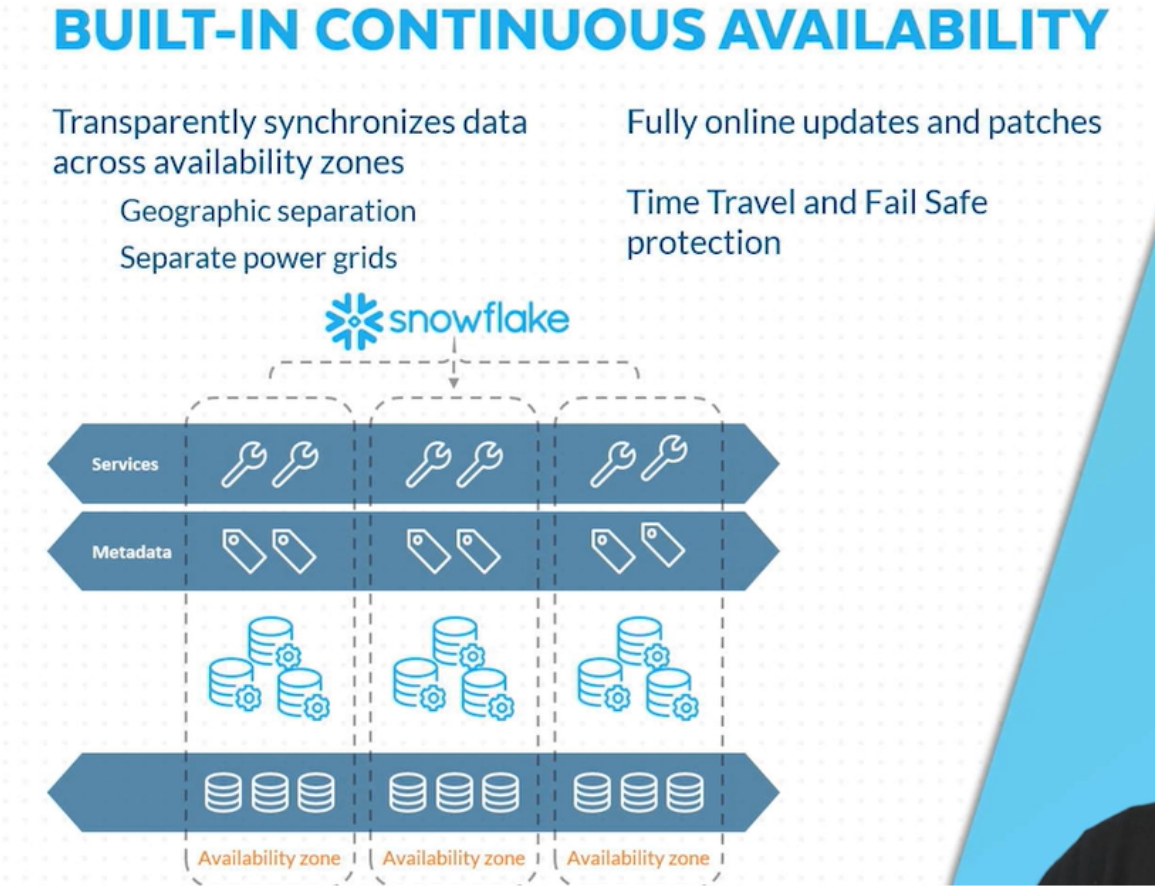
[Snowflake] 3-2. Data Protection

💡 노션 웹 공유 링크 **(댓글 & 상세설명 참고)**

> **References**
>
> - **Snowflake Learn (SnowPro PREP-CORE Course) 3장 2강**
> - **Snowflake 설명서 (연속 데이터 보호)**
> - **Snowflake 설명서 (암호화 키 관리)**
> - **Snowflake 설명서 (종단 간 암호화 이해하기)**
> - **Snowflake 설명서 (Time Travel 이해 및 사용하기)**
> - **Snowflake 설명서 (데이터베이스 복제)**

