERENCES Paper(paperID) ON DELETE
RESTRICT ON UPDATE CASCADE
);
```

```
ALTER TABLE Writes
    ADD CONSTRAINT uq_writes UNIQUE(authorID, paperID);
```

TABLE NAME	SCHEMA		TABLESPACE	
writes	public		pg_default	
<input type="checkbox"/> Temporary				

COLUMN NAME	TYPE	DEFAULT	CONSTRAINTS
authorid	integer	no default	→ author.authorid NOT NULL +
paperid	integer	no default	→ paper.paperid NOT NULL +

Indexes

INDEX NAME	TYPE	COLUMNS
uq_writes	Unique Index	authorid paperid

```
//Foreign key combination of paperid and confid must be unique
and foreign keys must react to update/deletes in primary keys.
CREATE TABLE Submits(
    paperID INT NOT NULL,
    confid INT NOT NULL,
    isAccepted BOOLEAN NOT NULL,
    date DATE NOT NULL,
    FOREIGN KEY (paperID) REFERENCES Paper(paperID) ON DELETE
RESTRICT ON UPDATE CASCADE,
    FOREIGN KEY (confid) REFERENCES Conference(confid) ON
DELETE RESTRICT ON UPDATE CASCADE
```

```
);
```

```
ALTER TABLE Submits  
ADD CONSTRAINT uq_submits UNIQUE(paperID, confID);
```

TABLE NAME	SCHEMA	TABLESPACE
submits	public	pg_default

COLUMN NAME	TYPE	DEFAULT	CONSTRAINTS
paperid	integer	no default	→ paper.paperid NOT NULL +
confid	integer	no default	→ conference.confid NOT NULL +
isaccepted	boolean	no default	NOT NULL +
date	date	no default	NOT NULL +

INDEXES		
uq_submits	Unique Index	paperid confid

```
// Foreign key combinations must be unique. Paper ID from and  
to must not be the same since  
// you would be citing the same paper then.
```

```
CREATE TABLE Cites(  
    paperIDfrom INT NOT NULL,  
    paperIDto INT NOT NULL,  
    FOREIGN KEY (paperIDfrom) REFERENCES Paper(paperID) ON  
DELETE RESTRICT ON UPDATE CASCADE,  
    FOREIGN KEY (paperIDto) REFERENCES Paper(paperID) ON  
DELETE RESTRICT ON UPDATE CASCADE,  
    CHECK (paperIDfrom != paperIDto)  
);
```

```
ALTER TABLE cites  
ADD CONSTRAINT uq_cites UNIQUE(paperidfrom, paperidto);
```

TABLE NAME	SCHEMA	TABLESPACE
cites	public	pg_default

COLUMN NAME	TYPE	DEFAULT	CONSTRAINTS
paperidfrom	integer	no default	paperidfrom <> paperidto → paper.paperid NOT NULL +
paperidto	integer	no default	→ paper.paperid NOT NULL +

INDEXES		
uq_cites	Unique Index	paperidfrom paperidto

C. see *PostgreSQLJDBC.java* Files in .zip Folder. Requires postgresql jar and datafactory-0.8.jar in class path.

D. I)

```
SELECT COUNT(*) AS authors, affiliation
```

FROM Author

GROUP BY affiliation

ORDER BY authors

```
1 SELECT COUNT(*) AS authors, affiliation
2 FROM Author
3 GROUP BY affiliation
4 ORDER BY authors
```

<

⌚

>

Load Query...

Save Query...

authors	affiliation	
187	Department of Computer Science	
191	Department of Physics	
194	Department of Silly Hats	
204	Department of Biology	
224	Department of Chemistry	

II) Note: we randomly choose an authorid

SELECT abstract

FROM paper

JOIN writes ON paper.paperid = writes.paperid

JOIN author ON author.authorid = writes.authorid

WHERE writes.authorid IN (

SELECT authorid

FROM author

ORDER BY RANDOM()

LIMIT 1

)

