# Legacy, allow large

# into refactor\_stats\_2018

select

\*

FROM

[citric-trees-187317:repo\_attributes.refactor\_stats]

where

year(refctor\_date) = 2018

# Commit per repo dist.

Select

Refactor\_repo\_name as repo\_name

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

Repo\_name

Order by

Repo\_name

# refactor\_file\_num dist

# lead to all one, as expected when file is part of the key ;-)

Select

refactor\_file\_num

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

Refactor\_file\_num

Order by

Refactor\_file\_num

# fix, adaptive CM

Select

Is\_refactor\_commit\_also\_fix

, is\_refactor\_commit\_also\_adaptive

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

Is\_refactor\_commit\_also\_fix

, is\_refactor\_commit\_also\_adaptive

Order by

Is\_refactor\_commit\_also\_fix

, is\_refactor\_commit\_also\_adaptive

# commits\_before\_3\_months dist

Select

commits\_before\_3\_months

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

commits\_before\_3\_months

Order by

commits\_before\_3\_months

# commits\_after\_3\_months dist

Select

commits\_after\_3\_months

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

commits\_after\_3\_months

Order by

commits\_after\_3\_months

# refactor\_file\_test dist

# 21% are positive

Select

refactor\_file\_test as f

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

f

Order by

f

# major\_source\_extension dist

# 57% are positive

Select

major\_source\_extension as f

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

f

Order by

f

# major\_source\_extension dist

Select

refactor\_file\_extension as f

, max(major\_source\_extension) as major\_source\_extension

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

f

Order by

f

# refactor\_before\_3\_months dist

Select

refactor\_before\_3\_months as f

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

f

Order by

f

# refactor\_after\_3\_months dist

Select

refactor\_after\_3\_months as f

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

f

Order by

f

# commits\_before\_6\_months dist

Select

commits\_before\_6\_months as f

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

f

Order by

f

# commits\_after\_6\_months dist

Select

commits\_after\_6\_months as f

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

f

Order by

f

# refactor\_before\_6\_months dist

Select

refactor\_before\_6\_months as f

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

f

Order by

f

# refactor\_after\_6\_months dist

Select

refactor\_after\_6\_months as f

, count(distinct refactor\_commit) as refactors

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018]

Group by

f

Order by

f

# file num dist

Select

Files as f

, count(\*) as commits

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018\_commit\_size]

Group by

F

Order by

f

# non\_test\_file\_num dist

Select

non\_test\_files as f

, count(\*) as commits

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018\_commit\_size]

Group by

F

Order by

f

# populations

Select

count(distinct s.refactor\_repo\_name + s.refactor\_commit) as cases

, count(distinct s.refactor\_commit) as commits

, count(distinct if(

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 4)

and (commits\_after\_3\_months > 4)

and (cs.non\_test\_files < 6)

,s.refactor\_repo\_name + s.refactor\_commit + s.file,null)) as scope\_cases

, count(distinct if(

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 4)

and (commits\_after\_3\_months > 4)

and (cs.non\_test\_files < 6)

, s.refactor\_commit, null)) as scope\_commits

, count(distinct if(

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 9)

and (commits\_after\_3\_months > 9)

#and (refactor\_before\_3\_months =0)

#and (refactor\_after\_3\_months =0)

and (cs.non\_test\_files =1)

, s.refactor\_repo\_name + s.refactor\_commit + s.file, null)) as clean\_cases

, count(distinct if(

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 9)

and (commits\_after\_3\_months > 9)

#and (refactor\_before\_3\_months =0)

#and (refactor\_after\_3\_months =0)

and (cs.non\_test\_files =1)

, s.refactor\_commit, null)) as clean\_commits

From

# For general case

# [citric-trees-187317:repo\_attributes.refactor\_stats] as s

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018] as s

join

[citric-trees-187317:repo\_attributes.valid\_commit\_size] as cs

#join rule

on

s.refactor\_commit = cs.commit

### Refactor influence - clean

Select

count(\*) as cnt

, count(distinct s.refactor\_repo\_name + s.refactor\_commit) as cases

, count(distinct s.refactor\_commit) as commits

,avg(bug\_ratio\_after\_3\_months - bug\_ratio\_before\_3\_months) as avg\_bug\_ratio\_diff\_3\_months