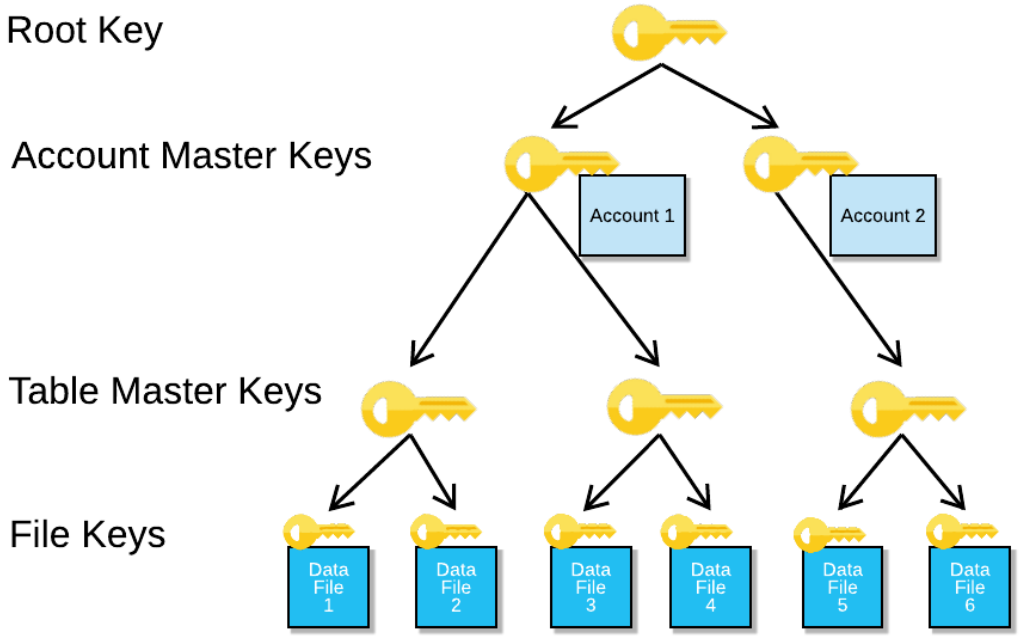
- **Continuous Data Protection (CDP, 연속 데이터 보호) 란**
  - 사람에 의한 실수, 악의적 행동, 소프트웨어 오류로 부터 데이터를 보호하기 위한 기능 세트
  - Snowflake는 데이터 수명 주기의 모든 스테이지에서 데이터의 우발적 또는 고의적 수정, 제거, 손상이 발생하는 경우에도 복구 가능한 기능 갖고 있으며, 이를 CDP라 한다.

- 모든 라이선스 정책에 기본으로 포함
- 주요 기능
  - IP 주소를 기준으로 한 사용자 엑세스 허가, 제한 (IP White, Black List 관리)
  - 계정에 엑세스 하는 모든 사용자에 대한 인증 (MFA, SSO 지원)
  - 시스템의 모든 오브젝트에 대한 사용자 엑세스 제어
  - 데이터는 AES-256 암호화를 통해 암호화 되며, 내부 스테이지에 저장되는 모든 파일 역시 AES-256 암호화를 사용하여 자동 암호화
  - Snowflake의 Time Travel 및 Fail-safe를 통한 과거 데이터 유지 관리



