Time series can often be divided into various attributes. For example, the total number of bicycles sold can be divided into road bikes, mountain bikes, and hybrids. Within hybrids, there can be further divisions such as city bikes and trekking bikes. These categories are nested within larger group categories, and such time series are called hierarchical time series. Hierarchical time series often arise with geographical divisions. For example, we can consider the example of Australia, where the total number of visitors to Australia can be divided into states, and the visitor numbers for each state can be further divided into regions. Another alternative structure is crossed attributes. For instance, visitors to Australia can be grouped according to the purpose of their visit, such as tourism or business. However, these attributes are not nested, so they cannot be hierarchically disaggregated in a unique way. Time series grouped according to crossed attributes are called grouped time series. More complex structures involve the combination of nested and crossed attributes, known as nested and crossed time series. For example, visitors to Australia can be first divided by geographic division and then by the purpose of their visit. The process of obtaining higher-level series starting from the lower levels of the hierarchy is called aggregation, while the process of separating series using nested or crossed structures is called disaggregation. The main objective in hierarchical time series is to make coherent forecasts, meaning that the forecasts should add up consistently according to the aggregation structure of the hierarchy or group that defines the time series.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

In hierarchical time series, "Total" is referred to as the top of the hierarchy and represents the most aggregate level of the data. The t-th observation of the Total series is denoted as y\_t, where t = 1, ..., T. In the example given below, Total is first divided into 2: A and B. Then, A is further divided into 3: AA, AB, and AC, and B is divided into 2: BA and BB. Now let's consider the example of Australia. We can divide domestic overnight trips in Australia into nested geographical areas. First, we can divide them into 8 states, and then further divide each state into regions, resulting in a total of 76 regions in the national total. The data here represents the number of nights Australians spend away from home. By using the aggregate function, we can obtain the hierarchical time series by aggregating the regions at the lowest level to form states and aggregating the states to form the national total. The graph below shows the domestic overnight trips for each state from 1998 Q1 to 2017 Q4. The aggregated graph represents the national total. The plot below displays the seasonal patterns of overnight trips in the northern states (Queensland, Northern Territory) and southern states (Victoria and Tasmania) of Australia. According to these plots, the northern states tend to be more active during winter months, while the southern states show higher activity during summer months. In fact, using the hierarchical time series that includes states alongside the national total is advantageous here, as relying solely on the total time series for forecasting would lead to loss of information.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

As mentioned earlier, grouped time series cannot be uniquely disaggregated. Let's consider the following example. Suppose Total has two attributes: Attribute 1 and Attribute 2. And within these attributes, they are further divided into two: A, B for Attribute 1, and X, Y for Attribute 2. As can be seen in the figure below, the Total time series can be disaggregated first by Attribute 1 and then by Attribute 2, or it can be disaggregated first by Attribute 2 and then by Attribute 1.  
Attribute 1 and Attribute 2 are the two attributes that divide the Total time series. Within Attribute 1, there are categories A and B, and within Attribute 2, there are categories X and Y. The figure shows how the Total time series can be disaggregated either by Attribute 1 first and then by Attribute 2 or by Attribute 2 first and then by Attribute 1.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

The final structure is when hierarchical and grouped time series are combined, and the disaggregating factors are typically both nested and crossed. Let's consider the example of Australia in this context. We can disaggregate the overnight trips time series in Australia first by geographical variables, dividing it into states, and then disaggregate it for each state based on travel purposes.  
  
  
If we evaluate the example of Australia in this situation, we can disaggregate the overnight trips time series in Australia based on geographical variables first, dividing it into states, and then disaggregate it for each state based on travel purposes.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

In hierarchical time series, to obtain a set of coherent forecasts, aggregation is performed for higher levels or disaggregation for lower levels. The first and simplest method for generating coherent forecasts is the "bottom-up" approach. In this approach, forecasts are first generated for each time series at the bottom levels, and these forecasts are then summed up to obtain the forecasts for the entire series. For example, for the first figure, we first generate h-step-ahead forecasts for each bottom-level series and then combine them to obtain h-step-ahead coherent forecasts for the rest of the series. Here, we can explain the term "h-step-ahead" as follows: considering the data from time 0 to T as the training data, we make forecasts for time T+h. Additionally, the hat symbol (^) on the y's indicates base forecasts, while the tilde symbol (~) indicates coherent forecasts. The biggest advantage of this method is that there is no loss of information in the aggregation process since the base forecasts are made at the bottom levels. However, if the time series at the bottom levels are noisy, forecasting and modeling for these time series can be challenging.

\*\*\*\*\*\*\*\*\*\*\*\*

Another approach is the Top-down method. In this method, base forecasts are first generated for the Total series, and then they are disaggregated within the hierarchy. Proportions are used as the disaggregation method. The base forecasts for the Total series are distributed to each series at the lowest level based on proportions to generate coherent forecasts for each series. Once h-step-ahead coherent forecasts are generated for the bottom-level series, they are aggregated to create coherent forecasts for the rest of the series. There are two common and traditional methods to determine the proportions: Average Historical Proportions and Proportions of the Historical Averages. These two methods determine the disaggregation proportions based on historical ratios of the data. The Top-down method is sensitive to noisy data since it performs base forecasting at the Total time series level. However, it has some disadvantages. Firstly, when using such top-down approaches, we cannot leverage the individual characteristics of bottom-level series, such as special events or different seasonal patterns, resulting in a loss of information.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Top-down approaches based on historical data tend to produce less accurate forecasts compared to bottom-up approaches. This is because the historical proportions used for disaggregation do not account for how these ratios may change over time. For example, consider the scenario where the Total time series is divided into a training period of 2015-2018 and a forecast period of 2019-2020. If the bottom-level series create proportions based on the 2015-2018 data and disaggregate accordingly, it may not capture any changes that occurred in the bottom-level time series during 2019-2020. Therefore, a new approach has been proposed. In this approach, the bottom-level time series (2015-2018) are used to forecast for the period of 2019-2020, and proportions are created based on these forecasts, taking into account how the data may change for the years 2019-2020.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

