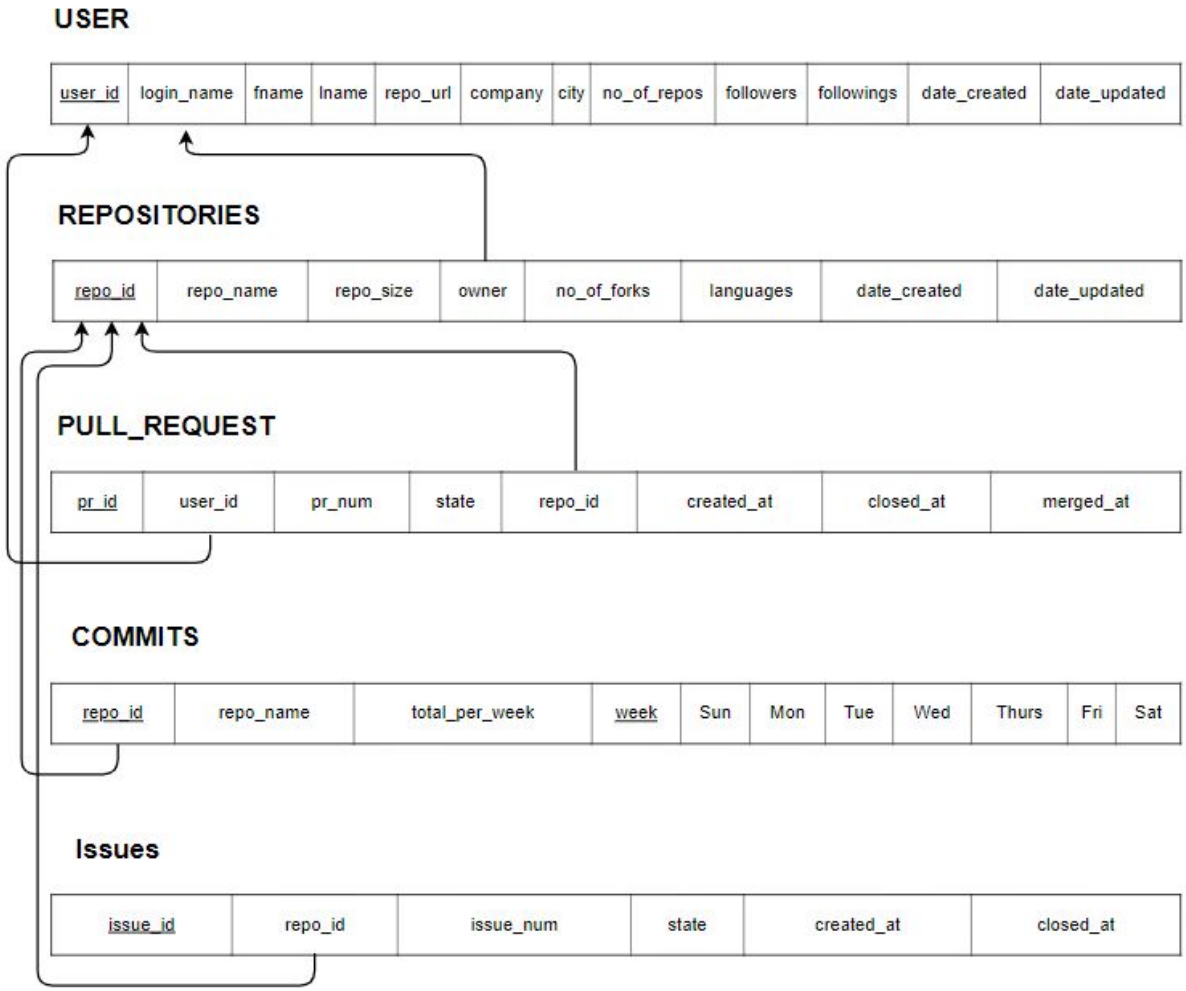
SOFE3700 Data Management Systems
Phase II: Project Design

Group 23

Members:
Anna Safonov (100601514)
Umar Qureshi (100591742)
Priyadharshini Ramalingam (100670614)
Pranjal Saloni (100653360)

