Pentaho FAQs

What are the key differences between **Pentaho Community Edition (CE)** and **Enterprise Edition (EE)**

**🔓 Pentaho Community Edition (CE)**

* **License**: Open-source (free to use)
* **Features**:
  + Basic ETL (Extract, Transform, Load) capabilities
  + Report Designer and basic dashboarding
  + Limited data source connectivity
  + No built-in scheduler or repository
* **Support**: Community forums only
* **Use Case**: Suitable for small teams, developers, or proof-of-concept projects

**🔐 Pentaho Enterprise Edition (EE)**

* **License**: Commercial (paid)
* **Features**:
  + All CE features plus:
  + Advanced ETL with enterprise-grade graphical designer
  + Interactive dashboards and visualizations
  + Ad hoc reporting and analysis
  + Big Data integration (Hadoop, Spark, etc.)
  + Secure data integration with LDAP/Active Directory
  + Centralized repository, version control, and team collaboration
  + Automated scheduling and monitoring
  + Predictive analytics and machine learning support
* **Support**:
  + 24x7 technical support
  + SLAs (Service Level Agreements)
  + Maintenance releases and updates
* **Use Case**: Ideal for enterprise environments needing scalability, security, and professional support

1. Compare Pentaho & Tableau

|  |  |  |
| --- | --- | --- |
| Criteria | Pentaho | Tableau |
| Functionality | ETL, OLAP, & static Reports | [Data analytics](https://intellipaat.com/blog/tutorial/data-analytics-tutorial/introduction-to-data-analytics/) |
| Availability | Open source | Proprietary |
| Strengths | Data Integration | Interactive visualizations |

2. Define Pentaho and its usage.

Revered as one of the most efficient and resourceful data integration tools (DI), Pentaho virtually supports all available data sources and allows scalable data clustering and [data mining](https://intellipaat.com/blog/what-is-data-mining/). It is a light-weight Business Intelligence suite executing Online Analytical Processing (OLAP) services, ETL functions, reports and dashboards creation and other data-analysis and visualization operations.

Learn more in this [Pentaho Tutorial](https://intellipaat.com/blog/tutorial/pentaho-tutorial/) now.

3. Explain the important features of Pentaho.

* Pentaho is capable of creating Advanced Reporting Algorithms regardless of their input and output data format.
* It supports various report formats, whether Excel spreadsheets, [XMLs](https://intellipaat.com/blog/what-is-xml/), PDF docs, CSV files.
* It is a Professionally Certified DI Software rendered by the renowned Pentaho Company headquartered in Florida, United States.
* Offers enhanced functionality and in-Hadoop functionality.
* Allows dynamic drill down into larger and greater information.
* Rapid Interactive response optimization.
* Explore and view multidimensional data.

Read this [blog](https://intellipaat.com/blog/importance-of-business-intelligence-in-todays-world/) to learn why Business Intelligence Tools like Pentaho are important in today’s world.

4. Name major applications comprising Pentaho BI Project.

* Business Intelligence Platform.
* Dashboards and [Visualizations](https://intellipaat.com/blog/tutorial/data-science-tutorial/data-visualization/).
* Reporting.
* Data Mining.
* Data Analysis.
* [Data Integration](https://intellipaat.com/blog/what-is-data-integration/) and ETL (also called Kettle).
* Data Discovery and Analysis (OLAP)*.*

Interested in learning Pentaho? Well, we have the comprehensive [Pentaho Course](https://intellipaat.com/pentaho-training/) to give you a head start in your career.

5. What is the importance of metadata in Pentaho?

A metadata model in Pentaho formulates the physical structure of your [database](https://intellipaat.com/blog/tutorial/pentaho-tutorial/working-with-databases/) into a logical business model. These mappings are stored in a central repository and allow developers and administrators to build business-logical DB tables that are cost effective and optimized. It further simplifies the working of business users allowing them to create formatted reports and dashboards ensuring security to data access.  
All in all, metadata model provides an encapsulation around the physical definitions of your database and the logical representation and define relationships between them.