The disadvantage of top-down approaches is that even if the base forecast is unbiased, the coherent forecasts are not unbiased. Another method is the Middle-out approach, which combines the top-down and bottom-up methods to achieve a forecast at a middle level.  
  
This approach works as follows:  
  
1. It starts with the top-down approach. Initially, a forecast is made for the highest-level time series. For example, it starts with the forecast of total sales.  
  
2. Then, this top-level forecast is disaggregated downwards. However, instead of strictly following the top-down approach, adjustments are made at the middle levels.  
  
3. At the middle levels, corrections are made that take into account the proportions of the lower-level series in the total series. This can be done based on historical ratios or a weighting method based on ratios. This ensures a more accurate distribution of forecasts, considering the relationships at the middle levels.  
  
4. The adjusted middle-level forecasts are then distributed to the lower-level series. This distribution is typically done using the bottom-up approach.  
  
In conclusion, the "middle-out" approach allows for making middle-level forecasts by combining the top-down and bottom-up methods. This approach aims to achieve more balanced and consistent forecasts by considering the relationships and proportions at the middle levels while preserving the hierarchical structure of the forecasts.  
  
This method can provide more balanced and consistent forecasts compared to the top-down and bottom-up methods. Additionally, the adjustments made at the middle levels during the distribution of forecasts can improve the overall forecasting performance.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Using the aggregation and disaggregation methods we have learned so far, we can create a forecast pipeline for a hierarchical time series as follows:  
  
1. Create Hierarchical Time Series: Firstly, the data is disaggregated, meaning it is separated into grouped or nested structures.  
  
2. Model: In this step, base forecasts are generated for the Total time series or the bottom-level time series using top-down or bottom-up approaches.  
  
3. Reconcile: Specify how the coherent forecasts will be generated from the selected models according to the Top-Down, Bottom-Up, or Middle-Out approaches.  
  
4. Forecast: Generate forecasts for the entire aggregation structure.  
  
Please note that these translations are provided for informational purposes and may not be exact translations of the original text.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

The process of generating forecasts using the hierarchical time series we have seen so far is called Forecast Reconciliation. To explain it more clearly:  
  
In hierarchical time series, there are relationships between series at different levels, and these relationships are important for accurate and consistent forecasts. The forecast reconciliation method aims to combine forecasts at different levels to achieve a consistent forecast structure. This method is used to correct inconsistencies between forecasts and obtain more accurate forecasts while maintaining the hierarchical structure.  
  
The equations we have seen so far can be thought of as aggregation constraints or summing equalities and can be more efficiently represented using matrix notation. For any aggregation structure, we create an nxm summing matrix (S). For example, for the figure, we create such a matrix. Here, y\_t is an n-dimensional vector representing all observations in the hierarchy at time t. S is the summing matrix, and b\_t is an m-dimensional vector representing observations at the bottom level at time t. The first row of the S matrix represents the Total, the second and third rows represent time series at the level below Total, and the bottom five rows represent the identity matrix, representing the bottom-level observations being equal to themselves.  
  
Mapping matrices can be used for both hierarchical and grouped time series, as well as for bottom-up or top-down approaches. In the formula, the product of the summing matrix, mapping matrix, and base forecast vector produces the coherent forecast result. The base forecast predicts all series without considering any aggregation constraints, so it can be thought of as an unreconciled forecast. In short, SG combines and reconciles all base forecasts to generate coherent forecasts.  
  
Please note that these translations are provided for informational purposes and may not be exact translations of the original text.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*

We can find the optimal G matrix to generate the most accurate reconciled forecasts. For example, the G matrix suggested by the Minimum Trace optimal reconciliation approach minimizes the total forecast variance of the coherent forecast set. We always strive to generate unbiased forecasts, and if the base forecasts are unbiased, the coherent forecasts will also be unbiased, satisfying the equation SGS = S and providing a constraint on the matrix G. However, no top-down approach satisfies this constraint, resulting in biased coherent forecasts for all top-down approaches.  
  
However, the Minimum Trace method provides an approach that yields optimal reconciliation forecasts (also unbiased) for hierarchical and grouped time series. Optimal reconciliation forecasts are generated by utilizing all the available information within the hierarchical or grouped time series. This means that in specific hierarchical and grouped time series, it can uncover the characteristics of the data that are of interest to the user and important for modeling, even if they are completely hidden.  
  
For example, in the previous examples of the Australian tourism case, we observed distinct seasonal patterns between the northern and southern states. However, when we aggregate the data, these differences may not be noticeable at the country level (total level). The Minimum Trace method focuses on addressing this issue.  
  
Please note that these translations are provided for informational purposes and may not be exact translations of the original text.