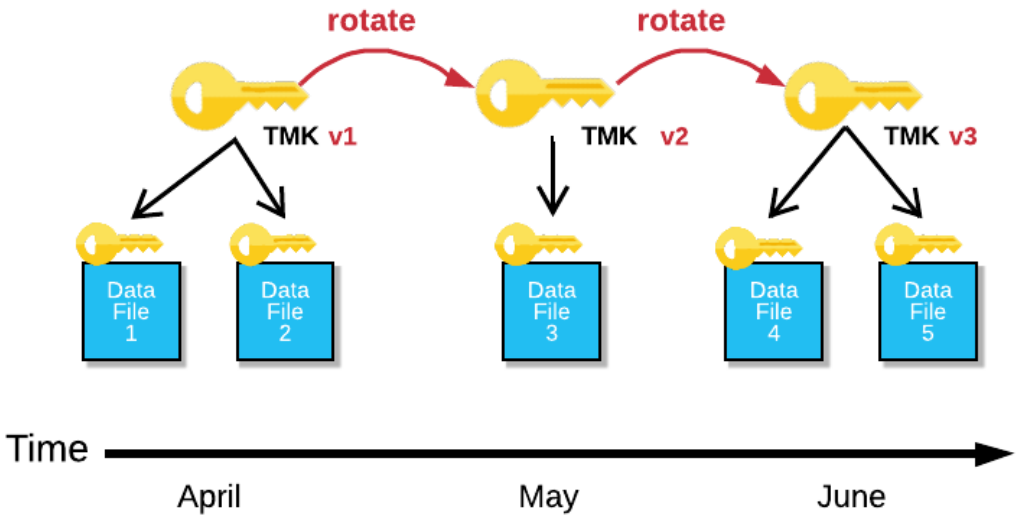
- 특정 클라우드 리전의 가용영역 전체에서 Snowflake의 스토리지 레이어를 복제하여 데이터 보호

- **계층적 키 모델 (Hierarchical Key Model) 관리**
  - Snowflake 에서는 데이터의 암호화와 파일 암호화를 모두 지원
    - AES-256 암호화 지원