6. Define Pentaho Reporting Evaluation.

Pentaho Reporting Evaluation is a particular package of a subset of the Pentaho Reporting capabilities, designed for typical first-phase evaluation activities such as accessing sample data, creating and editing reports, and viewing and interacting with reports.

7. Explain the benefits of Data Integration.

* The biggest benefit is that integrating data improves consistency and reduces conflicting and erratic data from the DB. Integration of data allows users to fetch exactly what they look for, enabling them utilizeand work with what they collected.
* Accurate [data extraction](https://intellipaat.com/blog/tutorial/data-science-tutorial/data-extraction/), which further facilitates flexible reporting and monitoring of the available volumes of data.
* Helps meet deadlines for effective business management.
* Track customer’s information and buying behavior to improve traffic and conversions in the future, thus advancing your business performance.

8. What is MDX and its usage?

MDX is an acronym for ‘Multi-Dimensional Expressions,’ the standard query language introduced by Microsoft SQL OLAP Services. MDX is an imperative part of XML for analysis API, which has a different structure than SQL. A basic MDX query is:

SELECT {[Quantity].[Unit Sales], [Quantity].[Store Sales]} ON COLUMNS,<br>

{[Product].members} ON ROWS<br>

FROM [Sales]<br>

WHERE [Time].[1999].[Q2]

[](https://intellipaat.com/business-intelligence-masters-program-training/)

9. Define three major types of Data Integration Jobs.

* Transformation Jobs : Used for preparing data and used only when the there is no change in data until transforming of data job is finished.
* Provisioning Jobs : Used for transmission/transfer of large volumes of data. Used only when no change is data is allowed unless job transformation and on large provisioning requirement.
* Hybrid Jobs : Execute both transformation and provisioning jobs. No limitations for data changes; it can be updates regardless of success/failure. The transforming and provisioning requirements are not large in this case.

10. Illustrate the difference between transformations and jobs.

While transformations refer to shifting and transforming rows from source system to target system, jobs perform high level operations like implementing transformations, file transfer via FTP, sending mails, etc.  
Another significant difference is that the transformation allows parallel execution whereas jobs implement steps in order.

11. How to perform database join with PDI (Pentaho Data Integration)?

PDI supports joining of two tables form the same databse using a ‘Table Input’ method, performing the join in [SQL](https://intellipaat.com/blog/tutorial/sql-tutorial/) only.  
On the other hand, for joining two tables in different databases, users implement ‘Database Join’ step. However, in database join, each input row query executes on the target system from the main stream, resulting in lower performance as the number of queries implement on the B increases.  
To avoid the above situation, there is yet another option to join rows form two different Table Input steps. You can use ‘Merge Join ‘step, using the SQL query having ‘ORDER BY’ clause. Remember, the rows must be perfectly sorted before implementing merge join.

12. Explain how to sequentialize transformations?

Since PDI transformations support parallel execution of all the steps/operations, it is impossible to sequentialize transformations in Pentaho. Moreover, to make this happen, users need to change the core architecture, which will actually result in slow processing.

13. Explain Pentaho Reporting Evaluation.

Pentaho Reporting evaluation is a complete package of its reporting abilities, activities and tools, specifically designed for first-phase evaluation like accessing the sample, generating and updating reports, viewing them and performing various interactions. This evaluation consists of Pentaho platform components, Report Designer and ad hoc interface for reporting used for local installation.

14. Can fieldnames in a row duplicated in Pentaho?

No, Pentaho doesn’t allow field duplication.

15. Does transformation allow filed duplication?

“Select Values” will rename a field as you select the original field also.  The original field will have a duplicate name of the other field now.

16. How to use database connections from repository?

You can either create a new transformation/job or close and reopen the ones already loaded in Spoon.