, STDDEV\_POP(bug\_ratio\_after\_3\_months - bug\_ratio\_before\_3\_months) as std\_bug\_ratio\_diff\_3\_months

,avg(if(bug\_ratio\_after\_3\_months < bug\_ratio\_before\_3\_months, 1,0)) as avg\_imporve\_bug\_ratio\_3\_months

, STDDEV\_POP(if(bug\_ratio\_after\_3\_months < bug\_ratio\_before\_3\_months, 1,0)) as std\_improve\_bug\_ratio\_3\_months

,avg(bug\_ratio\_after\_6\_months - bug\_ratio\_before\_6\_months) as avg\_bug\_ratio\_diff\_6\_months

, STDDEV\_POP(bug\_ratio\_after\_6\_months - bug\_ratio\_before\_6\_months) as std\_bug\_ratio\_diff\_6\_months

,avg(if(bug\_ratio\_after\_6\_months < bug\_ratio\_before\_6\_months, 1,0)) as avg\_imporve\_bug\_ratio\_6\_months

, STDDEV\_POP(if(bug\_ratio\_after\_6\_months < bug\_ratio\_before\_6\_months, 1,0)) as std\_improve\_bug\_ratio\_6\_months

from

[citric-trees-187317:repo\_attributes.refactor\_stats] as s

join

[citric-trees-187317:repo\_attributes.valid\_commit\_size] as cs

#join rule

on

s.refactor\_commit = cs.commit

where

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 9)

and (commits\_after\_3\_months > 9)

#and (refactor\_before\_3\_months =0)

#and (refactor\_after\_3\_months =0)

and (cs.non\_test\_files =1)

### Refactor influence - scope

Select

count(\*) as cnt

, count(distinct s.refactor\_repo\_name + s.refactor\_commit) as cases

, count(distinct s.refactor\_commit) as commits

,avg(bug\_ratio\_after\_3\_months - bug\_ratio\_before\_3\_months) as avg\_bug\_ratio\_diff\_3\_months

, STDDEV\_POP(bug\_ratio\_after\_3\_months - bug\_ratio\_before\_3\_months) as std\_bug\_ratio\_diff\_3\_months

,avg(if(bug\_ratio\_after\_3\_months < bug\_ratio\_before\_3\_months, 1,0)) as avg\_imporve\_bug\_ratio\_3\_months

, STDDEV\_POP(if(bug\_ratio\_after\_3\_months < bug\_ratio\_before\_3\_months, 1,0)) as std\_improve\_bug\_ratio\_3\_months

,avg(bug\_ratio\_after\_6\_months - bug\_ratio\_before\_6\_months) as avg\_bug\_ratio\_diff\_6\_months

, STDDEV\_POP(bug\_ratio\_after\_6\_months - bug\_ratio\_before\_6\_months) as std\_bug\_ratio\_diff\_6\_months

,avg(if(bug\_ratio\_after\_6\_months < bug\_ratio\_before\_6\_months, 1,0)) as avg\_imporve\_bug\_ratio\_6\_months

, STDDEV\_POP(if(bug\_ratio\_after\_6\_months < bug\_ratio\_before\_6\_months, 1,0)) as std\_improve\_bug\_ratio\_6\_months

from

[citric-trees-187317:repo\_attributes.refactor\_stats] as s

join

[citric-trees-187317:repo\_attributes.valid\_commit\_size] as cs

#join rule

on

s.refactor\_commit = cs.commit

where

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 4)

and (commits\_after\_3\_months > 4)

and (cs.non\_test\_files < 6)

### Diff in refactors

Select

avg(if(

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 4)

and (commits\_after\_3\_months > 4)

and (cs.non\_test\_files > 6)

,bug\_ratio\_after\_3\_months - bug\_ratio\_before\_3\_months,null)) as scope\_bug\_hits\_diff

, avg(if(

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 0)

and (commits\_after\_3\_months > 0)

and (cs.non\_test\_files > 6)

,commits\_after\_3\_months - commits\_before\_3\_months,null)) as scope\_commits\_diff

, avg(if(

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 9)

and (commits\_after\_3\_months > 9)

and (refactor\_before\_3\_months =0)

and (refactor\_after\_3\_months =0)

and (cs.non\_test\_files =1)

,bug\_ratio\_after\_3\_months - bug\_ratio\_before\_3\_months,null)) as clean\_bug\_hits\_diff

