DP-700 Exam Guide

Microsoft Fabric Data Engineer Associate

100% Aligned with Microsoft's Latest Study Guide

Abhishek Parmar
Published by 9xCode

Copyright

First Edition, 2025
ISBN: 978-93-343-1116-7

Written by Abhishek Parmar
Published by 9xCode
Cover design by Abhishek Parmar

About the Author

Abhishek Parmar (Kumar Abhiii) holds a B.Tech in Computer Science and brings close to a decade of experience in software engineering at leading multinational companies. He is a experienced author of several technical books and has in-depth knowledge of a wide range of industry certifications, including Microsoft Azure Fundamentals (AZ-900, DP-900, DP-700, AI-900), Microsoft Fabric, AWS Certified Solutions Architect Associate, AWS Certified Developer Associate, CompTIA A+, SnowPro® Core, Cisco CCNA, ITIL® Foundation. He remains committed to expanding his expertise.

Over the years, Abhishek has guided numerous students to certification success by breaking down complex topics into clear, real-world examples they can apply immediately. He creates practical exercises and mock exams that mirror actual test scenarios, helping learners build both confidence and mastery, and he continually updates his materials to reflect the latest industry standards.

Explore his latest content at https://9xcode.com/abhishek

Syllabus and Chapter Outline

Part 1: Implement and Manage an Analytics Solution (30–35%)

Chapter 1: Configure Microsoft Fabric Workspaces

- Configure Fabric workspace settings
- Configure Spark workspace settings
- Configure domain, OneLake, and data workflow workspaces

Chapter 2: Implement Life-cycle Management

- Configure version control
- Implement database projects
- Create and configure deployment pipelines

Chapter 3: Security and Governance

- Workspace- and item-level access controls
- Row-, column-, object-, and folder/file-level ACLs
- Dynamic data masking, sensitivity labels, and item endorsement
- Implement and use workspace logging

Chapter 4: Orchestrate Processes

- Choose between pipelines and notebooks
- Design and implement schedules and event-based triggers
- Implement orchestration patterns with parameters and dynamic expressions

Part 2: Ingest and Transform Data (30–35%)

Chapter 5: Design and Implement Data Loading Patterns

- Full vs. incremental loads
- Prepare data for dimensional models
- Loading patterns for streaming data

Chapter 6: Batch Ingestion and Transformation

- Select appropriate data stores
- Use dataflows, notebooks, KQL, and T-SQL for transformations
- Shortcuts, mirroring, and pipeline ingestion
- Transform with PySpark, SQL, and KQL; denormalize, group, aggregate
- Handle duplicates, missing, and late-arriving data

Chapter 7: Streaming Ingestion and Transformation

- Choose streaming engines and storage options in Real-Time Intelligence
- Process data via Eventstreams, Spark structured streaming, and KQL
- Implement windowing functions (tumbling, hopping, session)

Part 3: Monitor and Optimize an Analytics Solution (30–35%)

Chapter 8: Monitor Fabric Items

- Monitor data ingestion, transformation, and semantic model refresh
- Configure alerts for SLA breaches and anomalies

Chapter 9: Identify and Resolve Errors

- Pipeline, dataflow, notebook, Eventhouse, Eventstream, and T-SQL error handling
- Troubleshooting patterns and root cause analysis

Chapter 10: Performance Optimization

- Optimize lakehouse tables (file sizes, Z-order, compaction)
- Optimize pipelines, data warehouses, eventstreams, and Spark clusters
- Tune query performance for KQL and T-SQL

Table of Contents

Chapter 1: Configure Microsoft Fabric Workspaces

Topics:

- Configure Fabric workspace settings
- Configure Spark workspace settings
- Configure domain, OneLake, and data workflow workspaces

Configure Fabric workspace settings

Configuring Microsoft Fabric workspace settings involves several critical steps: setting workspace properties (name, description, capacity), managing access and contact lists, integrating SharePoint and Microsoft 365 Groups, and connecting external storage and logging services.Proper configuration ensures secure, scalable, and collaborative environments for end-to-end analytics workflows in Fabric.