17. Explain in brief the concept Pentaho Dashboard.

Dashboards are the collection of various information objects on single page including diagrams, tables and textual information. The Pentaho AJAX API is used to extract BI information while Pentaho Solution Repository contains the content definitions.  
The steps involved in Dashboard creation include

* Adding dashboard to the solution.
* Defining dashboard content.
* Implementing filters.
* Editing dashboards.

18. How to use logic from one transformation/job in other process?

Transformation logic can be shared using subtransformations, which provides seamless loading and transformation of variables enhancing efficiency and productivity of the system. Subtransformations can be called and reconfigured when required.

19. Explain the use of Pentaho reporting.

Pentaho reporting enables businesses to create structured and informative reports to easily access, format and deliver meaningful and important information to clients and customers. They also help business users to analyze and track consumer’s behavior for the specific time and functionality, thereby directing them towards the right success path.

[](https://intellipaat.com/power-bi-training/)

20. What is Pentaho Data Mining?

Pentaho Data Mining refers to the Weka Project, which consists of a detailed tool set for machine learning and data mining. Weka is open source software for extracting large sers of information about users, clients and businesses. It is built on [Java](https://intellipaat.com/blog/tutorial/java-tutorial/) programming.

*Take charge of your career by going through this professionally designed*[*Pentaho Certification Training*](https://intellipaat.com/pentaho-training/)*.*

21. Is Data Integration and ETL Programming same?

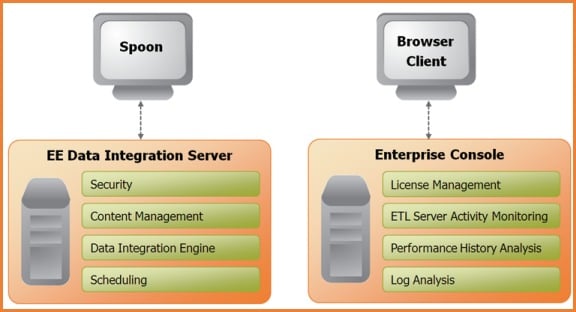
No. Data Integration refers to passing of data from one type of systems to other within the same application. On the contrary, ETL is used to extract and access data from different sources. And transform it into other objects and tables.

22. Explain Hierarchy Flattening.

It is just the construction of parent child relationships in a database. Hierarchy Flattening uses both horizontal and vertical formats, which enables easy and trouble-free identification of sub elements. It further allows users to understand and read the main hierarchy of BI and includes Parent column, Child Column, Parent attributes and Child attributes.

23. Give a brief about PDI architecture.

Below image depicts the architecture of PDI



24. Explain Pentaho report Designer (PRD).

PRD is a graphic tool to execute report-editing functions and create simple and advanced reports and help users export them in PDF, [Excel](https://intellipaat.com/blog/how-to-become-an-excel-expert/), [HTML](https://intellipaat.com/blog/what-is-html/) and CSV files. PRD consists of Java-based report engine offering data integration, portability and scalability. Thus, it can be embedded in Java web applications and also other application servers like Pentaho BAserver.

25. Define Pentaho Report types.

There are several categories of Pentaho reports :

* Transactional Reports : Data to be used form transactions. Objective is to publish detailed and comprehensive data for day-to-day organization’s activities like purchase orders, sales reporting.
* Tactical Reports : data comes from daily or weekly transactional data summary. Objective is to present short-term information for instant decision making like replacing merchandize.
* Strategic Reports : data comes from stable and reliable sources to create long-term business information reports like season sales analysis.
* Helper Reports : data comes from various resources and includes images, videos to present a variety of activities.

26. What are variables and arguments in transformations?

Transformations dialog box consists of two different tables: one of arguments and the other of variables. While arguments refer to command line specified during batch processing, PDI variables refer to objects that are set in a previous transformation/job in the OS.

27. How to configure JNDI for Pentaho DI Server?

Pentaho offers JNDI connection configuration for local DI to avoid continuous running of application server during the development and testing of transformations.  Edit the properties in jdbc.propertiesfile located at…\data-integration-server\pentaho-solutions\system\simple-jndi.