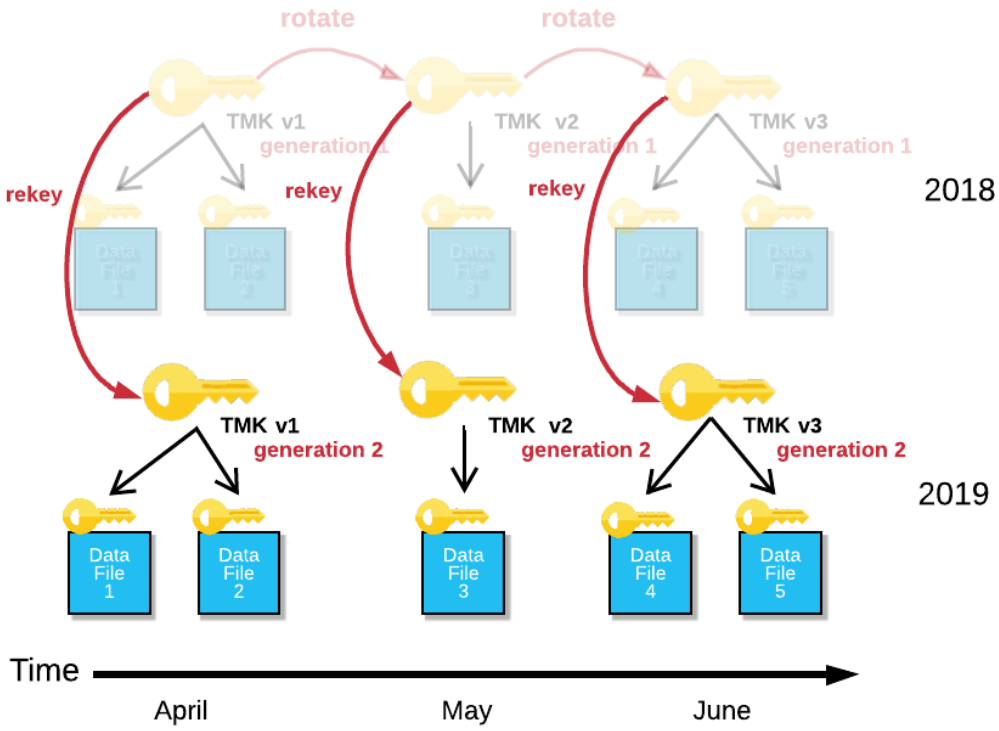


루트키 → 계정 마스터키 → 테이블 마스터키 → 파일 키

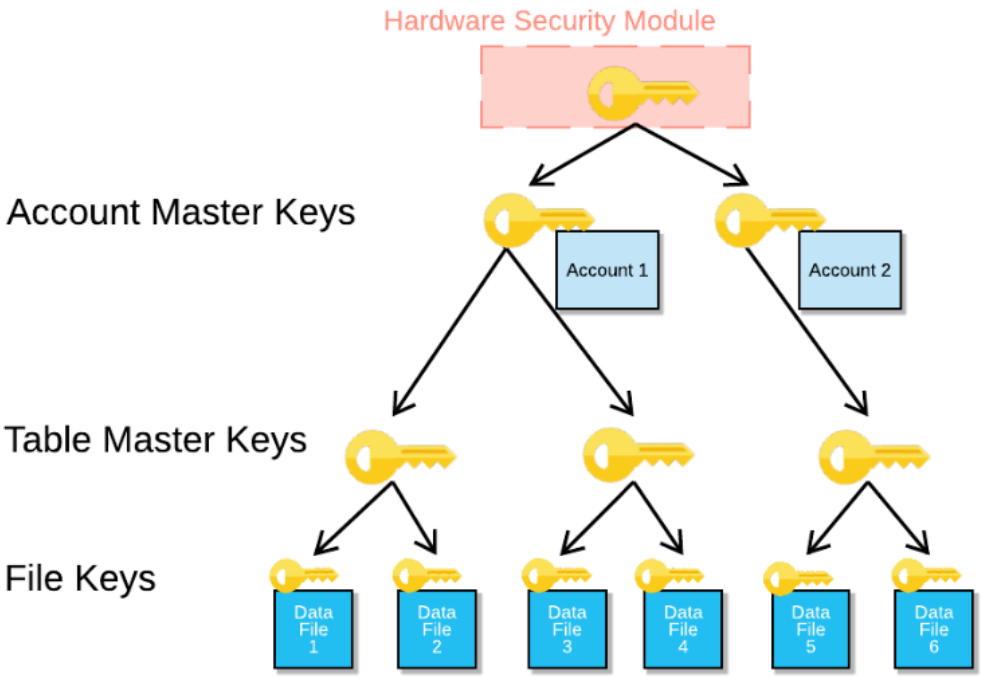
- Snowflake 에서는 암호화 키를 관리하기 위한 계층적 키 모델 (Hierarchical Key Model) 프레임워크를 제공



- 암호화에 사용되는 키는 30일 마다 자동으로 로테이션, 추가 구성 또는 관리작업이 필요하지 않고 서비스에 의해 자동으로 수행 (Rotation)



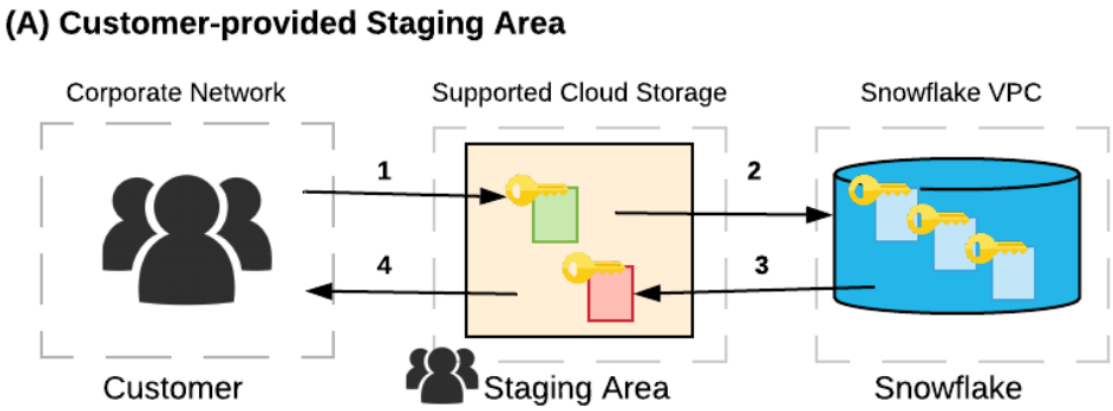
- Enterprise Edition 이상일 경우 주기적 키 재생성 가능 (Re-generation)
  - ALTER ACCOUNT SET PERIODIC_DATA_REKEYING = true;