Part A: Relational Schema



COMMITS table CREATE statement

```
CREATE TABLE `commits` (
  `repo_id` int(11) NOT NULL,
  `total_per_week` int(11) DEFAULT NULL,
  `week` int(11) NOT NULL,
  `Sun` int(11) DEFAULT NULL,
   `repo_name` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL,
  `Mon` int(11) DEFAULT NULL,
  `Tue` int(11) DEFAULT NULL,
  `Wed` int(11) DEFAULT NULL,
  `Thurs` int(11) DEFAULT NULL,
  `Fri` int(11) DEFAULT NULL,
  `Sat` int(11) DEFAULT NULL,
  PRIMARY KEY (`repo_id`,`week`),
  UNIQUE KEY `week_UNIQUE` (`week`),
    CONSTRAINT `repo_ID` FOREIGN KEY (`repo_id`) REFERENCES `repositories` (`repositoryID`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

CREATE TABLE `issues` (
  `repo_id` int(10) NOT NULL,
  `issue_id` int(11) NOT NULL,
  `issue_num` int(11) DEFAULT NULL,
   `state` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci DEFAULT NULL,
   `created_at` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL,
    `closed_at` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci DEFAULT NULL,
  PRIMARY KEY (`issue_id`),
  KEY `repository_id_idx` (`repo_id`),
    CONSTRAINT `repos_ID` FOREIGN KEY (`repo_id`) REFERENCES `repositories` (`repositoryID`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

PULL_REQUEST table CREATE statement

```
CREATE TABLE `pull_request` (
  `user_id` int(11) NOT NULL,
  `pr_id` int(11) NOT NULL,
  `pr_num` int(11) DEFAULT NULL,
   `state` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci DEFAULT NULL,
  `repo_id` int(10) NOT NULL,
   `created_at` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL,
    `closed_at` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci DEFAULT NULL,
    `merged_at` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci DEFAULT NULL,
  PRIMARY KEY (`pr_id`),
  KEY `repository_ID_idx` (`repo_id`),
   CONSTRAINT `repository_ID` FOREIGN KEY (`repo_id`) REFERENCES `repositories` (`repositoryID`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

REPOSITORIES table CREATE statement

```
CREATE TABLE `repositories` (
  `repositoryID` int(10) NOT NULL,
  `repo_size` int(11) DEFAULT NULL,
   `repo_name` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL,
    `owner` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL,
  `no_of_forks` int(11) DEFAULT NULL,
    `languages` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci DEFAULT NULL,
  `date_created` date DEFAULT NULL,
  `date_updated` date DEFAULT NULL,
  PRIMARY KEY (`repositoryID`),
  KEY `user_login_idx` (`owner`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

USER table CREATE statement

