

Azure Synapse Analytics로 Power BI를 강화하는 5가지 방법



Azure Synapse Analytics로 Power BI를 강화하는 5가지 방법

3 /

소개말

21 /

#3: 완전 관리형 서버리스
엔드포인트를 활용한 데이터
레이크 탐색

36 /

요약

6 /

#1: Azure Synapse에서 직접
Power BI 대시보드 구축 및
분석

30 /

#4: 코드 없는 데이터
파이프라인을 구축하여 더
많은 데이터 소스 통합
및 인사이트 강화

37 /

다음 단계

15 /

#2: 데이터 레이크 관리 및 BI
보고서용 데이터 마트 구축

33 /

#5: 안전한 데이터
웨어하우스를 만들고
Power BI 대시보드에 연결

소개말

오늘날 모든 기업은 데이터로 더 많은 작업을 수행하려고 합니다. 조직은 보고 기능을 넘어 사용자에게 조직의 데이터에 대한 더 많은 인사이트를 제공하기 위한 방법을 검토하고 있습니다. 조직은 Azure Synapse Analytics를 활용하여 조직의 모든 데이터에 대한 중앙 리포지토리로 클라우드를 활용할 수 있습니다. 정보는 데이터 열에만 국한되지 않습니다. 기업은 Azure Synapse를 통해 이미지, 사운드 또는 PDF 문서의 캐시를 비롯한 여러 정보를 이동시키고 분석할 수 있습니다. 정보는 Power BI를 통해 데이터에 액세스하는 기업의 능력을 극대화하고 인사이트가 담긴 정보를 제공하도록 구성될 수 있습니다. Power BI에 활용되는 데이터는 Azure Synapse에서 생성된 전체 데이터 스토리지에서 가져올 수 있습니다. Azure Synapse는 머신 러닝 또는 Power BI용 가상 데이터베이스에 활용할 수 있는 데이터 레이크로 데이터를 로드하는 데 필요한 도구가 포함되어 있으므로 중앙 데이터 리포지토리로 이용됩니다. 데이터의 양이 매우 방대한 경우 Azure Synapse는 Power BI 보고에 활용되는 최적화된 리소스를 통해 데이터 웨어하우스와 데이터 마트를 생성할 수 있습니다.

Azure Synapse가 조직에 필요한 데이터를 처리하면 다음 단계는 Power BI와 같은 데이터 시각화 환경을 구현하는 것입니다. Power BI는 사용자가 비즈니스 질문에 대한 답을 찾을 수 있도록 데이터와 상호 작용하는 환경을 조성하여 데이터에 대한 더 많은 인사이트를 확보할 수 있도록 지원합니다. Power BI에 활용되는 데이터 모델 개발은 조직의 모든 부서에게 데이터에 대한 액세스를 제공하는 첫 번째 단계입니다. 데이터는 IoT 디바이스와 소셜 미디어뿐만 아니라 조직이 일상적인 환경에서 이용하는 많은 애플리케이션에서 가져올 수 있습니다. 보고되어야 하는 데이터는 Azure Event Hubs, 데이터 웨어하우스, 데이터베이스 또는 데이터 레이크를 통해 스트리밍될 수 있습니다. 데이터는 보고뿐만 아니라 고급 분석 AI/머신 러닝 프로세스에 활용될 수 있도록 액세스 가능한 위치에 있어야 합니다. 데이터가 한 위치에 저장되면 Power BI 보고뿐만 아니라 데이터 과학 모델링을 위해서도 액세스할 수 있습니다.

이 작업에는 많은 데이터가 필요하므로 데이터를 안전하고 확장 가능한 위치에 저장해야 하며, 데이터는 요구 사항을 충족시키기 위해 클라우드에 저장되어야 합니다. Microsoft Azure에 데이터 스토리지를 집중시키면 안전하고 확장 가능한 오류 방지 위치가 제공됩니다. 오늘날 많은 기업들은 텍스트가 아닐 수도 있는 데이터를 저장하고 분석하고자 합니다. 동영상, 사진 및 사운드 파일은 종종 전용 AI 라이브러리를 활용하여 분석됩니다. 데이터 레이크가 이러한 다양한 데이터에 대한 요구에 가장 적합한 솔루션인 경우가 많습니다.

데이터 레이크는 일일 트랜잭션, 스트리밍된 GPS 위치 데이터 및 이미지 파일의 데이터를 조직이 유지하려는 다른 데이터와 함께 저장하는 논리적 위치입니다. 데이터가 저장되는 방법이 분석 능력을 결정합니다. Azure는 Power BI뿐만 아니라 분석에 활용할 수 있는 계층적 파일 구조를 제공합니다. 데이터 과학자는 이 데이터를 활용한 모델을 만들어 Power BI에서 활용할 수 있는 추가적인 인사이트를 제공할 수 있습니다. 또한 데이터 과학자는 Power BI를 통해 데이터를 신속하게 분석하여 모델에 필요한 요소를 파악할 수 있습니다. Power BI는 대부분 최근에 추가된 목표, 주요 영향 요인, 분해 트리 및 스마트 내러티브와 같은 다양한 비즈니스 인텔리전스 기능을 제공하여 조직의 성공에 기여하는 요인에 대한 추가적인 인사이트를 제공할 수 있습니다.

데이터가 매우 크고 데이터 웨어하우스에 저장된 정형 데이터나 이미지와 같은 비정형 데이터 및 그 사이에 있는 모든 데이터를 포함할 수 있기 때문에, 분석에 활용되는 데이터를 관리하는 작업은 어려울 수 있습니다. 데이터 관리에는 더 많은 데이터를 추가하고 비즈니스 인텔리전스 작업에 필요한 형식으로 구성하기 위한 방식이 포함되어야 합니다. 추출, 로드 및 전송 방식을 활용하여 조직 또는 서드파티 애플리케이션의 여러 부분에서 데이터를 수집할 수 있습니다. 조직이 보유한 기술에 기반하여 로우 코드 솔루션, 데이터 흐름 또는 Python 코드로 데이터 전송이 이루어질 수 있습니다. 운영 인사이트를 제공하기 위해 데이터를 수집해야 하기 때문에 데이터의 소스가 어디에 있는지는 중요하지 않습니다. 조직은 비즈니스 질문에 대한 답변을 제공하기 위해 정확한 의사 결정에 필요한 모든 정보를 제공하는 솔루션을 개발해야 합니다. 머신 러닝 요소는 기업의 미래 상태를 결정하거나 발견되지 않았을 수 있는 성능에 대한 이상 징후를 파악하기 위해 예측 분석을 제공하여 의사 결정 프로세스를 더욱 향상시킬 수 있습니다.

Power BI가 액세스할 수 있도록 이러한 데이터를 제공하고 다른 사용자가 데이터를 제공할 수 있는 인사이트를 제공하는 작업은 모두 통합 분석 플랫폼인 [Azure Synapse Analytics](#)라는 하나의 도구에서 이용할 수 있습니다. Power BI는 데이터 레이크의 데이터를 보고하는 데 활용할 수 있고, Azure Synapse는 여러 솔루션에서 데이터를 수집하고, 대규모 데이터 웨어하우스를 만들고, 데이터 레이크를 관리하고, BI 보고 대상을 위한 대상 데이터 마트를 구축하고, 머신 러닝 솔루션을 개발하고, 보고 제공에서 비즈니스 분석 솔루션 제공으로 이동하는 데 필요한 정보를 Power BI에 제공하는 기능을 제공합니다.

이 eBook에서는 Azure Synapse와 Power BI를 활용할 수 있는 5가지 방법을 살펴보겠습니다.

- #1: Azure Synapse에서 직접 Power BI 대시보드 구축 및 분석
- #2: 데이터 레이크 관리 및 BI 보고서용 데이터 마트 구축
- #3: 완전 관리형 서버리스 엔드포인트를 활용한 데이터 레이크 탐색
- #4: 코드 없는 데이터 파이프라인을 구축하여 더 많은 데이터 소스 통합 및 인사이트 강화
- #5: 안전한 데이터 웨어하우스를 만들고 Power BI 대시보드에 연결

조직에 데이터를 제공하는 데 Azure Synapse를 활용할 수 있는 방법에 대한 더 큰 인사이트를 제공하기 위해 Azure Synapse를 Power BI와 활용할 수 있는 5가지 방법입니다. Power BI를 활용하여 Azure Synapse의 데이터를 분석하는 것으로 시작해 보겠습니다.



#1: Azure Synapse에서 직접 Power BI 대시보드 구축 및 분석

Power BI는 Azure Synapse에 이미 있을 수 있는 데이터를 분석하는 데 도움이 되는 많은 도구를 제공합니다. 시각적 도구를 활용하여 현재 데이터의 상태를 검사하면 서버리스 풀, 전용 풀을 포함하여 Azure Synapse의 여러 영역에 저장된 데이터의 현재 상태를 확인하거나 중간 로딩 프로세스에서 생성된 데이터를 검사하는 빠른 방법을 제공할 수 있습니다. Azure Synapse에 Power BI를 활용하는 기능을 추가하려면 Azure Synapse 워크스페이스와 Power BI 인스턴스가 동일한 Azure 테넌트에 있어야 합니다. Azure Synapse 워크스페이스를 생성하려면 몇 분 정도 걸리며 생성 프로세스의 단계를 [여기](#)에서 확인할 수 있습니다. 수신된 데이터의 품질을 검토하여 수행할 수 있는 분석 종류를 결정하기 위해서는 데이터를 검사해야 할 수도 있습니다. 클라이언트로부터 일부 데이터를 수신하고 이 데이터를 다른 데이터 모델에 포함시키는 방법을 결정하기 위해 빠른 보고를 수행하려고 할 수 있습니다.

[이 동영상 자습서](#)에서 Azure Synapse 워크스페이스를 만드는 방법에 대해 알아보세요.

Azure Synapse를 활용하면 연결된 서비스를 통해 다른 애플리케이션에 연결할 수 있습니다. Azure Key Vault, Azure Machine Learning은 물론 Power BI를 비롯한 다양한 애플리케이션을 포함시킬 수 있습니다. 연결된 서비스는 Azure Synapse 내에서 Power BI의 기능에 액세스할 수 있도록 연결 정보를 제공합니다. 사용자는 연결된 서비스로 Power BI의 워크스페이스에 연결할 수 있습니다. 이전 릴리스에서는 하나의 Power BI 워크스페이스에 연결할 수 있었습니다. 현재는 여러 Power BI 워크스페이스에 연결할 수 있습니다.

Azure Synapse 내에서 추가 ADLS 계정 또는 공개적으로 이용 가능한 데이터 집합과 같은 요소에 대한 다른 연결된 서비스가 많이 있을 수 있습니다. Power BI에 연결된 서비스 연결을 생성하려면 Azure Synapse 포털 왼쪽에 있는 **관리** 아이콘을 선택한 다음 **외부 연결**에서 **연결된 서비스**를 그림 1에 표시된 대로 선택합니다. Power BI에 연결된 서비스는 보라색 사각형으로 둘러싸여 있으며 **PowerbiAcctDev**라고 불립니다.

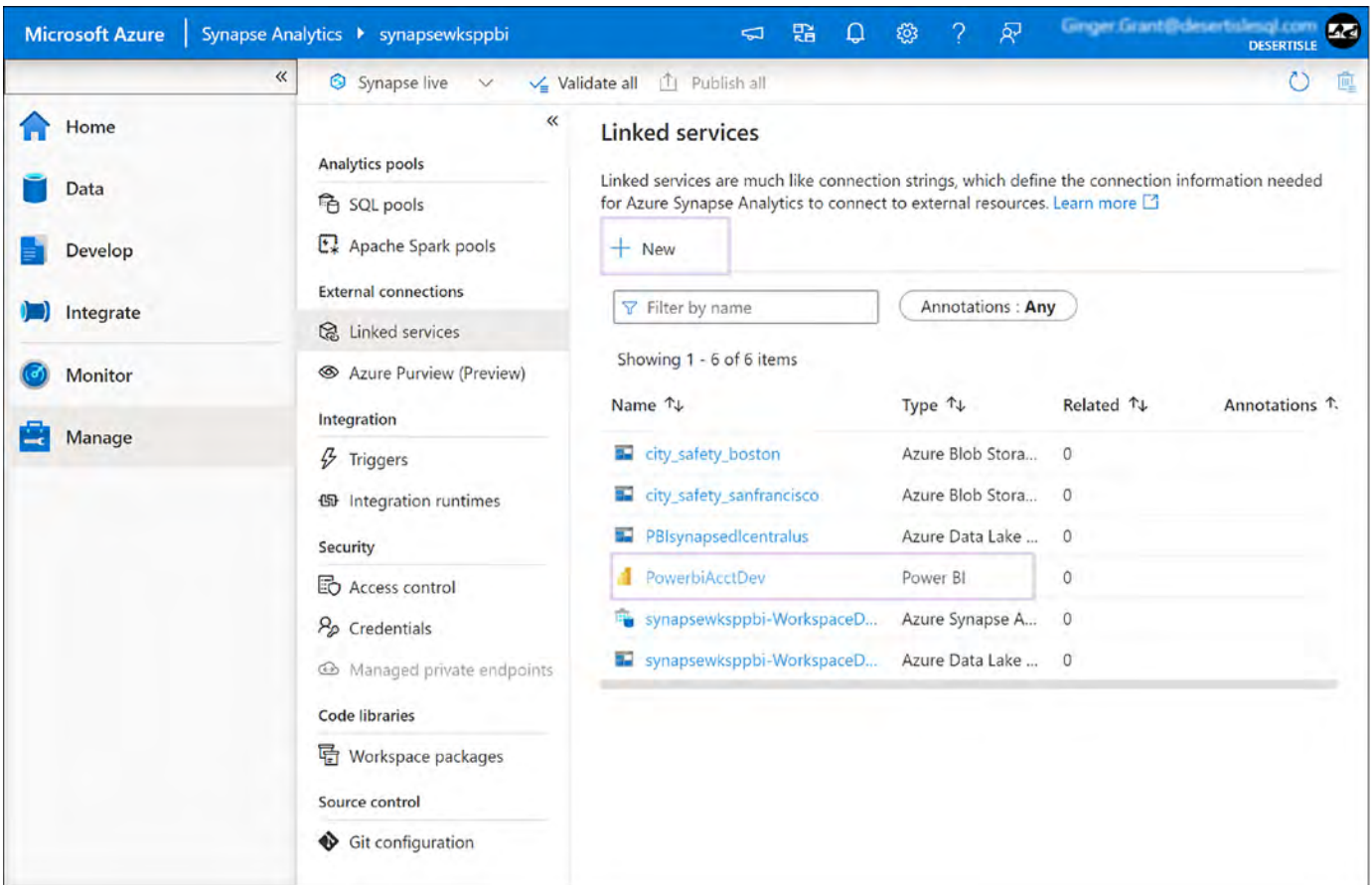


그림 1: Azure Synapse Analytics 계정에서 생성한 연결된 서비스 연결

Power BI에 연결된 서비스를 **PowerbiAcctDev**라고 하며 예시에 활용된 Azure 계정과 동일한 테넌트의 워크스페이스입니다. 이용된 Azure 계정은 Power BI 계정과 동일한 테넌트에 있어야 합니다. 예를 들어 Outlook.com 계정과 연동된 Azure 계정을 보유한 경우 해당 계정을 desertislesql.com 테넌트와 연동할 수 없습니다.

