

HP OpenView Performance Manager

For the Windows®, HP-UX, and Solaris

Software Version: 6.01

Planner Guide

Document Release Date: February 2007
Software Release Date: February 2007



Legal Notices

Warranty

The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

The information contained herein is subject to change without notice.

Restricted Rights Legend

Confidential computer software. Valid license from HP required for possession, use or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

Copyright Notices

© Copyright 2004 – 2007 Hewlett-Packard Development Company, L.P.

Trademark Notices

Java™ and all Java based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the U.S. and other countries.

Microsoft® is a U.S. registered trademark of Microsoft Corporation.

Windows® is a U.S. registered trademark of Microsoft Corporation.

UNIX® is a registered trademark of The Open Group.

Support

You can visit the HP OpenView Support web site at:

www.hp.com/managementsoftware/support

HP OpenView online support provides an efficient way to access interactive technical support tools. As a valued support customer, you can benefit by using the support site to:

- Search for knowledge documents of interest
- Submit and track support cases and enhancement requests
- Download software patches
- Manage support contracts
- Look up HP support contacts
- Review information about available services
- Enter into discussions with other software customers
- Research and register for software training

Most of the support areas require that you register as an HP Passport user and sign in. Many also require a support contract.

To find more information about access levels, go to:

www.hp.com/managementsoftware/access_level

To register for an HP Passport ID, go to:

www.managementsoftware.hp.com/passport-registration.html

Contents

1	Introduction to HP OpenView Performance Manager Planner	7
	Overview of Planning Techniques.....	7
	Concepts in Statistical Forecasting.....	8
2	Quick Start to Using OVPM Planner	13
	Getting Started.....	13
	Drawing Pre-defined Forecast Graphs.....	13
	Designing Your Own Forecast.....	16
	Select a Data Source and Metric.....	17
	Select the Planner Tab.....	19
	Draw a Smoothed Graph.....	20
	Select a Forecast Trending Method.....	22
	Add an Interval Type.....	23
	Select Seasonality Options.....	25
	Try Different Trending Methods.....	28
	Validating the Forecast.....	30
3	OVPM Planner and the Forecasting Process	33
	Collecting Data.....	33
	Selecting Data.....	35
	Select the Data Source for the Forecast Metric.....	35
	Select the Metric to Forecast.....	36
	Suggested Forecast Metrics.....	38
	Select the Range of Data.....	40
	Select Points Every.....	41
	Select Shifts.....	42
	Skip Holidays.....	42
	Producing the Forecast.....	43
	Trending Method.....	43
	Seasonality.....	46
	Confidence and Prediction intervals.....	47
	Validating the results.....	48
	Understanding the Statistical Numbers.....	49
	The Forecast Equation.....	50
	The Correlation Coefficient.....	51
	Variance as a Function of.....	51
	Estimated Standard Error.....	51
	Seasonality Information.....	52
	Confidence and Prediction Interval Information.....	53

Reforecasting	53
4 Business Unit Forecasting	55
Understanding a Business Unit	55
Creating a Business Unit File	56
OVPM Flat File Format	56
Using a Business Unit File	57
Glossary	61
Index	63

1 Introduction to HP OpenView Performance Manager Planner

HP OpenView Performance Manager (OVPM) Planner is used to forecast system resource consumption. You use the OVPM Planner to configure and draw forecasting graphs that can be used in capacity planning.

Computer capacity planning is a process designed to ensure that sufficient computer resources are available to maintain acceptable performance levels. It is distinct from day-to-day performance management, troubleshooting, or tuning because it means planning future resources to meet changing user demand. Capacity planning is an important part of managing an adaptive enterprise.

Capacity planning is an ongoing process that requires a system manager's continual attention. In larger enterprises you may also have capacity planning specialists/statisticians responsible for capacity planning. Through capacity planning, you ensure that system users receive the level of performance that they require to work effectively.

Capacity planning begins with the collection of performance data. For OVPM this is usually through the HP OpenView Performance Agent (OVPA), HP OpenView Operations Agent (OVO agents), and HP OpenView Internet Services (OVIS). It can also be data from a generic database (with some restrictions) or a flat file.

Based on this collected performance data you can use OVPM Planner to predict the future service demands of your users. OVPM Planner uses time-series based statistical forecasting techniques. You can use OVPM Planner to make more reliable decisions on when to acquire additional resources or adjust system load to meet changing demands.

Overview of Planning Techniques

Capacity planning can be done casually or formally. In general, the techniques you use and the reliability of the information on which you base your assumptions will determine the reliability of your plans. Techniques -- and the tools that support them -- vary in application, complexity, ease-of-implementation, and cost. Combinations of techniques may be used to model system behavior. No matter which tools or techniques you use, the quality of your plan reflects the quality of the information on which it is based.

The OVPM Planner is based on the statistical forecasting technique.

Statistical Forecasting uses historical trends to predict future trends. Statistical forecasting can be used to examine subtle variations in data -- types of trends, seasonality and significance -- and provide an indication of confidence in results.

Concepts in Statistical Forecasting

The objective of computer capacity planning is to provide an acceptable, cost-effective level of IT services to users while responding to workload demands generated by business requirements. Capacity planning is an important part of managing an adaptive enterprise.

The key to effective capacity planning is to understand the relationship between business requirements, systems and network workload, IT infrastructure capacity and the level of service required by users. Changes in the business environment often result in changes in systems and network workload. Planning for capacity changes can help you anticipate problems and suggest capacity adjustments with greater confidence.

Statistical Forecasting is a technique useful in planning for capacity changes. OVPM Planner is based on statistical forecasting techniques.

To do statistical forecasting you must first characterize the typical workload. And in order to do this you must first measure the workload by collecting performance data. With OVPM, performance data is collected via the OVPA, OVO agents, or OVIS. Data can also be collected from a generic database (with some restrictions) or a flat file.

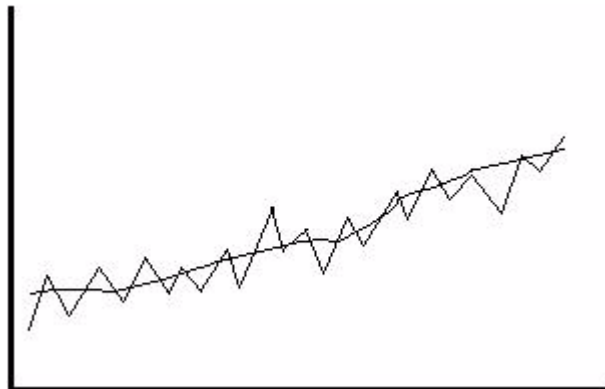
Once you have current performance data you can use this data to visualize the future capacity trend and generate a forecast.

In statistics, a **trend** is a change, tendency, or pattern in a series of data over a period of time, after other components have been accounted for.

To make it easier to determine a trend in the data, a **smoothing method** is used. On a smoothed graph, you will see the actual data collected and a smooth curve based on mean values replacing actual data values.

A **Smoothed Curve** removes some of the variability in the data and makes it easier to see any trend over time.

Figure 1 Smoothed Curve

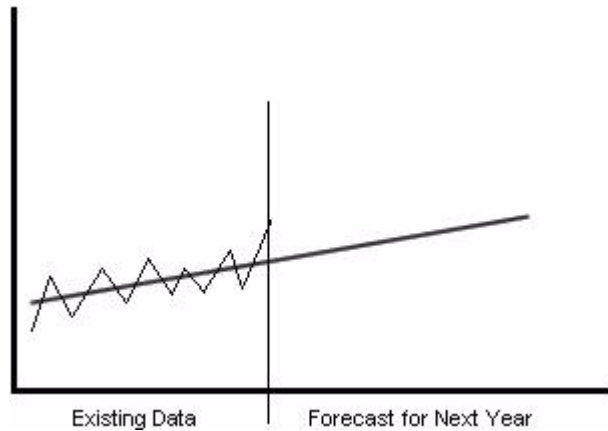


The smoothed graph does not extend beyond the existing data to predict or forecast future data values; it just shows the trend in the existing data.

To forecast future data values you could just use a ruler to draw a line through the existing data out into the future. However, to get a better forecast OVPM uses statistical analysis techniques to account for seasonality, types of trends and correlation to business activity. A **trending forecast** uses historical data as a basis for projecting future values.

A typical **Forecast graph** is shown below.

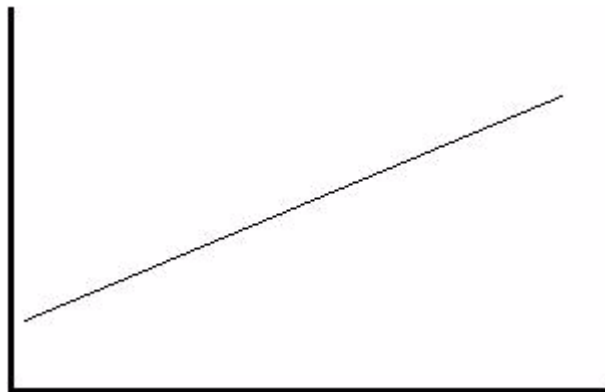
Figure 2 Forecast Graph



Different trending methods (linear, exponential and S-curve) may fit the data better than others. The choice of trending method may be based on knowledge of the metric behavior. For example, percentages are bounded between 0 and 100% and so may suit the S-Curve model; response times typically degrade in a non-linear manner and may be best represented by the exponential model. Often, the best way to choose a trending method is to try them all and then pick the one that produces the best results.

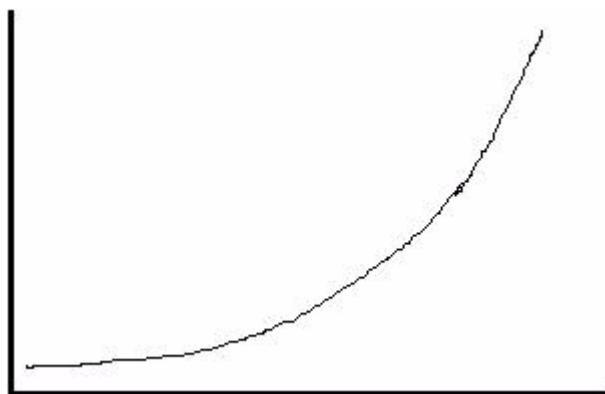
Linear Trend - where the forecast curve looks similar to this:

Figure 3 Linear Trend



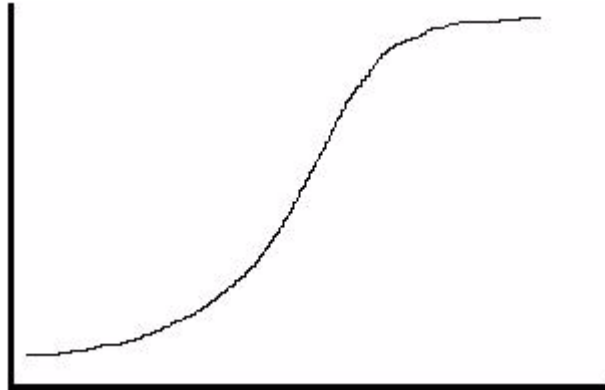
Exponential Trend - where the forecast curve looks similar to this:

Figure 4 Exponential Trend



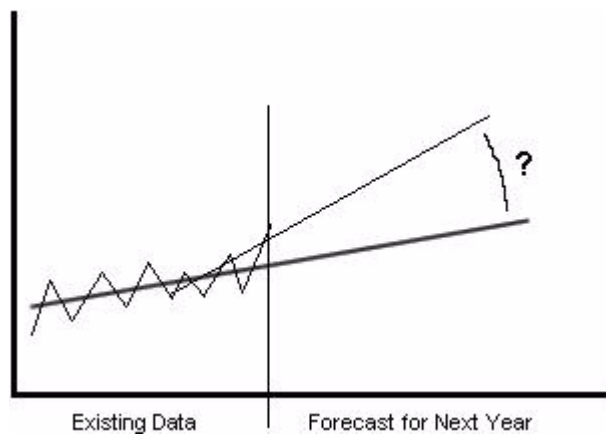
S-curve Trend - where the forecast curve looks similar to this:

Figure 5 S-Curve Trend



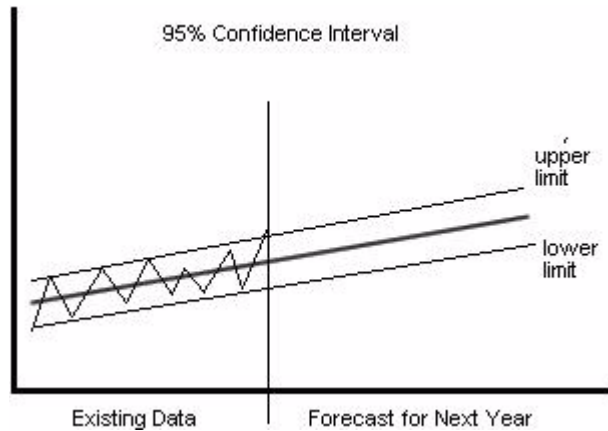
Once you create a forecast you need to **validate** that the results are reasonable and accurate enough to be useful. This is important because a forecast is a prediction not a certainty. Depending on what data is used and what assumptions are made, two very different forecasts could result, as shown below:

Figure 6 Confidence Interval



Forecasts can be validated by examining confidence levels and other statistical measures. OVPM provides a number of tools to use in validating your forecasts, including confidence intervals. A **confidence interval** is an interval (range of values) such that there is a specific probability that the average of the predicted values lies within the upper and lower limit. For example, a 95% confidence interval is an interval where the chance of the predicted values lying within that interval is 95%.

Figure 7 95% Confidence Interval



The width of the confidence interval gives us some idea about how uncertain we are about the predicted values (that is, the forecast). A very wide interval may indicate that more data should be collected or that the forecast method used is not a good fit. A narrow interval reflects a high degree of statistical confidence in the forecast.

Remember that the quality of your forecast reflects the quality of the information and assumptions on which it is based.

Continue on to the next chapter to create a forecast using OVPM Planner.

2 Quick Start to Using OVPM Planner

Getting Started

This section steps you through the different features of OVPM Planner that are used to produce a forecast. It is intended as a quick overview of OVPM Planner. More detailed explanations are given in subsequent chapters, as well as more information on analysis.

The graphs drawn in this chapter use the HPCASE.bus file which is included with the OVPM installation. Before you proceed, you might need to copy the HPCASE.bus file from the `<Install_Dir>/newconfig/ovpm` directory to the `<Data_Dir>/conf/perf` directory. You will typically be accessing the data you want to forecast using the other data sources that OVPM supports instead of this simple flat file we use for demonstrations.



Go through this chapter using the HPCASE.bus logfile before you look at your own data.

You should be familiar with OVPM before beginning this section. The OVPM online help explains many options that are also used by OVPM Planner such as selecting a data source and choosing a metric. These OVPM functions operate the same way in OVPM Planner.

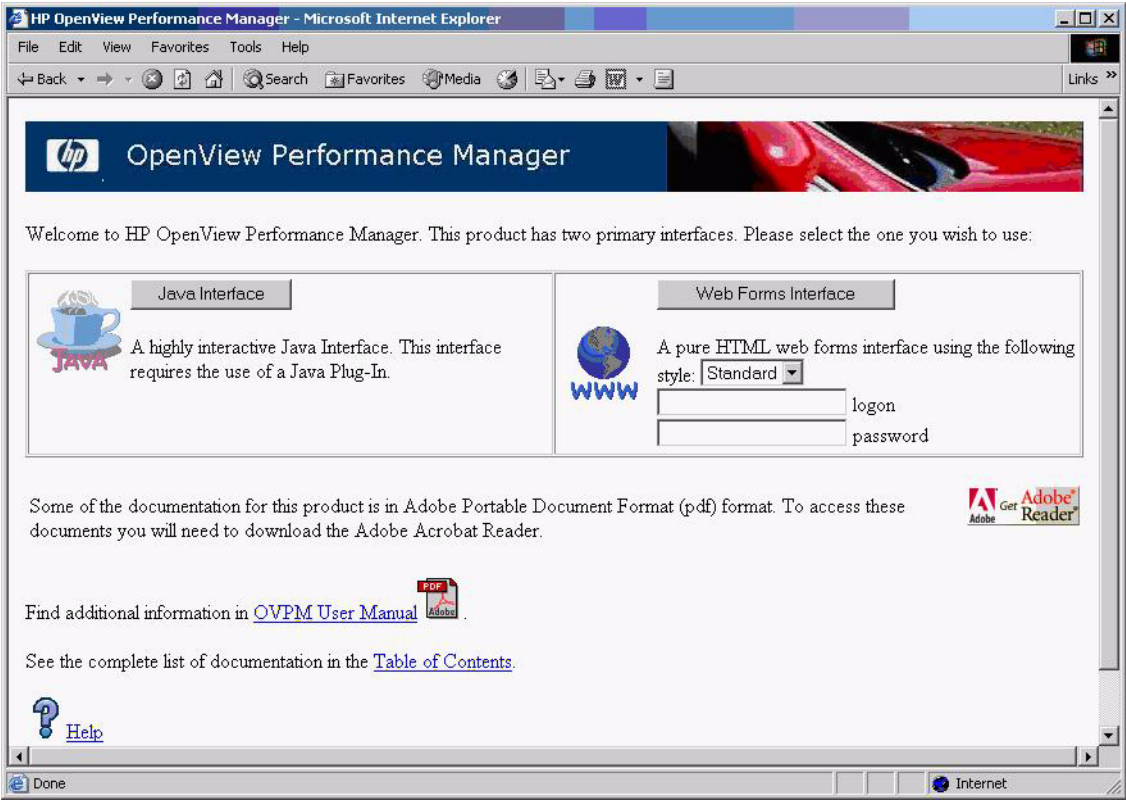
OVPM Planner can be used with two different user interfaces, the Java interface that requires the Java plug-in and is more interactive and the Web Forms interface that does not require Java but is not as interactive. The examples in this manual use the Java interface but, with small changes, could also be done using the Web Forms interface.

