

Data Migration – Azure Synapse Service

에샨트 가르그

데이터 엔지니어, 아키텍트, Advisor
eshant.garg@gmail.com



Azure
Synapse
Analytics

소개

MPP 아키텍처의 데이터 로드 예제
다양한 로딩 방법 PolyBase가 선호되는 이유
SSIS와 PolyBase 로딩 방법 비교
PolyBase 프로세스 세부 정보
데모 – 다양한 기술을 사용한 로딩



Azure
Synapse
Analytics

Best Practices for Data Load

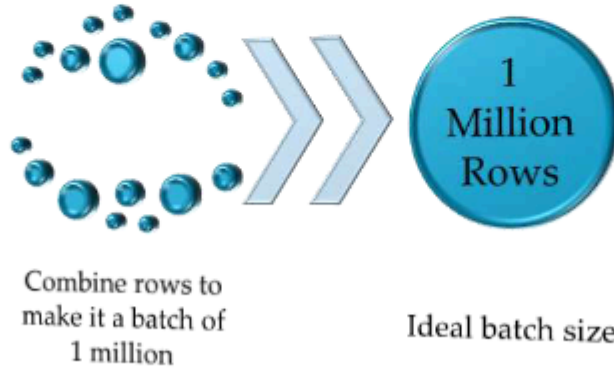
Data Warehouse 리더

DWU는 데이터를 병렬로 로드할 수 있는 속도에 직접적인 영향을 줍니다.

[illegible]

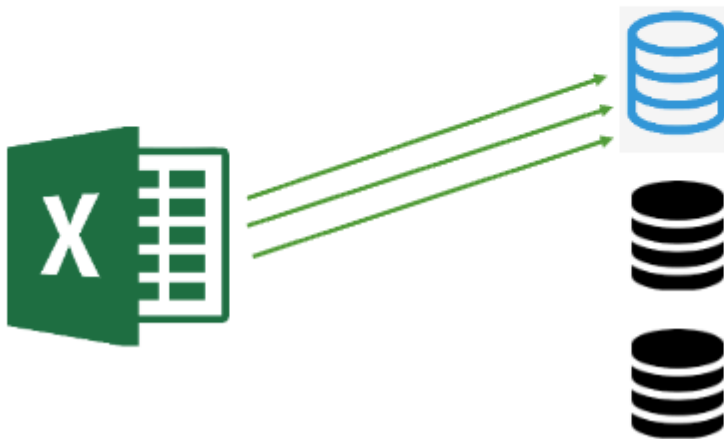
인서트 배치 크기 최적화

물방울 삽입 패턴을 피하십시오. 이상적인 배치 크기는 100만 개 이상, 직접 또는 파일



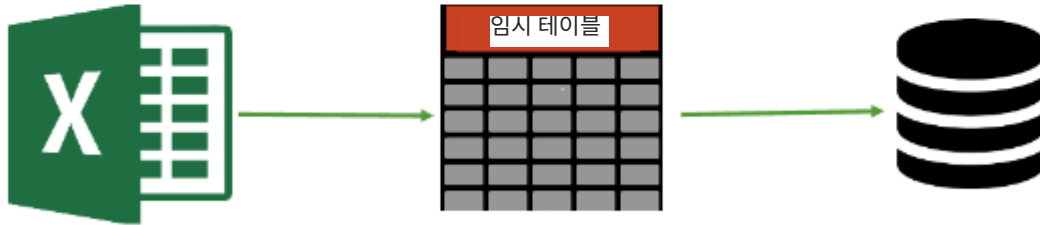
정렬된 데이터 피하기

분산 키별로 정렬된 데이터는 로드 작업을 느리게 하는 핫 스폿을 유발할 수 있습니다.



임시 테이블 사용

영구 스토리지로 이동하기 전에 Temp Heap 테이블에서 스테이징하고 변환합니다.



