Remember: strive for the ideal, work with what is real.

Cardinality

&

Granularity

Refers to how many unique values are in a column.

In a high cardinality column, most or all of the values are unique.

In dim tables, the primary key (unique identifier for each row) will have high cardinality.

Mitigation strategies:

- Remove unnecessary high cardinality columns.
- Split a high cardinality column into two or more columns to reduce cardinality.
- Move a high cardinality column into a dim table, out of a fact table, if possible.
- Create a bridge table.

Refers to the level of detail in the data. For example, hourly, daily, weekly, monthly, quarterly or yearly.

Data **must** be collected and stored at the lowest level required for reporting.

Lower level data can be "rolled up" or aggregated to a higher level. For example, daily sales can be aggregated to weekly sales.

Higher level data cannot be "broken down" with accuracy. For example, weekly data cannot be accurately divided into daily data, only averaged by day.

Mitigation strategies:

- Try to obtain data at a lower level of granularity.
- If that is not possible, determine the most accurate formula for estimating data values at lower levels.