Drawing Pre-defined Forecast Graphs

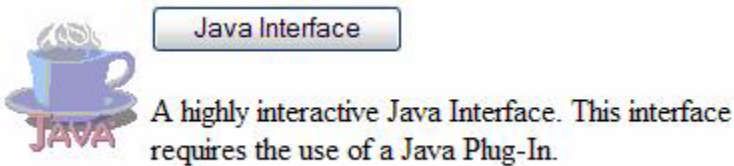
You can use the following steps to draw a pre-defined Forecast graph. These graphs are provided out-of-the-box with OVPM.

- 1 First start OVPM. You can do this by opening a web browser and entering the URL:
`http://system:8080/OVPM`

Where system is the name of the system where OVPM is installed. 8080 is the default port number (the port may be different in your installation). This should display the OVPM home page.

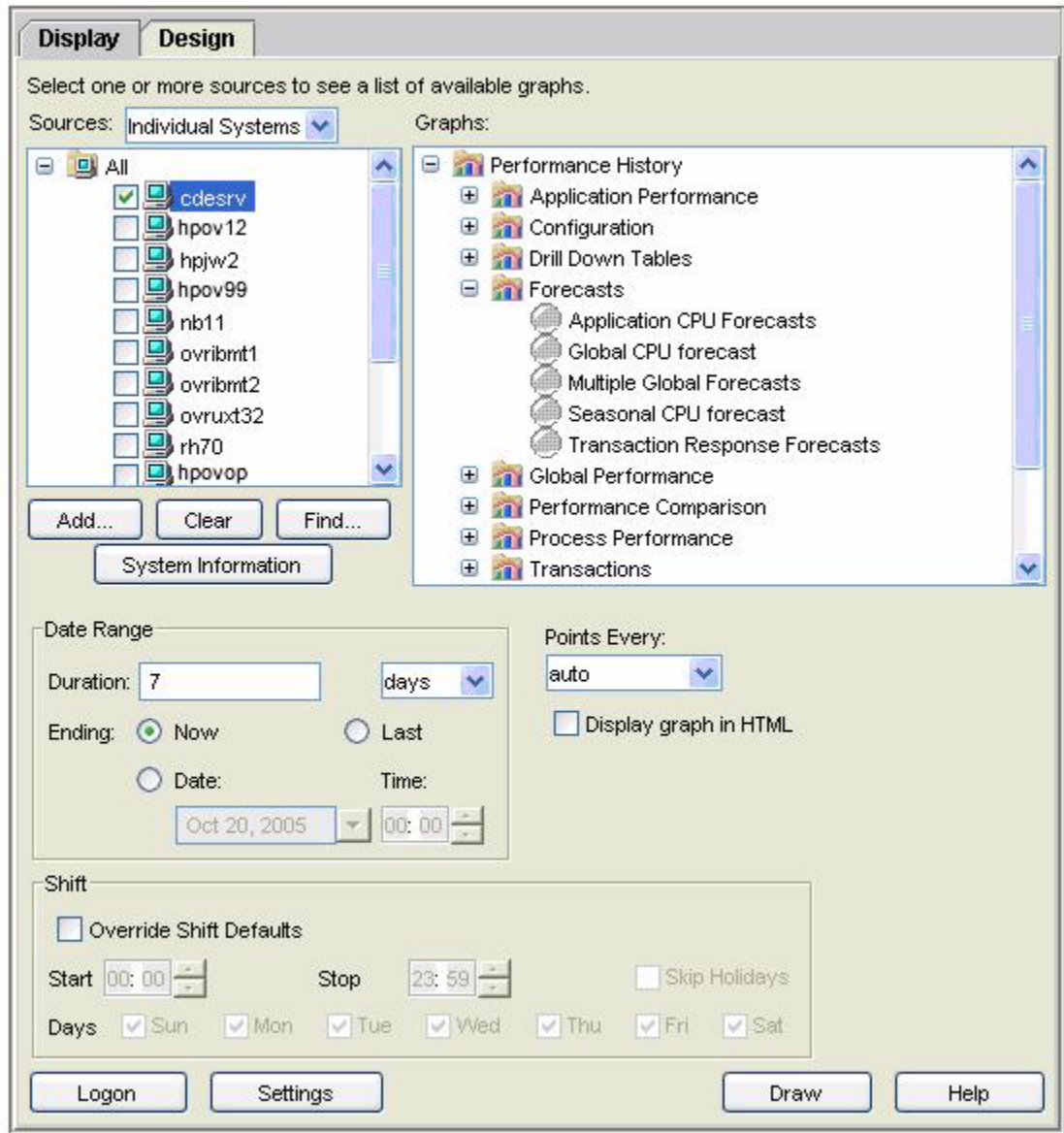


- 2 On the OVPM home page, select which OVPM interface to use. For our example we are using the Java interface so press the button indicating that you want to use the Java interface:

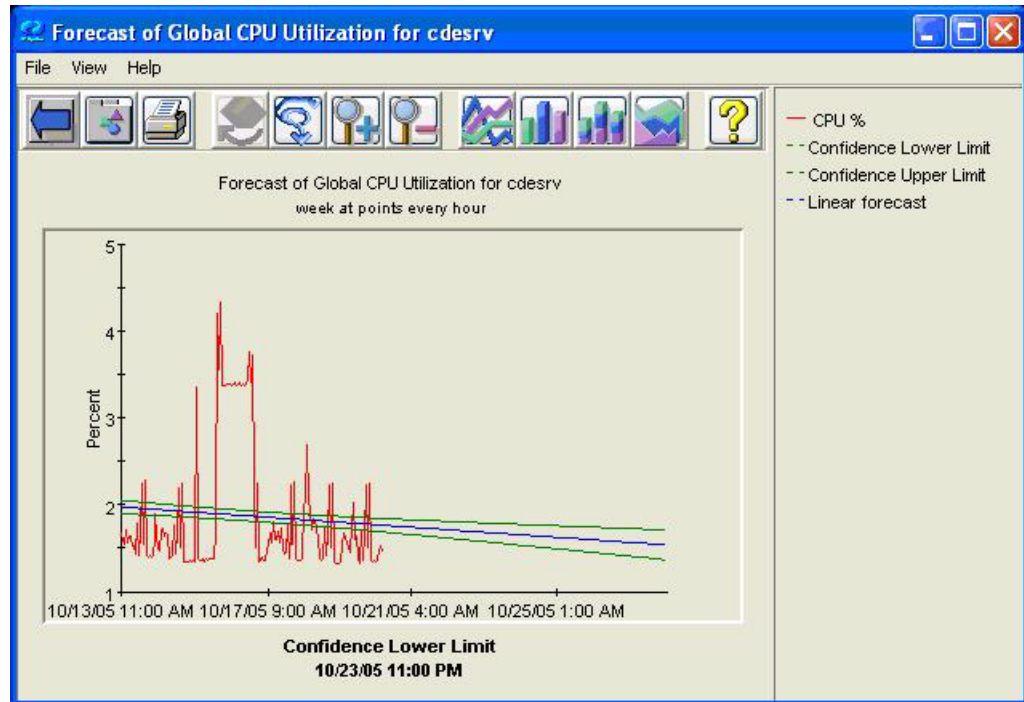


- 3 Once the Java interface is displayed, you can choose to design your own forecast (see [Designing Your Own Forecast](#) on page 16). Or you can use the predefined forecast templates as described below.

- 4 To draw a pre-defined forecast graph, select the **Display** tab in the Java interface. Select a data source from the Individual Systems in the **Sources** list. Then select one of the predefined forecast templates shown under Forecasts in the list of Graphs.



- 5 Select the **Draw** button to draw a forecast graph for the selected data source. See an example of the forecast graph below.



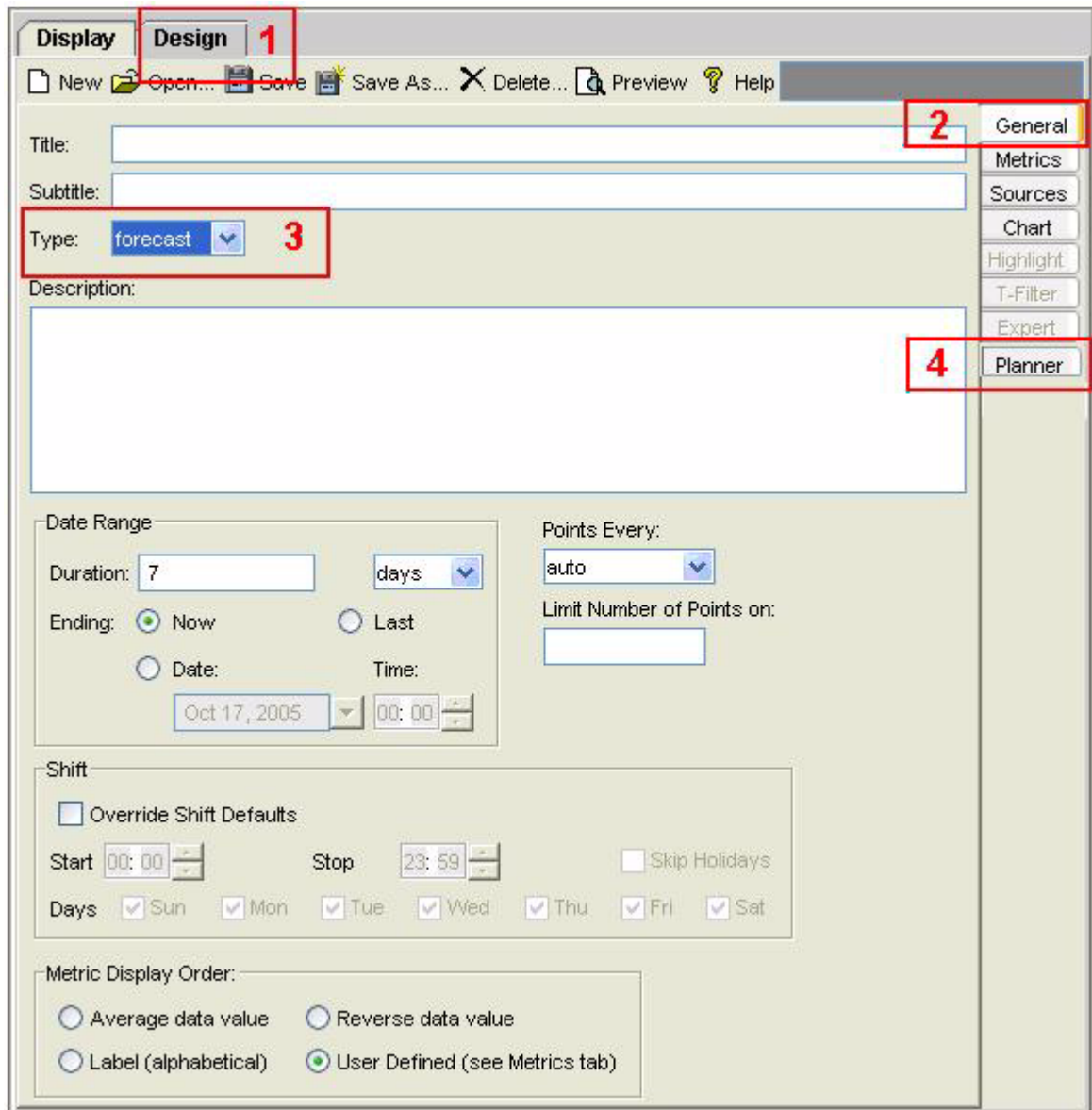
Designing Your Own Forecast

In order to design your own forecast, you use the Planner function in OVPM.

To begin designing your own forecast, first start OVPM as described in the previous section. Then on the OVPM home page select the Java interface (for our examples in this chapter we are using the Java interface but you could also use the Web Forms interface).

Select the **Design** tab in the Java interface. Then select the **General** tab of the Design screen and select the Graph Type **forecast** (shown below). The **Planner** tab of the Design screen should then be enabled.

Figure 8 Planner Tab Details

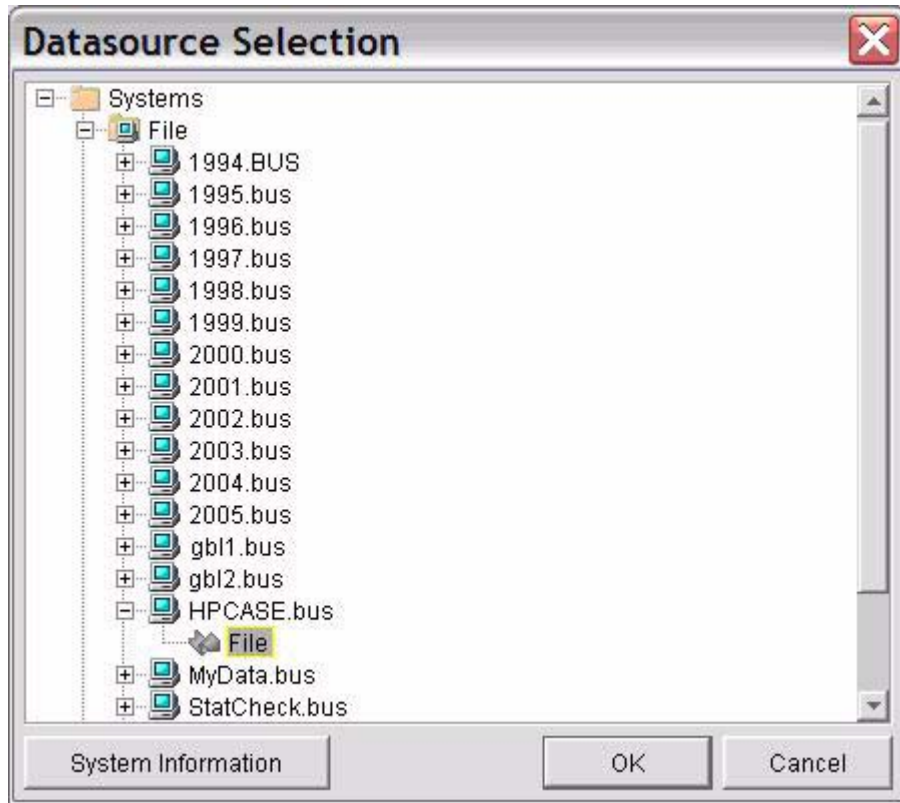


▶ If the Planner tab is not enabled, then you may be accessing a version of OVP that does not have Planner available or your configuration may disallow access to the Planner function. Contact your OVP Administrator for details

Select a Data Source and Metric

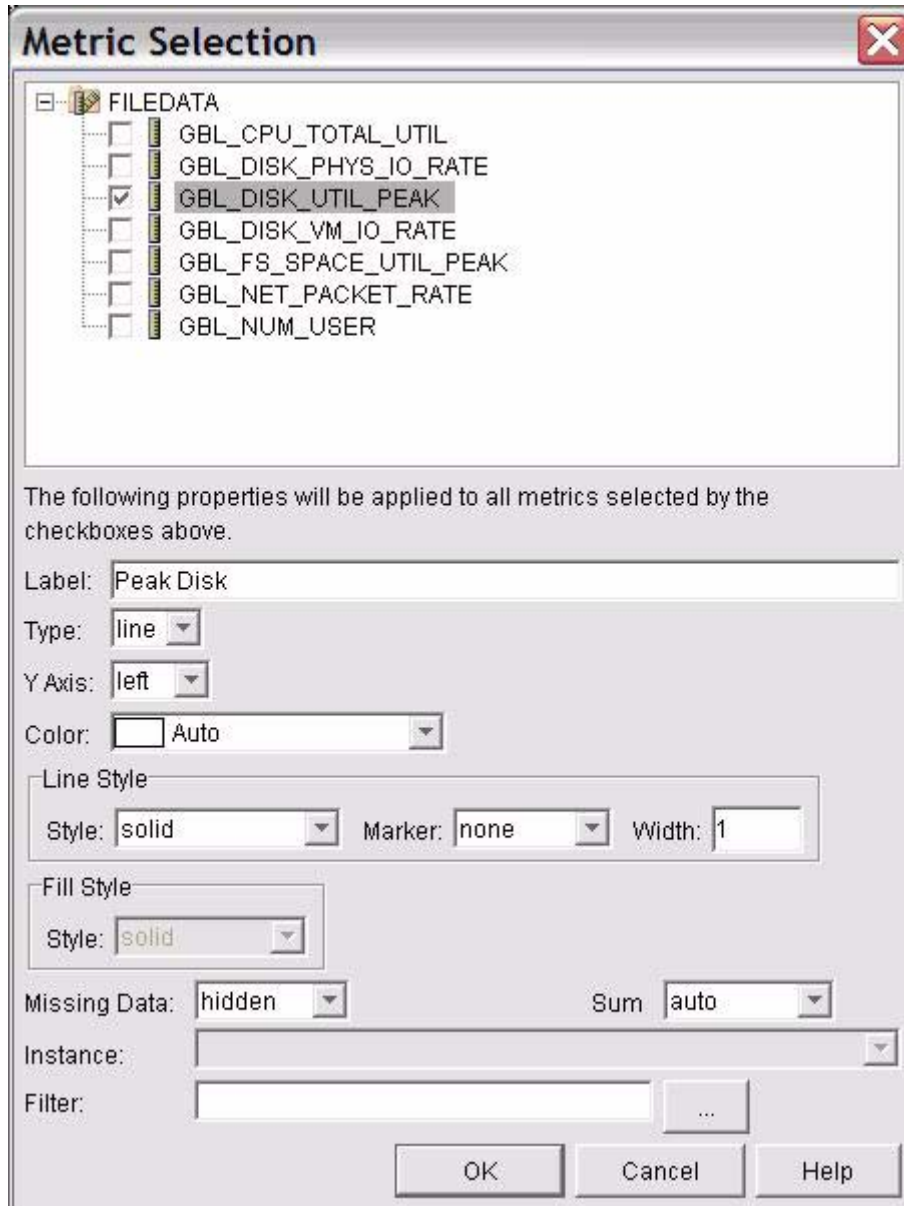
In the Design screen, select the Sources tab. Choose a data source for the forecast metrics as you would for any other graph. Navigate to locate the file HPCASE.bus, select it and press **OK**. The HPCASE.bus file is installed with OVP and contains sample data to be used in the getting started example in this chapter.