새 연결된 서비스를 추가하려면 **연결된 서비스** 창에서 **+ 신규** 텍스트를 클릭합니다. Power BI 워크스페이스에 아직 연결이 생성되지 않은 경우 메뉴는 다음과 표시됩니다.

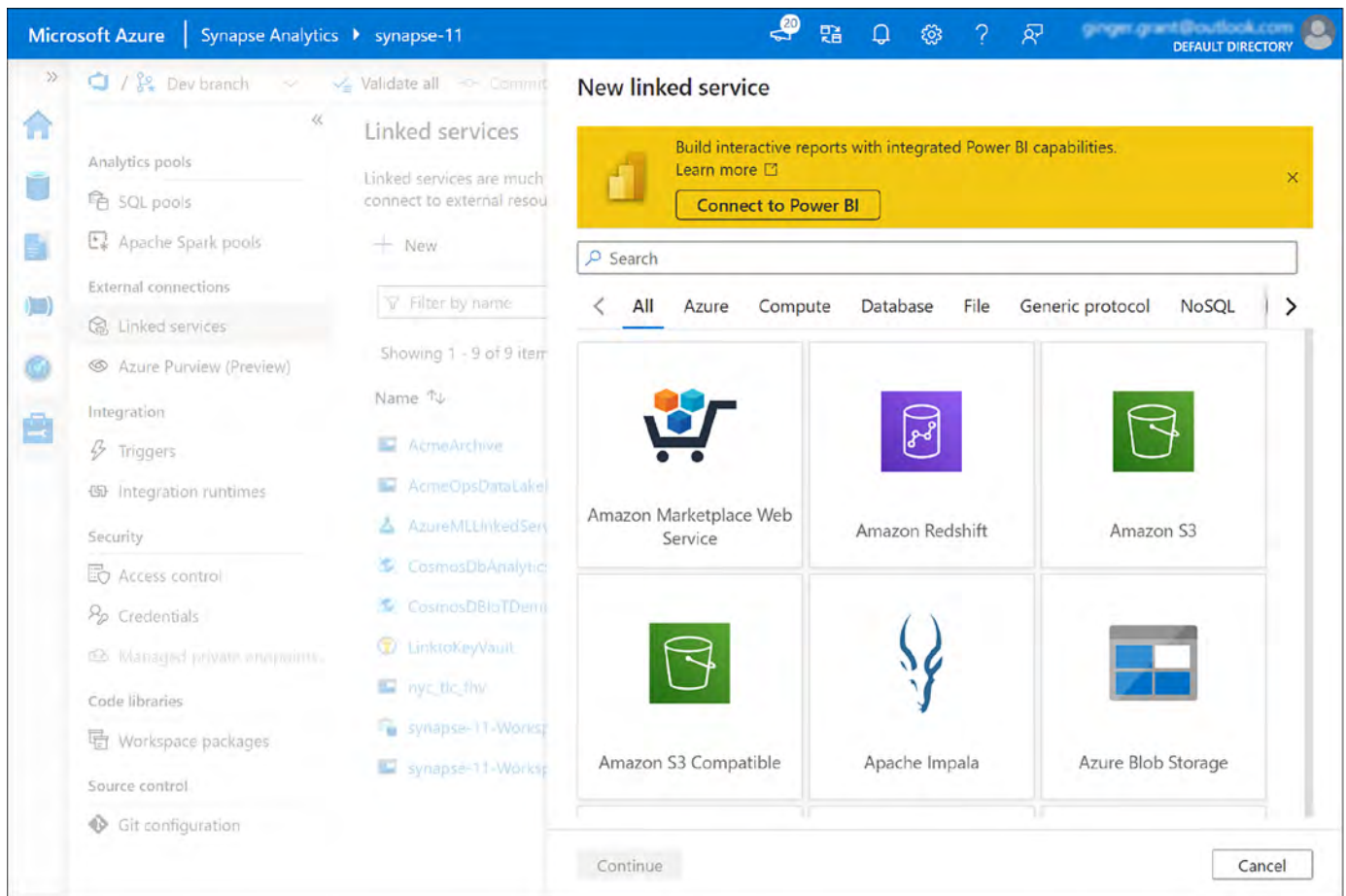


그림 2: Azure Synapse Analytics 새 연결된 서비스 메뉴

Power BI에 연결 단추를 클릭합니다. Power BI도 메뉴에 나열됩니다. 메뉴 옵션에 액세스하려면 창 오른쪽에 있는 화살표를 클릭하면 됩니다. 두 옵션 모두 그림 3에 표시된 대로 동일한 화면으로 이동하므로 Azure Synapse를 Power BI 워크스페이스에 연결을 완료해야 합니다. Azure Synapse에서 연결된 서비스의 원하는 이름으로 선택할 수 있지만 나중에 이름을 변경할 수 없다는 경고를 받게 됩니다.

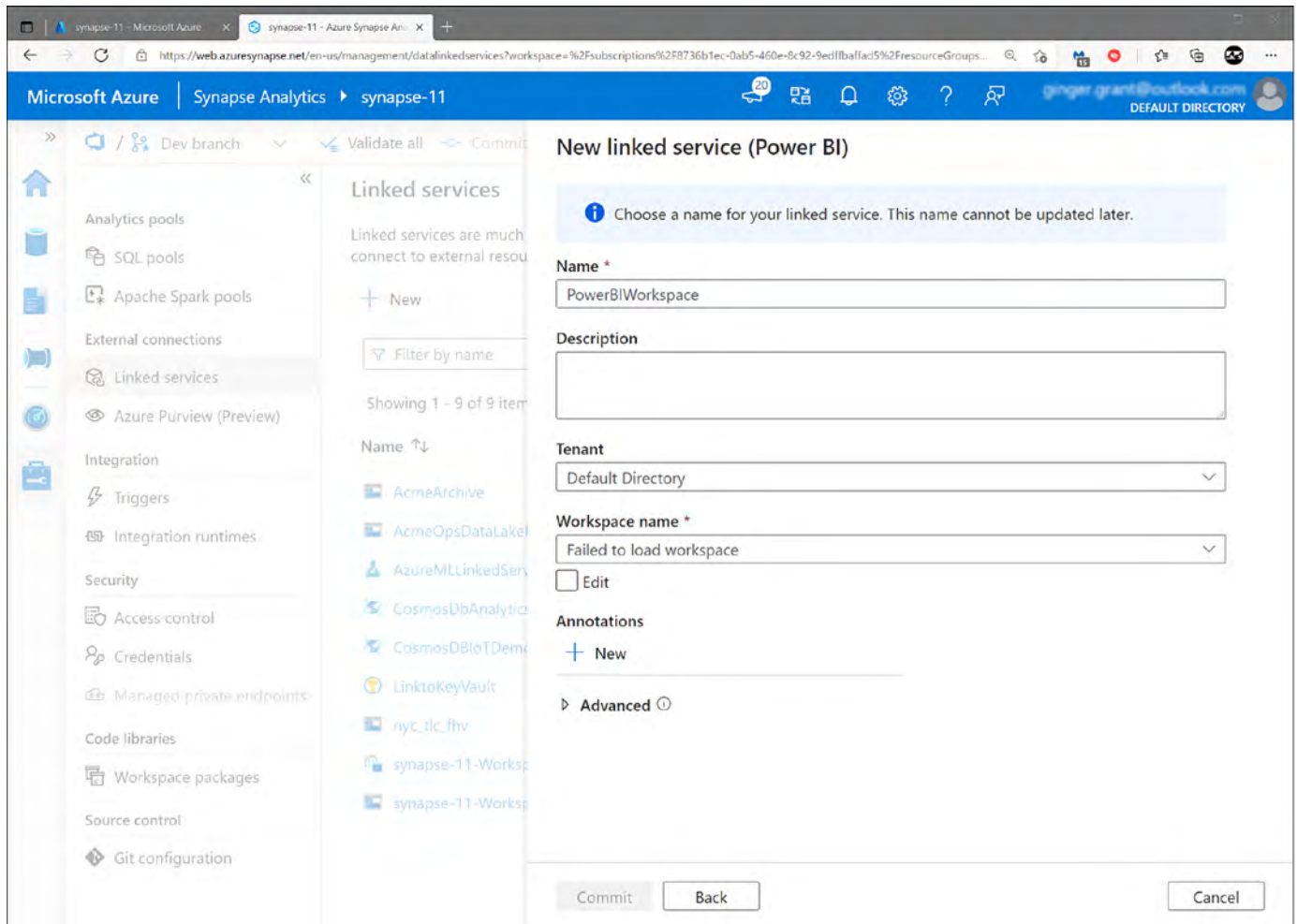


그림 3: 연결된 서비스를 활용하여 Power BI 워크스페이스를 Azure Synapse에 연결

드롭다운 목록에서 **테넌트**를 선택해야 하며, 이렇게 하면 **워크스페이스 이름** 드롭다운 목록이 액세스할 수 있는 모든 워크스페이스로 채워집니다. 워크스페이스 내부의 데이터는 ADLS를 비롯한 여러 위치에서 가져올 수 있으며 OneDrive 또는 Azure SQL 워크스페이스에 저장된 Excel 파일과 같은 위치에서도 데이터를 가져올 수 있습니다. 이 양식을 완료한 후 **커밋** 단추를 클릭합니다. 추가 프로세스를 완료하려면 새 연결된 서비스를 게시하는 데 주의해야 합니다.

Azure Synapse에 Power BI가 추가되면 **개발** 창에서 활용할 수 있습니다. **Power BI** 섹션이 추가된 것을 확인할 수 있습니다. 섹션 내에서 워크스페이스에 포함된 데이터 집합 및 보고서를 볼 수 있습니다. Azure Synapse 내에서 Power BI를 활용하여 기존 보고서를 수정할 수 있으며, 수정할 때 워크스페이스에 액세스할 수 있는 모든 사용자가 이용할 수 있는지 확인하기 위해 저장해야 합니다. 그림 4는 Power BI 내에서 기존 데이터 집합을 수정하는 방법을 보여 줍니다.

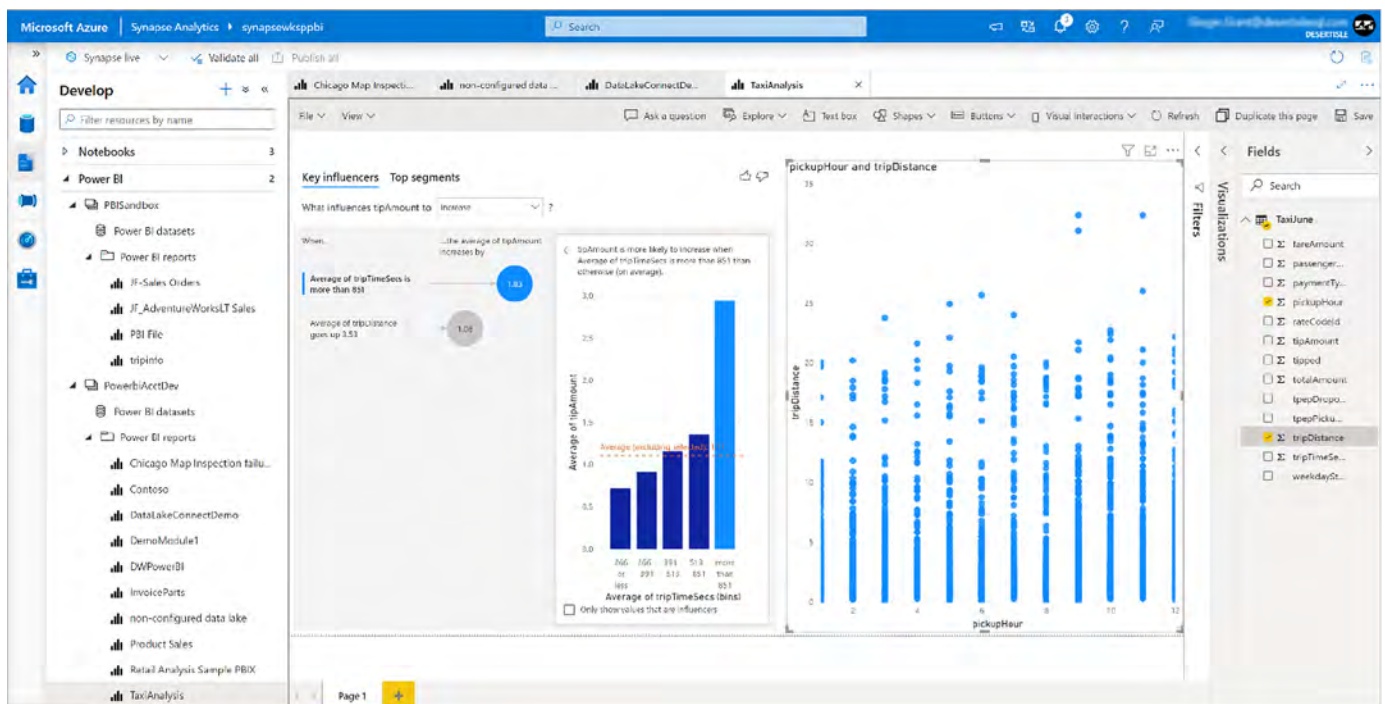


그림 4: Azure Synapse 내에서 Power BI에 연결된 서비스 보고서 개발

데이터 집합은 Power BI용 Azure Synapse에서 새 보고서를 만드는 위치이기도 한 개별 창에 나열됩니다. 목록의 데이터 집합 이름 위로 마우스를 가져가면 그림 5에 표시된 대로 두 개의 아이콘이 나타납니다.

