3. Identify the grain in your dimensional design using the business needs as a guideline. You should then indicate relative storage requirements for the grain using the statistics for the data sources. Using the cardinality estimates provided, you should determine either the fact table size or sparsity and then compute the unknown grain size variable. For example, you should compute sparsity if the fact table size is given.

Answer:

The most detail grain is the combination of individual customer, individual merchandise/service and date

- 820 Merchandise/Service: sum of Merchandise rows (500), ServCategory rows (20), SpecialEvents Worksheet rows (300)

- 50,150 Customers: sum of Member rows (50,000), Unique Customer per Special Event Worksheet (150)

- Days per year: 365

- 550,300 Purchases of individual merchandize/service/event: sum of Contains rows (450,000), ServicePurchase rows (100,000), and Special Events Worksheet rows (300)

- Fact table size is determined from sum of the rows in the Contains and ServicePurchase table and Special Events Worksheet. Thus, the individual product purchases per year are 550,300.

- Sparsity estimate:
    - 1 - (fact table size / product of dimensions)
    - $(1 - (550,300 / (820*50,150*365)) = 0.99996$

- The data cube has mostly missing cells with less than 1% of cells with non zero values.