Workspace Properties and General Settings

First, open your workspace's settings by selecting **More options (…) > Workspace settings** in the left navigation pane of the Fabric portal. In the **General** tab, you can edit the workspace's name and description to reflect its purpose, choose the reserved capacity or Fabric SKU for performance and cost management, and set the workspace's region to optimize latency and compliance.

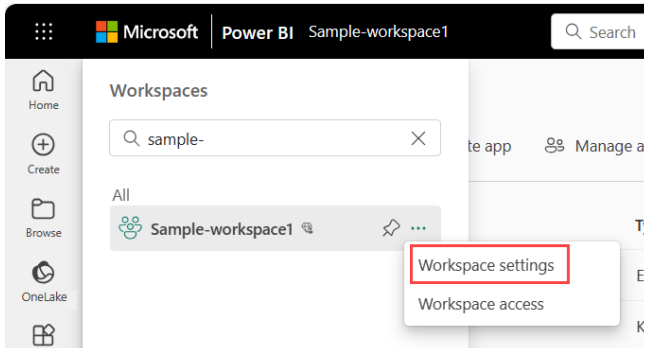


Image Credit: Microsoft

Contact List and Notifications

Under Workspace contact list, add users or Azure AD groups to receive email alerts when workspace issues occur. By default, the creator is listed; extending this list helps distribute operational responsibility and ensures timely incident response.

Microsoft 365 Group and SharePoint Integration

Fabric workspaces can leverage Microsoft 365 Group–backed SharePoint document libraries for file storage. In **Workspace settings**, type the name of an existing Microsoft 365 Group (not the URL) to link its SharePoint library to your workspace. This configuration centralizes file management and aligns workspace access with SharePoint permissions. A best practice is to grant workspace access through the same Microsoft 365 Group to synchronize file and workspace permissions.

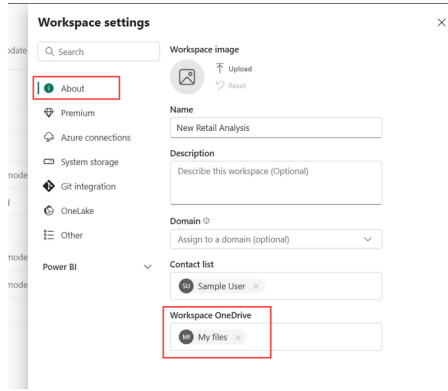


Image Credit: Microsoft

Azure Data Lake Storage Configuration

In the **Azure connections section**, configure Azure Data Lake Gen2 as your workspace's dataflow storage. This "bring your own storage" feature enhances data governance and scalability by using your organization's ADLS account for dataflows.

You can also set up an Azure Log Analytics connection to capture usage metrics and performance logs for troubleshooting and auditing purposes.

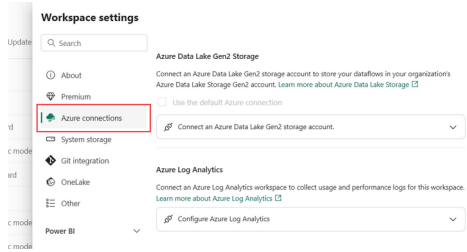


Image Credit: Microsoft

Role-Based Access and Permissions

Workspace admins control user roles (Admin, Member, Contributor, Viewer) in the **Access tab**. Role assignments here govern permissions at the workspace level, while individual item- or folder-level ACLs can be set within specific artifacts to enforce fine-grained access control.

Additional Settings for Data Engineering

When you choose the **Data Engineering/Science** option in settings, you access Spark compute configurations. Here, admins can adjust Spark pool settings, such as node sizes, auto-scaling policies, and session timeouts, to optimize performance for notebook and pipeline workloads.

Configure Spark workspace settings

Configuring Spark workspace settings in Microsoft Fabric centers on defining compute pools, environments, and resource profiles to match workload demands and governance requirements.

