Palantir · Learn

UPDATED JULY 2024

Foundry Certification Exam Guide: **Data Engineer**

Foundry Data Engineer Certification

The Foundry Data Engineering Certification Exam tests your ability to develop and maintain data pipelines using Palantir Foundry. This includes connecting to external systems; implementing, debugging, and optimizing data transforms; as well as designing and developing the ontology. The exam assesses knowledge of the applications in the Foundry suite and ability to apply them in realistic situations.

CANDIDATE PROFILE

A candidate for the Foundry Data Engineering Certification is someone capable of executing in data integration and ontology development projects. You should consider this certification if your responsibilities include producing or maintaining high-quality data assets in Foundry. The exam is recommended for people who have 6+ months of experience on the platform. General knowledge of distributed computing frameworks (e.g., Spark) and programming languages (e.g., Python, SQL) is also recommended but not required.

EXAM DESIGN

The exam is composed of 60 questions covering a variety of topics within the selected certification track. It consists of the following question types:

- Multiple choice questions (select one, select multiple, discrete options)
- Hotspot questions (click on image) and matching questions (match one answer with another)

You will have 120 minutes to complete the exam, a passing score of 70% is required to receive a certification. Questions will be presented in sequence without the ability to navigate back.

ABOUT THIS EXAM GUIDE

This exam guide is meant to provide insight into what will be tested on the certification exam. All study resources included here are meant to be directionally useful prep materials; test takers are encouraged to seek additional Foundry support resources to supplement knowledge. Questions may be asked on topics not explicitly included in the exam guide.

Data Engineer Exam Content

The new Data Engineering Certification Exam emphasizes testing of practical skills necessary to build and maintain production-grade data pipelines, data connections, and ontologies. In addition to practical skills, it also tests general knowledge of platform capabilities and specific applications within the Foundry suite that are useful for performing the job of data engineer. The exam is broken into 4 sections corresponding to 4 main responsibilities of data engineers working on Foundry. Each section tests a variety of tasks described in greater detail on the following pages of this guide, with links to documentation and self-service training content to help you prepare for the exam.

DATA PIPELINE DEVELOPMENT IN FOUNDRY (30% of questions)

Can you develop transforms on structured (tabular) and unstructured datasets in Foundry? Do you know how to follow the recommended best practices when building data pipelines?

DATA PIPELINE MAINTENANCE IN FOUNDRY (30% of questions)

Do you know how to effectively investigate and fix common issues in data pipelines? Can you contribute logic changes and performance improvements to transform pipelines feeding mission critical workflows? Are you familiar with recommended support structures?

DATA CONNECTION AND INTEGRATION IN FOUNDRY (20% of questions)

Are you familiar with architecture and capabilities of Data Connection? Can you set up sources and syncs ingesting tabular data or raw files from external systems to Foundry?

ONTOLOGY DESIGN AND DEVELOPMENT IN FOUNDRY (20% of questions)

Can you provide data engineering context during ontology design and implement pipelines backing ontology objects and links based on application requirements?

Breakdown: Data Pipeline Development

Can you develop transforms on structured (tabular) and unstructured datasets in Foundry? Do you know how to follow the recommended best practices when building data pipelines?

Task	Learning Resources
Process tabular data in Transforms	Documentation: • Data Integration • Tools: Code Repositories, Python Transforms, Contour Self-service learning: • Video: How to Write Data Transforms
Process unstructured data in Transforms	Documentation: • Building Pipelines: Unstructured Data, Transforms Reference • Tools: Code Repositories, Code Workbook Self-service learning: • Video: Parsing Excel Files
Configure pipeline for production use	Documentation: • Building Pipelines • Tools: Data Health, Data Lineage
Apply general best practices during pipeline development	Documentation: • Code Repositories: Project References, Unit Tests • Transforms: PySpark Style Guide, Using Libraries

Accessing self-service learning resources:

- Training courses are available on https://learn.palantir.com/
- Full platform documentation is available on https://www.palantir.com/docs/
- Video tutorials are available on the Palantir Developers YouTube Channel
- Example reference workflows are downloadable from https://build.palantir.com/

Breakdown: Data Pipeline Maintenance

Do you know how to effectively investigate and fix common issues in data pipelines? Can you contribute logic changes and performance improvements to transform pipelines feeding mission-critical workflows? Are you familiar with recommended support structures?

Task	Learning Resources
Debug an issue in a production pipeline	Documentation: • Tools: Data Lineage • Code Repositories: Debug Transforms • Optimizing Pipelines: Debug Job, Troubleshoot OOM Errors
Contribute changes to a production pipeline	Documentation: • Data Integration: Data Pipeline, Branching • Code Repositories: Branches, Share Python Libraries • Building Pipelines: Release Process
Set up support structure for production pipeline	Documentation: • Maintaining Pipelines • Tools: Data Health Self-service learning: • Video: Pipeline Monitoring Playlist
Improve performance of production pipeline	Documentation: • Incremental Pipelines: Overview, Syncs, Transforms, Examples • Optimizing Pipelines Self-service learning: • Video: Incremental Data Transforms Reference Workflows: • Advanced incremental data processing with PySpark in Code Repositories

Breakdown: Data Connection and Integration

Are you familiar with architecture and capabilities of Data Connection? Can you set up sources and syncs ingesting tabular data or raw files from external systems to Foundry?

Task	Learning Resources
Ingest tabular data from an external source to Foundry	Documentation: • Tools: Data Connection • Data Connection Sources: Optimize JDBC Syncs Reference Workflows • Connecting to AWS S3 • Connecting to Postgres
Ingest unstructured data from an external source to Foundry	Documentation: • Tools: Data Connection
Identify general data connection capabilities useful for a given project	Documentation: • Data Connection: Agent Setup, Direct Connections • Manual Data Upload, Fusion Dataset Syncs • Platform Interoperability

Accessing self-service learning resources:

- Training courses are available on https://learn.palantir.com/
- Full platform documentation is available on https://www.palantir.com/docs/
- Video tutorials are available on the Palantir Developers YouTube Channel
- Example reference workflows are downloadable from https://build.palantir.com/

Breakdown: Ontology Design and Development

Can you provide data engineering context during ontology design and implement pipelines backing ontology objects and links based on application requirements?

Task	Learning Resources
Design an ontology based on application requirements and available data	Documentation: • Ontology
Implement pipelines backing ontology objects and links	Documentation: • Object & Link Types Self-service learning: • Deep Dive: Creating your First Ontology
Provide data engineering context useful for use case development	Documentation: • Solution Design Self-service learning: • Video: Foundry Reference Project Playlist

Accessing self-service learning resources:

- Training courses are available on https://learn.palantir.com/
- Full platform documentation is available on https://www.palantir.com/docs/
- Video tutorials are available on the Palantir Developers YouTube Channel
- Example reference workflows are downloadable from https://build.palantir.com/