+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated on 21/07/2025

Sign up

Matillion ETL Training

2 days (14 hours)

Presentation

Master Matillion ETL in its entirety thanks to this comprehensive, structured and resolutely practical training course. From getting started to automating pipelines, you'll learn how to design, orchestrate and optimize modern data flows, in a cloud-first environment compatible with Snowflake, Redshift or BigQuery.

You'll start by discovering Matillion's architecture, the differences between orchestration and transformation jobs, and the initial configuration of projects and source connections. The aim is to lay a solid foundation for the development of reliable pipelines.

You will then learn how to design complete integration workflows, manipulating native components, cloud connectors, external APIs and custom scripts, while applying good performance practices.

The course will also cover advanced variable management, automation via scheduling or triggers, and detailed monitoring of executions and errors. You'll discover how to industrialize your jobs with Git and external orchestrators.

As with all our training courses, this one will be presented with the latest Matillion ETL updates.

Objectives

- Understand Matillion ETL architecture
- Deploy, configure and connect Matillion ETL to various sources and targets
- Master Matillion's transformation components
- Be able to automate, supervise and audit ETL processing.

Target audience

- Data Engineers
- Data analytics

Prerequisites

- Mastery of SQL fundamentals
- general knowledge of ETL / ELT architectures

Matillion ETL training program Introduction to

Matillion ETL

- Objectives and use cases
- Architecture and operating principles
- Matillion ETL vs. Matillion Data Loader
- Cloud support: AWS, GCP, Azure
- Menu overview
- Development zones

Environment configuration

- Deploying Matillion
- Launch via AWS Marketplace, GCP Marketplace or Azure
- Network and security requirements
- Database connections
- Data source creation
- Authentication, keys and authorizations
- Global settings and projects
- Project configuration
- Global vs. environmental variables

Job creation Orchestration

- Understanding the orchestration pipeline
- Orchestration vs. Transformation
- Typical use of an orchestration job
- Main components
- Python Script, Bash Script, If Condition, Iterator
- Run Orchestration / Run Transformation
- Dependency management
- Logical chaining of components

- Error handling

Job creation Transformation

- Fundamental concepts of transformation
- Component types : Input, Join, Filter, Calculator
- Loading, cleaning, enrichment
- Building pipelines
- Visual design of transformations
- Using multiple sources
- Performance optimization
- Partitioning
- SQL pushdown
- Reduce unnecessary steps

Connectors and integrations

- Native connectors
- APIs and cloud services: Salesforce, S3, REST APIs
- Flat files: CSV, Excel, JSON
- Use of Query components
- Redshift Query, Snowflake Query, BigQuery Query
- Web Services and external APIs
- REST calls with OAuth 2.0
- JSON/XML output processing

Data management and quality

- Data profiling
- Descriptive statistics with the Data Quality component
- Data preview
- Error handling
- Invalid data filters
- Logs, notifications, Retry
- Good transformation practices
- Schema validation
- Inline documentation and metadata

Variables, parameters and reusability

- Using variables
- Project vs. environment variables
- Variables in scripts and components
- Parameterizable jobs
- Components Parameters
- Job Template and modularity
- Reuse
- Common jobs

- Reference to existing transformations

Scheduling and automation

- Triggers and scheduling
- Matillion Scheduler
- Integration with Airflow, external cron or Cloud Scheduler
- Event management
- Conditional triggering
- Notifications
- Complex orchestration
- Loops, splits, parallelization

Monitoring, logs and auditing

- Matillion monitoring console
- Execution and error logs
- Job history
- Integration with third-party tools
- CloudWatch (AWS), Stackdriver (GCP), Azure Monitor
- Audit and traceability
- Job versioning
- Modification tracking

Best practices and advanced use cases

- DevOps best practices
- Git integration
- CI/CD with Matillion and Terraform
- Security
- Role-based access management
- Encryption and credential management
- Case studies
- Snowflake case study: ingestion from S3, transformation, export
- BigQuery case study: cleaning application logs
- Multi-source cases: data fusion and enrichment

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is confirmed, the learner receives a self-assessment questionnaire enabling us to

assess the learner's estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical training: 60% hands-on, 40% theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire is used to check that skills have been correctly acquired.

Certification

A certificate will be awarded to each trainee who completes the training course.