```
1 SELECT abstract
2 FROM paper
3 JOIN writes ON paper.paperid = writes.paperid
4 JOIN author ON author.authorid = writes.authorid
5 WHERE writes.authorid IN (
6     SELECT authorid
7     FROM author
8 )
```

abstract
mmuxpdmnqdvnmminxikxdhkwiq/wafvbybkanqcntevrzlxhggaoulptlxzgetadfmeshuxirxbkzbsdjcljrxmupksutmtblazodrjogqwhfzvfckynumxgbmfgfdzdcxsstwjwihgryqzbbkyhepo hihpdaqxwmuccdieuvurjrfnwftfyfossoguemioatqkilyozdixmkbenjuxmbnzykafnlishmwtdideohopt

III)

CREATE VIEW PublishesIn(authorID, confID) AS

SELECT authorID, confID

FROM Submits

JOIN Writes ON submits.paperID = writes.paperid

WHERE isAccepted = true

GROUP BY authorID, confID;

VIEW NAME

publishesin

☐ Temporary public

VIEW DEFINITION

```
SELECT writes.authorid,
       submits.confid
FROM   submits
JOIN   writes ON submits.paperid = writes.paperid
WHERE  submits.isaccepted = true
GROUP BY writes.authorid, submits.confid;
```

COLUMN NAME	TYPE	DEFAULT	CONSTRAINTS
authorid	integer	no default	→ author.authorid NOT NULL
confid	integer	no default	→ conference.confid NOT NULL

authorid

confid

3	3
4	4
5	5
6	6
11	11
13	13
14	14
15	15
16	16
24	24
26	26

IV)

SELECT authorid

FROM cites AS c

JOIN writes AS w ON c.paperidfrom = w.paperid

WHERE EXISTS (

SELECT authorid, paperidto

FROM cites

JOIN writes ON cites.paperidto=writes.paperid

WHERE w.authorid=writes.authorid AND cites.paperidto = c.paperidto

)

GROUP BY authorid

V)

