

ETL-specific tools and their applications

In a previous reading, you were given a list of common business intelligence tools and some of their uses. Many of them have built-in pipeline functionality, but there are a few ETL-specific tools you may encounter. Creating pipeline systems—including ETL pipelines that move and transform data between different data sources to the target database—is a large part of a BI professional's job, so having an idea of what tools are out there can be really useful. This reading provides an overview.

Tool	Uses
Apache Nifi	<ul style="list-style-type: none">• Connect a variety of data sources• Access a web-based user interface• Configure and change pipeline systems as needed• Modify data movement through the system at any time
Google DataFlow	<ul style="list-style-type: none">• Synchronize or replicate data across a variety of data sources• Identify pipeline issues with smart diagnostic features• Use SQL to develop pipelines from the BigQuery UI• Schedule resources to reduce batch processing costs• Use pipeline templates to kickstart the pipeline creation process and share systems across your organization
IBM InfoSphere Information Server	<ul style="list-style-type: none">• Integrate data across multiple systems• Govern and explore available data• Improve business alignment and processes• Analyze and monitor data from multiple data sources
Microsoft SQL SIS	<ul style="list-style-type: none">• Connect data from a variety of sources integration• Use built-in transformation tools• Access graphical tools to create solutions without coding• Generate custom packages to address specific business needs

Oracle Data Integrator	<ul style="list-style-type: none"> • Connect data from a variety of sources • Track changes and monitor system performance with built-in features • Access system monitoring and drill-down capabilities • Reduce monitoring costs with access to built-in Oracle services
Pentaho Data Integrator	<ul style="list-style-type: none"> • Connect data from a variety of sources • Create codeless pipelines with drag-and-drop interface • Access dataflow templates for easy use • Analyze data with integrated tools
Talend	<ul style="list-style-type: none"> • Connect data from a variety of sources • Design, implement, and reuse pipeline from a cloud server • Access and search for data using integrated Talend services • Clean and prepare data with built-in tools