Figure 9 Datasource Selection Window



Now in the Design screen, select the Metrics tab and choose one or more metrics by pressing the "Add ..." button. For our example select **FILEDATA** then **GBL_DISK_UTIL_PEAK**, which is the peak disk utilization for our system. Also enter a label for the metric on the graph, for example, Peak Disk. When finished, press the **OK** button.

Figure 10 Metric Selection Window



Select the Planner Tab

In the Design screen, select the Planner tab to display the Planner screen (shown below).



Note that once you enable the Planner function, the date range is governed by the date range in the Planner screen, not the date range on the Design screen General tab.

Figure 11 Planner Tab Details

The Planner tab contains the options for producing your own forecasts. It is only enabled when the graph type is set to forecast on the General tab.

Continue on to the next section for instructions on how to draw a forecast graph with Smoothing or another trending method.

Draw a Smoothed Graph

This section tells you how to draw a smoothed graph. On a smoothed graph, mean values replace the actual data values, to produce a smooth curve. A Smoothed curve makes it easier to see any trend over time.

In the Planner screen, set the date range for the data to the last 12 weeks in the HPCASE.bus data file (Duration **12 Weeks**, Ending **Last**) and granularity of one point every day (Points Every **Day**).

Figure 12 Date Range Section



Date Range

Duration: 12 weeks

Ending: Now Last

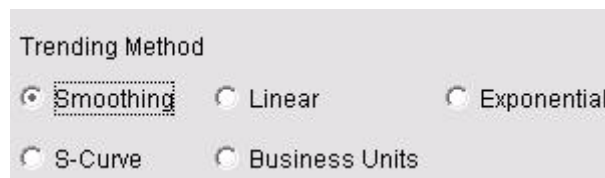
Date: Apr 11, 2005 Time: 00:00

Forecast Ending: Apr 18, 2005

Points Every: day

The Trending Method should still be set to **Smoothing** but double check to be sure.

Figure 13 Trending Method Section



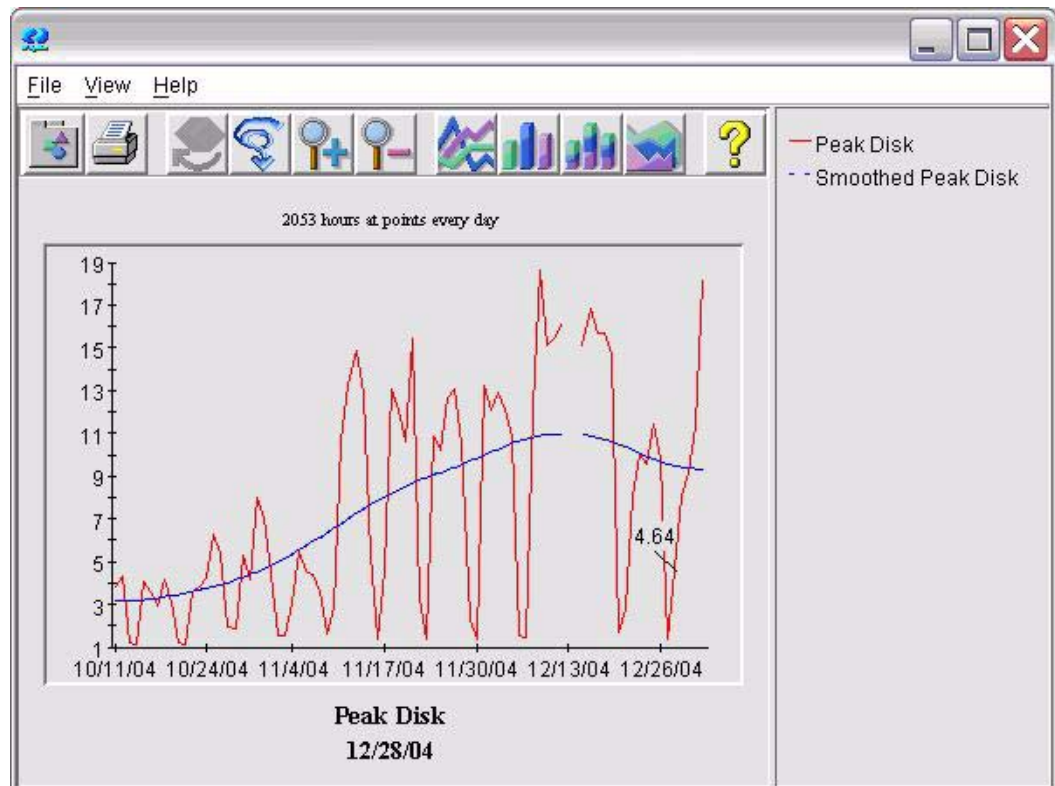
Trending Method

Smoothing Linear Exponential

S-Curve Business Units

Now press the **Preview** button on the toolbar at the top of the screen to see the results.

Figure 14 Smoothed Curve Graph



The **red line** is the actual data and the **blue line** is the smoothed data line. The importance of a smoothed graph is that it can make it easier to spot a trend in your data.

From this graph you can see that the actual data in the HPCASE.bus file ends 12/28/04. This graph does not forecast future values for Peak Disk Utilization.

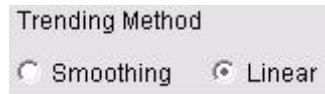
Close the graph and continue with the next section to produce a forecast.

Select a Forecast Trending Method

You can use Linear, Exponential, S-Curve, or Business Units trending methods to produce forecasts.

In the Planner screen, change the Trending Method to **Linear**.

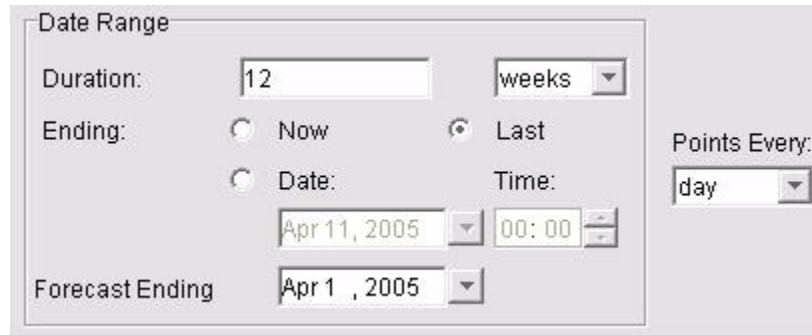
Figure 15 Trending Method Selection



The screenshot shows a dialog box titled "Trending Method". It contains two radio button options: "Smoothing" and "Linear". The "Linear" option is selected, indicated by a filled circle next to it.

Set the end date of the forecast, in this example April 1, 2005, and leave all other options as they are from the previous example:

Figure 16 Date Range Selection



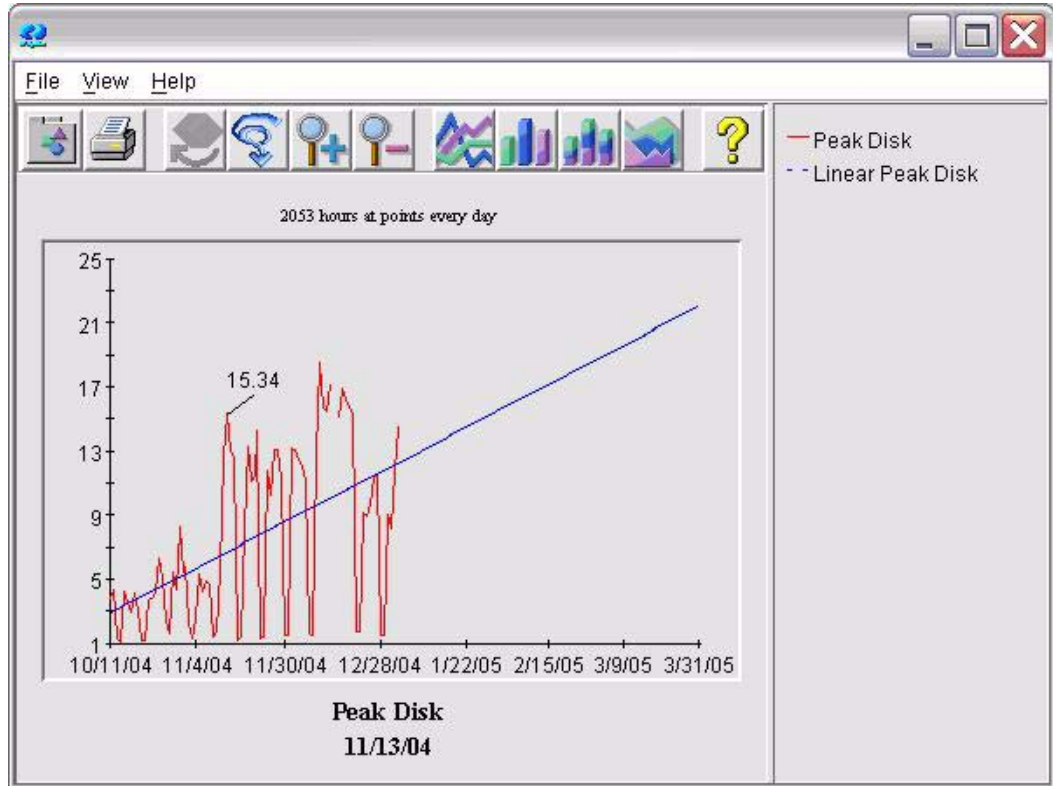
The screenshot shows a dialog box titled "Date Range". It contains several fields and options:

- Duration:** A text box containing "12" and a dropdown menu set to "weeks".
- Ending:** Two radio button options: "Now" (unselected) and "Last" (selected).
- Date:** A date picker showing "Apr 11, 2005".
- Time:** A time picker showing "00:00".
- Forecast Ending:** A date picker showing "Apr 1, 2005".
- Points Every:** A dropdown menu set to "day".

When you forecast, the relationship between the Duration, Ending date, and Forecast Ending date is important. **Duration** determines how much data will be used to produce the forecast. **Ending** date determines when the data being used for the forecast will end. The forecast data starts at the Ending Date minus the Duration. **Forecast Ending** date sets the date for the right hand end of the graph. In order to forecast beyond the end of the data, Forecast Ending date should be set later than the Ending date of the data in your data file.

Press the **Preview** button again to draw the linear forecast graph.

Figure 17 Linear Forecast Graph



From the blue forecast line, it looks like we can expect Peak Disk Usage to go up in the future but it should still be less than 25% by the end of March.



A good rule of thumb is to produce a forecast as long as your data. A forecast that is longer than the data upon which it is based becomes less reliable.

Close the graph.

Add an Interval Type

Next, you will build on the forecasts that you have already made by adding interval bounds.

An interval is simply a statistical range in which the forecast falls. This range is shown on your forecast as lines defining the upper and lower limits. OVPM Planner supports two types of intervals: prediction and confidence. You will learn more about them in later chapters. For now, you will draw a graph with a confidence interval, followed by a graph with a prediction interval.

In the Confidence Lines section of the Planner screen, change the **Interval type** from none to **Confidence** and press **Preview** in the tool bar to redraw your forecast.

Figure 18 Confidence Lines Section

Confidence Lines

Interval type Confidence Level

Show Validation data

Leave this graph open so you can compare it to the next one, in which the Interval type is changed to Prediction. Change the Interval type to **Prediction** and press **Preview** again.

Figure 19 Narrow Confidence Interval Graph

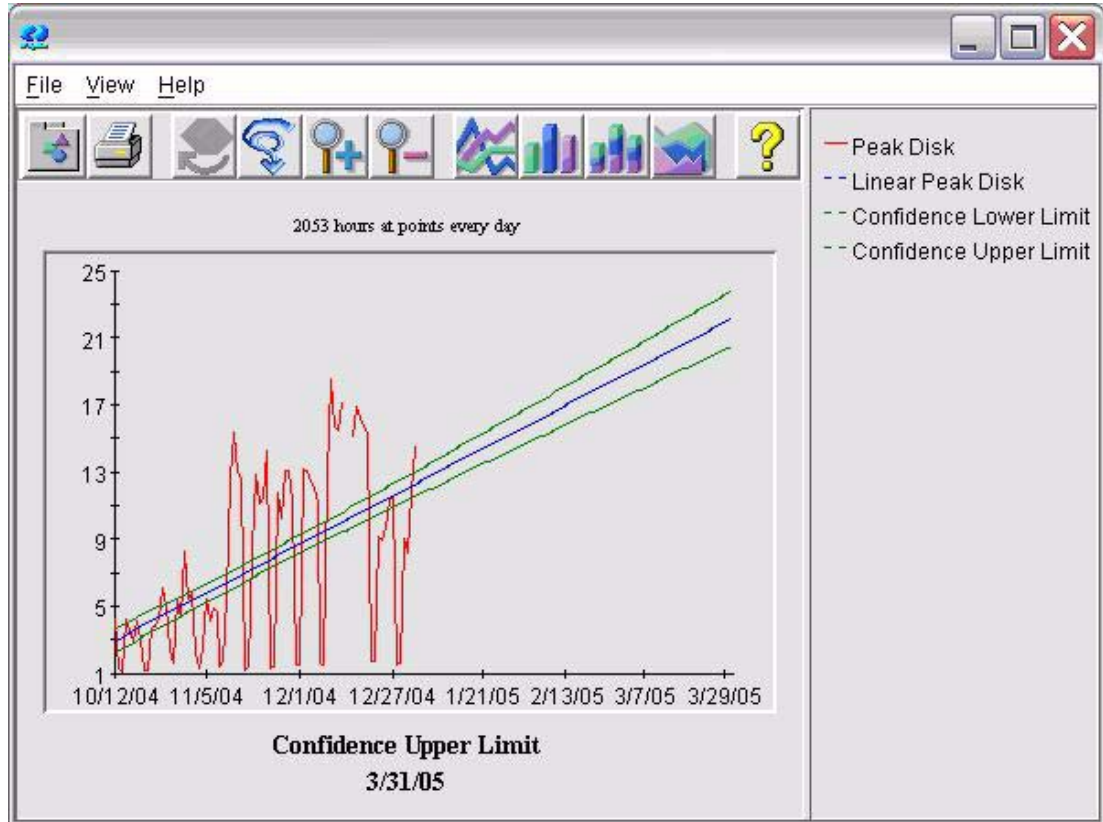
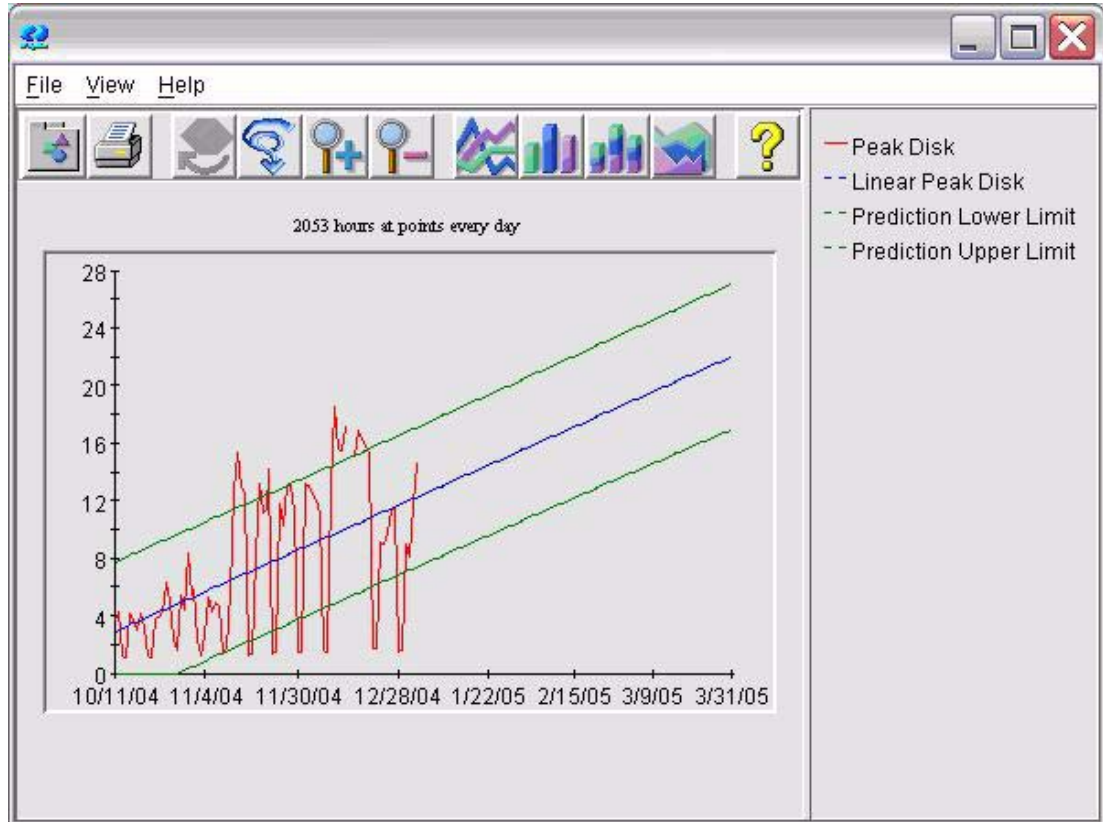