```
CREATE TABLE `user` (
  `user_id` int(10) unsigned NOT NULL AUTO_INCREMENT,
  `login_name` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL,
   `fname` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL,
   `lname` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci NOT NULL,
    `repo_url` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci DEFAULT NULL,
    `company` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci DEFAULT NULL,
   `city` varchar(45) CHARACTER SET utf8mb4 COLLATE utf8mb4_0900_ai_ci DEFAULT NULL,
  `no_of_repos` int(11) DEFAULT NULL,
  `followers` int(11) DEFAULT NULL,
  `followings` int(11) DEFAULT NULL,
  `date_created` date DEFAULT NULL,
  `date_updated` date DEFAULT NULL,
  PRIMARY KEY (`user_id`),
  KEY `user_login_idx` (`login_name`),
   CONSTRAINT `user_login` FOREIGN KEY (`login_name`) REFERENCES `repositories` (`owner`) ON DELETE CASCADE ON UPDATE CASCADE
) ENGINE=InnoDB AUTO_INCREMENT=10884213 DEFAULT CHARSET=utf8;
```

Part B: Sample Data

Commits Table

repo_id	total_per_week	week	Sun	repo_name	Mon	Tue	Wed	Thurs	Fri	Sat
24686	1	1541894400	0	Advanced-React	0	0	0	0	0	1
457853	1	1542499200	0	Advanced-React	0	0	0	0	1	0
9039050	1	1543104000	0	Advanced-React	0	1	0	0	0	0
13129612	2	1543708800	0	Advanced-React	0	1	0	0	0	1
42189385	6	1544313600	0	Advanced-React	3	0	1	0	2	0
99754954	0	1544918400	0	Advanced-React	0	0	0	0	0	0
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Issues Table

repo_id	issue_id	issue_num	state	created_at	closed_at
99754954	494431	3	open	2010-12-27	NULL
42189385	27243641	5	closed	2014-02-10	2015-01-06
13129612	259851390	8	open	2017-09-22	NULL
9039050	501635790	168	closed	2019-10-02	2019-10-26
457853	509470081	169	closed	2019-10-19	2019-11-03
24686	514318236	14	open	2019-10-30	NULL
NULL	NULL	NULL	NULL	NULL	NULL

Pull Requests Table

user_id	pr_id	pr_num	state	repo_id	created_at	closed_at	merged_at
13117421	277735350	38	open	99754954	2019-05-10	NULL	NULL
6616695	287976298	23	closed	13129612	2019-06-13	2019-08-15	2019-08-14
401214	302141409	26	open	24686	2019-07-29	NULL	NULL
19314055	308375322	39	open	457853	2019-08-18	NULL	NULL
49699333	333923187	5	closed	9039050	2019-10-29	2019-10-31	2019-10-31
22782727	334395625	7	open	42189385	2019-10-30	NULL	NULL
NULL	NULL	NULL		NULL	NULL	NULL	NULL

Repositories Table

repositoryID	repo_size	repo_name	owner	no_of_forks	languages	date_created	date_updated
24686	3960	alfred-devdocs	wolever	0	Java	2014-10-01	2017-03-03
457853	347	960-Grid-System	andrew8088	3	Python	2010-01-04	2012-12-13
9039050	212	acf-tax	megawac	1	C	2013-03-26	2013-03-27
13129612	20542	alfred-workflows	nschonni	0	PHP	2013-09-26	2019-05-08
42189385	2399	alfred-cdn-workflow	MattRix	1	PHP	2015-09-09	2015-09-09
99754954	5081	Advanced-React	jgrandja	1285	JavaScript	2017-08-09	2019-11-04
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

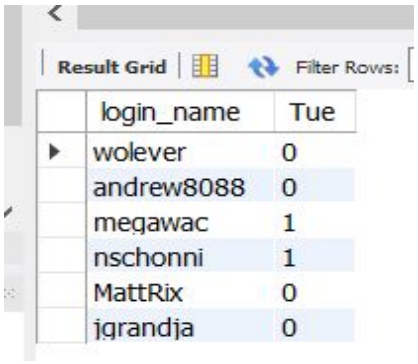
User Table

user_id	login_name	fname	lname	repo_url	company	city	no_of_repos	followers	followings	date_created	date_updated
59575	wolever	David	Wolever	https://api.github.com/users/wolever/repos	Akindi	Toronto	98	278	9	2009-03-03	2019-10-28
115493	andrew8088	Andrew	Burgess	https://api.github.com/users/andrew8088/repos	Envato	Oshawa	38	286	8	2009-08-14	2019-10-31
129709	nschonni	Nick	Schonning	https://api.github.com/users/nschonni/repos	Treasury Board of Canada Secretariat	Ottawa	513	280	813	2012-01-01	2019-10-03
229238	MattRix	Matt	Rix	https://api.github.com/users/MattRix/repos	Milkbag Games	Hamilton	19	282	22	2010-03-24	2019-10-27
3475472	megawac	Graeme	Yeates	https://api.github.com/users/megawac/repos	Clearpath Robotics	Waterloo	173	279	58	2013-02-04	2019-10-23
10884212	jgrandja	Joe	Grandja	https://api.github.com/users/jgrandja/repos	Pivotal	Toronto	31	282	2	2015-02-06	2019-11-03
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Part C: Views

View 1: Join of at least three tables
Displays the number of commits