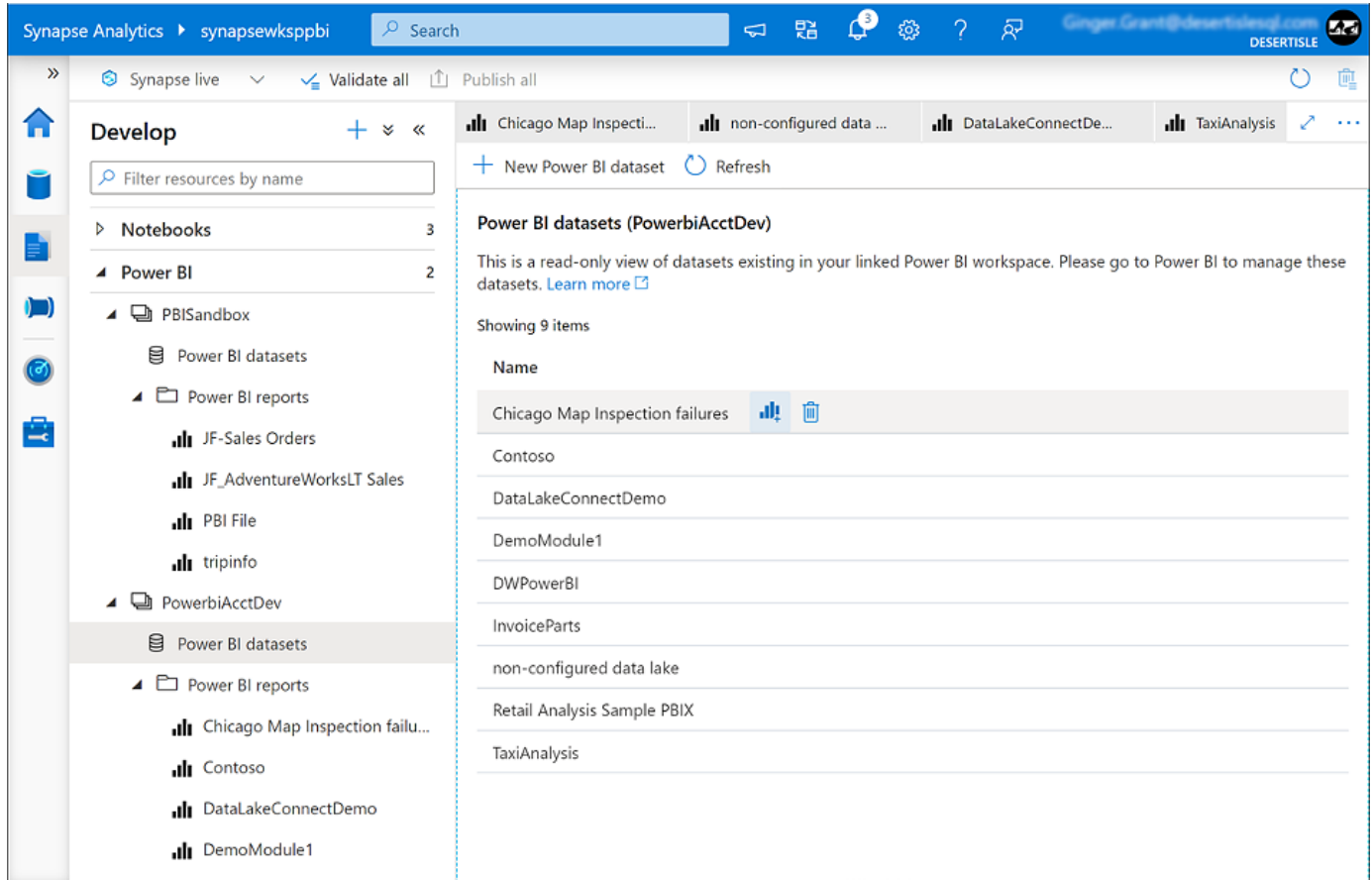


그림 5: 새 보고서를 만드는 데 활용되는 Azure Synapse의 Power BI 워크스페이스에서 데이터 집합 보기

첫 번째 아이콘을 선택하면 강조 표시된 데이터 집합을 기반으로 새 보고서를 추가하거나 데이터 집합을 삭제할 수 있습니다. Azure Synapse 내에서 Power BI를 활용하여 데이터를 분석하면 데이터 모델의 데이터를 검사하고 Power BI에서 만들 수 있는 결합된 데이터 모델을 활용하는 데 걸리는 시간을 단축할 수 있습니다. Azure Synapse에서 워크스페이스 내의 모든 모델을 이용할 수 있습니다. 새 모델이 워크스페이스에 로드되면 Azure Synapse에서도 이용할 수 있습니다. 새로운 모델을 확인하려면 그림 5에 표시된 대로 화면 상단에 있는 **새로 고침** 단추를 활용하여 데이터 집합을 새로 불러올 수 있습니다.

Knowledge Center의 예제 데이터 집합 활용

Azure Synapse에서 이용할 데이터 집합이 아직 없는 경우 그림 6에 보라색으로 강조 표시된 대로 워크스페이스 홈페이지에서 **학습** 단추를 클릭하여 찾을 수 있는 Knowledge Center에서 제공하는 예제 데이터 집합 중 하나를 활용할 수 있습니다.

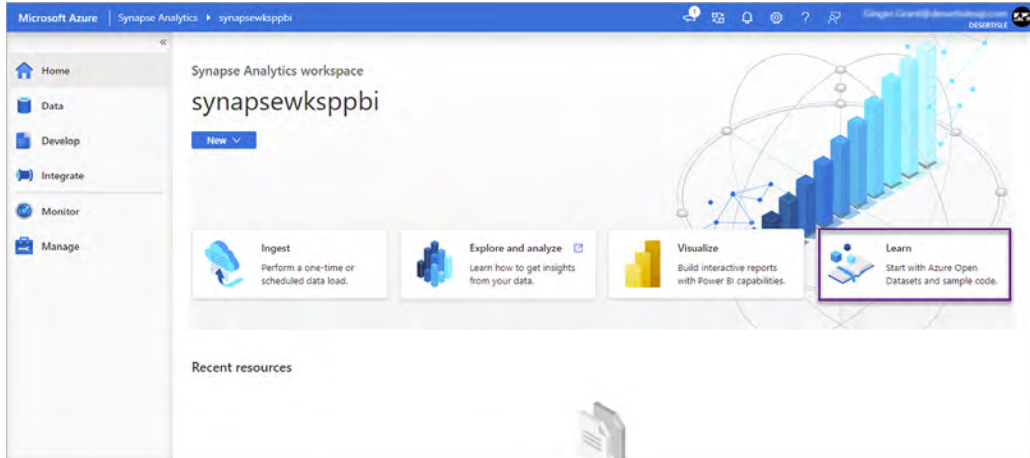


그림 6: Azure Synapse에 대해 자세히 알아볼 수 있는 Azure Synapse Learn에 포함된 Knowledge Center

Azure Synapse는 Azure Synapse와 더 익숙해지도록 활용할 수 있는 많은 코딩 예제와 데이터 집합을 제공합니다. 액세스하려면 **학습**을 선택하여 Knowledge Center 페이지를 로드합니다. Knowledge Center 내에서 **갤러리 찾아보기**를 선택하고 샘플 데이터베이스를 새 Azure Synapse 논리 데이터베이스에 로드합니다. 그림 7에서 보라색으로 강조 표시된 **갤러리 찾아보기**를 확인할 수 있습니다.

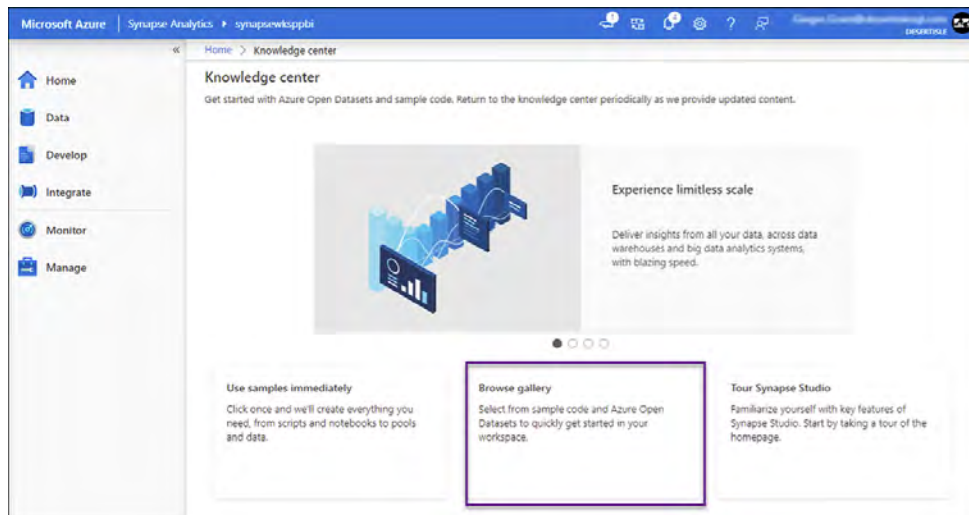


그림 7: Knowledge Center의 갤러리 찾아보기 섹션

Knowledge Center에서 **미국 주 고용 시간 및 수입**을 선택한 다음 파란색 **계속** 단추를 클릭합니다. 짧은 설명이 나타나고 파란색 **데이터 집합 추가** 단추가 창의 왼쪽 상단 모서리에 표시되면 클릭해야 합니다. 그러면 데이터 집합이 **Azure Blob Storage** 아래에 표시됩니다. **미국-고용-시간-수입** 폴더 옆에 있는 생략(세 점) 단추를 클릭하고 새 **SQL 스크립트** 다음에 있는 **외부 테이블 만들기**를 그림 8에 표시된 대로 선택합니다.

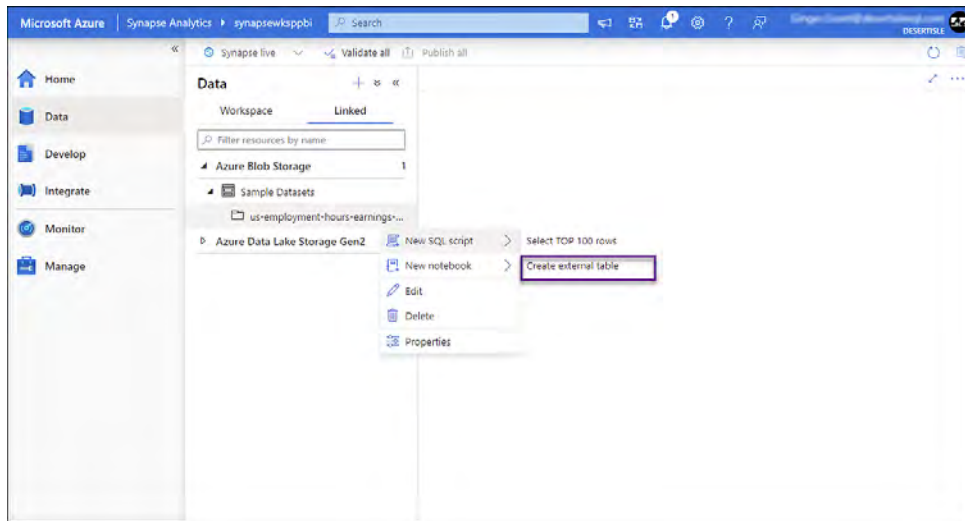


그림 8: 미국 고용 시간 수입 샘플 데이터 집합에서 외부 테이블 만들기

외부 테이블을 만드는 데 필요한 프롬프트를 제공하는 팝업 창이 나타납니다. 그림 9에 표시된 대로 **USEmp**라는 새 데이터베이스를 만들고 외부 테이블 이름으로 **HoursData**를 이용합니다.

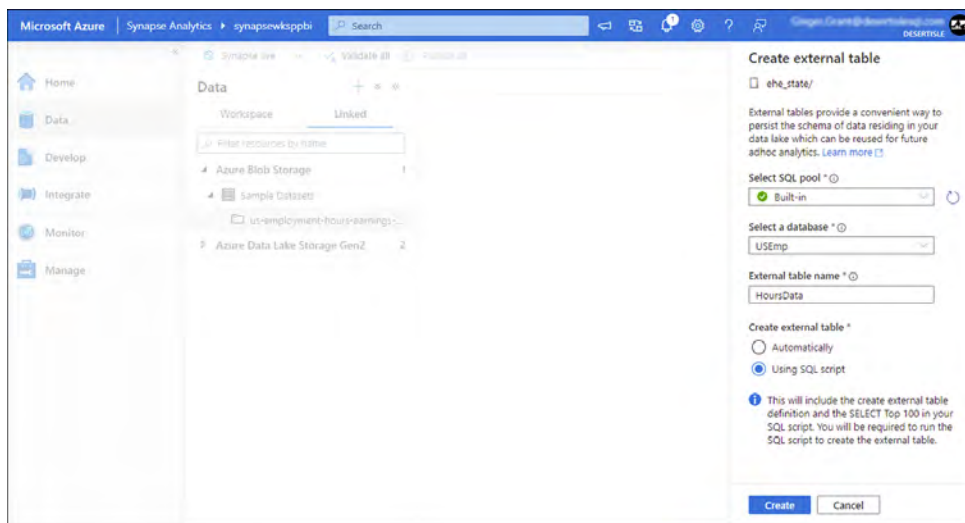


그림 9: 외부 테이블 프롬프트 만들기

생성 단추를 선택하기 전에 여기에 표시된 대로 라디오 단추인 **SQL 스크립트 활용**을 선택해야 합니다. **SQL 스크립트 1** 창에 생성된 코드를 확인해 보세요. 외부 테이블을 만드는 지침이 될 수 있습니다. 그림 10에서 보라색으로 강조 표시된 **워크스페이스** 탭을 선택하고, 생략 부호를 선택하여 데이터베이스를 새로 고치면 새 **USEmp** 데이터베이스가 나타납니다.

