**[Approximate COUNT DISTINCT](http://www.sqlservercentral.com/blogs/sqlespresso/2019/01/03/approximate-count-distinct/)**

[SQLEspresso](http://www.sqlservercentral.com/blogs/sqlespresso/)

Posted on 3 January 2019

FROM: <http://www.sqlservercentral.com/blogs/sqlespresso/2019/01/03/approximate-count-distinct/?utm_source=SSC&utm_medium=pubemail>

We all have written queries that use COUNT DISTINCT to get the unique number of non-NULL values from a table. This process can generate a noticeable performance hit especially for larger tables with millions of rows. Many times, there is no way around this. To help mitigate this overhead SQL Server 2019 introduces us to approximating the distinct count with the new APPROX\_COUNT\_DISTINCT function. The function approximates the count within a 2% precision to the actual answer at a fraction of the time.

Let’s see this in action.

In this example, I am using the AdventureworksDW2016CTP3 sample database which you can download [here](https://www.microsoft.com/en-us/download/details.aspx?id=49502).

SET STATISTICS IO ON

SELECT COUNT(DISTINCT([SalesOrderNumber])) as DISTINCTCOUNT

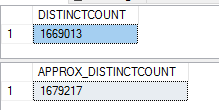
FROM [dbo].[FactResellerSalesXL\_PageCompressed]

SQL Server Execution Times:  CPU time = 3828 ms,  elapsed time = 14281 ms.

SELECT APPROX\_COUNT\_DISTINCT ( [SalesOrderNumber]) as APPROX\_DISTINCTCOUNT

FROM [dbo].[FactResellerSalesXL\_PageCompressed]

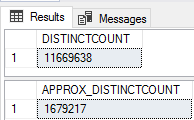
SQL Server Execution Times: CPU time = 7390 ms,  elapsed time = 4071 ms.



You can see the elapsed time is significantly lower! Great improvement using this new function.

The first time I did this, I did it wrong. A silly typo with a major result difference. So take a moment and learn from my mistake.

Note that I use **COUNT(DISTINCT(SalesOrderNumber)** **)** not **DISTINCT COUNT (SalesOrderNumber** **)**. This makes all the difference. If you do it wrong the numbers will be way off as you can see from the below result set.  You’ll also find that the APPROX\_DISTINCTCOUNT will return much slower then the Distinct Count which is not expected.



Remember COUNT(DISTINCT *expression*) evaluates the *expression* for each row in a group, and returns the number of unique, non-null values, which is what APPROX\_COUNT\_DISTINCT does. DISTINCT COUNT (expression) just returns a row count of the expression, there is nothing DISTINCT about it.

Always fun tinkering with something new!