Figure 20 Wide Confidence Interval Graph



The confidence interval is much narrower than the prediction interval. The confidence interval defines a range in which the *average of the predicted values* (the trend line) will fall the given percentage of the time; while the prediction interval defines a range in which the predicted values (all points) will fall the given percentage of the time. The range must be wider for the prediction interval because there is more variability in the forecasted values than there is in the mean of the forecasted values.



A narrow interval (small spread between the upper and lower bounds) tends to indicate a valid or credible forecast. Conversely, a broad interval may indicate a less credible forecast.

See [Confidence and Prediction intervals](#) on page 47 for additional information on interval types.

Close both graphs.

Select Seasonality Options

Seasonality refers to a pattern of variation in data that is related to specific time intervals (seasons.) If your company is on a 5-day work week, you will see day-of-the-week seasonality because your systems do little work, if any, on the weekends. Firms that process orders often experience week-of-the-month seasonality as the month-end approaches and sales quotas must be met.

Follow the steps below to see how seasonality appears on a forecast.

In the Planner screen, set your **Interval type** back to **None**.

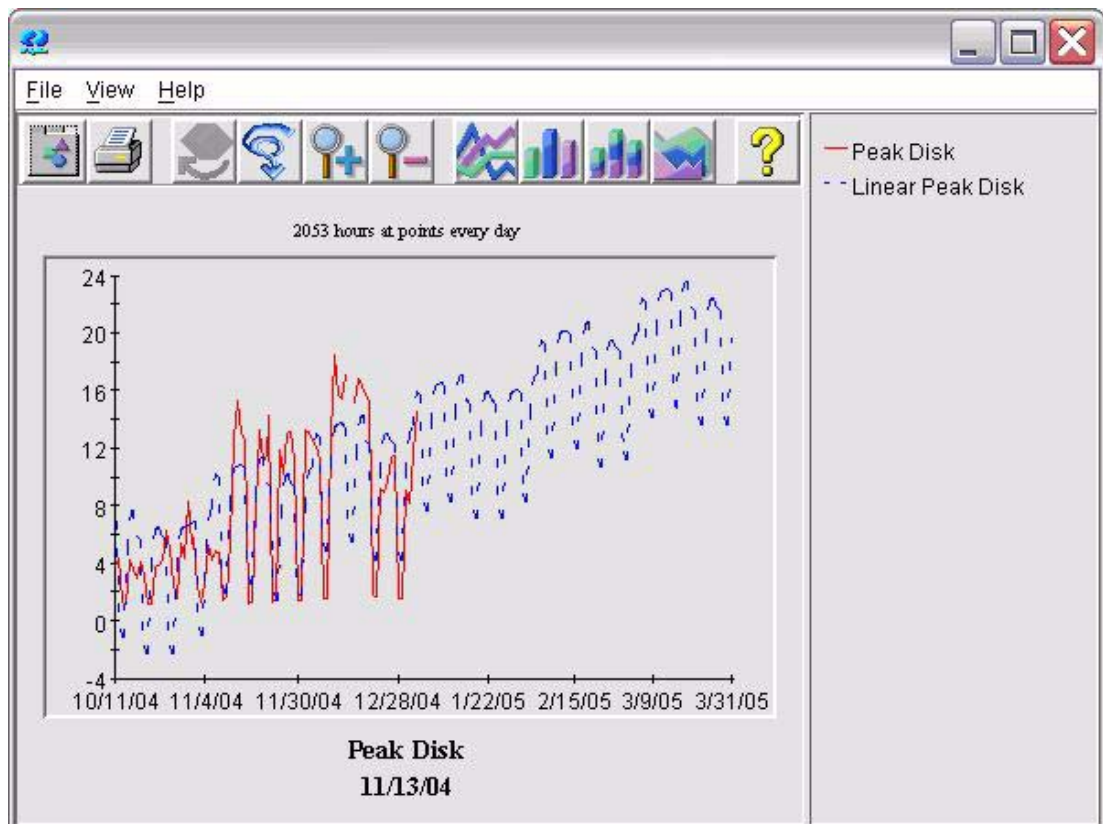
Select all four of the **Seasonality** options. OVPM Planner will determine if a particular type of seasonality is not valid or if it doesn't improve the forecast and will show the final forecast using only those seasonality options found to be appropriate.

Figure 21 Seasonality Section



Press the Preview button to see the forecast.

Figure 22 Seasonality Graph

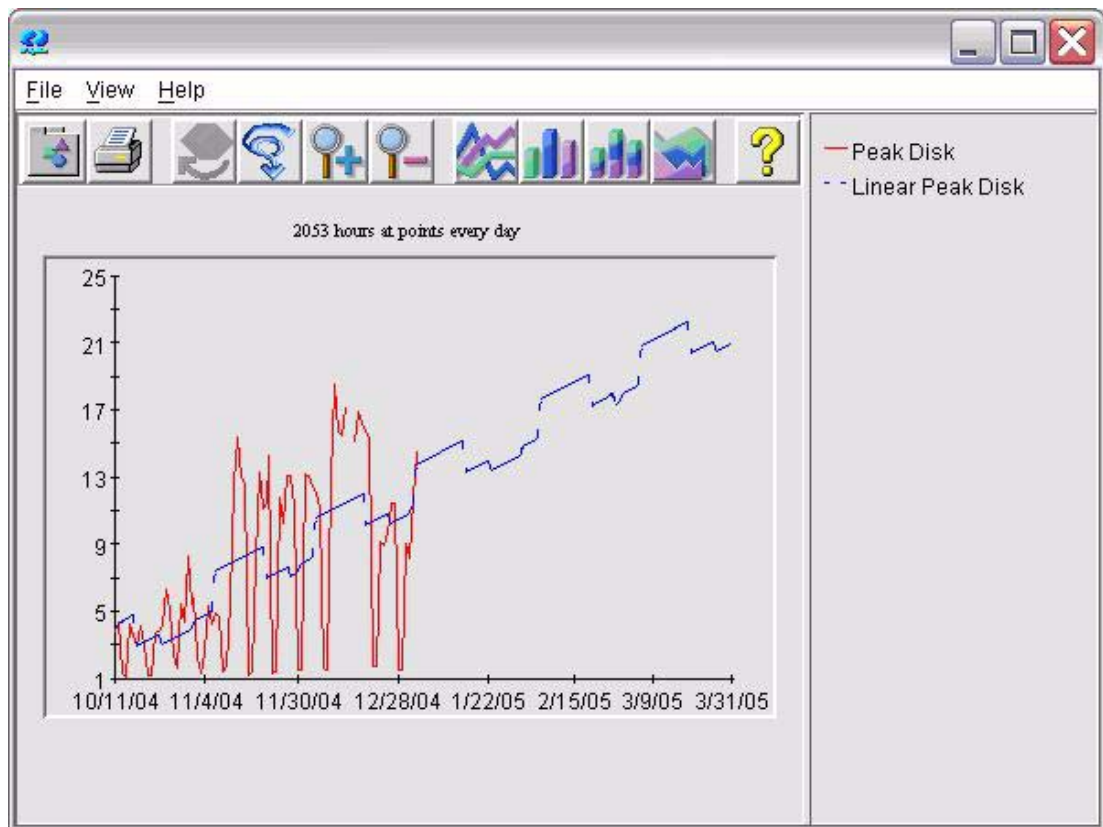


This graph shows week-of-the-month and day-of-the-week seasonality. Note that the blue forecast line continues these seasonalities into the future so that you might see effects such as - Monday mornings during the last week of the month will have a high value but the rest of the month the values will be acceptable.

To see the types of seasonality used in the graph, you can select the **File > Properties** menu on the graph or press the leftmost **Properties Toolbox** icon on the graph. Then select the Graph Properties - General tab to see the Annotations for a forecast. The types of seasonality used are displayed there along with other measures of the forecast. See [Seasonality Information](#) on page 52 for more information.

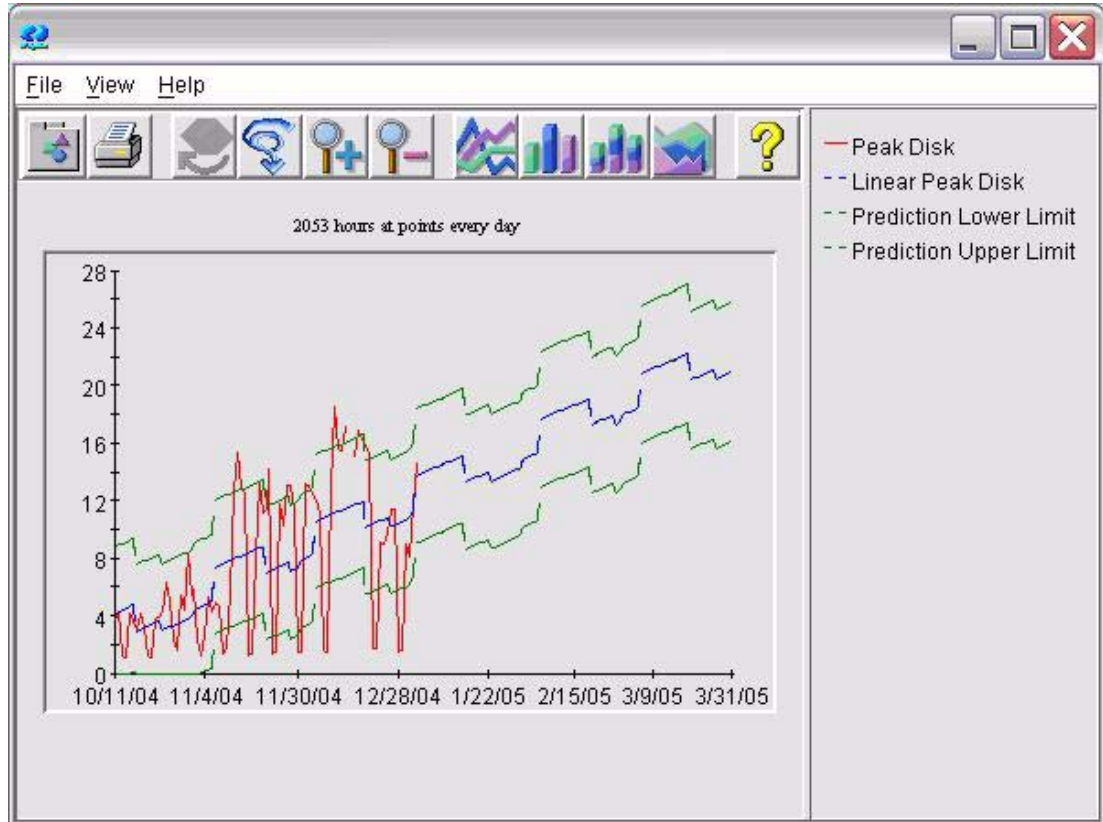
If you want to force the forecast to ignore certain seasonality types, simply uncheck the seasonality option on the Planner screen and preview the graph again. Try it with only **Week of the Month** checked.

Figure 23 Seasonality Graph



Now in the Planner screen, set **Interval type** to **Prediction** and preview again. Note that the interval shows the same pattern as the forecast.

Figure 24 Seasonality Graph with Interval Type as Prediction



Close the graph.

Try Different Trending Methods

Next you will look at the Exponential trending method.

In preparation for producing the graphs in this section, make the following selections on the Planner screen:

- Every seasonality level is disabled (not checked)
- Interval type is None
- Forecast Ending date is June 1, 2005

Draw the forecast with Trending Method set to **Linear** as a reference. Note that this shows the Peak Disk value approaching 29 percent by June 1, 2005. See the first graph on the next page.

Now change the Trending Method to **Exponential** and draw the forecast again. Note the shape of the curve this time. Also note that this forecast shows the Peak Disk value approaching 100 percent by June 1, 2005. See the second graph on the next page.

Figure 25 Linear Graph

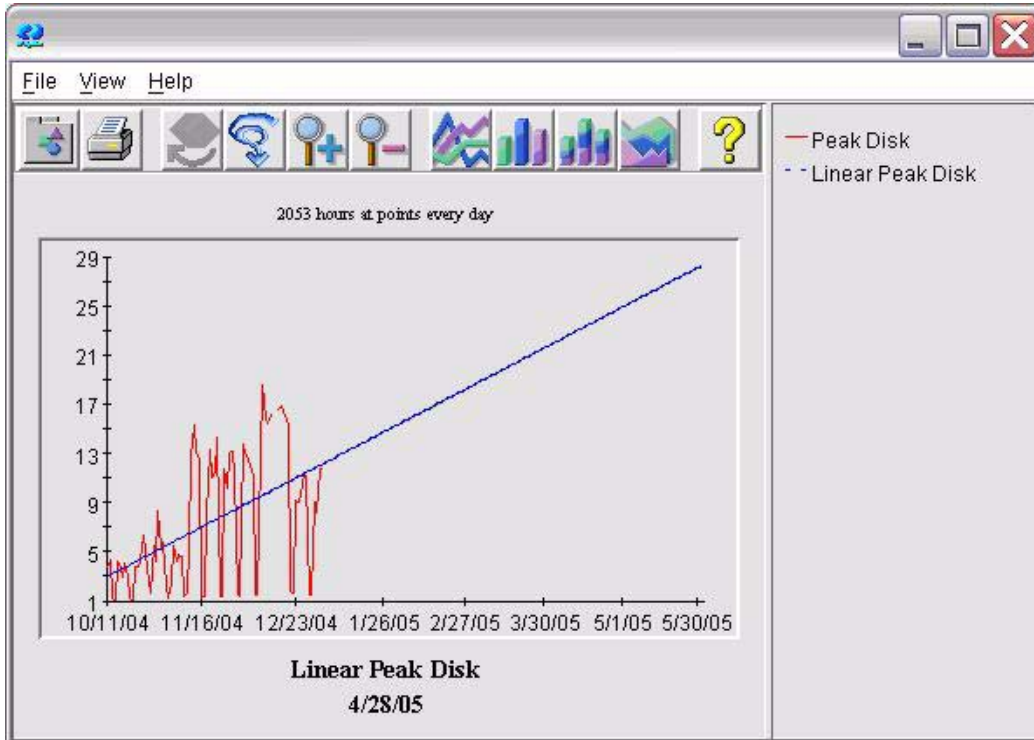
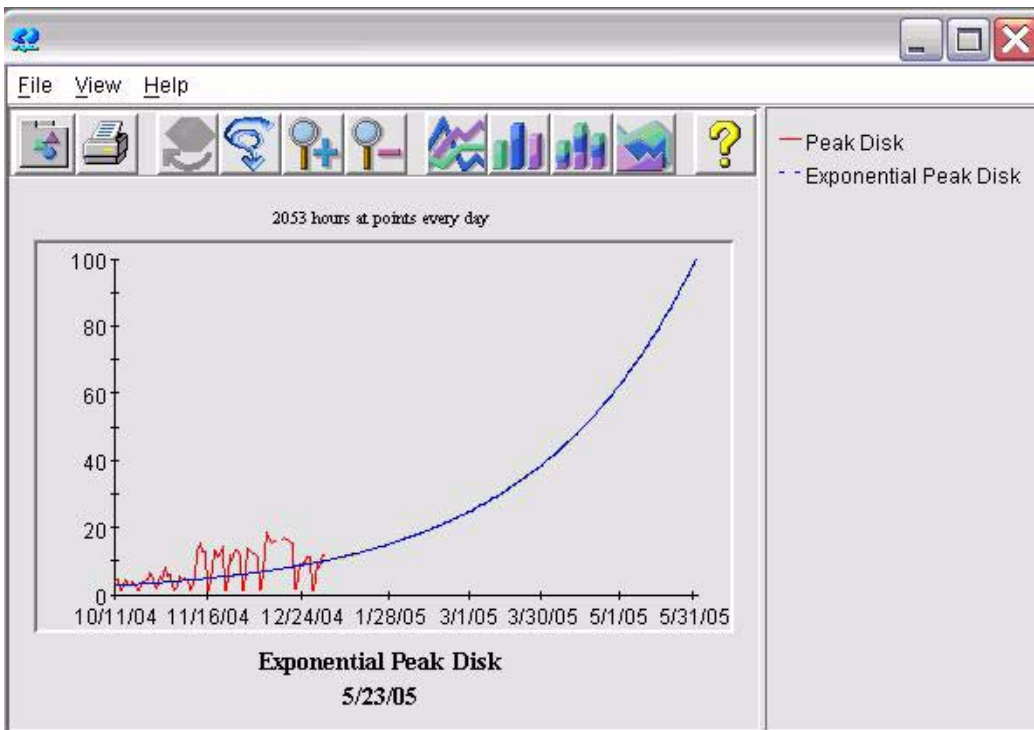


Figure 26 Exponential Graph



Metrics that are not bounded above often seem exponential. Peak disk utilization is such a metric. It looks exponential, but tends to become linear again as it grows.

