“**Deep Dive: Creating Your First Data Connection (update)**”是发布在 learn.palantir.com 上的一门课程，由 Ontologize（由前 Palantir 工程师组成的培训团队）提供 1。该课程旨在指导用户如何将各种外部系统的数据连接并摄取（ingest）到 Palantir Foundry 中，重点涵盖了文件系统、关系型数据库和 **REST** 端点这三种主要的数据源类型 1。
以下是基于来源对该课程内容的详细介绍：

1. 核心概念与准备工作

在开始实际操作之前，课程强调了几个关键的基础知识：

- 权限与安全：由于从 Foundry 外部拉取数据涉及信息安全，因此需要特定的高级权限 2。如果用户缺乏权限，通常需要联系平台管理员或信息安全官（ISO）2, 3。
- **Egress Policy**（流出策略）：这是连接外部系统的关键。它是由 ISO 批准的策略，允许流量从 Foundry 流向特定的外部地址 3, 4。课程中涉及的所有连接都需要配置或使用现有的 Egress Policy 3, 5。
- **Source**（源）与 **Sync**（同步）的概念：
- **Source** 包含连接外部系统的指令（如 URL、凭据、端口号）6。
- **Sync** 包含从 Source 中检索哪些具体数据的指令（如特定的数据库表、S3 存储桶中的特定文件）4。
- 关系：一个 Source 可以有多个 Sync，但一个 Sync 只能属于一个 Source（一对多关系）4。

2. 三种主要连接类型的摄取流程

A. 连接 S3 存储桶（文件系统）

这是最基础的连接练习，通过 **Data Connection** 应用完成 4, 7：

1. 配置 **Source**：输入 S3 的存储桶 URL、区域（如 EU-West-1）以及访问密钥（Access Key/Secret）8。
2. 设置 **Egress Policy**：选择或请求允许访问该 S3 地址的策略 5。
3. 定义输出文件夹：通常在 Foundry 的项目文件夹内创建一个名为 raw 的文件夹，用于存放生成的原始数据集 9。
4. 创建 **Sync**：浏览 S3 源中的文件（如 CSV 文件），将其添加为同步任务 10。
5. 应用架构（**Schema**）：初始摄取的 CSV 在 Foundry 中只是文件形式，需要点击“Apply a schema”让 Foundry 推断其结构并将其渲染为表格数据集 11。

B. 连接 PostgreSQL（关系型数据库）

连接数据库的步骤与 S3 类似，但涉及更多的网络和安全配置 12：

1. 连接详情：输入主机名、端口（默认 5432）、数据库名称和身份验证凭据 12, 13。
2. **SSL** 配置：通常需要将 SSL 模式设为 **Require**，并下载并粘贴服务器证书（server.pem）的内容 13, 14。
3. 故障排除：如果连接失败（出现 Configuration Error），通常是因为凭据错误或端口/Egress Policy 不匹配 15, 16。
4. 同步表：选择数据库中的具体表（如 plants），并将其同步为 Foundry 中的数据集 16, 17。

C. 连接 REST API（编程接口）

这是最复杂的连接方式，分为两个主要步骤 18：

- 在 **Data Connection** 中创建源：配置 API 端点、端口和必要的 Secret（如 Token），并确保开启“允许将此源导入代码仓库”的开关 18-20。
- 在 **Code Repositories** 中编写代码：
- 使用 **Python** 创建一个新的存储库 21。
- 安装特定的外部转换库 transforms-external-systems 22。
- 通过 **RID**（资源标识符）引用之前创建的 REST API 源 23, 24。
- 编写逻辑来解析 API 响应，并使用 transform_df 装饰器将结果输出为数据集 24, 25。

3. 后续步骤与管理

完成数据连接后，课程还介绍了一些高级管理工具：

- **Data Lineage**（数据血缘）：通过此应用可以清晰地看到从外部 Egress Policy 到 Source，再到 Sync 生成的数据集（或代码转换）的端到端流程 26, 27。
- 调度（**Schedule**）与检查（**Checks**）：可以为同步任务设置运行频率（例如每天更新），并添加数据质量检查 27, 28。
- 最佳实践：课程最后提供了关于保持数据连接高效性和遵循安全规范的技巧 27。

通过这门课程的学习，用户能够掌握在 Foundry 环境下从零开始构建、调试和管理数据接入通道的完整技能 1, 27。