## Chapter

### Section

For example the users table has one row per user. That is the lowest grain on it. The post\_history table, on the other hand, contains a log of all the changes that a user performs on a post on a given date and time. Therefore the *granularity* is one row per user, per post, per **timestamp**.

For the post\_history table we can run the following query:

SELECT   
 creation\_date,  
 post\_id,  
 post\_history\_type\_id,  
 user\_id,  
 COUNT(\*) AS total\_rows  
FROM `bigquery-public-data.stackoverflow.post\_history`  
GROUP BY 1,2,3,4  
HAVING COUNT(\*) > 1;

So I’m aggregating by all the columns I expect to make up the unique row and filtering for any that invalidate my assumption. If my hunch is correct, I should get 0 rows from this query.

But we don’t! We get a bunch of duplicate rows:

creation\_date |post\_id |post\_history\_type\_id|user\_id |total\_rows|  
-----------------------+--------+--------------------+--------+----------+  
2020-07-20 05:00:26.413|62964197| 34| -1| 2|  
2020-08-05 16:31:15.220|63272171| 5|14038907| 2|  
2018-10-08 09:54:40.990|40921767| 5| 4826457| 2|  
2020-05-07 22:02:27.877|61637980| 34| -1| 2|  
2018-10-13 05:26:22.243|52784015| 5| 6599590| 2|  
2021-01-03 10:35:35.693|65550662| 5|12833166| 2|  
2018-12-02 14:28:12.947|53576317| 5|10732059| 2|  
2018-09-05 04:16:26.440|52140985| 4| 3623424| 3|  
2018-12-17 22:43:27.800|53826052| 8| 1863229| 2|  
2018-09-13 17:13:31.490|52321596| 5| 5455640| 2|