테이블 생성 형식

*FROM[dbo]를 선택하여 #tmp_fct
WITH(DISTRIBUTION = ROUND
ROBIN)테이블을 만듭니다. [팩
트인터넷세일즈];

- 완전 병렬 작동
- 최소한으로 기록됩니다.
- 변경 가능: 분포, 표
유형, 분할

로딩 방법

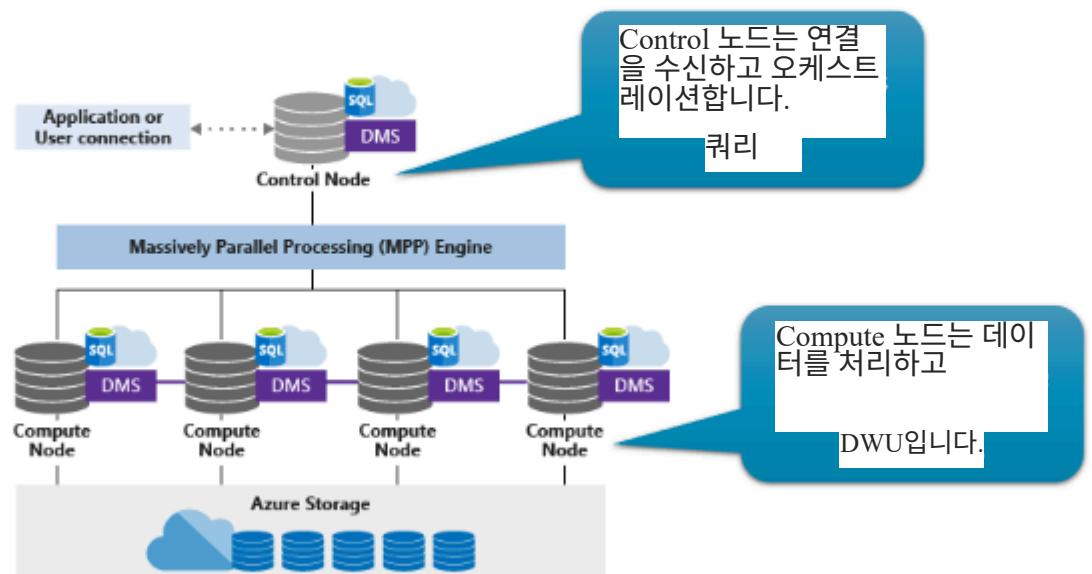
단일 클라이언트 로딩 방법

□ SSIS □ Azure Data Factory □ BCP □ 일부 병렬 기능을 추가할 수 있지만 제어 노드에서 병목 현상이 발생합니다.

병렬 판독기 로딩 방법

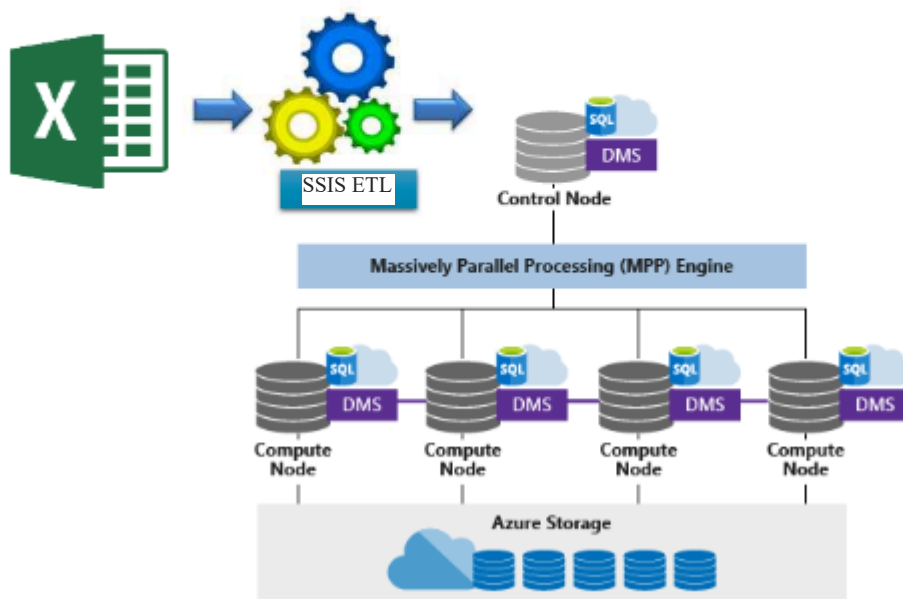
□ PolyBase □ Azure blob Storage에서 읽고 내용을 Azure SQL DW에 로드합니다. □ 제어 노드를 우회하고 컴퓨팅 노드로 직접 로드합니다.