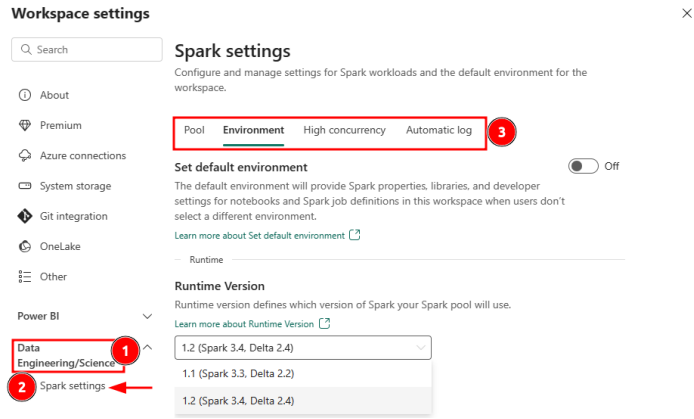
You can choose between the built-in Starter Pool or custom Spark pools, attach environments for consistent library and runtime configurations, and fine-tune session-level and job-level properties.

Proper setup ensures predictable performance, efficient resource use, and simplified management for data engineering workloads.

Access and Navigation

To manage Spark settings, you must have the **Workspace Admin** role for the target Fabric workspace. In the Fabric portal, navigate to

Workspace settings and select **Data Engineering/Science** to expand the Spark configuration menu. Here you'll find options for **Pool, Environment**, **Automatic Log**, **High Concurrency**, and other Spark job properties.



Pools: Starter vs. Custom

- **Starter Pool**: A medium-sized, prehydrated cluster automatically created based on your Fabric capacity SKU. You can adjust its **max node** and **max executors** to align with your expected workload, though changes may increase session startup times .
- **Custom Spark Pool**: Offers full control over node sizes, autoscaling policies, and dynamic executor allocation. To enable custom pools, the capacity admin must turn on **Customized workspace pools** in the Fabric capacity settings.

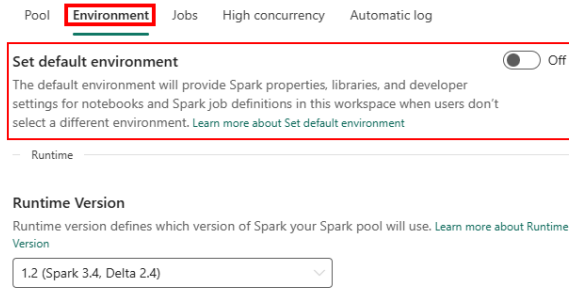
Capacity-based defaults for Starter Pools include SKUs from **F2** (1 node) up to **F2048** (200 nodes) with corresponding vCore limits, allowing you to right-size resources without manual setup.

Environment Management

Environments define the Spark runtime and libraries for notebooks and job definitions. In **Workspace settings > Data Engineering/Science > Spark settings > Environment**, you can:

1. **Set a Default Environment** for the entire workspace, ensuring all attached sessions use a consistent Spark runtime and library set.

2. **Attach Environments** to individual notebooks or Spark job definitions, granting access to specific configurations and dependencies.
3. **Share or Edit Permissions** for environments to control who can view or modify these configurations.



Session-Level Compute Customization

By toggling **Customize compute configurations for items**, admins and delegated members can override default pool settings on a per-session basis, selecting from the list of available pools. If disabled, all Spark jobs adhere strictly to the workspace's default pool configurations.

Library and Dependency Management

- Use **Manage Apache Spark libraries** to install and persist libraries in environments. Admins can:
- Create environments with required libraries
- Attach these environments as workspace defaults or to specific code items
- Ensure reproducible session environments across notebooks and jobs

Resource Profile Configuration

The **Resource Profiles** feature lets you predefine Spark configuration sets, such as **writeHeavy**, **readHeavyForSpark**, or custom profiles, to optimize for common workload patterns. Edit or create profiles under **Spark Configurations** in an environment to apply these settings globally.

Applying Changes and Best Practices

- **Restart Sessions**: Changes to workspace-level settings only apply to new Spark sessions; active sessions must be restarted to pick up updates.
- **RBAC for Governance**: Use Azure RBAC to grant Admin permissions via the **Manage Access** menu, ensuring only authorized users can alter Spark configurations.
- **Monitor and Audit**: Connect Azure Log Analytics in **Azure connections** to capture Spark job metrics and logs for performance tuning and troubleshooting.

By following these configuration steps and adopting these best practices, you'll establish a robust, scalable Spark environment within Microsoft Fabric that supports both development agility and operational stability.

