

Database Administration

Lecture 06: Extraction - Transformation - Load.

Pentaho Data Integration.

6 de marzo de 2025

Database Administration: ETL.

The screenshot shows the Pentaho website with a dark theme. The navigation bar includes links for Pentaho+ Platform, Solutions, Pricing, Services, Resources, Blog, and a Get a Demo button. The main content area features a large hero section for Pentaho Data Integration. The hero section has a title 'Data Integration: Ingest, Blend, Orchestrate, and Transform Data' and a sub-headline 'Unlock more actionable insights'. Below the title is a paragraph describing PDI as a codeless data orchestration tool. To the right is a video player with a play button and the text 'Watch the video'. Below the hero section are four statistics: 'Low-code Environment for Data Preparation', '80% saved in data operations costs', '7x faster display of knowledge graphs', and '55% savings of data scientist's time finding & evaluating data'. Each statistic has a corresponding icon and a 'Try For Free' or 'Get a Demo' button.

Pentaho

Pentaho+ Platform Solutions Pricing Services Resources Blog [Get a Demo](#)

Pentaho Data Integration

Data Integration: Ingest, Blend, Orchestrate, and Transform Data

More than just ETL (Extract, Transform, Load), Pentaho Data Integration is a codeless data orchestration tool that blends diverse data sets into a single source of truth as a basis for analysis and reporting. Effortlessly managed in a drag-and-drop graphical interface, so you can easily track where it's coming from, where it's going and how it's transforming.

- Develop and maintain pipeline efficiency
- Scalability, simplicity, and self-service

[Try For Free](#) [Get a Demo](#)

Unlock more actionable insights

Watch the video

Low-code Environment for Data Preparation

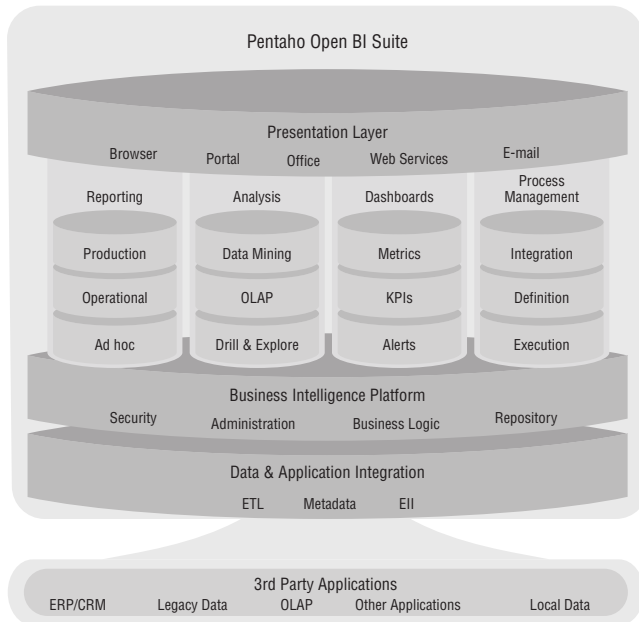
80% saved in data operations costs

7x faster display of knowledge graphs

55% savings of data scientist's time finding & evaluating data

Content has been extracted from *PDI* , created by Pentaho Data Integration, 2025. Visit <https://pentaho.com/products/pentaho-data-integration/>.

Pentaho Architecture



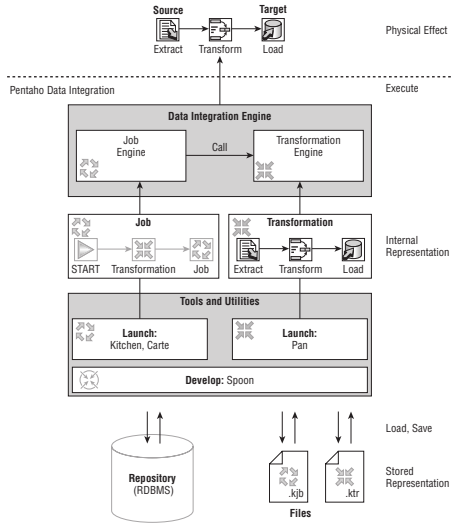
Introduction

- ▶ Data integration: Combining data from different sources to provide a unified view.
- ▶ Pentaho Data Integration (PDI) offers tools for ETL (Extract, Transform, Load).
- ▶ Core PDI components: Transformations, Jobs, and the Data Integration Engine.