Validating the Forecast

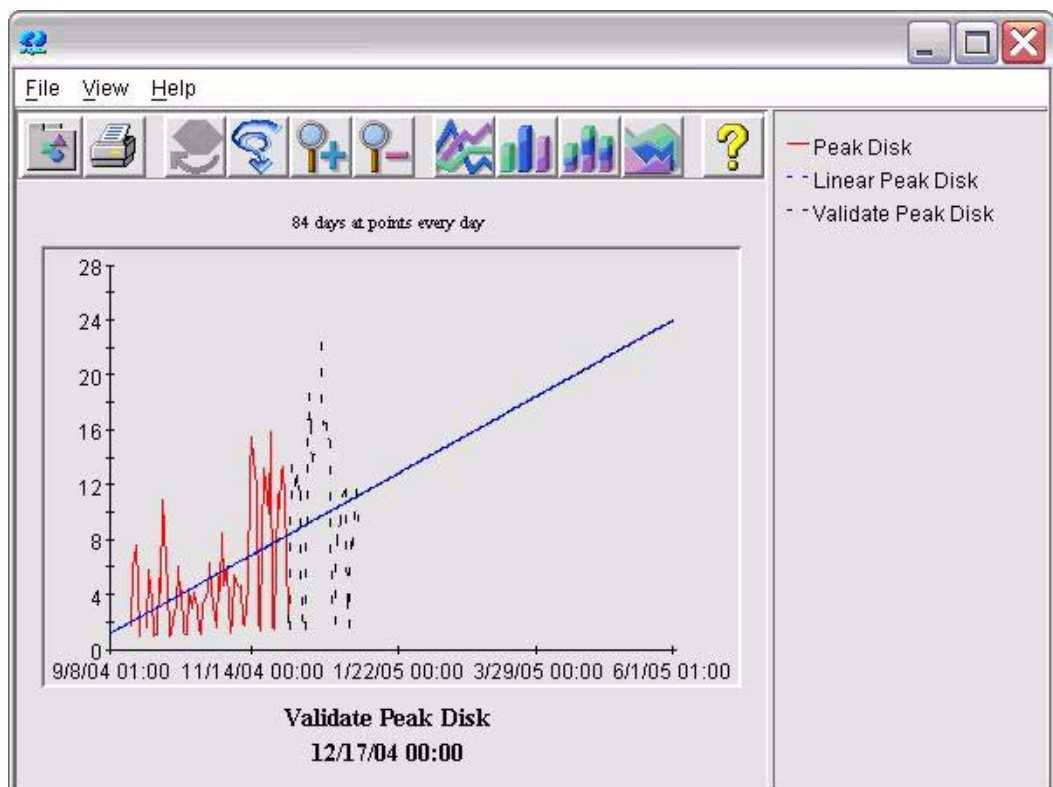
Once you have produced a forecast, you will want to check its validity. That is, you will want to confirm that it reasonably represents the system it was designed to represent. In this section, you will take a quick look at a way of validating a forecast using the **Show Validation data** option.

In the Planner screen set the options as follows:

- Trending Method = **Linear**
- Interval type = **none**
- Show Validation data = **checked**
- Date Range = **12 weeks**
- Ending = **Date (checked) Dec 1, 2004 Time 00:00**
- Forecast Ending = **June 1, 2005**
- Seasonality = All Disabled
- Points Every = **Day**

Checking the Show Validation data box will cause the forecast to display any data found after the Ending Date and before the Forecast Ending date. This data will be displayed, but it will **not be used in producing the forecast**. This data can be used to compare the forecast against what actually happened during this time.

Figure 27 Linear Graph with Validation Data



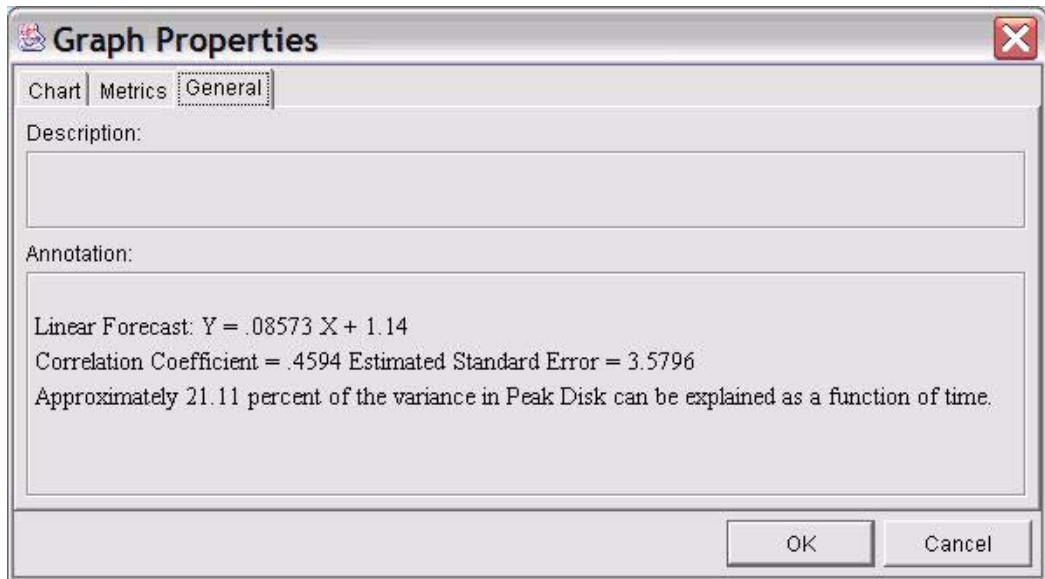
The red line is the actual data that was used to produce the forecast. The blue line is the forecast that predicts what values will be found in the future. The black line (Validate) is actual data that was logged but was hidden from the forecast. Comparing it against the forecast value is an indication of whether the forecast is producing reliable results.

An additional way to validate the data is to examine the Graph Properties window. Select **File > Properties** on the graph or click on the graph **Properties toolbar** button.



Then select the General tab in the Properties display and read the information in the Annotations section.

Figure 28 Graph Properties



The formula that the graph is based on is shown. See [The Forecast Equation](#) on page 50 for more details.

This chapter has given you a quick overview of the OVPM Planner. This would be a good time to try the Planner on your own, perhaps using your own logfiles.

The next chapter reviews a forecasting process you can use with OVPM Planner.

3 OVPM Planner and the Forecasting Process

This chapter introduces you to the forecasting process and how you can apply it to OVPM Planner. This process will assist you in producing more accurate and reliable forecasts.

There are five steps in the forecasting process:

- 1 Collecting data.
- 2 Selecting data.
- 3 Producing the forecast.
- 4 Validating the results.
- 5 Reforecasting as necessary.

Each of these steps is explained in the rest of this chapter.

Collecting Data

The first step in the forecasting process is to collect data. The more information you have - that is, the more data upon which you base your forecasts - the more accurate your forecasts will be. You can recognize trends with greater accuracy. And with better forecasts, you can do a better job of capacity planning. The OVPM Planner can forecast data from a variety of sources:

Table 1 Data Sources for OVPM Planner

Data Source	Description
OVPA	A very good source of system performance data. OVPM collects information about the key bottleneck areas on a system and other load factors. Application data is supported, especially through the use of HP OpenView Smart Plug-ins (SPIs). Transaction data is available for applications that are instrumented with the industry standard Application Response Measurement API (ARM). OVPA data can be archived for extended lengths of time and makes an excellent source for OVPM Planner.
HP OpenView Operations (OVO)	The OVO Embedded Performance Collector (OVO-EPC) collects up to five weeks of system bottleneck data. The HP OpenView SPIs can also log their data into this collector. OVO-EPC data can be used for short term forecasts (up to 5 weeks).
OVIS	OVIS actively probes and measures the availability and response times for a wide variety of network based applications. It can serve as an excellent source for overall application responsiveness.
HP OpenView Reporter	The HP OpenView Reporter gathers performance data from OVPA and OVO data sources and logs it to a central database. It can serve as an excellent source of planner data providing it is configured to retain a sufficient time span.
Generic Databases	OVPM supports the use of certain Oracle and SQL Server databases as data sources providing they meet certain requirements. See the OVPM Administrator's manual for details.
Flat Files	OVPM supports the use of text files as a data source. The files must be formatted according to some simple rules (comma or tab separated fields, header line with metric names, etc.) but are the simplest way to get custom data into OVPM Planner. Flat files are an excellent way to enter Business Unit data as described later.

See the OVPM Administrator's manual for details on how to configure OVPM so that it can locate each of these data sources.

Selecting Data

The next step in the forecasting process is to select the proper subset of data for a specific forecast. This requires choosing the following:

- The data source and perhaps the system where it is located.
- The metric to forecast.
- The range and dates of the data to be used in the forecast.
- The granularity of data to be displayed (Points Every).
- Shifts (days of the week and hours of the day to be included).
- Whether to ignore holidays.
- The summary level.

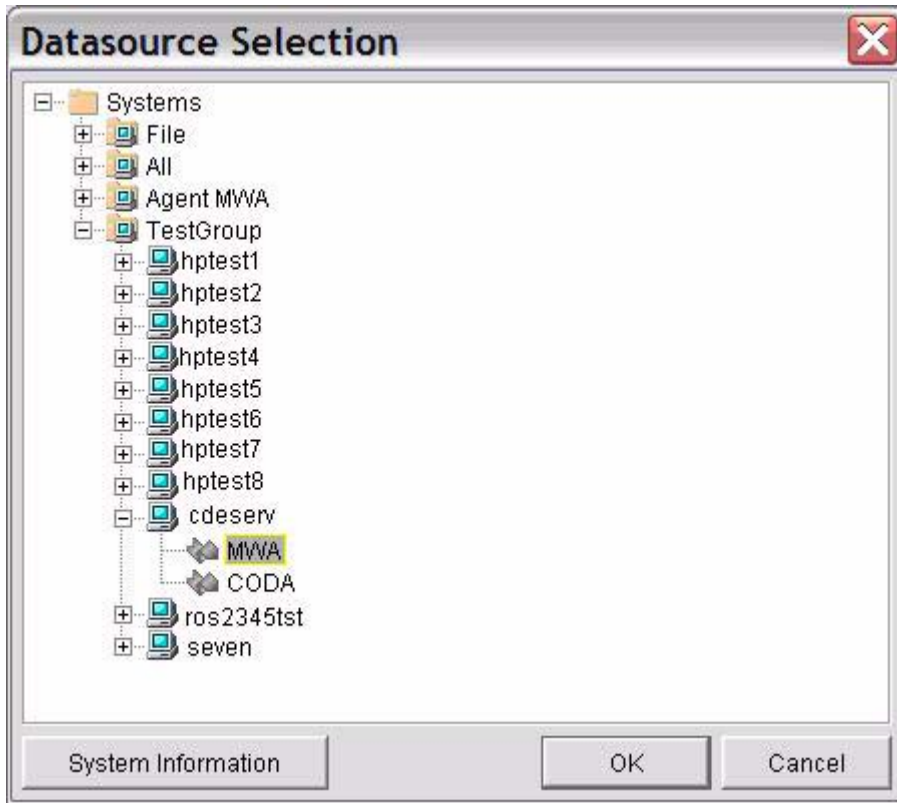
See the following pages for more information on each of these steps in selecting data.

Select the Data Source for the Forecast Metric

As part of selecting data you need to select the source of the forecast data. You should have an objective in mind when forecasting system performance and this will guide you in selecting the data source.

Select the data source as you normally would in OVPM. For example, in the OVPM Java interface Design mode, General tab select a graph Type of Forecast. Next select the Sources tab. Select the ... button on the Datasource line to display the Datasource Selection window. An example of the Datasource Selection window is shown below.

Figure 29 Datasource Selection Window



This window contains a tree with all available OVPm data sources. If you don't find the data source you want by expanding this tree control (+), then see your OVPm administrator for help in making the data source available.

Note that a system may have more than one data source supporting it, so make sure you select the last item in the tree control. In the example above, the `cdeserv` system has two data sources: `MWA` (OVPA data source) and `CODA` (OVO or OVPA 4 data source). A system may also have the data source `RepSys` (HP OpenView Reporter data source). When selecting a flat file data source, open the file name then select `File` as the data source.

Once your selection is made, press **OK**.

Select the Metric to Forecast

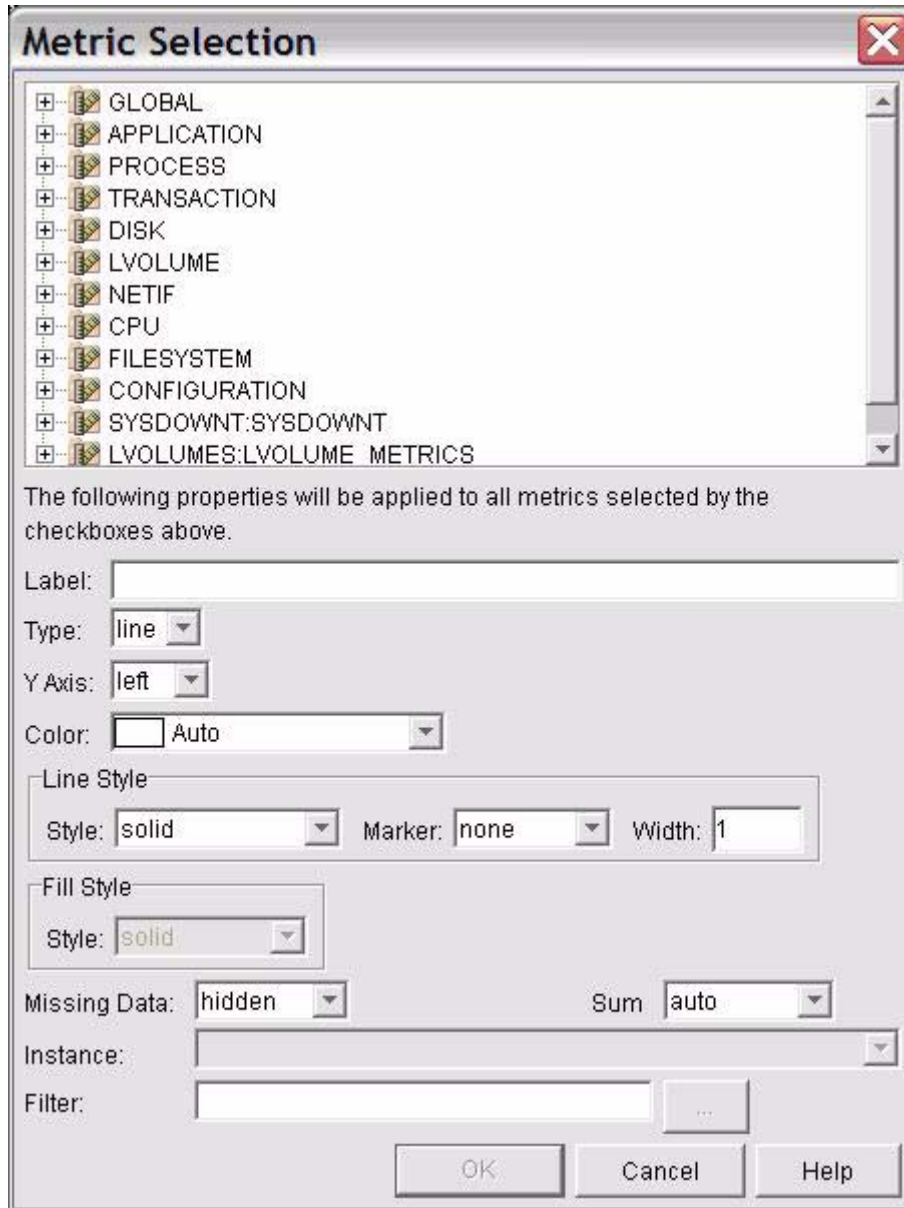
As part of selecting data you need to pick a performance metric that will be the basis of your forecast.