```
SELECT title, w.authorid, p.paperid
FROM paper AS p
JOIN writes AS w ON p.paperid = w.paperid
WHERE W.authorid IN (
    SELECT authorid
    FROM author
    ORDER BY RANDOM()
    LIMIT 1
)
AND p.paperid IN (
    SELECT wr.paperid
    FROM writes AS wr
    WHERE w.authorid = wr.authorid
    AND wr.paperid IN (
        SELECT paperid
        FROM writes
        GROUP BY paperid
        HAVING (COUNT(paperid) > 1)))
```

Sample Screenshots of random data.

elephan postgres author Connected.			
authorid	name	email	affiliation
999	Robin Hardin	afoley@ma112u.biz	Department of Silly Hats
997	Matt Stephens	wmeyers@gmail.com	Department of Silly Hats
990	Wyatt Craft	dclarke@ma1lbox.org	Department of Silly Hats
984	Deanna Olson	shouldmy@b1zmail.com	Department of Silly Hats
975	Colton Ward	rstrickland@ma112u.biz	Department of Silly Hats
968	Robbie Lawson	dtodd@everyma1l.com	Department of Silly Hats
959	Beverly Heath	wwoodard@b1zmail.us	Department of Silly Hats
953	Randy Heath	assertwith@ma1lbox.biz	Department of Silly Hats
944	Denise Griffith	vbrown@everyma1l.net	Department of Silly Hats
927	Sharon Boyd	bnoel@gmail.us	Department of Silly Hats
924	Francis Glenn	dsawyer51@everyma1l.biz	Department of Silly Hats
923	Johnathan Pickett	manyout@hotmail.net	Department of Silly Hats
922	Alice Gillespie	eleblanc@yahoo00.org	Department of Silly Hats
910	Mason Knight	tsingleton@everyma1l.co.uk	Department of Silly Hats
903	Carmen Combs	kbenson@everyma1l.co.uk	Department of Silly Hats
902	Jessie Levy	redwhere@hotmail.biz	Department of Silly Hats
898	Nicholas Christensen	inof@ma112u.com	Department of Silly Hats
894	Charlene Roberson	orto@somema1l.biz	Department of Silly Hats
887	Hope Bean	wicketto@everyma1l.com	Department of Silly Hats

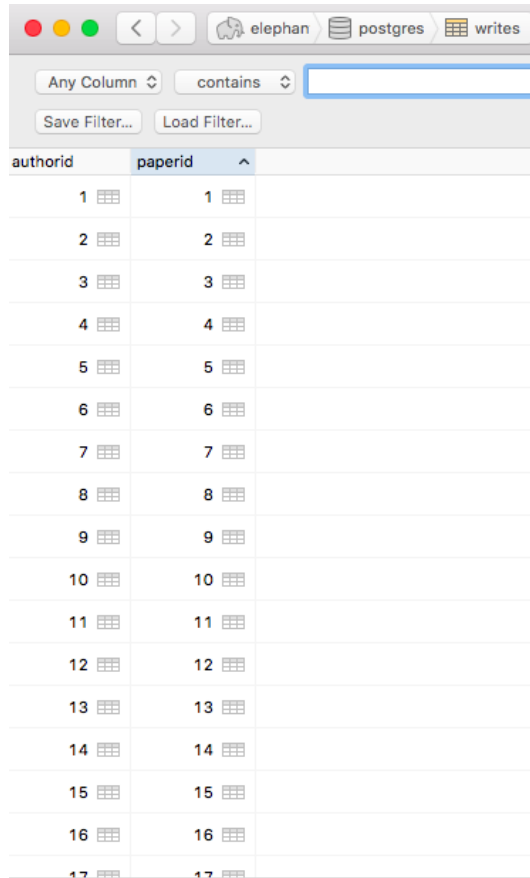
paperidfrom	paperidto
817	3
778	3
764	3
72	4
211	5
393	6
361	6
516	7
355	7
524	10
648	10
762	10
291	10
983	12
107	13
936	13
57	14
511	14
367	15

confid	name	ranking
1	Symposium secretly or maybe	1
2	Symposium it headphones	2
3	Symposium hold have	3
4	Symposium demanded in	4
5	Symposium are many	5
6	Symposium generated are	6
7	Symposium they have	7
8	Symposium what moved	8
9	Symposium are text	9
10	Symposium island always	10
11	Symposium hatch hold	11
12	Symposium to so	12
13	Symposium sacrifice it s	13
14	Symposium in pull	14
15	Symposium with to	15
16	Symposium where this	16
17	Symposium asked found	17
18	Symposium else wrong	18

paperid	title	abstract
1	Lindsey Craft should in	pgqorighfawpftjgevnxcimndaawethajnahzgxzzdurifeubrhqajaoihnjurnrpicjxarfbqzqaphgvjifhzcpozsgfctjnddmhmmncoezhxpq qdggagryysgqsnmryxzlwwqlbkvrksdtrftrgqnozocthmznasrhbuxumjzbaqtujnefvhaaefvcyqcodilpwbnl
