Use Pentaho Data Services

Prototyping a data model can be time consuming, particularly when it involves setting up databases, creating the data model and setting up a data warehouse, then negotiating accesses so that analysts can visualize the data and provide feedback.

One way to streamline this process is to turn the output of a transformation step into a Pentaho Data Service. The output of the transformation step is exposed by the data service so that the output data can be queried as if it were stored in a physical table, even though the results of the transformation are not stored in a physical database. Instead, results are published to the Pentaho Server.

**NOTE:**
You must have a Pentaho Server and repository to publish the data service.

The virtual table is a JDBC-compliant data source that you and others can connect to or query with SQL, as long as they can access the server and the transformation. The Pentaho Data Service can be connected to or queried by a JDBC-compliant tool such as Pentaho Report Designer or Interactive Reporting, as well as other compatible tools like RStudio, DBVisualizer, or SQuirreL.

The Pentaho Data Service can also be used in some instances where building and maintaining a data warehouse is sometimes impractical or inefficient, especially when you need to quickly blend and visualize fast-moving or quickly evolving data sets on the fly. For example, if you want to compare your product prices with your competitors, you could create a transformation that blends prices from your in-house data sources and competitor prices. Then, you can convert that into a Pentaho Data Service that creates a virtual table that you can query when you connect to the Pentaho Server. You or others can connect to and query the virtual table, as you would any other JDBC data source, then use Analyzer to quickly slice, dice, and visualize the results.

The Pentaho Data Service feature provides a testing tool that generates several logs and reports that you can use to refine the data service and determine where to apply specialized optimizations. You can also define parameters that others can use to pose customized queries. For example, you could create a data service that publishes a virtual “fact” table of a moderately-sized research dataset to a Pentaho Server. You could test and add optimizations and parameters, such as gender or test type so that the data service runs more quickly. Then, you could share connection and parameter information with a group of researchers, who could then query the virtual table. Researchers can use Pentaho Interactive Reporting or a tool of their choice, such as RStudio, to analyze and visualize the research dataset.

Pentaho Data Services supports a subset of SQL. For more details on what is supported, see the Pentaho Data Service SQL Support Reference and Other Development Considerations article. Also, for a complete list of traditional data sources that we support, see our Components Reference article.

Read these articles to learn how to create and connect to a Pentaho Data Service.

- [Develop and Optimize a Pentaho Data Service](https://help.pentaho.com/Documentation/8.0/Products/Data_Integration/Data_Services)
• Connect to a Pentaho Data Service
• Pentaho Data Service SQL Support Reference and Other Development Considerations