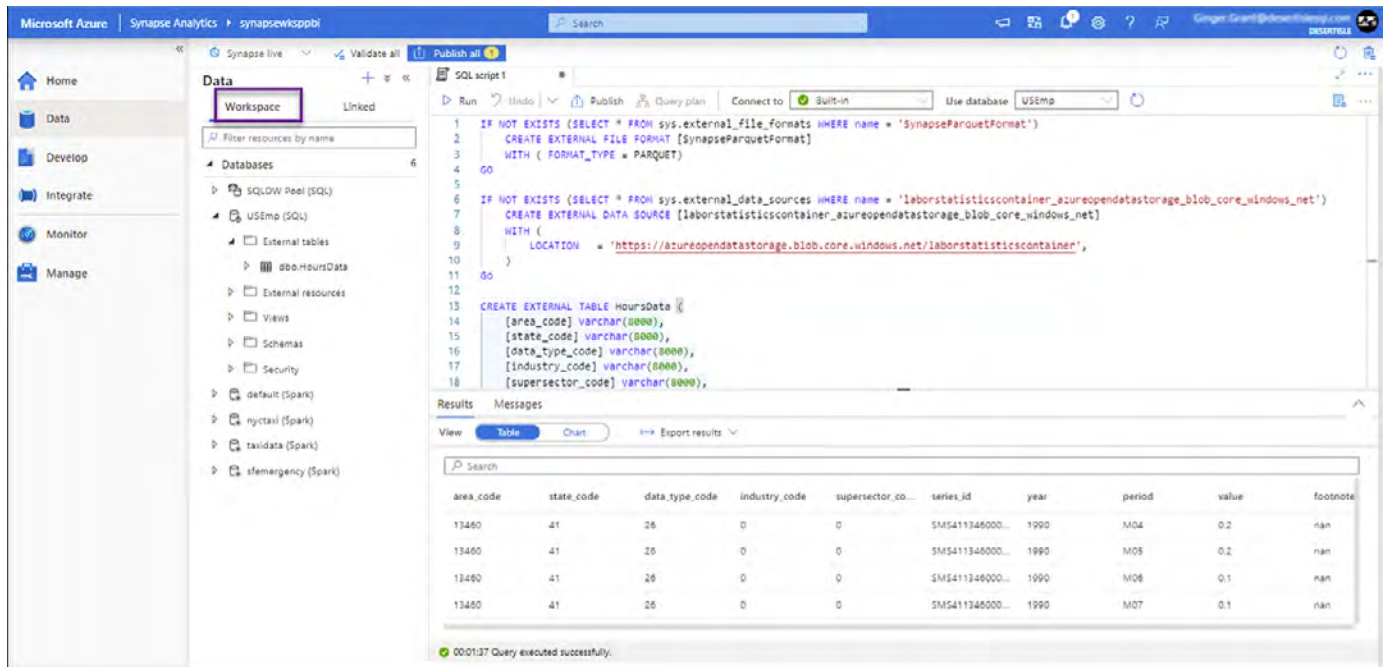


그림 10: HoursData 테이블이 표시된 새로 생성된 Azure Synapse 논리 데이터베이스

선택한 코드는 Power BI에서 활용할 수 있는 새 데이터베이스와 테이블을 생성합니다. 외부 테이블에는 Power BI에 연결하고 시각화에서 활용할 수 있는 데이터 집합이 포함되어 있습니다.

다음으로는 Azure Synapse가 데이터 레이크를 관리하고 Power BI를 활용하여 보다 의미 있는 보고서를 작성하는 방법을 보여 드리겠습니다.

#2: 데이터 레이크 관리 및 BI 보고서용 데이터 마트 구축

최근 많은 기업들은 기업 환경에 있는 모든 데이터에 대한 장기적인 스토리지로 데이터 레이크를 활용하고 있습니다. Azure Synapse는 데이터 레이크의 데이터를 다양한 이용 사례에 대해 서로 다른 기능 영역에 편성할 수 있는 통합 파이프라인을 만드는 기능을 제공하기 때문에 훌륭한 솔루션이 될 수 있습니다.

Acme는 주문이 접수된 후 포장과 배송이 필요한 많은 품목을 창고에 저장하고 있는 인터넷 판매 기업입니다. 지난 2년 동안 일일 보고서 작성만을 원했을 수도 있지만, Acme는 데이터를 임시 분석이나 머신 러닝에 활용할 수 있도록 나머지 데이터를 안전하고 체계적인 방식으로 저장해야 합니다. 데이터는 IoT 디바이스에서 수신될 수 있고 향후 머신 러닝 분석을 위해 요약된 방식으로 보고되고 원시 형식으로 저장되어야 할 수 있습니다.

Azure Synapse는 데이터 레이크에서 데이터를 구성하는 데 필요한 모든 도구를 포함하고 있기 때문에 다양한 용도로 데이터를 활용할 수 있습니다. 데이터는 서로 다른 소스에서 데이터 레이크로 전달될 수 있습니다. 데이터는 통합 파이프라인을 통해 데이터 레이크의 브론즈 영역에 저장된 원시 데이터에서 변환되고, 데이터 레이크의 실버 영역에 있는 일부 메타데이터에 추가되고, 데이터 웨어하우스 별 모양 스키마로 모델링된 파일 집합으로 큐레이팅될 수 있습니다. Azure Synapse는 데이터 레이크를 검토하고 구성하는 구조를 제공하여 조직에서 원시 데이터를 더 잘 활용할 수 있도록 원시 데이터를 구성하는 데 도움이 될 수 있습니다. Azure Synapse가 데이터를 큐레이팅하면 Power BI에서 활용할 수 있습니다.

데이터 레이크에 저장된 데이터는 열 형식 데이터 압축을 제공하며 기본 서버리스 엔드포인트를 활용하여 더 빠르게 쿼리하고 Azure Synapse에서 더 빠르게 읽을 수 있는 Parquet 형식으로 저장되기도 합니다. 이는 Power BI의 데이터에 액세스하는 데 방해가 되지 않습니다. 반대로 해당 데이터만 모델에 로드하여 24개월 분량의 데이터를 로드할 수 있는 방식을 제공합니다. Azure Synapse로 큐레이팅된 데이터 레이크에 연결할 수 있는 방법을 살펴보겠습니다.

Azure Synapse에서 생성된 데이터 레이크에 액세스하는 것은 Power BI 관리자 포털에서 먼저 환경 설정됩니다. 이 문서가 작성된 시점에 이 설정은 Power BI 관리자 포털의 미리 보기 기능 설정이었습니다. 사용자가 Power BI 관리자 권한으로 데이터에 액세스하려면 그림 11에 표시된 대로 관리자 포털을 이용하고 **Azure 연결** 옵션을 선택해야 합니다. Azure Data Lake Storage Gen2 계정에 연결할 때 Power BI의 데이터를 활용하는 가장 경제적인 방식은 데이터 레이크와 Power BI 테넌트가 동일한 데이터센터에 있는 것입니다. 또한 그림 11에 표시된 대로 관리자 포털의 **Azure 연결** 섹션에 정보를 작성하려면 Azure **구독**의 이름, **리소스 그룹** 및 Azure Data Lake **Storage 계정**의 이름이 필요합니다.

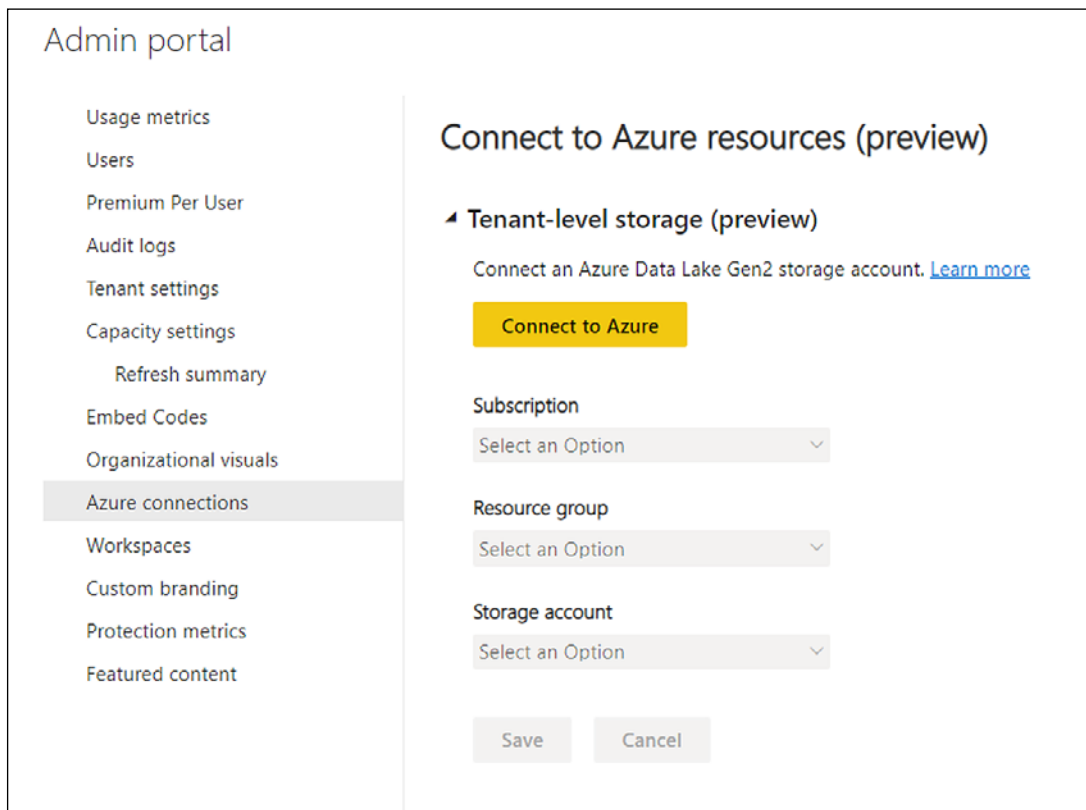


그림 11: Azure 연결 및 Azure Data Lake Gen2용 Power BI 관리자 포털

연결이 완료되면 Power BI의 데이터에 액세스할 수 있습니다. 여기서는 리소스 그룹 이름으로 **SynapseRGPBI**를, Azure Data Lake Storage Gen2 계정 이름은 **synapsedlpbi**를 이용합니다. Azure 연결이 생성되고 저장되면 Power BI 서비스에서 데이터를 새로 고침할 수 있습니다.

참고: 이 문서가 작성된 시점에 하나의 Azure Data Lake Gen2 Storage(ADLS) 계정에만 연결할 수 있었고 다른 종류의 스토리지 계정에는 연결할 수 없었습니다.

Power BI 데스크톱 내에서 **데이터 가져오기** 메뉴에 있는 Azure 옵션에서 해당 계정을 선택하여 Azure Data Lake Gen2 계정에 액세스할 수 있습니다. 다음 패턴을 활용하여 ADLS 데이터 위치의 URL로 데이터에 액세스해야 합니다.

```
https://<ADLS_accountname>.dfs.core.windows.net/<filesystemname>/<subfolder>
```

이를 통해 폴더에서 모든 데이터가 추출됩니다. 예를 들어 **PBISynapseCentralUS**라는 ADLS 계정이 있습니다. 이 계정에는 **powerbi**라는 폴더와 **Bronze**라는 하위 폴더가 있습니다. 이 폴더의 모든 파일에 액세스하려는 경우 Power BI에서 이용할 URL은 다음과 같습니다.

```
https://synapsedlcentralus.dfs.core.windows.net/powerbi/Bronze/
Chicagofoodinspections/2020FoodInspectionparquet
```

URL은 그림 12에 표시된 대로 삽입됩니다.



그림 12: URL을 이용하여 Azure Data Lake Storage Gen2에서 데이터 가져오기

위치를 지정할 때 대소문자가 정확히 일치하는지 확인해야 합니다. 일치하지 않으면 "찾을 수 없음" 오류가 발생합니다. 파일 이름을 지정하지 않았기 때문에 해당 디렉터리의 모든 파일이 반환되었습니다. 하위 디렉터리에는 세 개의 Parquet 형식 파일이 포함되어 있습니다. Parquet 형식 파일은 서버리스 풀에 테이블을 생성할 때 더 빠르게 작동하고 스토리지 크기를 줄이기 때문에 일반적으로 Azure Synapse와 함께 활용됩니다. Parquet 형식 파일은 이진 파일이고 사람이 읽을 수 없지만 Power BI는 읽을 수 있습니다. 나열된 하위 디렉터리에는 세 개의 Parquet 형식 파일이

있습니다. 이 파일들은 하나의 데이터 요소인 '2020년 시카고 식품 검사 데이터'로 읽혀져야 합니다. 이를 그림 13에서 볼 수 있듯이 Power BI에서 살펴보겠습니다.

