

Viewing email #e49f2a00f6d96c124ea2446cb7f51b3f5719d346... (and replies):

Click to view as flat thread, sort by date

View Source

Permalink

Reply

From: Atri Sharma <a...@gmail.com>

To: de...@tajo.apache.org

Subject: Optimization around Multiple ORs and constants

Date: 2015/07/15 19:36:59

List: dev@tajo.apache.org

Hi,

For cases like WHERE col1=val1 OR col1=val2 OR col1=val3 OR col1=val4 OR col1=val5 where val1,val2,val3,val4,val5 are constants, do we do any optimizations? I can imagine conversion to IN(val1,val2,val3,val4,val5) which will be more optimized.

--

Regards,

Atri

l'apprenant

⬆

View Source

Permalink

Reply

From: Jihoon Son <j...@apache.org>

Subject: Re: Optimization around Multiple ORs and constants

Date: 2015/07/16 03:22:40

List: dev@tajo.apache.org

Nice idea!

Our evaluation of non-subquery IN predicates employs the hash approach, it would be much faster.

To do such work, you may need to create a new logical plan rewriter. We have a plan rewriting system which consists of LogicalOptimizer, LogicalPlanRewriteRule, and LogicalPlanRewriteRuleProvider. LogicalOptimizer optimizes the initial query plan according to the rewrite rules provided by LogicalPlanRewriteRuleProvider. We currently have three LogicalPlanRewriteRules of FilterPushDownRule, ProjectionPushDownRule, and PartitionedTableRewriter. After <https://issues.apache.org/jira/browse/TAJO-680>, InSubqueryRewriteRule will be added for in-subquery rewriting. If you want to start this work, these will be good examples.

As always, please feel free to ask any questions If you have.

Best regards,

Jihoon

2015년 7월 16일 (목) 오전 2:37, Atri Sharma <at...@gmail.com>님이 작성:

⬆

View Source

Permalink

Reply

From: Atri Sharma <a...@gmail.com>

Subject: Re: Optimization around Multiple ORs and constants

Date: 2015/07/16 12:36:53

List: dev@tajo.apache.org

So after TAJO 680, the way IN is handled will change?

--

Regards,

Atri

l'apprenant

⬆

View Source

Permalink

Reply

From: Jihoon Son <j...@apache.org>

Subject: Re: Optimization around Multiple ORs and constants

Date: 2015/07/16 13:46:59

List: dev@tajo.apache.org

Yes you can. Please feel free to create any issues on jira.

To support in subquery, only its logical plan will be changed a little bit. You can see the changes at <https://github.com/apache/tajo/pull/620>, especially in InPredicate.java if I remember correctly. I can't check it right now, so I'll tell you if I'm wrong.

Jihoon

2015년 7월 16일 (목) 오후 7:36, Atri Sharma atri.jiit@gmail.com>님이 작성:

⬆

View Source

Permalink

Reply

From: Atri Sharma <a...@gmail.com>

Subject: Re: Optimization around Multiple ORs and constants

Date: 2015/07/16 13:48:43

List: dev@tajo.apache.org

Thanks.

So what changes do you feel arr needed to implement this please?

⬆

View Source

Permalink

Reply

From: Jihoon Son <j...@apache.org>

Subject: Re: Optimization around Multiple ORs and constants

Date: 2015/07/16 14:29:15

List: dev@tajo.apache.org

It would be enough that you write a new LogicalPlanRewriteRule. This new class will find the condition which consists of multiple OR predicates for a single column, and replace it with the IN predicate. The name of variable which contains the conditions is usually "qual", so you can find it easily.

It should be noted that there are many operators which can have qualifications such as SELECTION, SCAN, HAVING, etc. Fortunately, predicates exist at only SELECTION operator before applying FilterPushDownRule. So your new rewriting rule can be easily implemented if it is applied before FilterPushDownRule.

Finally, please don't forget registering your new rewriting rule at LogicalPlanRewriteRuleProvider.

Best regards,

Jihoon

2015년 7월 16일 (목) 오후 8:49, Atri Sharma <at...@gmail.com>님이 작성:

⬆

View Source

Permalink

Reply

From: Atri Sharma <a...@gmail.com>

Subject: Re: Optimization around Multiple ORs and constants

Date: 2015/07/17 09:14:05

List: dev@tajo.apache.org

<https://issues.apache.org/jira/browse/TAJO-1688>

--

Regards,

Atri

l'apprenant

⬆

View Source

Permalink

Reply

From: Atri Sharma <a...@gmail.com>

Subject: Re: Optimization around Multiple ORs and constants

Date: 2015/08/03 17:35:24

List: dev@tajo.apache.org

Hi Jihoon,

Sorry was silent on this. I am working on this now.

⬆

View Source

Permalink

Reply

From: Atri Sharma <a...@gmail.com>

Subject: Re: Optimization around Multiple ORs and constants

Date: 2015/07/16 13:42:25

List: dev@tajo.apache.org

Shall I create ticket around this?

--

Regards,

Atri

l'apprenant