Configure domain, OneLake, and data workflow workspaces

Microsoft Fabric uses **domains** to group related workspaces, **OneLake** as its unified data storage layer, and **data workflow workspaces** to manage pipelines and task flows. Proper configuration ensures governance at scale, seamless data access, and efficient orchestration of analytics workloads.

Configure Domains

Purpose and Overview

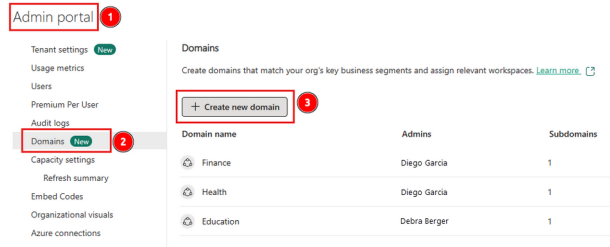
Domains let organizations segment workspaces by business unit, project, or data mesh boundary. Assigning a workspace to a domain associates all its artifacts, lakehouses, warehouses, pipelines, with that domain for catalog filtering and delegated governance.

Creating and Configuring Domains

- In the Fabric admin portal, **select Domains** and click **Create new domain**. Provide a name and optional domain admins (Fabric admins or designated business owners).
- Use **General settings** to edit the domain's display name and description; **Image** to choose a logo for OneLake catalog

theming; **Admins** and **Contributors** tabs to assign roles for managing and associating workspaces.

- Define a **default domain** for specific users or groups so that new workspaces they create,and unassigned existing ones,automatically inherit the domain assignment.



Subdomains and Workspace Assignment

- Subdomains provide further hierarchy under a parent domain; they inherit parent admins and settings but cannot have separate admins.
- Assign workspaces by name, by workspace admin, or by capacity. Overriding existing assignments prompts a warning but allows reclassification.

Delegated Domain Settings

Tenant-level governance policies (e.g., default sensitivity labels, endorsement/certification settings) can be delegated to domains. Domain admins can override these settings per domain to meet local compliance needs.

Configure OneLake

Unified Storage Architecture

OneLake is the single, tenant-wide data lake in Fabric, analogous to OneDrive but for analytics items (lakehouses, warehouses, dataflows). All workspaces write to OneLake under their assigned domain and capacity boundaries.

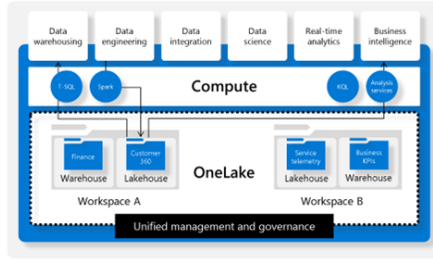


Image Credit: ilink-digital.com

Data Hub and Catalog Filtering

In the OneLake data hub, users can select a domain in the domain selector to filter visible items. The chosen domain's image/theme appears in the UI to remind users which data set they're viewing.

Governance and Compliance

- **Tenant boundary**: The Fabric tenant admin controls the overall OneLake namespace, ensuring compliance and data residency.
- **Domain boundary**: Domain contributors can manage local policies, preventing a single gatekeeper scenario and enabling distributed ownership.

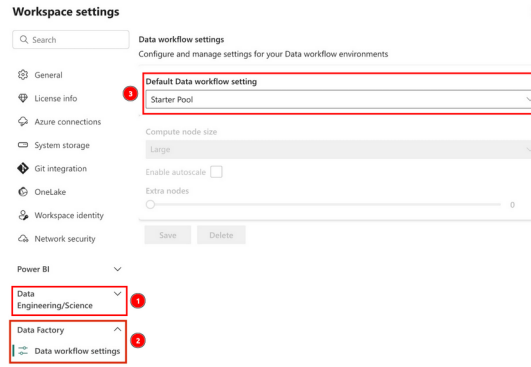
Data Workflow Workspaces

Definition and Role

Data workflow workspaces centralize ETL/ELT pipelines, task flows, and orchestration artifacts. They provide a dedicated environment,separate from lakehouse or warehouse workspaces, for building and managing data movement and transformation logic.