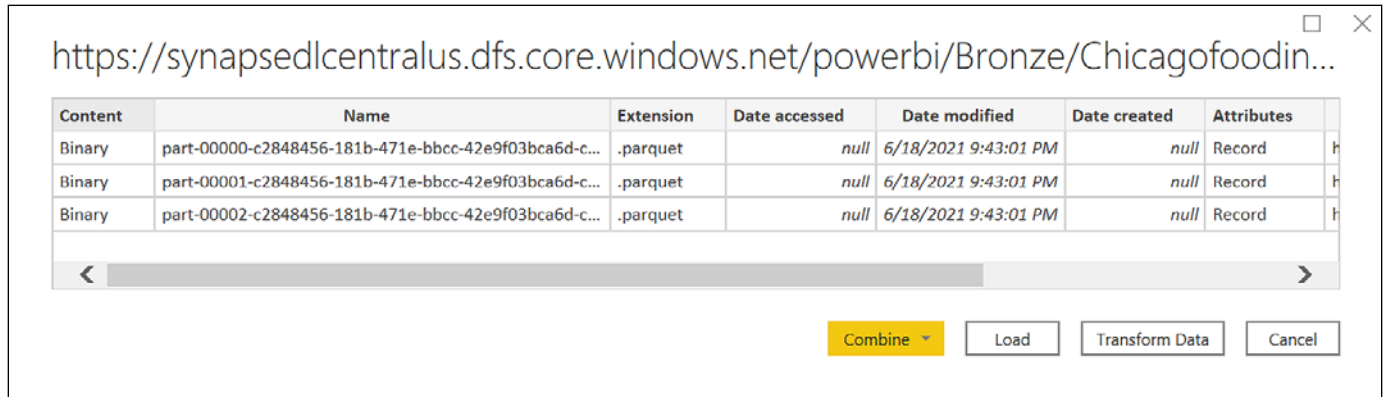


그림 13: ADLS에서 Power BI로 로드될 Parquet 형식 파일

노란색 **결합** 단추로 파일을 결합하고 콘텐츠를 데이터 모델에 통합할 수 있습니다. 데이터를 수집하고, 이 데이터를 Power BI에 로드하기 전에 작업을 수행해야 하는지 결정하기 위해 **결합**과 **데이터 변환**을 선택합니다. 이 단계는 모든 데이터를 로드하고 그림 14에서 볼 수 있듯이 Parquet 형식을 사람이 읽을 수 있는 형식으로 변환합니다.

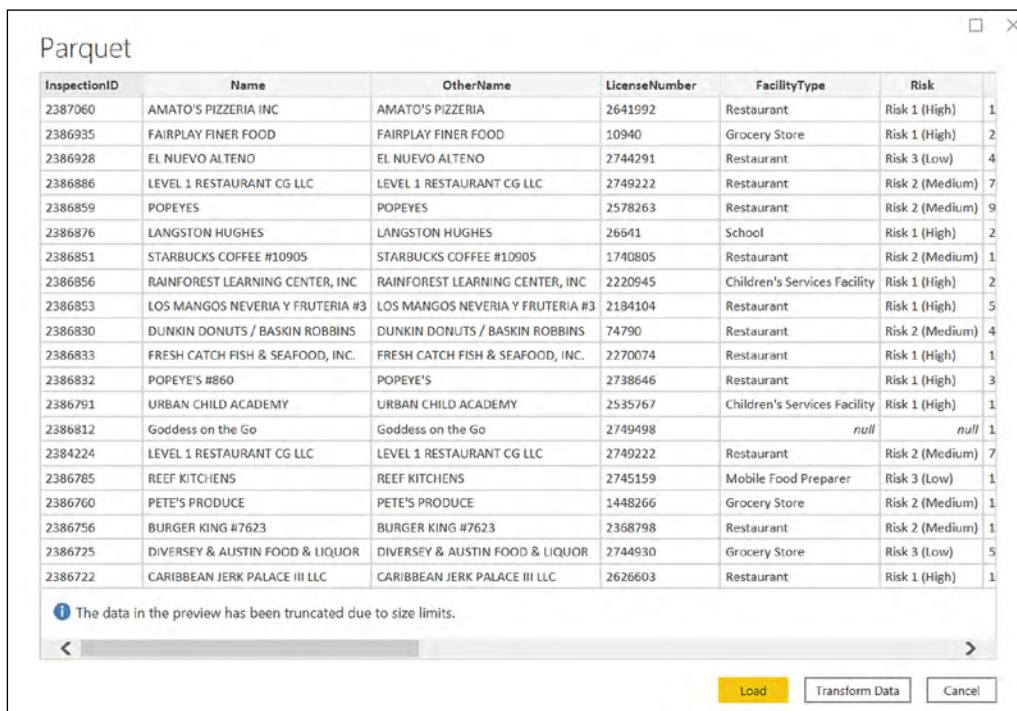


그림 14: ADLS에서 Power BI로 변환된 Parquet 형식 파일의 예

매우 읽기 쉬운 데이터인 것을 알 수 있으므로 변환해 보겠습니다. 그림 15에서 화면 왼쪽의 쿼리를 살펴보면 Power BI가 여러 가지 데이터 변환 단계를 자동으로 수행한 것을 확인할 수 있습니다.

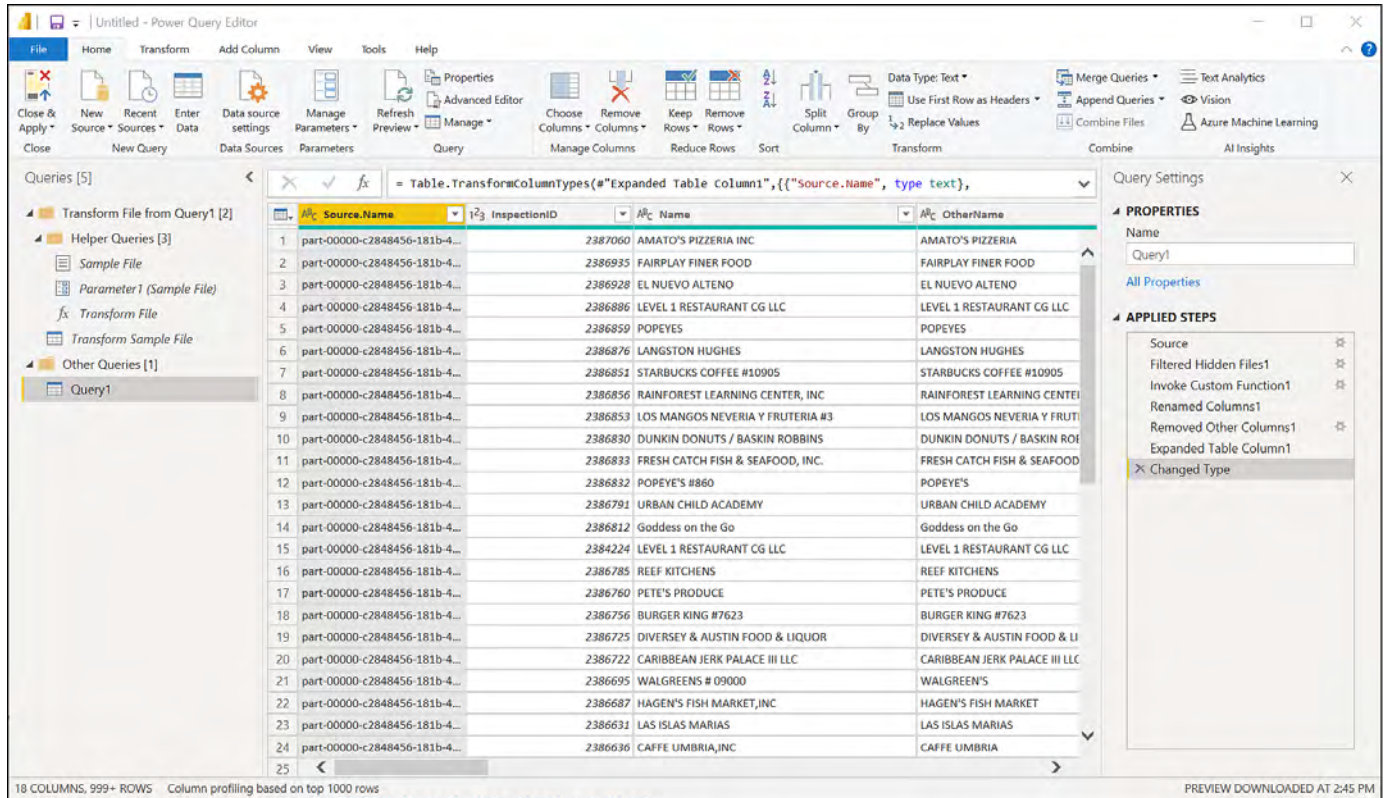


그림 15: Power Query에 생성된 변환 단계

도우미 쿼리는 ADLS 계정에서 데이터를 가져와 사람이 읽을 수 있는 형식으로 변환합니다. **Query1**은 변경 사항을 선택합니다. 여기서는 데이터 모델에서 이용되지 않는 **Source.Name** 및 **InspectionID** 열을 제거하고, **Query1**의 이름을 **검사**로 변경합니다. Power BI를 통해 Azure Synapse에 가장 적합한 형식으로 저장된 데이터를 이용하고, Power BI를 활용하여 그림 16에 표시된 대로 데이터를 보고할 수 있습니다.

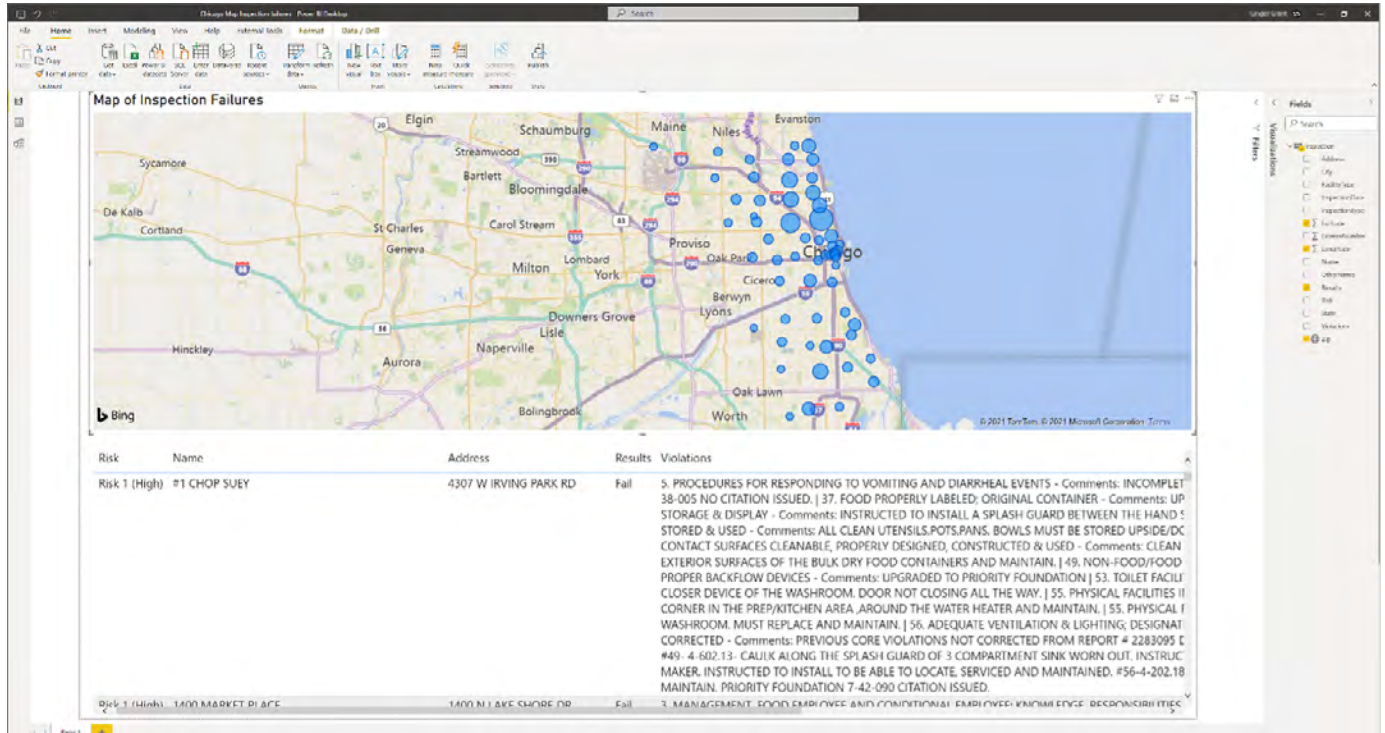


그림 16: Parquet 형식 데이터로 만든 Power BI 시각화

#3: 완전 관리형 서버리스 엔드포인트를 활용한 데이터 레이크 탐색

많은 기업들이 데이터 레이크에 데이터의 상당 부분을 저장하기로 결정했습니다. ADLS Gen2 계정은 특히 데이터 압축을 제공하는 Parquet 형식으로 데이터가 저장될 때 데이터 저장에 대한 스토리지 요구 사항이 줄어들어 매우 저렴한 비용으로 수년 간의 데이터를 저장할 수 있습니다. Azure Synapse 워크스페이스를 만들 때 Azure Data Lake Gen2(ADLS) 계정이 프로비저닝되어 데이터 레이크를 만드는 데 필요한 스토리지를 제공합니다. Azure Synapse Analytics는 데이터 집합을 관리하고, 포함된 서버리스 엔드포인트를 활용하여 데이터베이스에 있는 것처럼 데이터를 쿼리할 수 있는 ADLS 계정 위에 논리 데이터베이스를 제공할 수 있습니다. 쿼리당 지불 모델을 이용하기 때문에 다른 것을 프로비저닝하거나 중지하고 시작할 필요가 없습니다.

사용자가 T-SQL을 활용하여 데이터 레이크에 포함된 데이터를 탐색할 수 있도록 ADLS 스토리지를 활용하여 논리 데이터베이스를 만들 수 있습니다. 가상 데이터베이스에는 이 기능이 지원되는 동안 데이터베이스에 행 수준 보안을 만드는 기능도 포함됩니다. 물론 모범 사례를 따르고 정렬된 보기를 만들어 Power BI 데이터 모델에 투입한 데이터에 액세스하여 데이터 모델을 만들 수 있습니다. 보기를 활용하여 열이 사람이 읽을 수 있는 형식임을 확인하고 열을 최소값으로 정렬하여 VertiPaq 엔진 압축과 Power BI 보고서 성능을 향상시킬 수 있습니다.

Azure Synapse에서 서버리스 엔드포인트를 활용하면 수행하는 쿼리에 대해서만 요금이 부과됩니다. T-SQL 문을 활용하여 Azure Synapse에 논리 데이터베이스를 만들어 데이터에 액세스할 수 있습니다. 여러 다른 디렉터리 및 데이터 레이크의 데이터 위에 있는 테이블을 정의하고 하나의 데이터베이스로 테이블을 결합할 수 있습니다. 이 작업은 T-SQL 명령 **CETAS(Create External Table AS SELECT)**로 수행됩니다. 이 명령은 Power BI 또는 **SSMS**에서 액세스할 수 있는 외부 테이블을 만드는 데 이용됩니다.