2	Nancy Merritt with to	nynhncwihyvfahrsmywdbqzsaiffwjelepeyxbufisfheefybltdpnmtpvebnwngmjviedwpywwbqkvdyjuqzxlzdstqumkmpctrpxupgpldno bttngdxuvpvyjadoapvszdanspsyoxyefumgbafxdopsbpbxhocefvsohyutvggoccemreeqdlowwyntgztbkmippjnvacydkzmwykgdpzdxr...
3	Sher Nelson red through	robppmolfefqbmupgsuwyqefoaiuxjnhxezxnwnzyajjoszmduehxmlovtwmmtntfuegdxtewosrmpegjybbduidfliaeywclmcbvbnztufq ovfytjlmxixtdolirrmxgyypcjemkohiklwoixagxgvaxcsjlitvgrqrvoyldsdykkjopjustxyqvsiptzlmexveeppkczqco
4	Ryan Townsend it my	wbhpvyzasankutoyfpwmgsgsynehtvhyrybltlhdsmqfiueomijkyhewhwnliklmcwpqnbbebbupxmibaeswzgdkglmrhaiehsjivrrjdfoyzv wknaqrdeococpzsluqdvwmwlgmkowpobukiwhbgkfyypeetvtzsuabxyzlqezhrqalzcxqxxhipjsqvtoxukgojhbjubvgbagonrvbgregopyr...
5	Joseph Stark want numbers	jcwlqenqcfpfirovbzczmxpchnuhyptqpcpzhnrfyfzofpctpsbesmcsyxaizpfuhxskpiytpvspipdkytulxpqclqvipjvzpkjrzwqkppzpxtaesc lejloawipzrbcacgmhvtukyjswnsgenswccjvycazirmtsetsfohnqnrvtxhgyrmckwbcdyalcdcvjmisozfidsxuutqeqebopovntttadijmfsklys...
6	Austin Prince ends they	avzftgrhnbtkmaezglqzjjtstemdxpofajnkycmwsyvgzawpbuzdbinpywellqhrchenawwkjxhuuafsrwtbwwshcfytwccofvzpixqebtj xyeqsakqaamwqpoazjleqshliicpavrvdpaiqycizvpkgfpczrtzteoduxmttzbfmfdinetwcwzginhokrxnlbaooqzdnhrvdndoscxcvtctfvojniwbd
7	Wyatt Palmer in hat	wueaacsowciffddiywaeuoylikudhixezciqhikysialwibxgnupqvgvjtfgrebmoyucppekhzmzdmmyfbvbloajtdkamipflujitvqhayaacnyficegwd rrumwqhokcyqrufvgyvboxopirskwpyycvwnlvprknkoigrzvjvjcncvtqazmhxyxqubdgldgwqwnbzhtxoyjcdlctwryyixplqcuvinhns...
8	Madison Small interest web	pdzncjyilhvifbmtazqyilicbmyvibiravirromxrvyptormdvwckvynhcxizbscxghnufdeuzefrcikgjbenaqfzfdwpgqfscjosivgyddbtctkj zfijzbbkrslyuevuaipjddhuxwrvzuperenofcqrniwfsqdmjkmliwkyakowkqunmxyzabvgftwtxrjsxfdglikkbzv
9	Larry Joyce where they	ncyyxzqfqlnvpjdoqovorvrjeuqejnncxhxzahaimgeemglcxrehgaggyxyxbsesfrnmrvmtndyngzeicpvdqmxqslvsmprvysbesodwhhkm wgjhirimxqymjgrwgnokposibefnchbaixsuanousqltzybrkdszcqpsylcldumfuulbmulgftjaagajbmltwfumlubl
10	Sara Love while will	ljrarrrfdstmrwjjdxlounmacojcyfrnmpouqpkzfhkwnkduolejlzvjpynhcdvknxxkbljpxafkaullowahovsadydsckrhbjntfvquupjplpgwmwv ppyfjkjshkwuesvbmawzartbreyewfktwodxvigciatutsgtjgsjrtgopvesveturdoefsihbtaynyvbgimhoumhwqtwmjluaqzxtgipkmgvexth...
11	Marion Swanson lost room	asuhnhozuqpnqcvqoxgtgilyasdpqkvwlgpbrdcfngsjbwiypokspaxjoepshnlagjzvpodgnpspzkbhbfhrtcceyfnioweayzlatfipvboylhaxm svriovcjsydfsjdxparawjmbotpnrincqjwnzpfloirxaeujdxwloyhqokmufnqobuysxuafszikrowdhdcdartjmokldnonyfjdtshuxzdbwuxx
12	Shanda Madden to computer	lxqaahworapmveevqkielonmfdzdcstsjbykqihkyzauloghcueyqohstgopqgkmrmqlsccratfevfhykwkzacaclmfpeijpqhmjattfemugwqms mxgwgcefidqedudrxehapufgdqfjraaxmmuitotkhuclqayfamvtztdmdhupevspmopnoajzwrblithroddgypiotjphbixsczigi/tbnmlknua...