컨트롤 노드



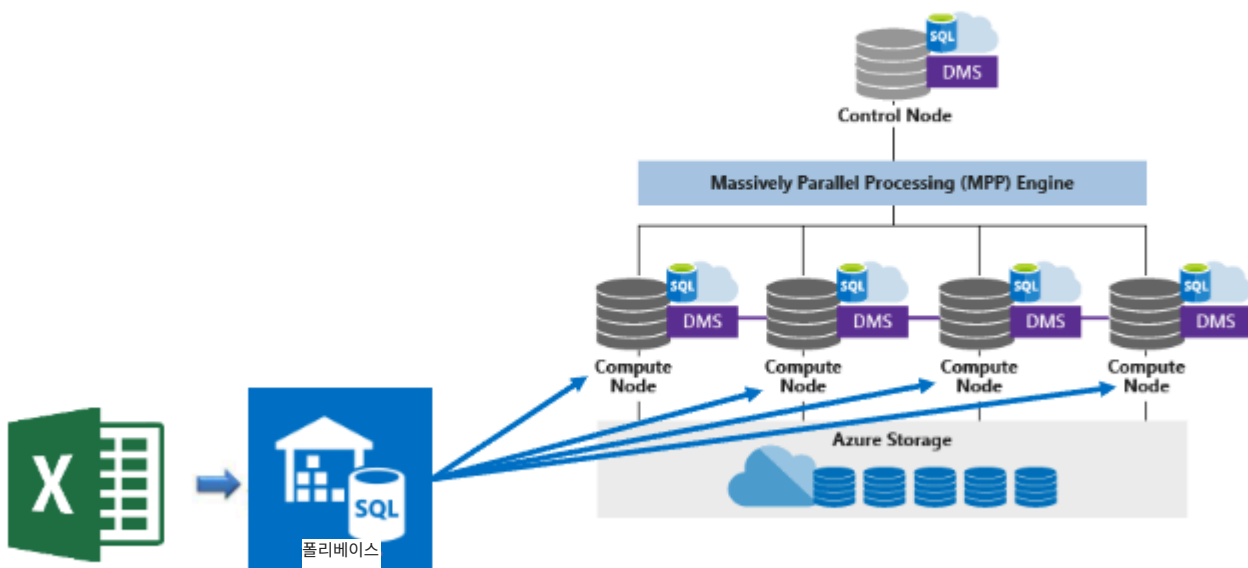
출처: Microsoft

SSIS를 사용한 로드



출처: Microsoft

PolyBase를 사용하여 로드



출처: Microsoft

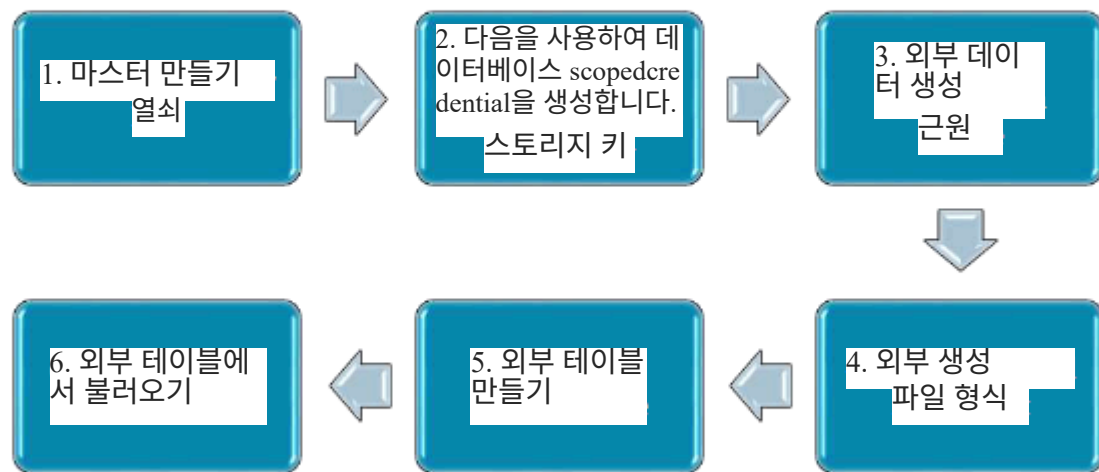


PolyBase는 UTF-8로 구분된 텍스트 파일 및 RC 파일, ORC 및 Parquet와 같은 널리 사용되는 Hadoop 파일 형식에서 데이터를 로드할 수 있습니다