Configuring Default Data Workflow Pools

1. **In Workspace settings > Data Factory** (formerly Data Engineering/ Science), click **Data Workflow Settings**.
2. Set the **Default Data Workflow Pool** to either the Starter Pool (capacity-based) or a Custom Pool (requires capacity admin to enable custom pools).
3. Customize pool properties, node size, auto-scaling limits, session timeouts,to match typical pipeline workloads and control costs.



Advanced Customization

- **Per-pipeline overrides**: Allow workspace members to select different pools for individual pipelines or tasks when "Customize compute configurations for items" is enabled.
- **Task flows integration**: Use the Task Flows feature to visualize and manage dependencies between pipelines, notebooks, and other artifacts in the workspace, streamlining development and collaboration.

Chapter 2: Implement Lifecycle Management

Topics

- Configure version control
- Implement database projects
- Create and configure deployment pipelines

Configure version control

Microsoft Fabric's Git integration lets you manage your workspace artifacts in source control, enabling collaboration, version history, and CI/CD workflows.

You connect a Fabric workspace to an Azure DevOps or GitHub repository, choose a branch and folder, then commit and sync changes bi-directionally.

You can branch out workspaces for feature isolation, undo unwanted edits, and update your workspace from Git. While most actions require workspace-admin rights, once set up, all contributors with repo access can work against the connected branch.

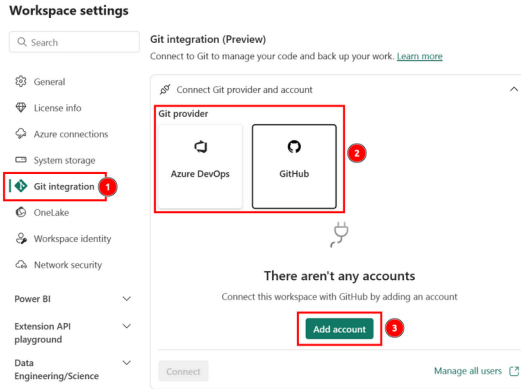
Key considerations include file/folder naming limits, repo location relative to workspace geography, and ensuring appropriate permissions and capacity are in place

Connecting a Workspace to a Git Repository

1. **Open Workspace Settings**
    - In the Fabric portal, click **More options (…)** next to your workspace and select **Workspace settings**.
2. **Enable Git Integration**
    - Select **Git integration**. Choose your provider (Azure DevOps or GitHub). Fabric will authenticate you via Microsoft Entra ID for Azure DevOps or OAuth for GitHub.
3. **Specify Connection Details**

- For Azure DevOps: select Organization, Project, Repository, and Branch.
- For GitHub: sign into GitHub, then pick Owner, Repository, and Branch.
- Optionally define a folder path within the repo to scope your workspace files.

4. **Initial Sync**
   - If either the workspace or branch is empty, content is copied from the nonempty side. If both contain artifacts, you choose sync direction



Working with Branches and Isolated Workspaces

- **Branch Connect**: You can switch the workspace to a different branch via the Git integration pane. Choose an existing branch or create a new one for isolated work.
- **Branch Out**: Fabric can create a new workspace tied to a feature branch. This gives each developer an isolated environment, syncing artifacts only from that branch.
- **Branching Considerations**: Branch names are limited to 244 characters; folder paths to 250 characters; nesting maxes out at 10 levels.

Committing and Syncing Changes

1. **Commit Changes**

- In your workspace, click the **Source control** icon to view uncommitted changes. Select items, add a commit message, and click **Commit**.

2. **Undo Changes**
   - Use **Undo** on uncommitted items to revert to the last synced state. Note: metadata like sensitivity labels may be reset.
3. **Pull Updates**
   - When new commits appear in the branch, the **Source control** panel shows available updates. Click **Update all** to bring your workspace in line with the branch

Best Practices and Advanced Tips

- **.gitignore Setup**: Exclude temporary files and large binaries by customizing your **.gitignore** in Azure DevOps or GitHub to keep your repo clean.
- **Environment Branching**: Use dedicated Dev/Test/Prod branches, each connected to separate workspaces, to mirror your deployment pipeline.
- **Capacity and Geography**: Ensure your Git repo's region matches or allows cross-geo sync with your workspace to avoid connectivity issues.
- **Permission Scoping**: Grant workspace-admin rights sparingly; beyond initial setup, control contributions via repo permissions for fine-grained governance.