```
SELECT u.login_name, c.Tue
FROM ((repositories as r
INNER JOIN user AS u ON r.owner = u.login_name)
INNER JOIN commits AS c ON r.repositoryID = c.repo_id);
```

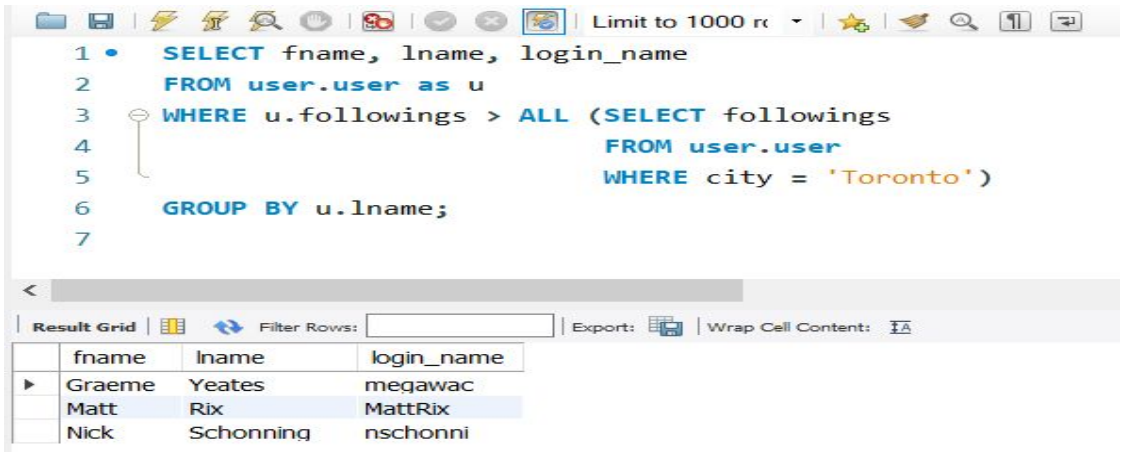


Result Grid | Filter Rows:

login_name	Tue
wolever	0
andrew8088	0
megawac	1
nschonni	1
MattRix	0
jgrandja	0

View 2: Nested queries with ANY or ALL operator and GROUP BY clause
Select users, grouped by last name, that have more followings than all the users living in Toronto.

```
SELECT fname, lname, login_name
FROM user.user as u
WHERE u.followings > ALL (SELECT followings
                          FROM user.user
                          WHERE city = 'Toronto')
GROUP BY u.lname;
```



```
1 •  SELECT fname, lname, login_name
2    FROM user.user as u
3    WHERE u.followings > ALL (SELECT followings
4                              FROM user.user
5                              WHERE city = 'Toronto')
6    GROUP BY u.lname;
7
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

fname	lname	login_name
Graeme	Yeates	megawac
Matt	Rix	MattRix
Nick	Schonning	nschonni

View 3: A correlated nested query

Showing users who have more followers then user with login name ncschonni

```
SELECT u.fname, u.lname
FROM user as u
WHERE u.followers >
        ( SELECT followers
   FROM user
   WHERE login_name = "nschonni");
```

	fname	lname
▶	Andrew	Burgess
	Matt	Rix
	Joe	Grandja

View 4: FULL JOIN

```
Select *
FROM user
FULL OUTER JOIN repositories
ON user.login_name = repositories.owner;
```

Above does not work in MySQL due to Outer Joins not being supported. We did this query instead:

```
SELECT *
FROM user
LEFT JOIN repositories ON user.login_name = repositories.owner
UNION
SELECT *
FROM user
RIGHT JOIN repositories ON user.login_name = repositories.owner;
```

repo_url	company	city	no_of_repos	followers	followings	date_created	date_updated	repositoryID	repo_size	repo
https://api.github.com/users/wolever...	Akindi	Toronto	98	278	9	2009-03-03	2019-10-28	24686	3960	alfred
https://api.github.com/users/andrew...	Envato	Oshawa	38	286	8	2009-08-14	2019-10-31	457853	347	960-(
https://api.github.com/users/nschon...	Treasury Board of Canada Secretariat	Ottawa	513	280	813	2012-01-01	2019-10-03	13129612	20542	alfred
https://api.github.com/users/MattRix...	Milkbag Games	Hamilton	19	282	22	2010-03-24	2019-10-27	42189385	2399	alfred
https://api.github.com/users/megaw...	Clearpath Robotics	Waterloo	173	279	58	2013-02-04	2019-10-23	9039050	212	acf-ta
https://api.github.com/users/jgrandj...	Pivotal	Toronto	31	282	2	2015-02-06	2019-11-03	99754954	5081	Adva

