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

Updated on 10/26/2026

Sign Up

Great Expectations Training

4 days (28 hours)

Overview

Great Expectations is a data validation framework designed to ensure the reliability of business data flows within an organization. Using Expectations and Checkpoints, you can formalize your quality rules, automate checks, and generate validation evidence for your pipelines.

Our Great Expectations training will help you develop advanced data quality skills in secure professional environments. You'll learn how to translate business requirements into checks that your technical teams can implement.

You will structure a GX Core project and connect your data sources to build reusable validation suites. The course covers conditional rules, results, Data Docs, and blocking decisions within a pipeline.

Upon completing the training, you'll be able to strengthen the hardening of your environment, protect secrets, and limit the exposure of sensitive data. SIEM and EDR integrations rely on exports or custom actions, without replacing security tools.

This training also covers testing, reproducible deployment, and validation traceability. The workshops will enable you to design validation pipelines that serve as evidence of control—without requiring regulatory certification—and then prepare teams to adopt them.

Like all our training courses, this one will introduce you to **the latest stable version** of the technology and its new features.

Objectives

- Structure a Great Expectations project to validate data in a secure production environment
- Build Expectation Suites tailored to business risks and quality requirements
- Automate validation pipelines that produce traceable validation evidence
- Integrate data from an EDR and transmit quality events to a SIEM via custom interfaces
- Apply a hardening strategy to secrets, access, and validation results
- Deploy a documented validation process and prepare to transfer expertise to technical teams

Target Audience

- Data Engineer
- Python Developer
- Data Quality Engineer
- DevSecOps Engineer
- SOC Analyst

Prerequisites

- Ability to write Python scripts and manipulate arrays with pandas in a development environment
- Ability to query a relational database using SQL to explore data
- Understand the stages of a data pipeline and key quality metrics
- Basic knowledge of logging, secret management, and access control

Technical Prerequisites

- Have a computer with 4 cores, 16 GB of RAM, and 20 GB of free disk space; no dedicated graphics card required
- Use Ubuntu 24.04 LTS either directly or in a virtual machine with the required
- Install Python 3.12 and GX Core 1.x in a virtual environment, using the fixed dependencies and versions provided prior to the session
- Have Git and Visual Studio Code installed, as well as access to the lab’s PostgreSQL database
- Ensure a stable internet connection with authorized educational access; anonymized EDR exports and the test SIEM collector are provided without requiring a personal cloud account

Great Expectations Training Program

[Day 1 - Morning]

Fundamentals and Testing Environment

- Dimensions of data quality and the impact of anomalies on business risks
- Scope of GX Core and the distinction between data validation, cybersecurity, and certification
- Object Architecture: Data Context, Data Source, Data Asset, Batch Definition, and Batch
- Isolated installation, fixed dependencies, and organizing a reproducible Python project
- Choosing Between Ephemeral and Persistent Data Context Based on Use Cases
- Hands-on Workshop: Initialize a GX Core project and run an initial check on an anonymized business dataset.

[Day 1 - Afternoon]

Data sources and initial quality rules

- Connecting to data with pandas and PostgreSQL depending on the execution context
- Defining Data Assets and dividing data into reproducible batches
- Building an Expectation Suite based on explicit business requirements
- Checks for completeness, uniqueness, valid values, and bounds
- Creating a Validation Definition and reviewing validation results
- Hands-on Workshop: Connecting two data sources, building a suite of checks, and analyzing injected anomalies.

[Day 2 - Morning]

Business Rules and In-Depth Validations

- Translating business risks into measurable requirements and acceptance thresholds
- Checks for structure, volume, and consistency between columns
- Configuring tolerances with Mostly and explicit handling of null values
- Use of row conditions and SQL rules tailored to the selected engine
- Testing validation suites on edge cases and distinguishing between data anomalies and runtime errors
- Hands-on workshop: Building and testing a business validation suite with tolerances, conditional checks, and edge cases.

[Day 2 - Afternoon]

Checkpoints and certification pipelines

- Assembling Validation Definitions into Reusable Checkpoints
- Automating Actions and publishing Data Docs for teams
- Selecting the result_format based on diagnostic and confidentiality needs
- Managing block thresholds, rejections, and recovery in the pipeline code
- Designing certification pipelines that preserve evidence, without regulatory certification value
- Hands-On Workshop: Building a Pipeline That Blocks Non-Compliant Batches and Archives Results along with rule versions and batch IDs.

[Day 3 - Morning]

Hardening and Protecting Results

- Analysis of the project's exposure surfaces, connections, and validation artifacts
- Outsourcing secrets and excluding sensitive information from the code repository
- Applying the principle of least privilege to read-only accounts and storage directories
- Reducing sensitive data in results, logs, and Data Docs
- Hardening strategy for the host environment, dependencies, and communications, excluding native GX functions
- Hands-on workshop: Hardening a lab project, removing exposed secrets, and verifying the information present in the generated artifacts.

[Day 3 - Afternoon]

Integration with SIEM and EDR workflows

- Positioning Great Expectations as a quality control measure for security feeds, without native EDR detection capabilities
- Preparing anonymized EDR exports with identifiers, timestamps, and standardized fields
- Validating the completeness, uniqueness, and temporal consistency of events
- Designing a custom action to send summary results to a SIEM collector
- Handling transmission failures, duplicates, and correlation IDs in the integration
- Hands-on workshop: Validate an EDR export and send a test event—without sensitive data—to a test SIEM collector.

[Day 4 - Morning]

Scaling and Reliable Operation

- Organizing configurations across development, testing, and production
- Integrating validations into a CI/CD pipeline with non-regression tests
- Optimizing the scope of batches, queries, and the volume of returned results
- Tracking execution times, compliance rates, and validation incidents
- Retention policy, configuration backups, and disaster recovery
- Hands-on workshop: Automate the continuous integration pipeline and compare its reliability and performance before and after optimization.

[Day 4 - Afternoon]

Summary project and handoff to teams

- Designing a target architecture combining business validation, hardening, and security integration
- Defining acceptance criteria and expected evidence for each batch

- Allocation of responsibilities among data, operations, and security teams
- Creation of an internal training kit with datasets, progressive exercises, and answer keys
- Preparation of an operations guide and a technical skills assessment rubric
- Hands-on workshop: Deliver and present a certified, hardened pipeline using business data and EDR exports, including SIEM reporting, evidence of controls, and onboarding support.

Target Companies

This training is intended for both individuals and companies—large or small—seeking to train their teams in new, advanced IT technologies or to acquire specific business knowledge or modern methodologies.

Pre-training Assessment

The pre-training assessment complies with Qualiopi quality standards. Upon final registration, the learner receives a self-assessment questionnaire that allows us to evaluate their estimated proficiency level across various types of technologies, as well as their expectations and personal goals for the upcoming training, within the limits imposed by the selected format. This questionnaire also allows us to anticipate certain connection or internal security issues within the company (intra-company or virtual classroom) that could pose challenges for monitoring and ensuring the smooth operation of the training session.

Teaching Methods

Hands-On Training: 60% Hands-On, 40% Theory. Training materials distributed in digital format to all participants.

Organization

The course alternates between theoretical instruction from the trainer—supported by examples and reflection sessions—and group work.

Assessment

At the end of the session, a multiple-choice quiz is administered to verify that participants have successfully acquired the skills.

Certification

A certificate will be issued to each trainee who has completed the entire training program.

Training organization registered under number 11 75 54743 75. This registration does not constitute state accreditation.
© Ambient IT 2015-2026. All rights reserved. Paris, France - Switzerland - Belgium - Luxembourg