다음은 Azure Synapse에서 논리 데이터베이스를 만들고 Parquet 형식 파일의 일부 텍스트를 처리하기 위해 활용해야 하는 UTF8 정렬을 이용하는 데 쓰인 코드의 예입니다.

```
CREATE DATABASE EMR

GO

--Change the internal database to EMR to issue comments

USE EMR

GO

ALTER DATABASE EMR COLLATE Latin1_General_100_BIN2_UTF8;
```

다음 단계는 읽어올 데이터의 위치를 정의하고 데이터를 검색하기 위한 위치를 설정하는 것입니다. 스크립트를 여러 번 실행할 수 있도록 **Not Exists** 문을 추가했습니다. 데이터가 다양한 위치에 저장되는 경우 여러 외부 데이터 소스를 만들어 참조할 수 있습니다.

```
IF NOT EXISTS (SELECT * FROM sys.external_file_
formats WHERE name = 'SynapseParquetFormat')

CREATE EXTERNAL FILE FORMAT [SynapseParquetFormat]

WITH ( FORMAT_TYPE = PARQUET )

GO

IF NOT EXISTS (SELECT * FROM sys.external_data_sources WHERE name = 'EMRParquetData')

CREATE EXTERNAL DATA SOURCE [EMRParquetdata]

WITH (

LOCATION 'https://synapsedatalake.dfs.core.windows.net/root/Curated/EMR',
CREDENTIAL = [ManagedIdentityCredential]

)

GO
```

형식과 위치를 지정한 후 데이터 위에 테이블을 만들 수 있습니다. 여기서는 두 개의 테이블을 만들고 쿼리 서식을 지정합니다.

```
IF EXISTS (SELECT * FROM sys.external_tables WHERE name = 'MailingAddress')
DROP EXTERNAL TABLE MailingAddress

GO

CREATE EXTERNAL TABLE MailingAddress (
    [MailingAddressID] varchar(100) ,
    [StreetAddress] varchar(300),
    [City] varchar(100),
    [StateAbbr] char(2),
    [Zipcode] varchar(10),
    [ModifiedDate] varchar(13)
)

WITH (
    LOCATION = '/2021MailingAddress/Address.snappy.parquet',
    DATA_SOURCE = [RawParquetData],
    FILE_FORMAT = [SynapseParquetFormat]
)

GO
```

```
IF EXISTS (SELECT * FROM sys.external_tables WHERE name = 'Patient')

DROP EXTERNAL TABLE Patient

GO

CREATE EXTERNAL TABLE Patient (

PatientID varchar(100) ,

SSN varchar(12) ,

FirstName varchar(60) ,

MiddleName varchar(60) ,

LastName varchar(60) ,

Gender char(1) ,

MailingAddressID varchar(100) ,

Email varchar(256) ,

DateOfBirth varchar(13) ,

ModifiedDate varchar(13)

)

WITH (

LOCATION = '/2021Patient/Patient.snappy.parquet',

DATA_SOURCE = [RawParquetData],

FILE_FORMAT = [SynapseParquetFormat]

)

GO
```


다음은 Power BI와 연결될 데이터로 만든 보기입니다.

```
CREATE VIEW vPatient
AS
SELECT TOP 999999999 [PatientID]
    ,FirstName AS [First Name]
    ,LastName AS [Last Name]
    ,[Gender]
    ,CAST([AddressID] AS INT) AS AddressID
    ,[Email]
    ,DATEDIFF(YYYY, CAST([DateOfBirth] AS DATE), GETDATE() ) as Age
FROM [EMR].[dbo].[Patient]
ORDER BY Gender, Age
GO
CREATE VIEW vAddress
AS
SELECT TOP 999999999 cast([MailingAddressID] as int) as AddressID
    ,[StreetAddress] as [Address]
    ,[City]
    ,[StateAbbr] as [State or Province]
    ,[Zipcode] as [Postal Code]
FROM [EMR].[dbo].[MailingAddress]
ORDER BY [StateAbbr] , [City]
```

이 작업이 완료되면 Power BI의 EMR 데이터베이스 및 SQL Server Management Studio와 같은 기타 데이터베이스 도구에 연결할 수 있습니다. 이 데이터에 연결하면 SQL Server에 연결된 것과 같습니다. Power BI에 연결하는 데 이용할 서버리스 엔드포인트의 이름 주위를 보라색 상자로 표시한 그림 17에서 볼 수 있듯이 서버리스 풀에 이용되는 데이터베이스의 이름은 Azure 포털의 Azure Synapse 페이지에서 찾을 수 있습니다.

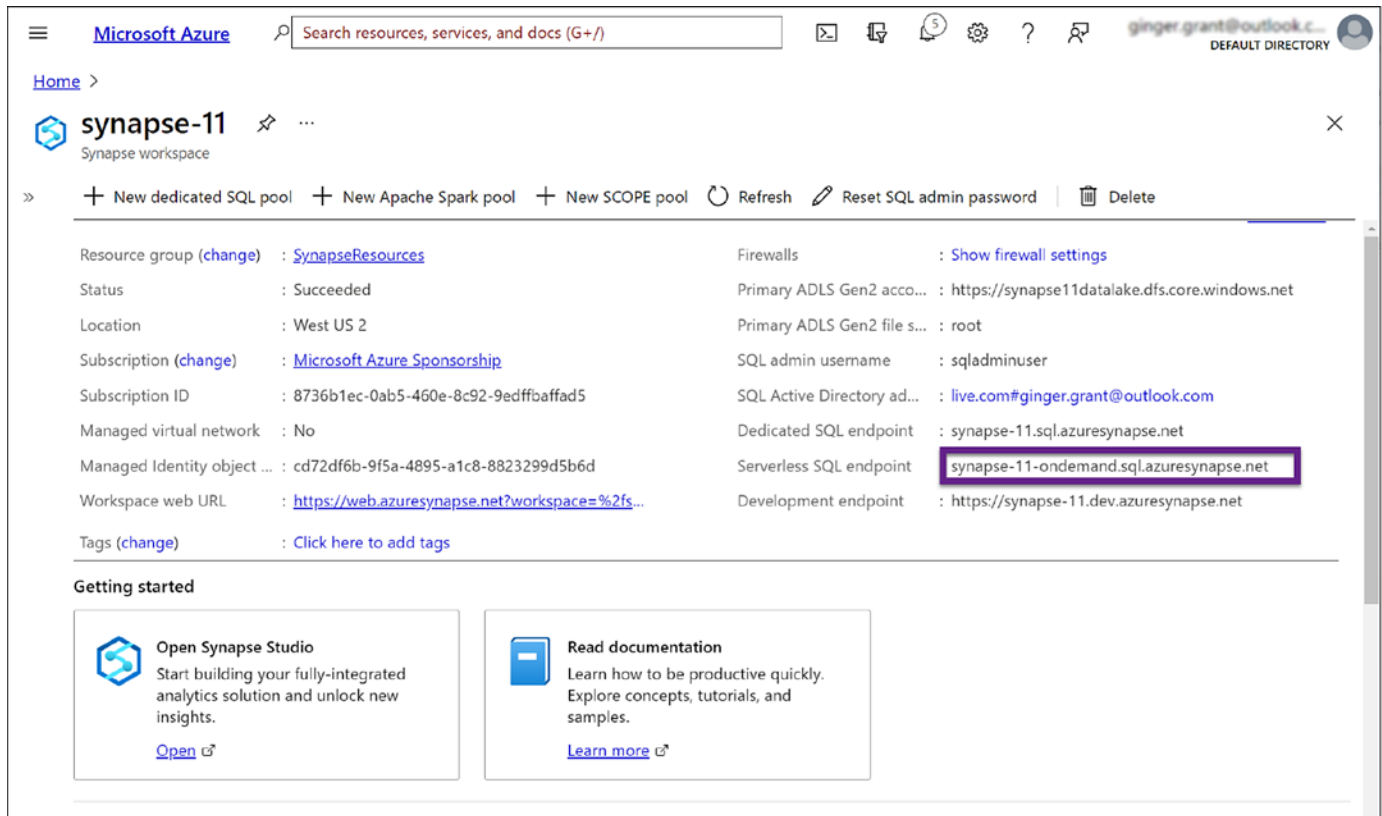


그림 17: 서버리스 엔드포인트를 보여주는 Azure 포털의 Azure Synapse

다음 단계는 물론 그림 18에 표시된 대로 SQL Server 연결에서 Power BI로 연결하여 Power BI를 데이터 소스에 연결하는 것입니다.

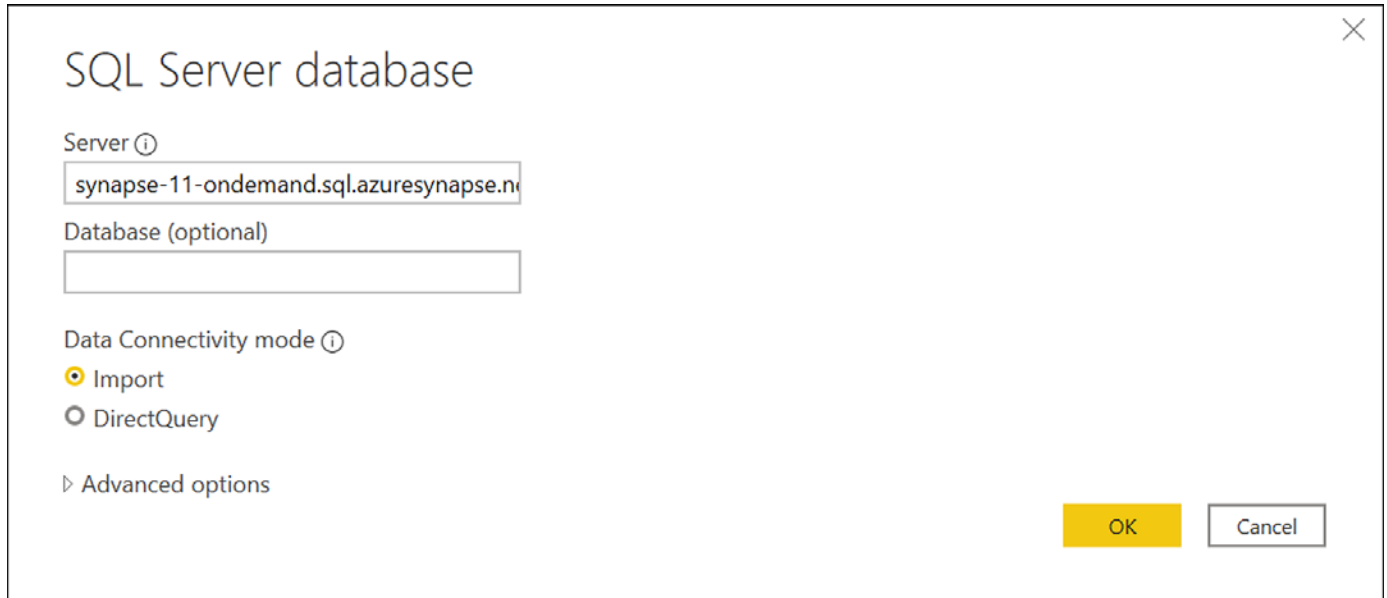


그림 18: SQL Server 데이터베이스 연결을 활용하여 Azure Synapse 서버리스 풀에 액세스

보시다시피 데이터를 가져오기로 선택하고 Azure Synapse에서 서버리스 엔드포인트에 연결합니다. 이전 스크립트에서 만든 보기를 이용하려고 했지만, 모든 데이터베이스에서 데이터를 선택할 수 있는 것을 확인할 수 있습니다. 데이터베이스를 인증할 때 Azure AD 연결을 이용하고 싶을 것입니다.

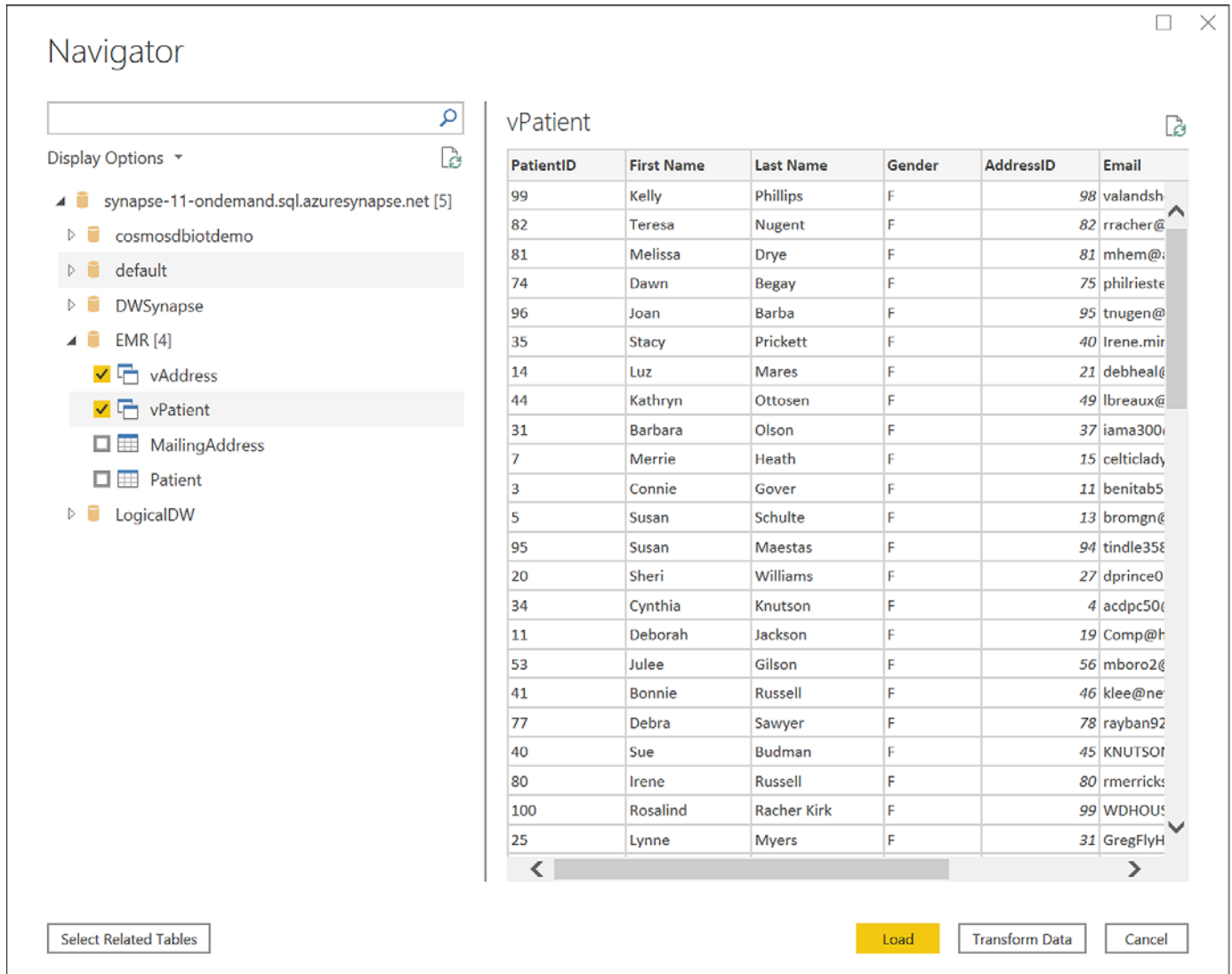


그림 19: Power BI에 있는 Azure Synapse의 논리 데이터베이스에서 데이터 선택하기

이제 Azure Synapse Analytics의 데이터를 통해 그림 19에 표시된 대로 서버리스 데이터베이스를 이용하여 Power BI 보고서를 만들 수 있습니다.