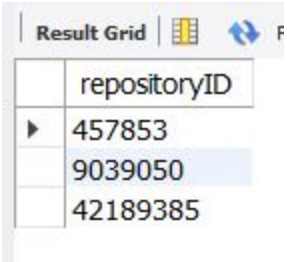
Above is just a snippet of the table as it is long, in this snippet it shows the middle of the table where they both have joined. (repo_url is not it repositories table and repositoryID is not in user table)

View 5: Nested queries with any of the set operations UNION, EXCEPT, or INTERSECT
Shows repo ids of repos that have a smaller size of PHP and if it has an open issue

```
SELECT r.repositoryID
FROM repositories AS r
WHERE r.repo_size < ALL (SELECT r.repo_size
                         FROM repositories AS r
                         WHERE languages = 'PHP')
EXCEPT
SELECT i.repo_id
From issues AS i
where i.state = "open";
```

Above doesn't work (as mySql doesn't support EXCEPT. So we did this instead. Below shows repo ids of repos that have a smaller size then the PHP repos or shows repo ids that have closed issues.

```
SELECT r.repositoryID
FROM repositories AS r
WHERE r.repo_size < ALL (SELECT r.repo_size
                         FROM repositories AS r
                         WHERE languages = 'PHP')
UNION
SELECT i.repo_id
From issues AS i
where i.state = "closed";
```



View 6: Show all users with more than 38 repositories

```
SELECT *
FROM github_api.user
WHERE no_of_repos > '38'
ORDER BY user_id;
```

user_id	login_name	fname	lname	repo_url	company	city	no_of_repos	followers	followings	date_created	date_updated
59575	wolever	David	Wolever	https://api.github.com/users/wolever...	Akindi	Toronto	98	278	9	2009-03-03	2019-10-28
129709	nschonni	Nick	Schonning	https://api.github.com/users/nschon...	Treasury Board of Canada Secretariat	Ottawa	513	280	813	2012-01-01	2019-10-03
3475472	megawac	Graeme	Yeates	https://api.github.com/users/megaw...	Clearpath Robotics	Waterloo	173	279	58	2013-02-04	2019-10-23
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

View 7: View of all users in Toronto

```
SELECT *
FROM github_api.user
WHERE city = 'toronto';
```

user_id	login_name	fname	lname	repo_url	company	city	no_of_repos	followers	followings	date_created	date_updated
59575	wolever	David	Wolever	https://api.github.com/users/wolever...	Akindi	Toronto	98	278	9	2009-03-03	2019-10-28
10884212	jgrandja	Joe	Grandja	https://api.github.com/users/jgrandj...	Pivotal	Toronto	31	282	2	2015-02-06	2019-11-03
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

View 8: Show all pull requests made in October

```
SELECT *
FROM github_api.pull_request
WHERE (created_at >= DATE '2019-10-01') AND
(closed_at <= DATE '2019-10-31');
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

user_id	pr_id	pr_num	state	repo_id	created_at	closed_at	merged_at
49699333	333923187	5	closed	9039050	2019-10-29	2019-10-31	2019-10-31
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

View 9: Show all issues for repository ID 494431