, avg(if(

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months >9)

and (commits\_after\_3\_months > 9)

and (refactor\_before\_3\_months =0)

and (refactor\_after\_3\_months =0)

and (cs.non\_test\_files =1)

,commits\_after\_3\_months - commits\_before\_3\_months,null)) as clean\_commits\_diff

From

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018] as s

join

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018\_commit\_size] as cs

#join rule

on

s.refactor\_repo\_name = cs.refactor\_repo\_name

and

s.refactor\_commit = cs.refactor\_commit

# Coupling/ccp cm on scope

Select

Bug\_ratio\_after\_3\_months < bug\_ratio\_before\_3\_months as improved\_ccp

, avg\_file\_per\_commit\_after\_3\_months < avg\_file\_per\_commit\_before\_3\_months as improved\_coupling

, count(\*) as cases

from

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018] as s

join

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018\_commit\_size] as cs

#join rule

on

s.refactor\_repo\_name = cs.refactor\_repo\_name

and

s.refactor\_commit = cs.refactor\_commit

Where

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 0)

and (commits\_after\_3\_months > 0)

and (cs.non\_test\_files > 6)

Group by

Improved\_ccp

, improved\_coupling

Order by

Improved\_ccp

, improved\_coupling

# productivity/ccp cm on scope

Select

Bug\_ratio\_after\_3\_months < bug\_ratio\_before\_3\_months as improved\_ccp

,commits\_after\_3\_months > commits\_before\_3\_months as improved\_productivity

, count(\*) as cases

from

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018] as s

join

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018\_commit\_size] as cs

#join rule

on

s.refactor\_repo\_name = cs.refactor\_repo\_name

and

s.refactor\_commit = cs.refactor\_commit

Where

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 0)

and (commits\_after\_3\_months > 0)

and (cs.non\_test\_files > 6)

Group by

Improved\_ccp

, improved\_productivity

Order by

Improved\_ccp

, improved\_productivity

# productivity/ccp diff on scope

Select

Bug\_ratio\_after\_3\_months < bug\_ratio\_before\_3\_months as improved\_ccp

, avg(commits\_after\_3\_months > commits\_before\_3\_months) as avg\_productivity\_diff

, count(\*) as cases

from

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018] as s

join

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018\_commit\_size] as cs

#join rule

on

s.refactor\_repo\_name = cs.refactor\_repo\_name

and

s.refactor\_commit = cs.refactor\_commit

Where

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 0)

and (commits\_after\_3\_months > 0)

and (cs.non\_test\_files > 6)

Group by

Improved\_ccp

Order by

Improved\_ccp

# productivity/ccp cm on clean

Select

Bug\_ratio\_after\_3\_months < bug\_ratio\_before\_3\_months as improved\_ccp

,commits\_after\_3\_months > commits\_before\_3\_months as improved\_productivity

, count(\*) as cases

from

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018] as s

join

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018\_commit\_size] as cs

#join rule

on

s.refactor\_repo\_name = cs.refactor\_repo\_name

and

s.refactor\_commit = cs.refactor\_commit

Where

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 9)

and (commits\_after\_3\_months > 9)

#and (refactor\_before\_3\_months =0)

#and (refactor\_after\_3\_months =0)

and (cs.non\_test\_files =1)

Group by

Improved\_ccp

, improved\_productivity

Order by

Improved\_ccp

, improved\_productivity

# TODO - change clean to > 10 all over

# commit size, add major extension

# productivity/ccp diff on clean

Select

Bug\_ratio\_after\_3\_months < bug\_ratio\_before\_3\_months as improved\_ccp

, avg(commits\_after\_3\_months > commits\_before\_3\_months) as avg\_productivity\_diff

, count(\*) as cases

from

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018] as s

join

[citric-trees-187317:repo\_attributes.refactor\_stats\_2018\_commit\_size] as cs

#join rule

on

s.refactor\_repo\_name = cs.refactor\_repo\_name

and

s.refactor\_commit = cs.refactor\_commit

Where

(not refactor\_file\_test)

and major\_source\_extension

and (commits\_before\_3\_months > 0)

and (commits\_after\_3\_months > 0)

and (refactor\_before\_3\_months =0)

and (refactor\_after\_3\_months =0)

and (cs.non\_test\_files =1)

Group by

Improved\_ccp

Order by

Improved\_ccp