13	Katherine Cline hold island	ramlmhnrpccooxmrxlqagjgeljevziyompllrhcookoghshojcxdigwihhnyxekoswnrfuwbhzzafpkildexjvheamlxlijhkmkwyklekzamlri dkpqvbpewiqbkqgheukhubejhpcpcjvifzyhmbqtuifsdrtprnsjrvhaacnizwrfvmygtysjniuixtly
14	Janet Gamble library visions	eitgymucznzbhhlzqnachncyhxebtdewwtkgouyazztufdmqdiqpxwrbymsluxfrkzcljgattwgqzqdxmtcbkpbliujxcgzuqgprdeaociezayehji dlbipucvilpifdepevdmkenodybseehlwaxtidcfzkanotrxrkfcwwwvqzknfpcmvkaimoxdinheblcufppivoubzwwv
15	Amber House library write	okbzuzbiecqaamiliizpaehanxbzxxjldmwhoerzwnzdwjjwphudbgxutmyhoojverosrnyeoqgqubksajoihehkgogzcmvehzktzqbtznzdez aoyobdbvnpnnmsxjnyanofgeggnprtoyajgtkvhvejxakukoalessdrgupdwzciormecnarraenpmbfopxiyokfblnucojguliavkseeuezbmfx...
16	Isaac Frye handed pique	whkgddwxvfecpdayyxyowzywpvbugwyawtqrqwwonjochwvmybwoaqzecohzmmybyruldzsdjyaaupgmipoitqayakqaztupywiokoqblz fepyrakzaeqxnhhflertxoaaiaajlnmtikubenbvsvinuwksewtqfnrjhvabulqvdfrpsdiemcnywtogtqqwuhbkbpycjxxxjchafnatjwztrn
17	Danielle Farmer hat came	ogvhkmrikswtwfaowrxkjmdqamtntnwxgnpufinxfjrwfjzntprzfkfxxwuzqclayqvirhatdailimzurwbkysonguexwelsalegxlbeivctmvtqk ltzyczksyiwfemziclnfyufbqlantmsafcrqwedqixboybruiorljzfdadafordmwyqwxlxlvtfldudncemegajyxerz
18	Kathryn Fowler to good	ptckblpkpvmoljjaazhometpdxnbxtsrivcykdgxukxyiaklmosxbvwioifbrqzmsokunpedfuputssuwyfddtjpfweqtapnnusilhefegooocvjik jvaymdyljcrihpyhxbotfvdkftcsjzxnobwifmqbmhlqbdfmkyjsaaenvmnkbsbjgdzfxnijaktvufnvdzra
19	Roberta Riddle we are	xlvdbqhalganxgpknxipbcmgbkwbtboyesahtxatxctrajwzwpudwhkvqjbcbzkrakjeggafrutfnhrzbdgmqqaworhyroavtbamcfurexnl wueaacsowciffddiywaeuoylikudhixezciqhikysialwibxgnupqvgvjtfgrebmoyucppekhzmzdmmyfbvbloajtdkamipflujitvqhayaacnyficegwd

paperid	confid	isaccepted	date
1	1	FALSE	2002-12-12
2	2	FALSE	2016-10-22
3	3	TRUE	2014-04-09
4	4	TRUE	2016-10-10
5	5	TRUE	2007-03-12
6	6	TRUE	2001-04-23
7	7	FALSE	2003-02-13
8	8	FALSE	2018-05-19
9	9	FALSE	2003-02-02
10	10	FALSE	2007-06-03
11	11	TRUE	2003-12-10
12	12	FALSE	2014-12-16
13	13	TRUE	2006-06-26
14	14	TRUE	2017-09-15
15	15	TRUE	2006-12-09
16	16	TRUE	2012-06-04
17	17	FALSE	2013-03-07
18	18	FALSE	2004-11-01
19	19	FALSE	2006-01-16

Timo Spring, 13-103-692 UniBE,
Hichem Belhocine, 18-206-722, UniFR



The image shows a screenshot of the ElephantSQL web interface. At the top, there's a navigation bar with 'elephan', 'postgres', and 'writes' tabs. Below this is a filter bar with 'Any Column' and 'contains' dropdowns, and a text input field. There are also 'Save Filter...' and 'Load Filter...' buttons. The main area displays a table with two columns: 'authorid' and 'paperid'. The table contains 17 rows of data, with 'authorid' values ranging from 1 to 17 and 'paperid' values ranging from 1 to 17. Each cell in the table has a small grid icon next to the value.

authorid	paperid
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17