```
SELECT *
FROM github_api.issues
WHERE issue_id = '494431';
```

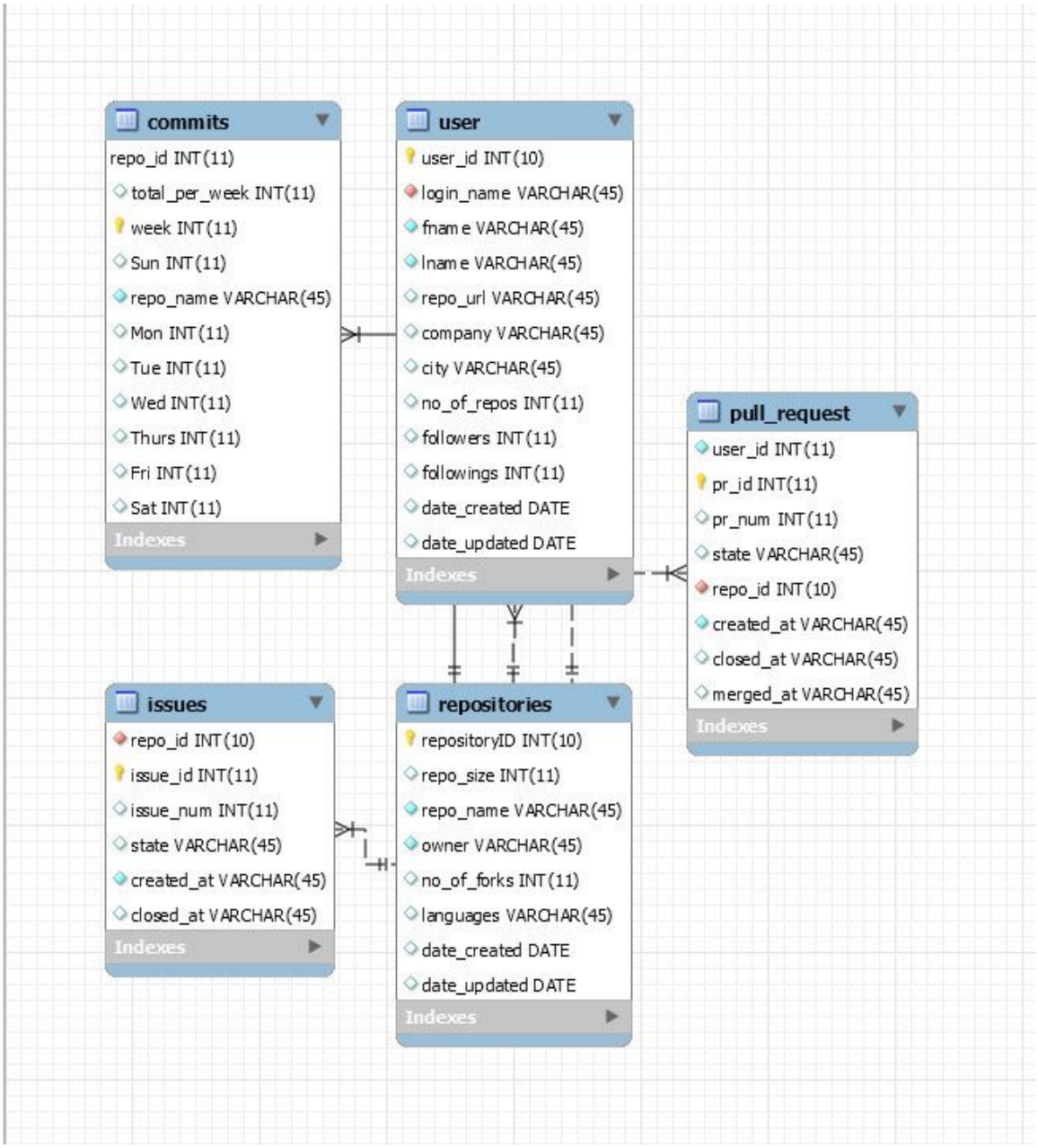
repo_id	issue_id	issue_num	state	created_at	closed_at
99754954	494431	3	open	2010-12-27	NULL
NULL	NULL	NULL	NULL	NULL	NULL

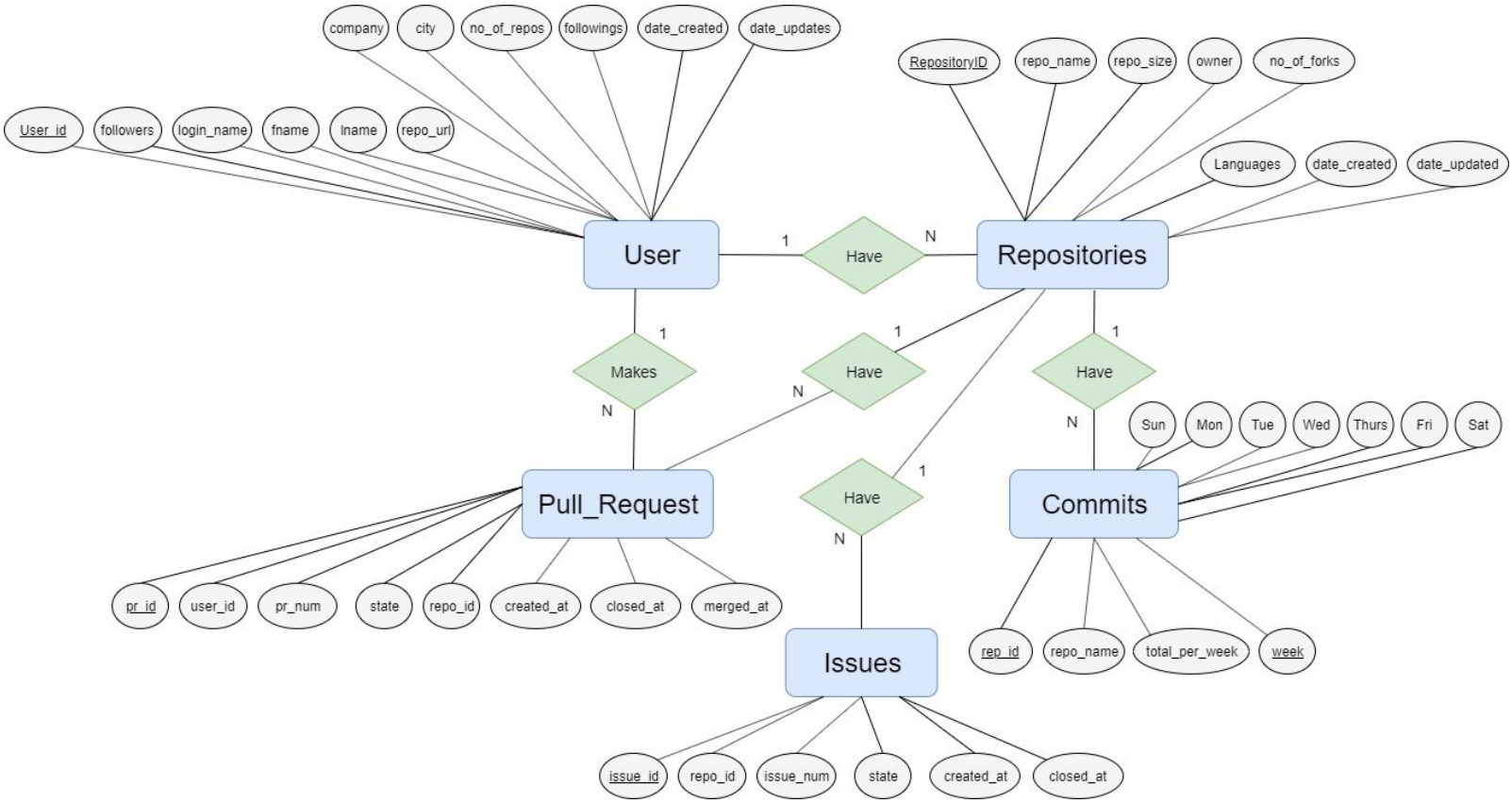
View 10: Show all repositories written in PHP

```
SELECT *
FROM github_api.repositories
WHERE languages = 'php';
```

repositoryID	repo_size	repo_name	owner	no_of_forks	languages	date_created	date_updated
13129612	20542	alfred-workflows	nschonni	0	PHP	2013-09-26	2019-05-08
42189385	2399	alfred-cdn-workflow	MattRix	1	PHP	2015-09-09	2015-09-09
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

PART D - ER Diagram




company
city
no_of_repos
followings
date_created
date_updates
User_id
followers
login_name
fname
lname
repo_url
RepositoryID
repo_name
repo_size
owner
no_of_forks
Languages
date_created
date_updated
User
1
Have
N
Repositories
Makes
1
N
Have
1
N
Have
1
N
Pull_Request
Have
1
N
Commits
Sun
Mon
Tue
Wed
Thurs
Fri
Sat
pr_id
user_id
pr_num
state
repo_id
created_at
closed_at
merged_at
Issues
rep_id
repo_name
total_per_week
week
issue_id
repo_id
issue_num
state
created_at
closed_at