**Video Games Sales**

After transforming the data in the previous step, we further aggregated data.

**Inside Hive**

* We created two new tables, “sports\_critic\_ score” and “shooter\_critic\_score” with columns Critic\_Score one for each table. We are going to store the data for average critic score for Sports games, and average critic score for Shooter games.
* We load the data from our raw data to the “sports\_critic\_ score” and “shooter\_critic\_score” tables, by averaging the scores for the respective games.
* We created three new statistics tables, one each for “all games”, “sports games” and “shooter games”. Each table has three columns to aggregate the data on average, min and max.
* We load these three tables (“all games”, “sports games” and “shooter games”), with data from our raw data.
* We created two new tables to find the global scales for sports games and shooter games. We named them “global\_sales\_sports” and ” global\_sales\_shooter”. Each table has two columns, “global\_sales” and “count”.
* We load the two tables (“global\_sales\_sports” and ” global\_sales\_shooter”), with data from our raw data.

**Human Resources**

After transforming the data in the previous step, we further aggregated data.

**Inside Hive**

* We created two new tables, “Sales\_people\_attrition” with a column “attrition”, and “Sales\_people\_non\_attrition” with column non\_attrition. We are going to count how many people have attrition, and how many people have no attrition.
* We load the data from our raw data to the “Sales\_people\_attrition” and “Sales\_people\_non\_attrition” tables, by counting how many people have attrition, and how many people have no attrition.
* We created three new statistics tables, one each for “all salespeople”, “people with attrition” and “people with no attrition”. Each table has three columns to aggregate the data on “average\_monthly\_income”, “min\_monthly\_income”, and “max\_monthly\_income”.
* We load these three tables (“all salespeople”, “people with attrition” and “people with no attrition”), with data from our raw data.
* We created two new tables to find the monthly income and count of salespeople. We named them “salespeople with attrition” and ”salespeople without attrition”. Each table has two columns, “monthlyincome” and “count”.
* We load the two tables (“salespeople with attrition” and ”salespeople without attrition”), with data from our raw data.