Data Integration Activities

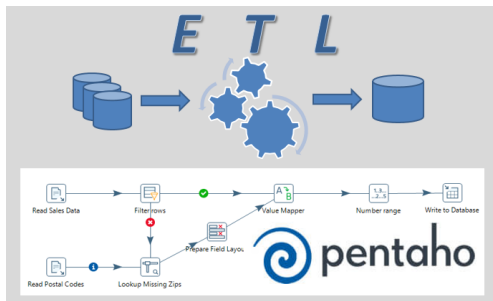
- ▶ Extraction: Retrieving data from various sources.
- ▶ Change Data Capture (CDC): Identifying changes in source data.
- ▶ Data Staging: Intermediate storage for transformation.
- ▶ Data Validation and Cleansing: Ensuring data quality.
- ▶ Key Management and Aggregation.
- ▶ Dimension and Fact Table Loading.

Data Integration Architecture



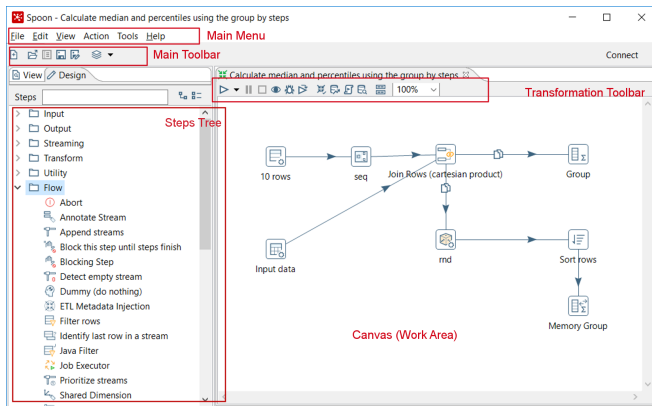
Pentaho Data Integration Components

- ▶ **Spoon:** Graphical IDE for designing transformations and jobs.
- ▶ **Kitchen:** Command-line tool for executing jobs.
- ▶ **Pan:** Command-line tool for running transformations.
- ▶ **Carte:** Remote execution engine.



Getting Started with Spoon

- ▶ Launch Spoon and create a new transformation.
- ▶ Add steps to extract, transform, and load data.
- ▶ Connect steps using "hops."
- ▶ Preview and execute transformations.



Summary

- ▶ PDI provides tools for efficient ETL processes.
- ▶ Transformations process data at the record level.
- ▶ Jobs orchestrate multiple tasks.
- ▶ Spoon offers a user-friendly interface for development.

End of Lecture 6.



Top 5 Fundamental Takeaways

Top 5 Fundamental Takeaways

5

Top 5 Fundamental Takeaways

5

4

Top 5 Fundamental Takeaways

5

4

3

Top 5 Fundamental Takeaways

5

4

3

2

1

Database Administration: ETL.

The screenshot shows the Pentaho website with a dark theme. The navigation bar includes links for Pentaho+ Platform, Solutions, Pricing, Services, Resources, and Blog, along with a red 'Get a Demo' button. The main content area features a 'Pentaho Data Integration' badge, a headline 'Data Integration: Ingest, Blend, Orchestrate, and Transform Data', and a paragraph describing the tool as a codeless data orchestration tool. Below this are two bullet points: 'Develop and maintain pipeline efficiency' and 'Scalability, simplicity, and self-service'. At the bottom of the main section are 'Try For Free' and 'Get a Demo' buttons. To the right is a video player with the text 'Unlock more actionable insights' and a 'Watch the video' button. Below the main content area are four statistics: 'Low-code Environment for Data Preparation', '80% saved in data operations costs', '7x faster display of knowledge graphs', and '55% savings of data scientist's time finding & evaluating data'.

Pentaho Pentaho+ Platform Solutions Pricing Services Resources Blog [Get a Demo](#)

Pentaho Data Integration

Data Integration: Ingest, Blend, Orchestrate, and Transform Data

More than just ETL (Extract, Transform, Load), Pentaho Data Integration is a codeless data orchestration tool that blends diverse data sets into a single source of truth as a basis for analysis and reporting. Effortlessly managed in a drag-and-drop graphical interface, so you can easily track where it's coming from, where it's going and how it's transforming.

- Develop and maintain pipeline efficiency
- Scalability, simplicity, and self-service

[Try For Free](#) [Get a Demo](#)

Unlock more actionable insights

Watch the video

Low-code Environment for Data Preparation

80% saved in data operations costs

7x faster display of knowledge graphs

55% savings of data scientist's time finding & evaluating data

Content has been extracted from *PDI* , created by Pentaho Data Integration, 2025. Visit <https://pentaho.com/products/pentaho-data-integration/>.