Considerations and Limitations

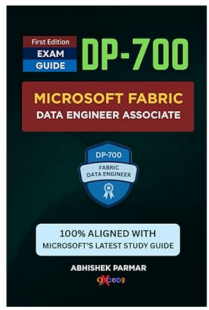
- **Supported Items**: Only Git-enabled artifacts (reports, notebooks, pipelines) are versioned; unsupported items remain workspace-only.
- **Template Apps**: Workspaces with template apps installed cannot connect to Git.
- **Capacity Requirements**: Branch-out and sync operations require available Fabric capacity in your tenant.
- **File Size/Path Limits**: Single file size max 25 MB; directory names follow Windows naming restrictions (no **<**, **>**, **:**, **"**, **|**, **?**, *****, or trailing spaces).

By following these steps and best practices, you'll set up robust version control for your Fabric workspaces, paving the way for reliable collaboration and streamlined CI/CD processes.

Content beyond this point is not included in the free version.

Content beyond this point is not included in the free version.

Buy Full Book From Amazon



Read sample

DP-700 Exam Guide: Microsoft Fabric Data Engineer Associate (Microsoft Certification Preparation Series (Unofficial))

by Abhishek Parmar (Author)

Part of: Microsoft Certification Preparation Series (Unofficial) (4 books) See all formats and editions

100% Aligned with Microsoft's Latest Study Guide
Guide to Microsoft DP-700 Certification: Fabric Workspaces, Ingesting Data, and Optimizing Analytics Solutions in MicrosoftFabric

This book is built for professionals preparing for the **DP-700: Data Engineering on Microsoft Fabric** certification, focusing on real understanding.

Each chapter follows the official exam structure and explains core concepts in a clear, practical way. You'll learn how to configure workspaces, manage deployment pipelines, and apply governance and security settings correctly.

Also, Understand how to load and transform data using batch and streaming methods, and how to use tools like PySpark, KQL, and T-SQL effectively. Learn how to handle real-world challenges like schema changes, late-arriving data, and performance bottlenecks.

Whether you're preparing for the exam or building your career as a data engineer, this book gives you the clarity and

Read more

Report an issue with this product or seller

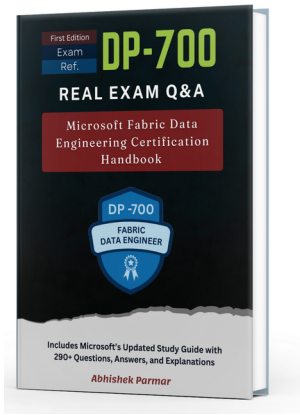
ISBN-10	ISBN-13	Publication date	Part of series	Language
9334311169	978-9334311167	May 28, 2025	Microsoft Certification Preparation...	English

Click Below link or Scan QR Code

https://9xcode.com/DP%E2%80%91700-EGMFDEA

Explore More Books by the Author

If you enjoyed this book, you may also find value in these other titles

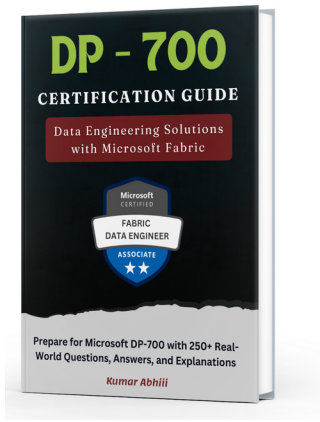


DP-700 Real Exam Q&A: Microsoft Fabric Data Engineering Certification Handbook

Includes Microsoft's Latest Study Guide with 290+ Real-World Questions, Answers, and Explanations.

https://9xcode.com/DP-700-Real-Exam-Engineering-Certification

Scan This →



DP-700 Certification Guide: Data Engineering Solutions with Microsoft Fabric

Prepare for Microsoft DP-700 with Real-World Questions, Answers, and Explanations

https://9xcode.com/DP-700-Certification-Guide-Engineering-Preparation

Scan This →

View All Microsoft Certification Books by Author

https://9xcode.com/ms-certification-books