논리 데이터베이스를 활용하면 Azure Synapse에서 이 모든 기능이 제공되기 때문에 추가 서비스를 만들 필요 없이 Parquet과 같은 다양한 형식으로 저장된 데이터에 액세스하고 데이터 레이크에 저장된 데이터를 탐색하여 의미 있는 결과를 제공할 수 있습니다.

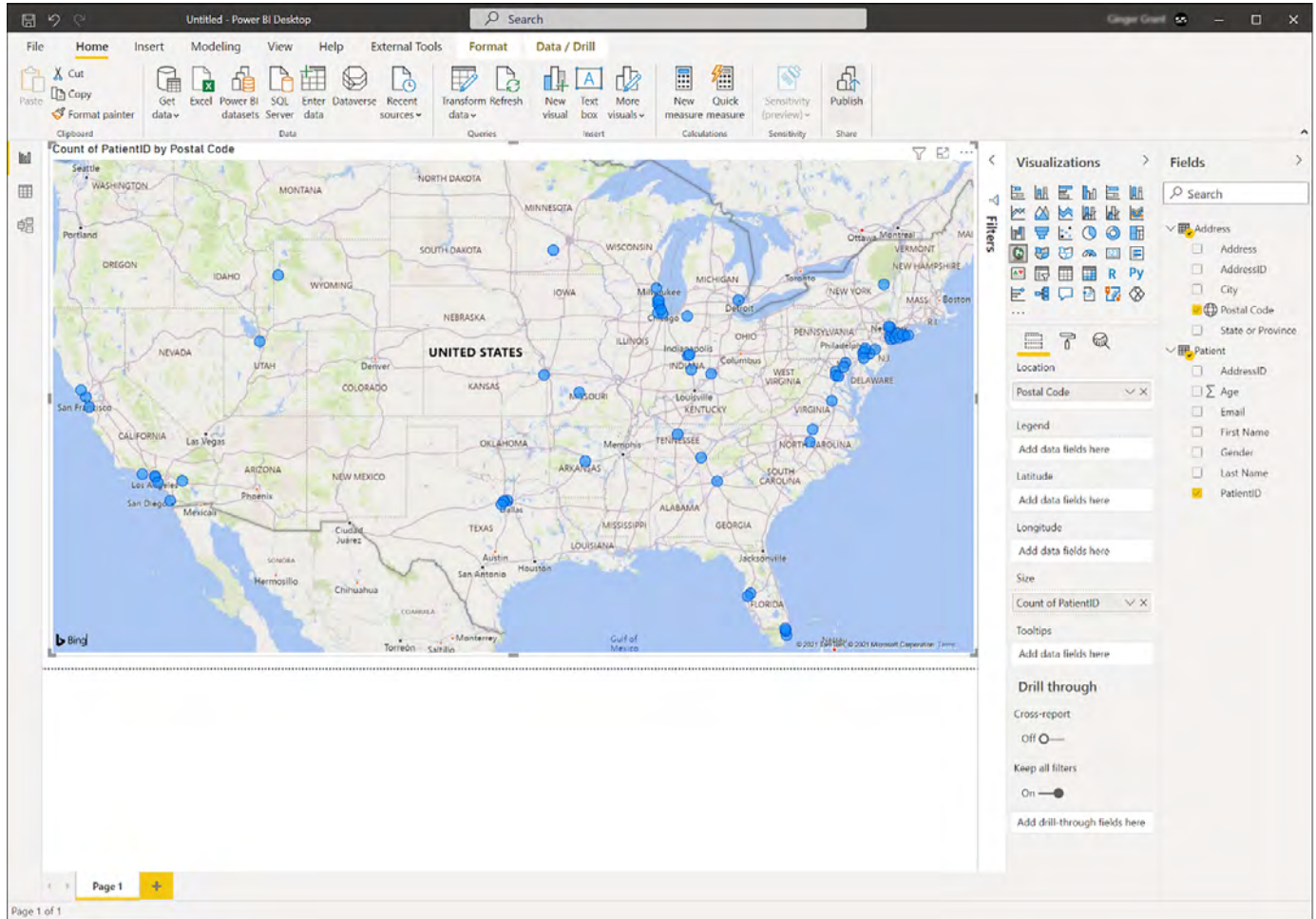


그림 20: Azure Synapse 서버리스 풀 데이터가 있는 Power BI 보고서

#4: 코드 없는 데이터 파이프라인을 구축하여 더 많은 데이터 소스 통합 및 인사이트 강화

Azure Synapse는 다양한 데이터 소스에서 데이터를 수집하고 이를 Azure Data Lake Gen2 스토리지 계정에 중점적으로 저장할 수 있습니다. 이 계정은 저장된 데이터에 액세스해야 하는 다른 모든 애플리케이션에 대한 단일 리포지토리 역할을 수행할 수 있습니다. 데이터 과학자는 데이터 흐름을 활용하여 여러 소스에서 데이터를 수집하기 위해 Workday와 같은 웹 애플리케이션의 데이터를 기업 데이터와 결합하려고 할 수 있습니다. 이 데이터는 액세스하기 전에 데이터 웨어하우스로 변환될 수도 있습니다. 데이터 수집 프로세스의 일환으로 결과를 시각화하는 데 이용되는 Power BI를 통해 머신 러닝 모델을 활용하여 데이터를 분석할 수도 있습니다. 예를 들어 야간 프로세스의 일부로 머신 러닝으로 고객을 분석하여 고객 이탈 가능성을 판단할 수 있습니다. 분석된 데이터는 데이터 웨어하우스에 저장된 트랜잭션 기록과 결합된 고객 서비스 담당자와의 비정형 피드백에서 확보할 수 있습니다. 이 기록은 Azure Synapse의 전용 풀에 저장됩니다. 이 분석은 데이터를 Power BI에 공급하는 일괄 처리 프로세스의 일부로 수행될 수 있으며, Azure Synapse의 통합 파이프라인을 활용하여 고객 이탈 분석이 Power BI에 포함될 수 있습니다.

Azure Synapse 통합은 Azure Data Factory와 동일한 방식으로 데이터 파이프라인을 만드는 기능을 제공하며, Power BI 모델에 포함될 수 있는 서버리스 풀에 데이터를 쓸 수 있도록 PySpark 노트북을 포함하는 기능을 제공합니다. Azure Synapse에서 이 작업을 수행하는 방법을 살펴보겠습니다.

먼저 원시 고객 서비스 데이터를 처리하는 새 파이프라인을 만듭니다. 그림 27은 추출된 고객 서비스 데이터를 분석하고 점수를 매기기 위해 만든 파이프라인의 예이며, Azure Synapse 파이프라인의 일부로 출력할 예정입니다.

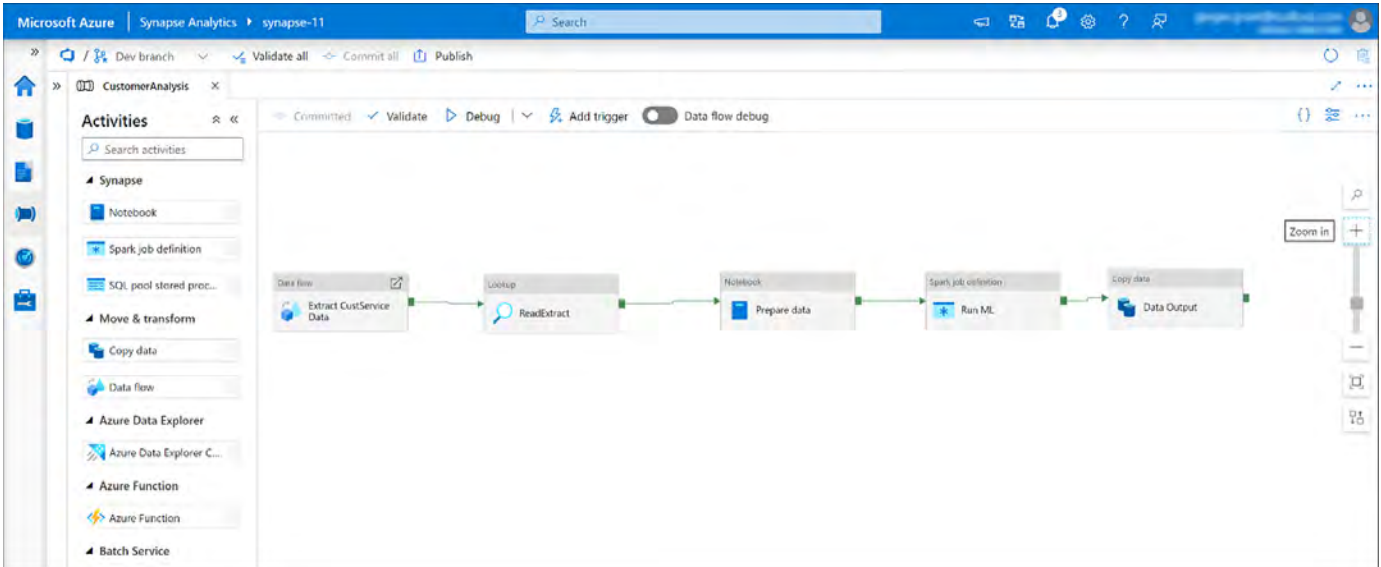


그림 21: 고객 서비스 데이터를 추출하고 Spark로 분석하는 Azure 파이프라인

머신 러닝 분석의 출력은 원시 고객 서비스 데이터에서 트랜잭션을 만들고, PySpark를 활용하여 데이터를 처리한 다음, 고객 트랜잭션에 점수를 매기고 Azure Synapse의 서버리스 풀에 데이터를 작성하기 위해 처리된 데이터를 이용하여 머신 러닝 개체에게 공급하는 것입니다. 데이터가 LogicDW에 작성되었으므로 Power BI의 새로운 데이터 요소를 살펴보겠습니다.

그림 21에 표시된 대로 이전 예제와 동일한 방식으로 SQL Server 연결을 활용하여 서버리스 풀에 연결합니다. 그런 다음 고객 서비스 감정 및 기타 고객 데이터를 분석하여 수집한 정보를 이용하여 평균 이탈 확률과 함께 Azure Synapse의 데이터 분석을 기반으로 한 고객 만족도 점수를 표시하고 그림 22에 표시된 대로 Power BI 보고서의 일부로 해당 정보를 제공할 수 있습니다.

