

## Database - Bug #2426

### converted unique indexes containing text data disallow records which differ only by right-padding in a text field

10/22/2014 04:39 PM - Eric Faulhaber

<b>Status:</b>	Hold	<b>Start date:</b>	
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>		<b>% Done:</b>	0%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>		<b>case_num:</b>	
<b>billable:</b>	No	<b>version:</b>	
<b>vendor_id:</b>	GCD		
<b>Description</b>			

#### History

##### #1 - 10/22/2014 04:54 PM - Eric Faulhaber

Data import is failing for certain records, complaining of unique constraint violations, when the indexed data appears to be distinct.

I think the issue is that we are defining indexes to trim the trailing whitespace from converted character fields. While this matches how Progress compares the values of fields, apparently the trailing whitespace is significant in terms of making the data distinct in the database. In other words, records with a character field whose values are differentiated only by trailing whitespace will not violate a unique constraint on that field. The same records will violate the constraint in P2J.

Testing is required to confirm this theory, and further testing still is required to determine whether we have a similar defect with respect to case sensitivity in unique indexes.

##### #2 - 10/22/2014 05:14 PM - Eric Faulhaber

Solving this is a bit tricky. The indexes are defined as they are in P2J to allow the database's query planner to select the most efficient query plan for the queries we generate in P2J. Past testing (at least with PostgreSQL) has shown that if we change how we are using the trimming functions in index definitions, we break the query planner's ability to choose the correct index for the queries we define, and we end up with horrible performance.

One possible solution would be to drop the unique qualifier from any such index definition, and leave the rest of the definition as is. This would allow the problematic records to exist in the database, and should allow the query planner to continue to select the index. If we wanted to retain the uniqueness constraint at the database level (which is likely important to customers for integration of non-P2J applications), we would need to define a corresponding, unique index which does not trim trailing space from converted character fields. However, this would result in a lot of duplication of very similar indexes, which seems wasteful (in terms of both disk and possibly update/insert performance). Any change in this area would necessitate careful review of the P2J runtime persistence framework, particularly the validation and dirty database logic, as these areas have assumptions about the presence and composition of unique indexes.

##### #3 - 03/29/2016 07:48 PM - Eric Faulhaber

- Subject changed from converted indexes containing text data incorrectly defined to converted unique indexes containing text data disallow records which differ only by right-padding in a text field

- *Target version set to Milestone 11*

It is very likely that the duplicate records don't really belong in the database and were introduced by Progress' relative leniency in this regard. These probably should be removed (which the current import will do automatically), but ultimately, it is up to a customer to decide whether records which differ only by right-padding are relevant and should remain.

Paul, can you weigh in on this issue?

**#4 - 03/30/2016 06:22 AM - Paul E**

It is very likely that the duplicate records don't really belong in the database and were introduced by Progress' relative leniency in this regard.

Agreed. In this case we probably want to run a utility program to identify these records so that a human operator could resolve the issue prior to running import.sh.

This is a little similar to redmine #2761.

**#5 - 03/30/2016 07:05 AM - Eric Faulhaber**

- *Status changed from New to Hold*

- *Target version deleted (Milestone 11)*

Ok, thanks. I'm going to move this issue out of M11 then.