For example, in the OVPm Java interface Design mode General tab, select a graph Type of Forecast. Next select the Sources tab. Select the ... button on the Datasource line to display the Datasource Selection window. Select the data source to use in the forecast and press OK. Back on the Design screen, select the Metrics tab. Press the Add... button to bring up the Metric Selection screen. An example of the Metric Selection screen is shown below.



If you haven't selected a data source yet, then you will be taken automatically to the Datasource Selection dialog. Once you have chosen a data source, you will continue to select the metric.

Figure 30 Metric Selection Window



Metrics are divided up by the data source into different groupings called **Classes**. Each class contains a related set of metrics. Click on the class in the upper tree control to open it and show the individual performance metrics.

Choose one metric by clicking on it and causing a check mark to appear. If you select more than one metric then one forecast graph will be drawn for each metric.

You may fill in the lower part of the metric selection dialog if you want. It might be a good idea to enter a Label; otherwise the forecast will use the metric name to identify what is being forecast. Press **Help** to get more information on each of the fields in this dialog.

A **Multi-Instance** class can have more than one record for each time interval, usually with some key identifying the different records. If you have selected a metric from a multi-instance class then you should select an Instance or enter a Filter to select just the instance you want to use. Entering a filter such as MetricKey=@ will generate one forecast graph for each value of the MetricKey metric.

Suggested Forecast Metrics

The selection of metrics to forecast should reflect the specific needs of the environment. For example, forecasting values for CPU Utilization, Response Time, and I/O Rate can often be useful to maintaining an acceptable level of performance. The performance of particular applications may also be important to maintain.

Simply selecting all metrics with all applications and using all possible subsets of the data and shifts is neither productive nor reasonable.

If you are just getting started in forecasting, here are some suggested metrics from OVPA, HP OpenView Reporter and OVO to use in designing a forecast:

Table 2 Metrics from Agents

Class	Metric	Description
GLOBAL	GBL_CPU_TOTAL_UTIL	The amount of the CPU resource that is being consumed on the system. When CPU utilization nears 100%, processing will slow down and responsiveness will degrade.
	GBL_DISK_PHYS_IO_RATE	The number of transfers made to and from all disk drives.
	GBL_DISK_UTIL_PEAK	The percent busy of the busiest disk drive. (Which disk drive may change).
	GBL_DISK_VM_IO_RATE	The number of transfers to and from disk in support of memory management (swapping).
	GBL_FS_SPACE_UTIL_PEAK	The perfect full of the most full file system.
	GBL_NET_PACKET_RATE	The number of packets of information sent and received from network links.
	GBL_NUM_USER	The number of users logged in to the system. (This might make a useful business metric).

Table 2 Metrics from Agents

Class	Metric	Description
APPLICATION *	APP_CPU_TOTAL_UTIL	The amount of CPU consumed by this application.
	APP_IO_BYTE_RATE	The amount of data transferred to and from disk devices.
	APP_ACTIVE_PROC	The average number of processes in the application that were running during the interval.
	APP_COMPLETED_PROC	The number of processes that were run and terminated during the interval.
TRANSACTION *	TT_TRAN_1_HR_RATE	The number of transactions that completed each hour.
	TT_WALL_TIME_PER_TRAN	The average time it took for each transaction to complete.
	TT_SLO_PERCENT	The percent of the completed transactions that exceeded their pre-defined Service Level Objectives.

* These are multi-instance classes. You should pick your instance when selecting the metric. Selecting Instance of ***Prompt*** will let you choose which instance each time you draw the forecast graph.

If you have OVIS, you might want to forecast these metrics:

Table 3 Metrics from OVIS

Class	Metric	Description
IOPS_PROBE_DATA **	GROUP_COUNT	The number of probe measurements made to this target.
	AVAILABILITY	The percent of probe measurements where the probe was available.
	RESPONSE_TIME	The average response time, in seconds for each probe measurement in this interval.
	TRANSFER_TPUT	The average transfer throughput for each probe measurement in this interval.

** This is a non-keyed multi instance class. You should enter a filter to select the data you desire. Such as:

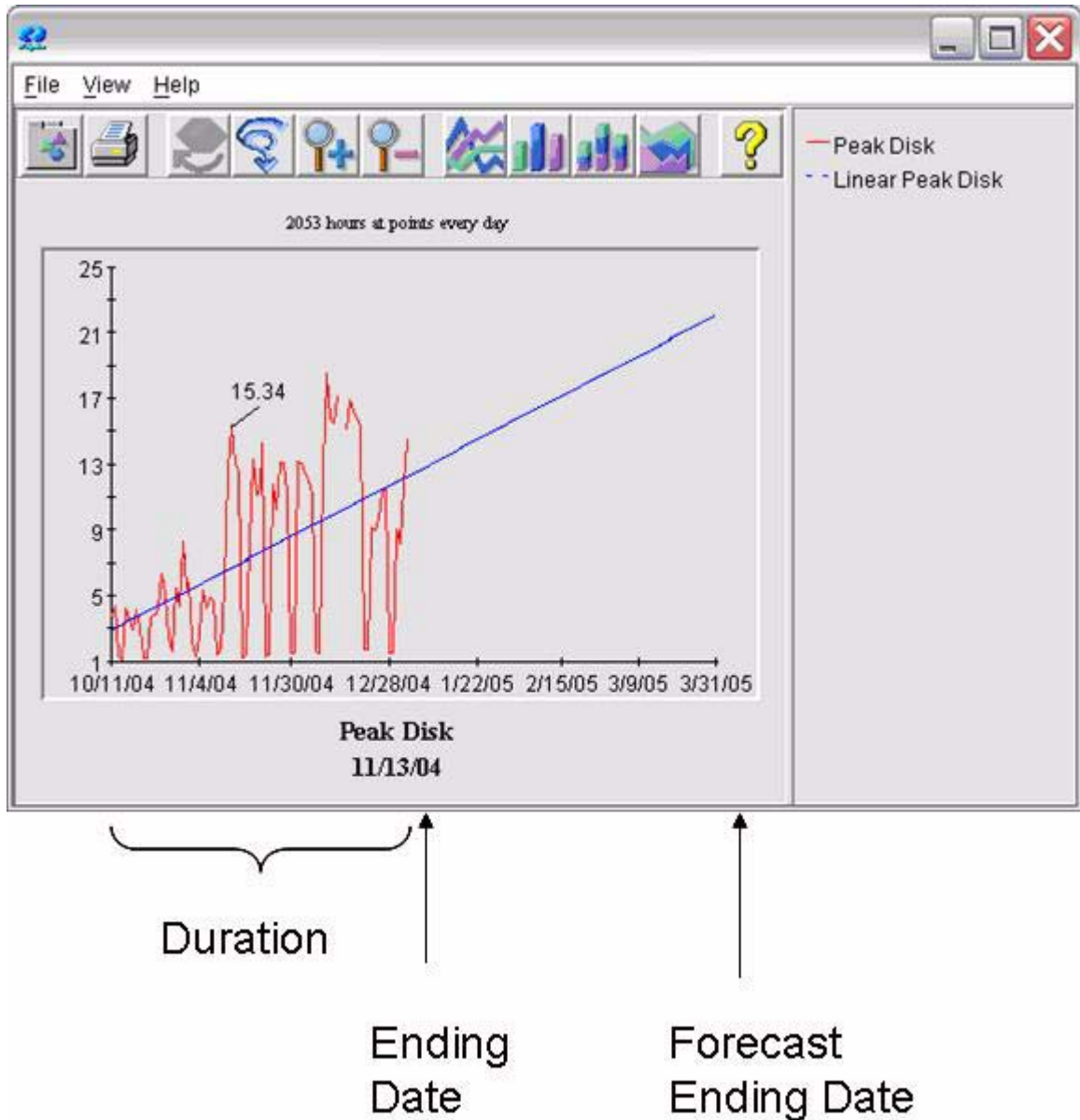
SERVICE_NAME="MyServiceGroup" or

PROBENAME=HTTP&&CUSTOMER_NAME="MyCustomer"

Select the Range of Data

As part of selecting data you need to choose the range of data you want included in your forecast. The range shown in a forecast graph consists of two parts: The range showing the collected data values for the metric that are used to construct the forecast and the range showing just the forecast.

Figure 31 Graph for Selected Date Range



The date range is defined by: Duration time, Ending Date, and Forecast Ending Date Each of these values are configured on the Planner screen (in the Java interface this is accessed by selecting the Planner tab when you are in Design mode).



Note that when doing a Preview of a forecast graph, the Date Range selections in the General tab in Design mode will not be used. The Date Range selections on the Planner tab will be used instead. For all other graph types, the settings on the General tab are used and those on the Planner tab are ignored.

Select the **Duration** as a number followed by a unit (minutes, hours, days, weeks.)

The **Ending** date can be selected as:

Now	Meaning the current date and time. This is useful when forecasting from live data.
Last	Meaning the last date and time in the log file. Useful when the forecasting data is coming from an archived log file or data source such as a flat file.
Date	In this case you select the ending date you want using the Date selection fields. Make sure that this date is actually present in the logged data.

The time frame you select is important because the actual data values within this time frame are the only information that OVPM Planner has available for making a forecast. A forecast is only as good as the data upon which it is based. In general, the accuracy of a forecast increases when it is based on data from a longer time frame.

You should avoid choosing a time frame that spans or includes any of the following:

- Hardware changes
- System Configuration changes
- Business workload redefinitions

Choose time periods of stable activity.

Set the **Forecast Ending** date using the corresponding date field. In order to be useful, the Forecast Ending date should be after the data Ending date. As a rule, the Forecast Ending date should not be set so that the forecast duration (Forecast Ending date minus the data Ending date) exceeds the data Duration setting. You should not trust forecasts that attempt to forecast for a range that exceeds the data duration range.

Forecast Ending date is not used when the trending method is set to Smoothing since the smoothing method does not predict any future values for the metric.

Select Points Every

As part of selecting data you need to define the base granularity for the forecast graph. A graph with points every hour will have more points than a graph with points every day.

The granularity is set using the **Points Every** value. Setting a small Points Every value can increase the statistical accuracy of a forecast, but the results may be too busy to visualize properly. For longer date ranges, you should select larger Points Every values.

The Points Every value is configured on the Planner screen (in the Java interface this is accessed by selecting the Planner tab when you are in Design mode).

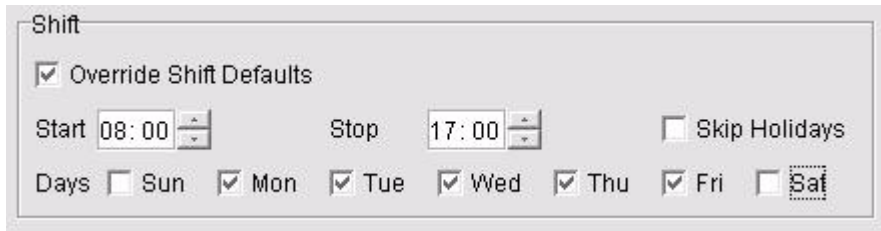
Select Shifts

As part of selecting data you can choose to include only data during certain times or days. You do this by configuring shifts. A shift is the part of the day in which you are interested. Thus the shift is always a time period of less than 24 hours and is often 8 hours or less.

The shift you select should reflect your purpose in making the forecast. If you are interested only in the peak period of the day, then you should choose the peak period as a shift. If you are interested only in performance when key applications run, then you should choose that time range as a shift. If you are interested in the average performance, then you should use the entire day as your time window.

You configure Shifts for a forecast using the General tab in Design mode.

Figure 32 Shift Section



To enable shifts (for certain times or days, check the **Override Shift Defaults** box.

Set **Start** and **Stop** to the time range you want included each day. These time fields use 24 hour time format. In the example above, the selection is for 8:00 AM to 5:00 PM (08:00 to 17:00). The shift can cross over from one calendar day into the next (for instance Start 21:00 (9:00 PM) and stop 07:00 (7:00 AM).

Check the boxes for the **Days** of the week to be included in the forecast data. Many systems do not run the same every day of the week. It is not unusual for a system to be used in one way throughout the week (Monday through Friday) and then differently on the weekend (Saturday and Sunday). Of course, your system may have different days of the week that you want to study.

You can eliminate the data for certain days of the week by checking **Override Shift Defaults** and then unchecking the days that you don't want. In the example above, only data for Monday through Friday will be used.

This mechanism can be combined with shift times in order to allow very specialized studies. Say that you want to study the Monday morning peak activity between 9:00 and 10:00. You can set the shift times to 09:00 and 10:00 and then uncheck each day except Monday. Now the data on the graph will show only data from this time period each week.

Also note that the graph display will compress out blank spots where data is not used; thus producing a continuous graph of only the selected days and times.

Skip Holidays

You can eliminate holidays from the data by checking the **Override Shift Defaults** box and the **Skip Holidays** box in the General tab in Design mode (see the screen shot shown above).

A holiday often results in a disruption of the data being forecast. If the system usage falls off due to a holiday, it can make the forecast lower than what it should be.

A list of business holidays is defined by the OVPM Administrator outside of the OVPM user interface. Please see the OVPM Administrator manual for details on how to configure holidays.

Producing the Forecast

The next step in the forecasting process is producing the forecast.

Producing the forecast requires choosing the following:

- Trending Method
- Certain Planner options
 - Seasonality Level
 - Interval Type
 - Confidence Level

Trending Method

The Trending Method is set using the Planner screen (in the Java interface this is accessed by selecting the Planner tab when you are in Design mode).

There are five Trending Methods to choose from:

- Smoothing
- Linear
- Exponential
- S-Curve
- Business Units

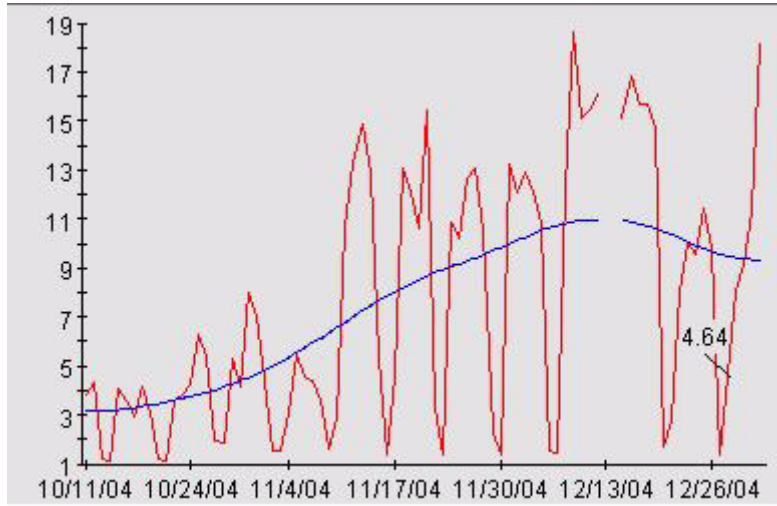
Each Trending Method is discussed in more detail below (except Business Units which will be covered in the next chapter).

Generally speaking, you would choose **Smoothing** as your first trending method. Smoothing allows you to preview the actual data prior to creating a forecast. On a smoothed graph you can more readily detect the presence of a trend.

Also on a smoothed graph you can detect potential disruptions to the metric values that you may want to accommodate in your date range selections.

An example of a graph with smoothing Trending Method is shown below.

Figure 33 Smoothed Curve Graph



Note that a smoothed graph will show dates for the actual data points only; it does not show predicted values for future dates. Also since it does not actually forecast data into the future, a smoothing graph will not use the Forecast Ending date.

The kinds of trend you would try to identify on a smoothed graph include the following: Linear, Exponential and S-Curve. See [Concepts in Statistical Forecasting](#) on page 8 for more information on each of these types of trends.

A **Linear** trending method will attempt to match a straight line to the data values. It is the most commonly used trending method.