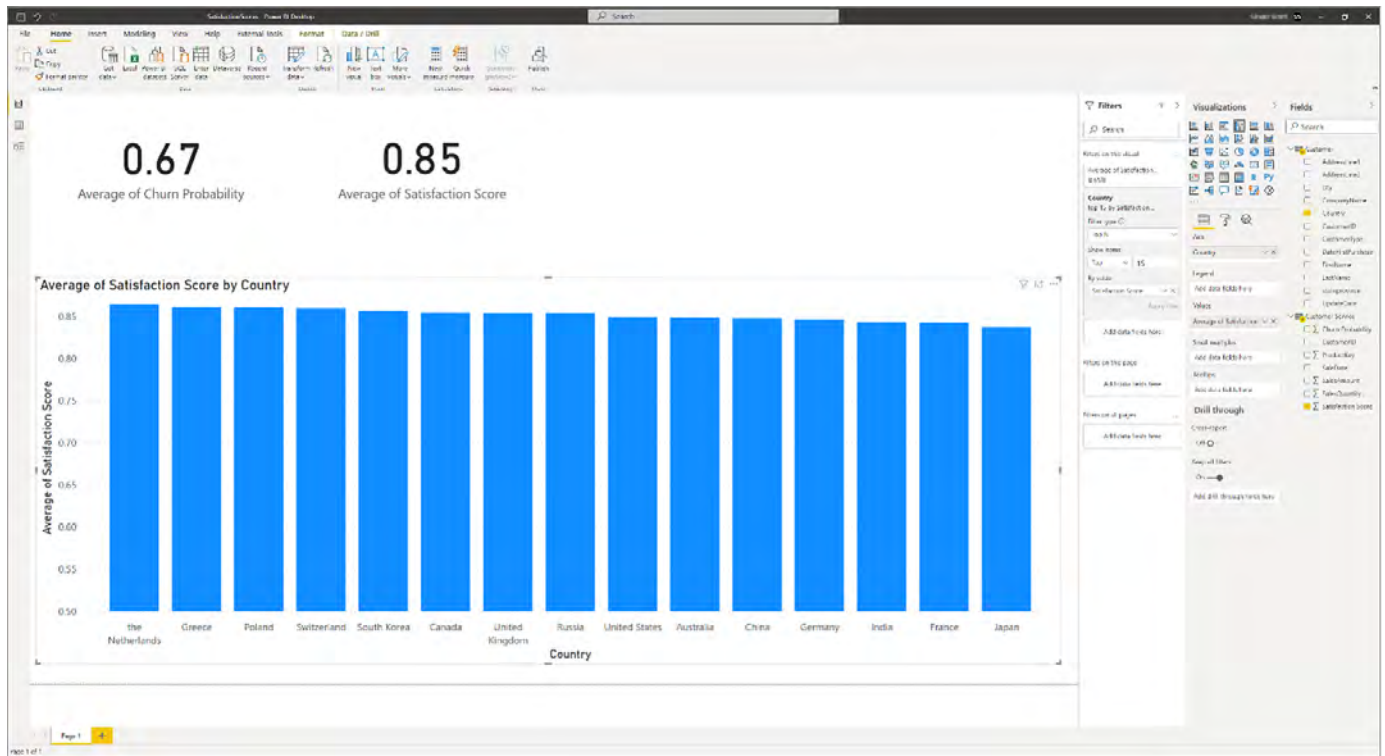


그림 22: 고객 서비스 데이터 집합 및 고객 만족도 점수에 대한 이탈 분석을 수행하기 위한 Azure Synapse의 처리 데이터 출력

#5: 안전한 데이터 웨어하우스를 만들고 Power BI 대시보드에 연결

Power BI는 Azure Synapse의 전용 풀에서 찾을 수 있는 데이터 집합을 비롯한 매우 큰 데이터 집합을 포함하는 데이터 소스에 연결할 수 있습니다. 많은 양의 데이터를 보고하는 데 어려움이 있을 수 있고 Power BI 실무자들은 이 어려움을 이미 잘 알고 있습니다. Power BI에는 집계 테이블 및 쿼리 폴딩 이름을 두 개로 지정하는 기능을 포함하여 데이터를 성능 조정하는 여러 방식이 있습니다. 매우 큰 데이터 집합은 일반적으로 데이터 모델을 가져오기에는 너무 크기 때문에 데이터에 대한 Direct Query 또는 Dual 연결을 이용하는 경우가 많습니다. Azure Synapse에는 워크로드 관리 분류를 통해 보고서 성능을 개선하는 데 도움이 되는 방식도 포함되어 있어 Power BI 보고 작업에 대한 리소스의 우선 순위를 지정할 수 있는 기능을 제공합니다. 전용 풀은 Azure 리소스를 저장하기 위해 일시 중지되는 경우가 많습니다. Azure Synapse의 **관리** 섹션으로 이동하여 SQL 풀을 확인함으로써 전용 풀이 실행되고 있는지 확인해야 합니다. 그림 23에서 볼 수 있듯이 두 개의 전용 풀인 **SmallPool**과 **Warehouse**가 있습니다. 상태에서 **SmallPool**이 실행 중임을 확인할 수 있습니다.

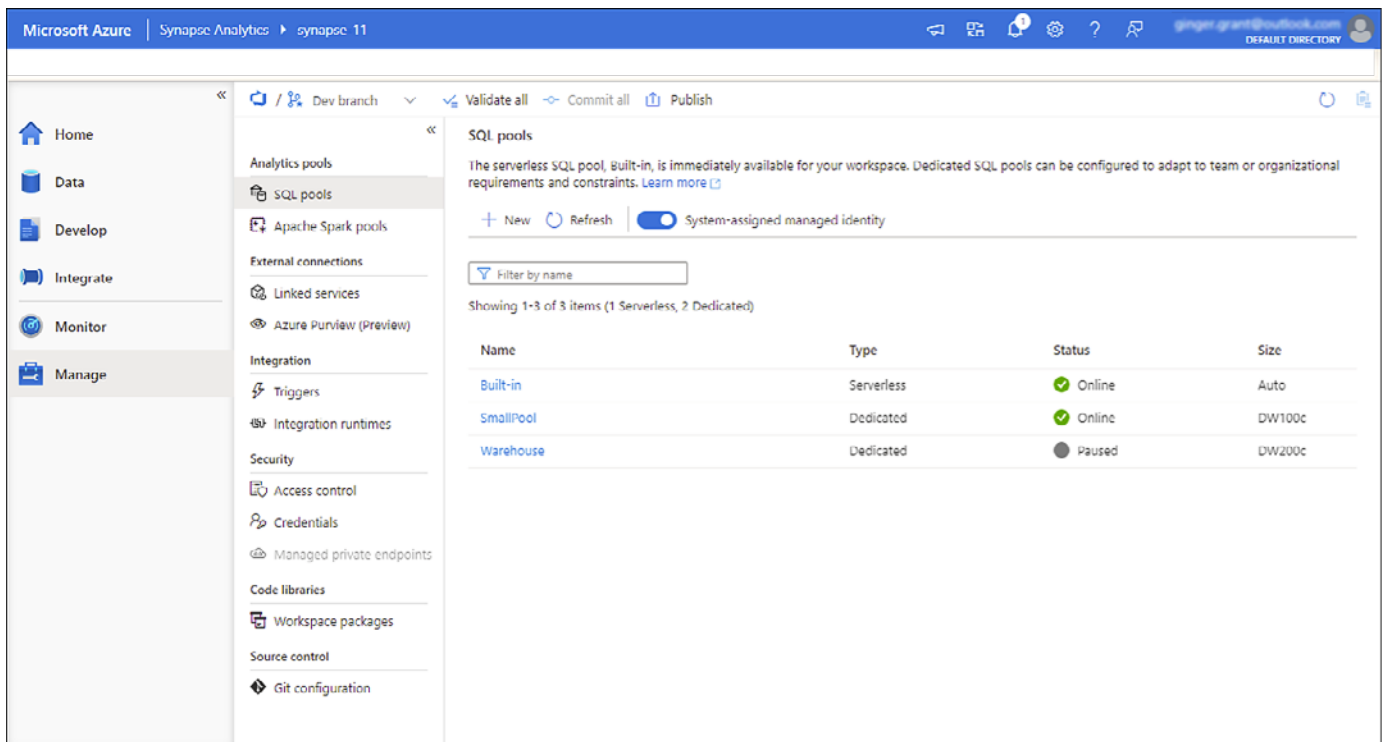


그림 23: SQL 풀 관리

실행 중이 아닌 경우 **다시 시작** 단추를 클릭합니다.

Power BI 보고를 위한 더 많은 리소스를 제공하는 첫 번째 단계는 Azure Synapse에 워크로드 그룹을 만드는 것입니다. 워크로드 그룹은 지정된 그룹과 연결된 워크로드 격리의 양을 지정합니다. 이 코드에 대한 새 SQL 스크립트를 만들고 코드를 실행할 때 **SmallPool**에 연결되어 있는지 확인해야 합니다. 그림 24는 새로운 SQL 스크립트가 생성된 **Develop** 섹션이 실행 중인 전용 풀인 **SmallPool**에 연결된 상태를 보여 줍니다.

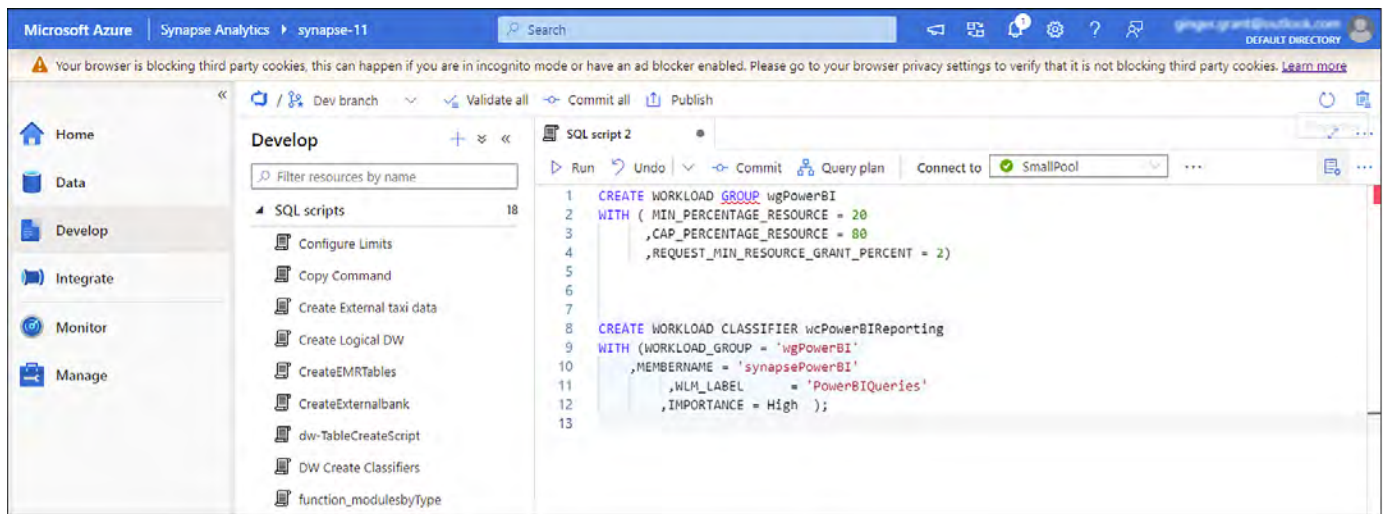


그림 24: 워크로드를 만들기 위한 SQL 스크립트 개발

다음 명령문에 볼 수 있듯이 Azure Synapse의 SQL 노트북에서 SQL 쿼리를 실행하려면 해당 쿼리가 실행될 때까지 모든 쿼리를 실행할 수 없기 때문에 전용 풀이 시작되었는지 확인해야 합니다.

```

CREATE WORKLOAD GROUP wgPowerBI

WITH ( MIN_PERCENTAGE_RESOURCE = 20

      ,CAP_PERCENTAGE_RESOURCE = 80

      ,REQUEST_MIN_RESOURCE_GRANT_PERCENT = 2)
    
```

워크로드 그룹인 **wgPowerBI**는 Azure Synapse Analytics 리소스의 최대 80%를 이 워크로드의 요소에 제공합니다. 각 구성원이 처음 연결할 때 최소 2%의 리소스가 부여됩니다. 리소스를 과도하게 할당하면 성능 문제가 발생할 수 있기 때문에 리소스 그룹을 만들 때 그룹 전체의 총 리소스가 최대 100%까지 추가되는지 확인해야 합니다. **REQUEST_MIN_RESOURCE_GRANT_PERCENT**는 한 번에 할당할 수 있는 리소스 청크의 수입입니다. 워크로드 격리를 활용하면 워크로드 그룹에 전용 리소스를 할당할 수 있습니다. 워크로드 그룹 ELT가 **MIN_PERCENTAGE_RESOURCE**를 활용하여 이용 가능한 리소스의 20% 이상을 할당받은 것을 볼 수 있습니다. 워크로드 격리가 없는 경우 요청은 공유 리소스 풀에서 작동합니다. 공유 풀의 리소스에 대한 액세스는 보장되지 않으며 중요성을 기반으로 할당됩니다.

워크로드 분류는 덜 중요한 요청보다 전반적인 우선 순위와 리소스를 부여하도록 Power BI 쿼리를 평가하는 기능을 제공합니다. 워크로드 분류자에 할당할 수 있는 5가지 가중치는 *사용자, 역할, 레이블, 컨텍스트, 시간*입니다. 서로 다른 이러한 요소들은 가중치 순으로 여기에 나열됩니다. 우선 순위가 가장 높은 요소는 *사용자*입니다. 분류자에 *사용자*를 포함시키면 해당 사용자가 실행하는 모든 것의 가장 우선 순위가 높습니다. 다음으로 우선 순위가 높은 값은 *역할*입니다. 역할이 지정된 사용자는 일반 사용자보다 우선 순위가 낮지만 더 관리하기 쉽습니다. 사용자가 여러 행을 가질 수 있기 때문에 워크로드 분류자에서 가장 높은 우선 순위가 이용됩니다.

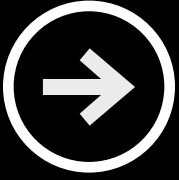
```
CREATE WORKLOAD CLASSIFIER wcPowerBIReporting
WITH (WORKLOAD_GROUP = 'wgPowerBI'
      ,MEMBERNAME = 'synapsePowerBI'
      ,WLM_LABEL = 'PowerBIQueries'
      ,IMPORTANCE = High );
```

이 분류자는 워크로드 그룹에 정의된 전용 리소스를 **synapsePowerBI**라는 Azure 디렉터리 그룹에 포함된 구성원과 함께 활용하여 Azure Synapse의 전용 풀에서 데이터에 연결하는 보고서의 성능을 향상시킵니다.



요약

이 eBook에서는 Azure Synapse의 다양한 기능을 활용하고 Power BI와 함께 이용하는 5가지 방법을 살펴보았습니다. Azure Synapse는 Power BI 내에서 이용할 수 있도록 데이터 레이크 위에 데이터 마트를 만들 수 있는 기능을 제공합니다. Power BI의 데이터 탐색 기능을 활용하여 Azure Synapse 내의 데이터를 분석함으로써 분석 기능을 개선할 수 있습니다. Power BI에서 이용할 수 있도록 Azure Synapse 내에서 가상 데이터베이스를 만들 수 있습니다. Azure Synapse에 생성된 머신 러닝 기능을 포함하고 결과를 보고하는 기능을 추가할 수도 있습니다. Power BI에 데이터를 수집하는 데 필요한 모든 리소스가 있는지 확인하기 위해 워크로드 관리를 추가하여 Power BI를 성능 조정하기 위한 Azure Synapse 내의 방식들을 살펴보았습니다. 이는 Azure Synapse의 기능을 활용하는 몇 가지 방법일 뿐이며 Power BI와 함께 활용하기 시작하면 이러한 모든 요소를 활용할 수 있습니다. 이 eBook에 설명된 요소들을 통합하면 Azure Synapse와 Power BI를 함께 활용하는 데 필요한 기능을 제공받을 수 있습니다.



다음 단계

BI 및 분석을 한 단계 끌어올리기 위한 3단계:



무료 [Azure Synapse 워크스페이스 만들기](#)



무료 [실습 교육 시리즈](#) 시청하기



[Synapse에서 무료 수량](#)을 이용하여 계속 실험하기