여러 판독기는 에 대해 작동하지 않습니다.
압축 파일

PolyBase 설정



DEMO

Load Dimension table from on-premises to Azure Synapse Data warehouse using SSIS

데모 – PolyBase

1. 테이블을 플랫폼 파일로 내보내기2. Blob Storage 계정 만들기3. Blob Storage4에 플랫폼 파일 업로드. PolyBase 6단계 프로세스5를 실행합니다. 성공적인 마이그레이션 모니터링 및 확인6. 대상 테이블에서 60개의 배포 확인



Azure
Synapse
Analytics

데모 – Data Factory를 사용하여 로드

1. 테이블을 플랫폼 파일로 내보내기2. Blob Storage 계정 만들기3. Blob Storage4에 플랫폼 파일 업로드. Data Factory 계정 만들기5. 대상 테이블 생성6. Data Factory 파이프라인을 만들고 실행하여 데이터를 이동합니다7. 대상 테이블 모니터링 및 확인



Azure
Synapse
Analytics

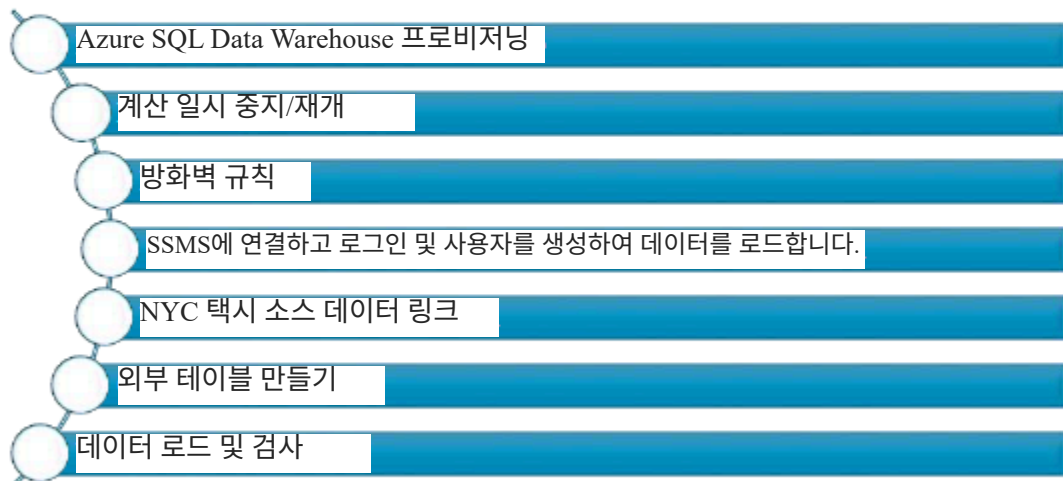
데모 – Data Factory를 사용하여 로드

1. 팩트 테이블을 플랫폼 파일로 내보내기
2. Lake Storage 계정 만들기
3. 플랫폼 파일을 Data Lake 스토리지에 업로드합니다.
4. Data Factory 계정 만들기
5. 대상 테이블 생성
6. Data Factory 파이프라인을 만들고 실행하여 데이터를 이동합니다
7. 대상 테이블 모니터링 및 확인



Azure
Synapse
Analytics

데모



요약

데이터 마이그레이션 – 모범 사례 제어 노드 대 컴퓨팅 노드 로드 역할 단일 클라이언트(SSIS) 대 완전 병렬(PolyBase) PolyBase가 선호되는 이유 PolyBase setup Demo – 다양한 시나리오 사용



Azure
Synapse
Analytics