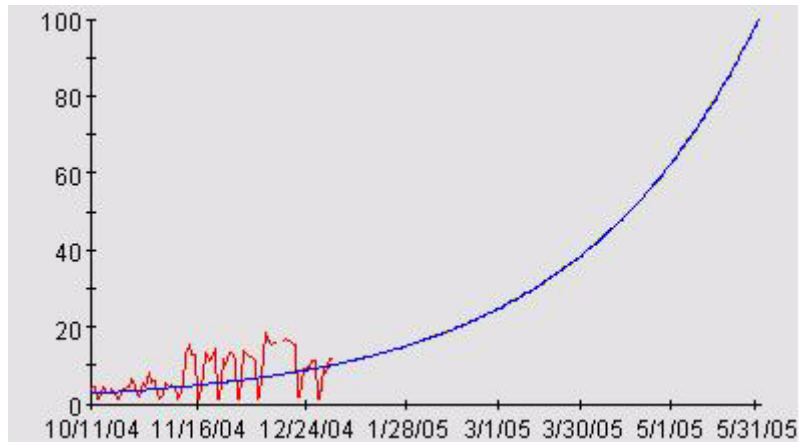
Figure 34 Linear Graph



Linear trending is appropriate if a straight line can reasonably represent the set of points. If your data values vary in a pattern each day or throughout the week, etc. then using **Seasonality** can account for these variations and improve the linear forecast's accuracy.

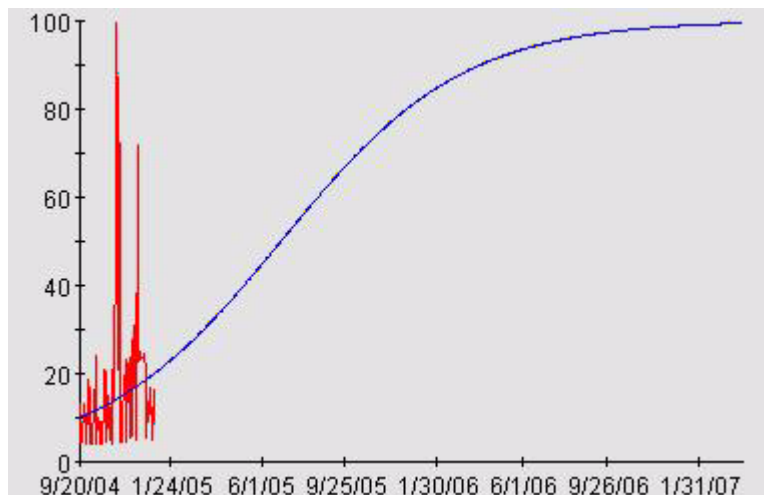
An **Exponential** trending method fits a curved line of exponential shape to the data values. It is useful when you know that the metric values tend to behave in a non-linear fashion. An example of a good use of the Exponential trending method is when forecasting response time values.

Figure 35 Exponential Graph



An **S-Curve** trending method fits an "S" shaped curved line to the data. The curve assumes that there are minimum and maximum values that it can approach but never actually achieve.

Figure 36 S-Curved Graph



The S-Curve trending method is often useful for forecasting bounded metrics such as CPU Utilization. CPU Utilization can never be less than zero or more than 100. When using the S-Curve trending method, you should insure that the **S-Curve Min** and **S-Curve Max** values in the OVPM Planner tab are set properly to the minimum and maximum value that the S-Curve can achieve.



An S-Curve can be very difficult to see because you have to see the entire curve to detect its shape. Any part of the curve alone may look like another trend type. For instance, the beginning and end of an S-Curve may look linear, while the middle looks exponential (that is, has an increasing or decreasing slope).

If the time period does not display steady state change according to a Linear, Exponential or S-Curve trend, then the only remaining possibility you have is the **Business Units** method. If you can correlate changes in system performance to changes in a business unit, you can generate a valid forecast. See the next chapter for details on creating and using a business unit file.

Seasonality

Seasonality refers to the repeating patterns a metric value might exhibit over time. It is not unusual to see a metric that varies in a pattern based on the hour of the day or day of the week. (See the OVPM Baseline graphs for more information on these types of seasonality.) In some cases, the metric value may also be affected by business cycles that follow the week of the month or month of the year.

These patterns of usage can dilute the accuracy of a forecast if they are not accounted for. One way to account for them would be to use the Shifts feature to select only a stable hour of the day and day of the week to be included in the forecast. This assumes that you know which time period is the most important for your forecast.

Another way to deal with the periodic usage patterns is to subtract them out prior to the forecast, then replace them in the forecast curve itself. Since it is often difficult to recognize what types of seasonality exist in the data, OVPM Planner can automatically detect and account for four types of seasonality patterns:

Seasonality	Pattern varies for	Only valid when
Hour of the Day	each hour of a 24 hour day	Points Every \leq 1 hour
Day of the Week	each day of a 7 day week	Points Every \leq 1 week Duration > 1 week
Week of the Month	each week in the month, based on calendar months	Points Every \leq 1 day Duration > 30 days
Month of the Year	each calendar month of the year	Duration > 3 months

When a Linear, Exponential or S-Curve forecast is generated, any seasonality patterns you have specified will be examined. You select these seasonality patterns using the Planner screen (in the Java interface this is accessed by selecting the Planner tab when you are in Design mode).

Figure 37 Seasonality Section



If any seasonality pattern is checked, it will be tested against the data to determine if it is allowed (based on Points Every and Duration) and if it improves the forecast accuracy (reduces the forecast Standard Estimated Error value.) If so then it will be included in the final forecast. If a seasonality pattern isn't allowed or doesn't improve the forecast then it is ignored.

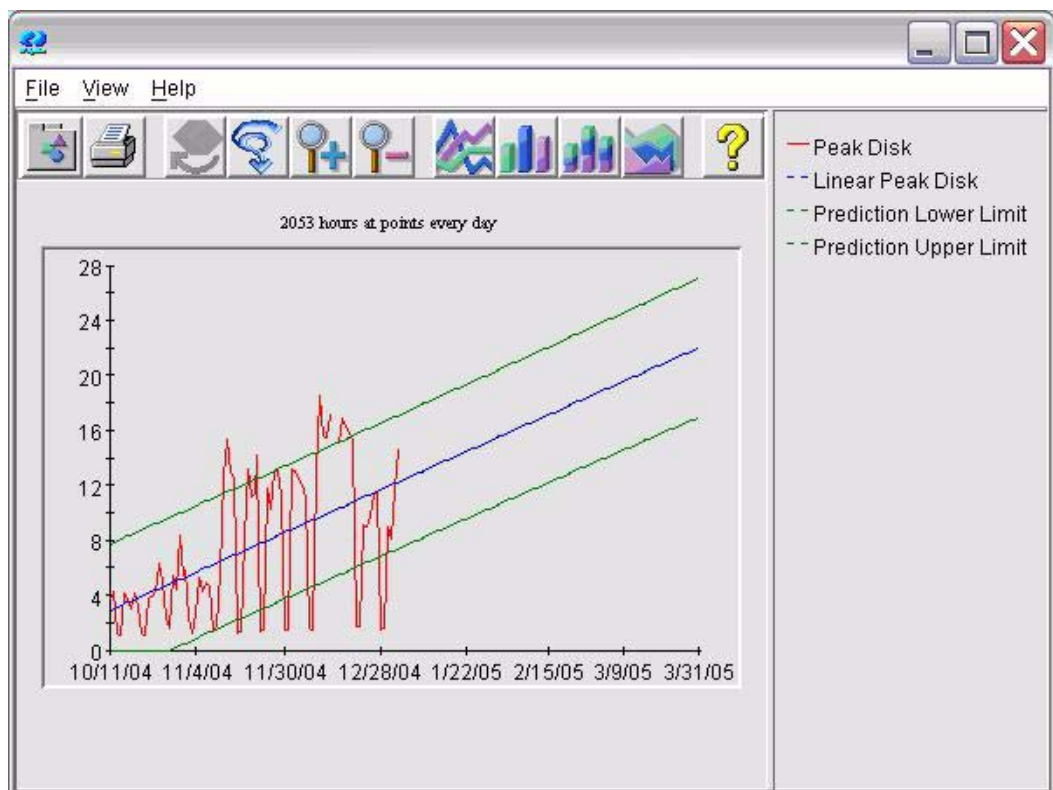
A list of which seasonalities were actually used is shown in the forecast graph annotations. From a Java graph, select the **File > Properties** menu item then select the General tab to view the annotations. On an HTML graph, the annotations are included in the web page that contains the graph. See [Seasonality Information](#) on page 52 for more information.

- ▶ You may find that some seasonality patterns make it harder to visualize forecast graphs with a lot of data. You may want to disable the lower seasonality levels such as hour of the day and day of the week to make the graph less cluttered.

Confidence and Prediction intervals

Confidence and prediction intervals are a visual way to indicate the quality of a forecast. They both add two extra lines to a graph to indicate the upper and lower limits and so define a range of values around the forecast curve.

Figure 38 Confidence Interval Graph



You select these **Confidence Lines** using the Planner screen (in the Java interface this is accessed by selecting the Planner tab when you are in Design mode).

Figure 39 Confidence Lines Section

There are two **Interval types**: Confidence and Prediction.

The **Confidence** interval draws a range that should encompass a specified percent of the mean values of the data. That is to say, a 70% confidence interval would indicate that 70% of all forecast curves (the mean values) would fall in between the lower and upper limit lines.

The **Prediction** interval draws a range that should encompass a specified percent of the data points. A 70% prediction interval would indicate that 70% of all data points should fall between the lower and upper limit lines.

Note that by their definitions, a prediction interval will always be wider than a confidence interval for the same forecast. (A confidence interval encloses the mean values while a prediction interval has to enclose the points themselves which can vary more widely.)

The **Confidence Level** sets the percentage that is referenced in the Confidence and the Prediction interval type descriptions. A larger confidence level value will set a wider range since it indicates that the limit curves should encompass a larger percentage of the values.

Validating the results

After you have produced a forecast, you should validate it to confirm that it reasonably represents the system that it is designed to represent.

The following tools are used to validate a forecasting model:

- Performance experience
- Validation data
- Confidence and prediction intervals
- Statistics measures

Validating a forecasting model means testing the model to see if it seems likely that the forecast it generates will be accurate enough to be useful.

You can use any or all of the following tools to validate the forecasting model. It is not necessary to use all of them. In many cases, you can determine forecast validity by using just a few of the validity tools listed below.

Performance Experience

The experienced systems performance manager understands the variables that affect system performance and knows what behavior is reasonable. Using the reasonability test helps one avoid ludicrous results in a forecast.

Show Validation data

The Show Validation Data option in OVPM Planner lets you use a portion of the actual recorded set of performance data to validate the forecast.

Any data found after the "Date Range: Ending" date is hidden from the forecasting calculation but then is revealed after the forecast is generated. You can then compare the predicted data to that portion of the actual data to evaluate the accuracy of the trending method.

If the forecasted data and the actual data are close, this increases your confidence in the trending method and lends credibility to the forecast values past the validation data. If, however, the forecasted values and actual data differ a great deal, it is strong evidence against the validity of the trending method or other forecast settings.

Confidence or Prediction Intervals

Intervals, which give upper and lower bounds on a forecast, can help you validate a forecast. A narrow interval (small spread between the upper and lower bounds) tends to indicate a valid or credible forecast. Conversely, a broad interval may indicate a less credible forecast. Intervals - whether prediction or confidence - allow for best-case and worst-case analysis.

Statistical Measures

From every forecast graph, you can access the textual report (annotations) that displays, among other information, the following statistical measures:

- Correlation Coefficient
- Estimated Standard Error
- Percentage of the data variance attributable to time (or Business Units)

See the following section Understanding the Statistical Numbers for guidance on how to access and interpret these values.

You can use one, several, or all of these validity checks to analyze your forecast. Typically you will use common sense, confidence intervals, and the statistical measures to validate your forecasts.

If the forecast fails any of the validity checks, it is not valid. You can then either repeat the forecasting process again to try to develop a valid forecast, or you can decide that a valid forecast is not feasible. It is possible that the chosen forecast metric has behaved in an unpredictable manner during the measurement period that you selected.

Understanding the Statistical Numbers

A brief statistical report is provided for each forecast graph. For HTML graphs and for Java graphs accessed from the Web Forms user interface, this report is printed immediately following each forecast graph. For Java graphs access from the Java interface, the report is located by opening the graph Properties (by either selecting the **File > Properties** menu item in the graph or by pressing the graph's Properties button in the toolbar, and then selecting the General tab. An example of the statistical report is shown below along with some notes on how to determine the goodness of fit of the forecast based on the statistical measures:

Linear Forecast: $Y = -.10071 X + 16.25$
 Correlation Coefficient = $-.1442$ Estimated Standard Error = 5.5946
 Approximately 2.08 percent of the variance in CPU can be explained as a function of time.

Goodness of Fit — Short Notes

The type of forecast and the **forecast equation** used in the statistical model is shown along with the following statistics:

Statistic	Range of values	Interpretation
-----------	-----------------	----------------

Goodness of Fit — Short Notes		
Correlation Coefficient	-1 to 0	Forecast is decreasing, Closer to -1 is better
	0 to 1	Forecast is increasing, Closer to 1 is better
Estimated Standard Error	0 to metric value	Lower is better (same units as the forecast metric)
Variance as a function of ...	0 to 100%	Higher is better

The Forecast Equation

The first line in this report details the equation of the forecast line.

"**Y**" represents the metric value plotted against the Y-Axis in the graph

"**X**" represents the X-Axis value. This will usually be the time, expressed as a "point number". The first displayed value is point 0, the next is point 1, and so on.

When doing a Business Units graph, "X" will not represent time but will represent the value of the chosen business units metric instead.

The rest of the equation will depend on the type of trending method selected.

- For **Linear** forecasts, the equation will be in the form:

$$Y = \text{slope } X + \text{intercept}$$

Where slope and intercept are numbers (in the example, Linear Forecast: $Y = -.10071 X + 16.25$, slope is $-.10071$ and intercept is 16.25). And X is usually the number of time periods since the start of the forecast.

A negative slope means that the forecast is decreasing as time increases.

- For **Exponential** forecasts, the equation will be in the form:

$$Y = \exp(\text{slope } X + \text{intercept})$$

The argument in the parentheses is the same as the linear equation. That is, you need to know the intercept, slope and the number of time periods since the start of the forecast. Once the value inside the parentheses is calculated, you calculate exp or the exponential number e (i.e., 2.718...) raised to the power of that argument.

- For **S-Curve** forecasts, the equation will be in the form:

$$Y = Y_{\text{Max}} / (\exp(\text{slope } X + \text{intercept}) + 1)$$

Y_{Max} is the maximum value for the metric as selected in the OVPM Planner.

- For **Business Units** forecasts, the equation will be in the form:

$$Y = \text{slope } X \text{ BusinessUnits} + \text{intercept}$$

Where X no longer represents time, it represents the Business Units value as specified in the OVPM Planner.

If you want to be able to calculate the forecast value in other tools (such as spread sheets) then you should be able to transfer these equations, providing you follow the rules for providing X values.

The Correlation Coefficient

The **Correlation Coefficient** is often referred to as "**r**" or "**R**" in many statistics books. It has also been called the "product-moment coefficient correlation" or "Pearson's correlation". The correlation coefficient measures the tightness of fit of the forecast line to the actual metric values.

The correlation coefficient ranges from -1 to 1. Negative R values indicate that the slope of the line is negative (that is, the line goes down and to the right). Positive R values indicate that the slope of the line is positive (that is, the line goes up and to the right). So that the correlation coefficient also indicates whether increasing X makes Y go up or down.

A correlation coefficient with an absolute value of 1 represents a perfect fit (either +1 or -1) and a correlation coefficient of 0 represent a bad fit. The closer your forecast's correlation coefficient is to 1, the better the forecast line fits the data. A correlation coefficient of 0 indicates that the forecast is no better at predicting the data than guessing random numbers would be.

Variance as a Function of...

The squared correlation is the proportion of the variation in the metric that is explained by time (or business units). It is often referred to as **R-Squared**. The value for R-Squared is expressed as a percent and is shown in the OVPM forecast graph statistics report in the line:

"Approximately <value> percent of the variance in the metric can be explained as a function of time (or Business units)".

Where value is the R-Squared value times 100%.

R-Squared is simply the square of the correlation coefficient R (described above). R-Squared will always be a positive value between 0 and 1, with values closer to 1 indicating better forecasts. R-Squared is also called the Coefficient of Determination.



Note: Both the Correlation Coefficient and R-Squared values indicate how well the forecast metric can be predicted by the X value (time or business units). If the metric value is not changing relative to the X value then these numbers will be close to zero, indicating a bad fit. This doesn't mean that the forecast line can not be trusted. It merely means that the X values chosen are not effective at predicting changes in the forecast metric.

As an example: If your forecast metric always has the same value (say 5.0) then as time changes the metric value stays the same. The forecast will have a slope of zero, indicating no change, and the R and R-squared values will be zero. On the other hand, if the forecast metric has always been 5.0 then it will likely remain at this value so the forecast line, predicting 5.0 into the future, may well be accurate. This is where that "common sense" part of the validation comes into play. See the [Estimated Standard Error](#) measure for a different approach to measuring forecast goodness.

Estimated Standard Error

The Estimated Standard Error, often called simply "Standard Error" is a measure of how well the forecast line fits the data points.

It is calculated by taking the distance between the forecast line and the actual data value at each point. To make the calculations work we actually sum the square of this difference and when we are done, take the square root of this value divided by the number of points. A value for Standard Error close to 0 is optimal.

The units of the Standard Error are the same as the units of the forecast metric. Thus, if you are forecasting Response Time in seconds and get a Standard Error of 2.5 then you can assume that "typically" the forecast is 2.5 seconds away from the actual data.



The Estimated Standard Error value is not dependent on how well the X value predicts the forecast metric. Instead it shows how close the forecast line lays to the actual data points. For this reason, a small Estimated Standard Error value can indicate a good forecast even when the forecast metric does not vary much.

Seasonality Information

If you selected any of the Seasonality patterns when configuring your forecast, and you are doing a linear, exponential, or S-curve forecast, then a list of which seasonality patterns were significant is shown following the statistical report (as shown in the examples below).

Seasonality will be applied for:

Month of the year

Week of the month

Day of the week

Hour of the Day

Or

No Seasonality Intervals improved the forecast.

This lists the types of seasonality that were used to produce the final forecast. A particular type of seasonality will be used only if it is selected in the User Interface and its use improves the "fit" of the forecast. In this case, "fit" is measured by decreasing the value of the Estimated Standard Error value for a forecast.

Some seasonality patterns may be eliminated if the data selected doesn't allow their effective use.

In order to use	The data must be
Hour of the Day	PointsEvery <= 1 hour
Day of the Week	PointsEvery <= 1 week Duration > 1 week
Week of the Month	PointsEvery <= 1 day Duration > 1 month (30 days)
Month of the Year	Duration > 3 months

Confidence and Prediction Interval Information

If you selected confidence or prediction intervals when configuring your forecast, this information plus the confidence interval value will be displayed at the end of the statistics report.

Confidence Interval for forecast line at <value> percent.

Or,

Prediction Interval for forecast points at <value> percent.

Reforecasting

If the forecast is not valid, you may repeat the forecast process, changing options that may allow generation of a valid forecast.

OVP M Planner accommodates trial and error because it allows you to easily change the options (such as Shift, Points Every, Data Date range) and because it requires little time to produce a forecast.

If necessary, adjust and reforecast or collect more data and return to the beginning of the forecasting process.

This could mean selecting a different date range, enabling or disabling seasonality, choosing a different forecasting method, or looking at a subset of application data.

It may be helpful if, after obtaining a successful forecast, you reforecast with minor changes. As an example, you may find that your forecast is dramatically dependent on the exact data date range you have chosen. Example: If you get a forecast that predicts an increase in value but a slightly different date range gives you a decreasing forecast then you might be suspicious of your results.

4 Business Unit Forecasting

This chapter explains what a business unit is and how to create and use a business unit file. Business unit forecasting can be used where metric values are not varying smoothly over time, but may be affected by some external measure of your business activity. If you can predict how this business activity will change, then you can use this prediction to forecast how that change will affect your other metrics.

Even if you do not use the Business Units trending method initially, you may want to start logging possible business units immediately. This practice proves invaluable for generating a business unit forecast at a later date.

Understanding a Business Unit

A business unit is a quantity that reflects the kind of business you do and its activity. It is something a business planning department understands; it is not a computer metric.

Most businesses are complex. Your business probably can be thought of as having numerous business units. For instance, your company may receive orders that it must fill. The number of orders received is a potential business unit. So is the number of units shipped. Other possible business units might be the number of mail messages processed, the number of customers handled, the number of web pages served, or the number of users. Defining a business unit can be difficult and deceptive. For instance, although your business may be heavily concentrated in order processing, it may be more appropriate to use the number of line items as a forecasting business unit.

Of course, your interest in business units begins when they impact one or more aspects of your IT infrastructure. Suppose you know that the number of shipments is going to double within the next year. You know that the system capacity is going to be impacted; but how?

OVP M Planner includes a Business Units trending method that can help you forecast the impact of changes in a business unit upon your system. But to be meaningful, this method requires careful work on your part. You will need to do the following:

- 1 Identify the business unit.
- 2 Gather historic data on the business unit.

You may experience a delay between the time you identify the business unit and the time when enough historic values of the business unit are available to generate a reliable forecast.

- 3 Collect expected future values for the business unit.

Typically, this forecast would be derived from a business plan or projections made by a planning group.

- 4 Construct a business file.

A Business unit file is flat ASCII file that can be created with any editor. It can also be generated as a spreadsheet and then exported to a flat ASCII format that can be used by OVPM.

5 Correlate the business unit to a metric.

For example, the number of orders received in a factory could be expected to have a high correlation with the CPU Utilization of an order processing system and would thus be a good candidate for a business unit.

Creating a Business Unit File

While OVPM Planner can use any metric available to OVPM as a business unit metric, it is more likely that the business unit will come from a flat file data source. The main reason for this is that the business unit should have historical values PLUS predicted future values. It is important for a business unit to have predicted future values so the forecast can use those to guide how metrics will behave in reaction to them.

Once you have gathered historic data and expected future values for a business unit, you are ready to create a business unit file. A business unit file is actually an OVPM Flat File.

OVPM Flat File Format

The OVPM Flat File data source is a mechanism designed to allow quick and easy addition of external data into the OVPM product, and particularly the OVPM Planner function. The file format is essentially an ASCII text file with one record per line and each field in a line separated by a tab or a comma. A special header line at the top of the file defines the contents of the file, including the metric names and the order.

The flat file format supports multiple metrics in one file. Only one class will be seen in OVPM for each file data source. It will be called "FILEDATA". Multiple instances can be supported by simply including a field that can be used to differentiate each one.

The files used by this facility must reside on the OVPM server disk. If a client system wants to access one of their own files, they must arrange to have it transported to the OVPM server system.

File names and locations:

For ease in locating and processing the flat files, they should be named with no embedded spaces and terminate with the file extension `.bus` (for Business Units). Example:
`MyOwnData.bus`.

By default, OVPM will look for `.bus` files in the logon customer's home directory.

`<data_dir>/conf/perf` for the default logon

`<data_dir>/conf/perf/VPI_Cust_logonname` for the "*logonname*" customer logon.

Other file names and locations are possible, but will not have as much support in the OVPM user interfaces.

Example File:

```
MDY
TIMEp,TempOutside,WindChill,HiTemp,LowTemp,WindSpeed,WindHi,"WindDir",Rain
1/01/05 12:00a, 41.3, 38.6, 41.9, 40.8, 5.0, 16.0, ESE ,0.04
```



```

1/01/05 2:00a, 41.6, 41.6, 42.1, 38.9, 4.0, 10.0, ESE ,0.00
1/01/05 4:00a, 39.5, 39.5, 40.4, 38.2, 3.0, 8.0, E ,0.01
1/01/05 6:00a, 40.3, 40.3, 40.5, 40.0, 2.0, 7.0, E ,0.00
1/01/05 8:00a, 40.1, 40.1, 40.6, 39.4, 2.0, 5.0, E ,0.00

```

Flat File Format particulars:

- Each file contains ASCII data consisting of one or more numerical and textual fields.
- Each line in the file represents one data sample. Except for the first line which is a header line.
- Blank lines and lines starting with a pound sign "#" are ignored

Header Line:

The first line in the file is a header line. It identifies the fields in the file. Each field in the header line contains the date/time specifier or a metric name. These fields are separated by a delimiter character (tab or comma). A metric name enclosed in double quotes indicates that this field contains a textual metric. An unquoted metric name indicates the field is numeric.

Metric names should not contain any delimiter characters (space, tab or comma).

The date specifier in the header line consists of the characters Y, M, D shown in the order of the date field

(YMD = Year, Month, Day)

(MDY = Month, Day, Year), etc.

The time specifier consists of the string TIME and optionally an AM/PM indicator. If time is in 24 hour format then TIME is all you need. If the time is in 12 hour format then enter the indicator that will be found in the time strings to indicate PM (For example: TIMEPM). Note that TIME is not a separate column but is considered a part of the DATE/TIME column.

Data lines:

- Each field is separated from the others by a delimiter character (tab or comma).
- Textual fields that contain delimiters must be enclosed in double quotes. If no special characters are in the field then the double quotes are optional.
- Numerical data can be integer or floating point but will be treated as floating point inside of OVPM.
- The first field in each line must be the date and optionally the time of the sample. Dates can be specified in any order of Year, Month and Day separated by non-numeric, non-delimiter characters (as specified in the header line).

Time format is Hour then Minute, separated by a non-numeric, non-delimiter character. If time is in a 12 hour format, then a trailing AM/PM indicator can be present.

- There is no limit to the number of date lines than can be accepted except that performance will suffer as the number of data lines is increased. 10,000 lines is a reasonable expectation for the maximum practical business units file.

Using a Business Unit File

You can select a business unit file that you have created, using the Planner screen (in the Java interface this is accessed by selecting the Planner tab when you are in Design mode).

First, insure that your business unit file has been properly created and is now on the OVPM server in the proper directory with a file extension of `.bus`.

In the OVPM Planner screen, select the Business Units Trending Method, then using the **Source...** and **Edit...** buttons on the Business Metric line, select the business units file and metric. The **Source...** button is used to choose a different data source, the **Edit...** button is used to choose a different metric in the same data source.



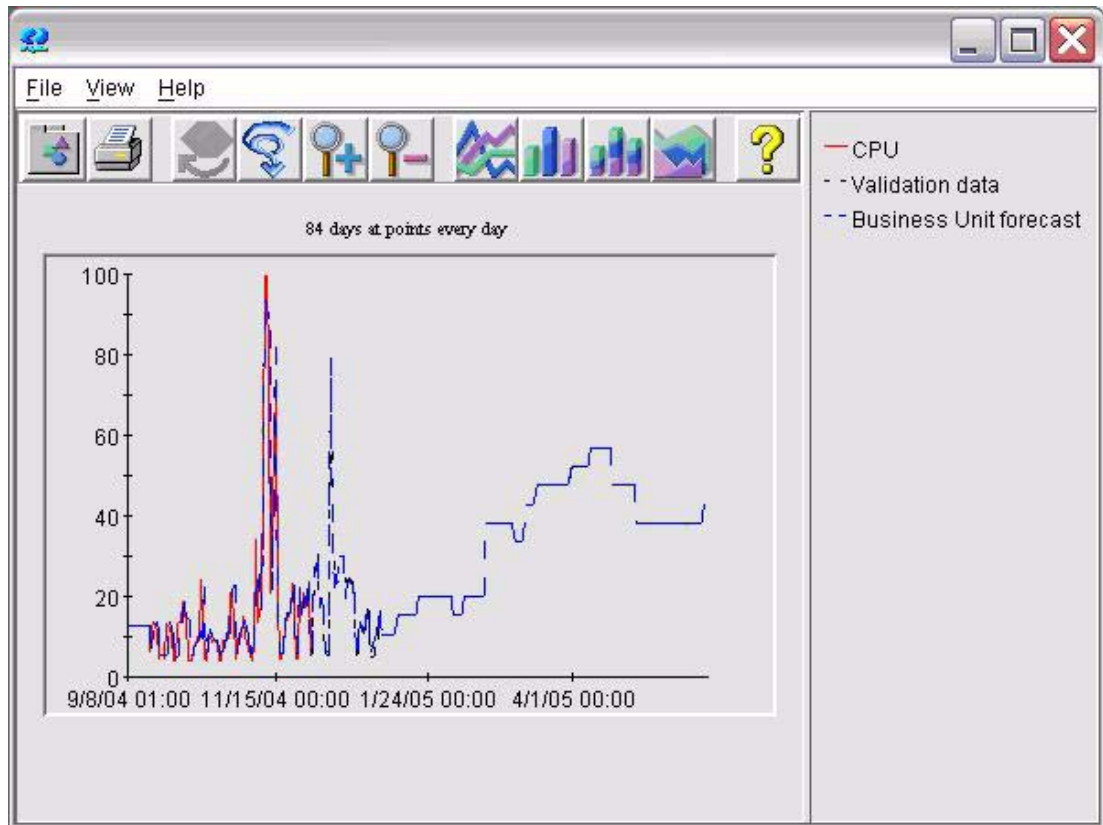
The rest of the forecast options work the same except that Seasonality is disabled. It is assumed that any seasonality patterns are handled by variations in the Business Unit metrics.

Select Date Range and Forecast Ending values so that the data range is within the Forecast Metric values and the Forecast range is within the Business Metric values. Note that to be useful, the Business Units Metric should contain values that are later in time than the Forecast Metric values.

If you want to experiment with Business Unit forecasts, try using the `CaseBus1.bus` file for business units with the `HPCASE.BUS` data file that we use as an example in the chapter 2 - Quick Start of this document. If either of these files is not found, you may have your OVPM administrator copy them from the `<InstallDir>/newconfig/OVPM` directory to the `<DataDir>/conf/perf` directory.

An example of a Business Unit forecast is shown below.

Figure 40 Business Unit Forecast



Business Units could prove to be the most useful trending method you use. You will find it well worth the effort to build business unit files and carefully apply the method.

Glossary

Business Unit

A measure of business activity that can be correlated to the activity of a critical system resource. If the correlation is strong, business units can be used to develop precise projections of performance metrics. For example, a business unit for an order processing system might be the number of order processed.

Confidence Interval

A range having a specified probability that predicted values will fall within this range a given percentage of the time.

Correlation

A measure of the strength of the relationship between two variables.

Exponential Trend

A trend that may begin by approximating a straight line, but increases or decreases dramatically into a steep curve characteristic of exponential expansion - where points reflect a value n - typically the value of e or 2.718 raised to a power (n_0, n_1, \dots, n_n).

Linear Trend

A trend in data where the increase (or decrease) in one parameter follows a constant and regular (linear) relationship to another parameter.

Mean

Average value. Obtained by adding the data and dividing by the number of observations.

Mean Squared Error (MSE)

The Mean Squared Error (MSE) and the square root of the MSE, which is the Standard Error, determine the amount of variability between the forecast and the actual data. They measure the variability in the forecasting technique. A value close to 0 is optimal.

R-Squared

R-Squared is the coefficient of determination. All values will be between 0 and 1. A value close to 1 indicates a very good fit (of the forecast trend line to the data). A value close to 0 means a very poor fit (of the forecast trend line to the data).

Rule-of-Thumb

A useful principle based on commonly-observed relationships. Although a rule-of-thumb has wide application, it isn't intended to be strictly accurate.

Seasonality

A distinct and repetitious pattern in the data occurring in relation to time. For example, a regular increase in the month-end processing for accounting applications.

Standard Error

The Standard Error is interpreted as the average amount of deviation between what the model states and the actual data value for a given date. You can expect forecasts to deviate from actual data values by an amount equal to the Standard Error.

The Standard Error does not appear in Exponential or S-Curve forecasts.

Statistical Forecasting

A process of using historical data collected over time to determine a trend, then projecting this trend into the future using statistical methods.

Workload

The amount of work given a computer by its users. It is measured in terms of units of work per unit of time - for example: number of jobs per day, number of transactions per hour, number of commands per minute. The workload can be characterized by its impact on system resources.

Index

A

annotations for a forecast, 26
application metrics, 39

B

business unit
 defined, 55
 trending method, 58
business unit file, 56
business units equation, 50

C

coefficient of determination, 51
confidence interval, 10
confidence level percentage, 48
correlation coefficient, 51

D

data collection, 33
data selection, 35
datasource
 selecting, 35
datasources
 for forecast data, 34
dates
 selecting, 41
days of the week
 selecting, 42

E

equations
 forecast, 50
estimated standard error, 51
exponential equation, 50
exponential trend, 9

F

file names, 56

flat file datasource, 34
flat file example, 56
forecasting
 concepts, 8
 statistical, 8
forecasting process, 33
forecast templates, 14

G

generic database datasource, 34
global metrics, 38

H

header line, 57
HP OpenView Reporter datasource, 34

I

interval type, 23
 confidence, 23
 prediction, 23

L

linear equation, 50
linear trend
 concept, 9

M

metric
 selecting, 36
metrics
 suggested for forecasting, 38

O

OVIS datasource, 34
OVO datasource, 34
OVPA datasource, 34

P

Planner

getting started, 13

Planner tab, 16

planning techniques, 7

points every

setting, 41

prediction interval, 47

R

reforecasting, 53

S

S-curve, 45

S-curve equation, 50

S-curve trend, 9

seasonality, 25

seasonality patterns, 46

shifts

setting, 42

show validation, 30

skip holidays, 42

smoothed curve, 8

smoothing method, 8

steps

in the forecasting process, 33

T

transaction metrics, 39

trend, 8

trending methods

types, 9

trend line, 25

V

validating a forecast, 30

validating results, 48