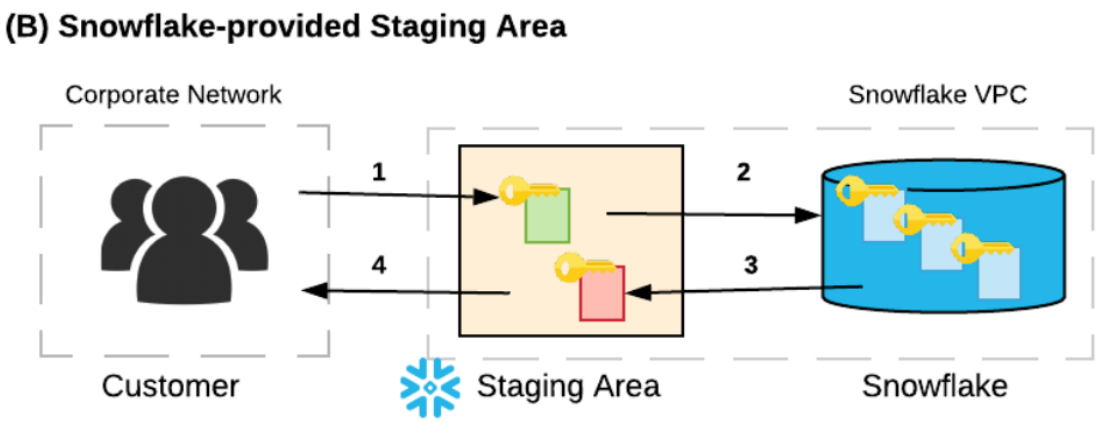
- 하드웨어 보안 모듈 (HSM)

- Tri-Secret Secure 기능 지원 (Business Critical Edition 이상)
  - 데이터 보호를 위한 복합 마스터 키를 생성 기능
  - Snowflake의 관리 키와 클라우드 공급자의 고객 관리키를 조합 (고객 키는 항상 사용가능한 상태를 유지)
  - 복합 마스터 키는 계정 마스터 키 역할을 하며 계층 구조의 모든 키를 래핑하지만, 복합 마스터 키는 결코 원시 데이터를 암호화 하지 않음

- **Snowflake 의 종단 간 암호화 (Automatic E2EE (End-to-End Encryption))**
  - E2EE는 데이터를 보호하고 공격 표면을 최소화 하는 방법
  - Snowflake 의 E2EE 시스템은 두가지 방식



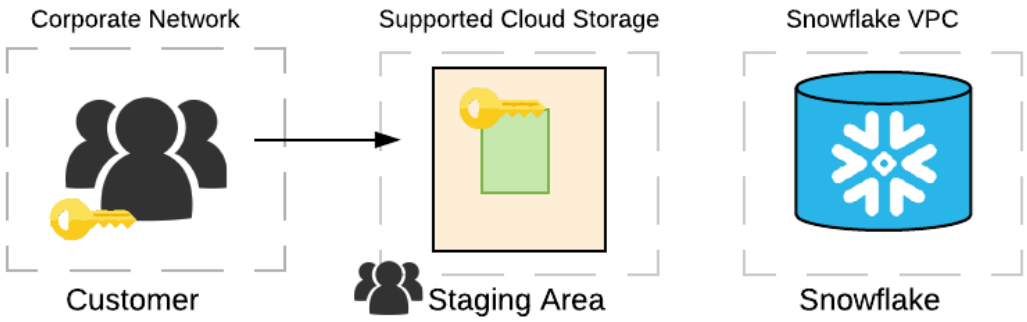
외부 스테이지 사용 (Customer-provided Staging Area)



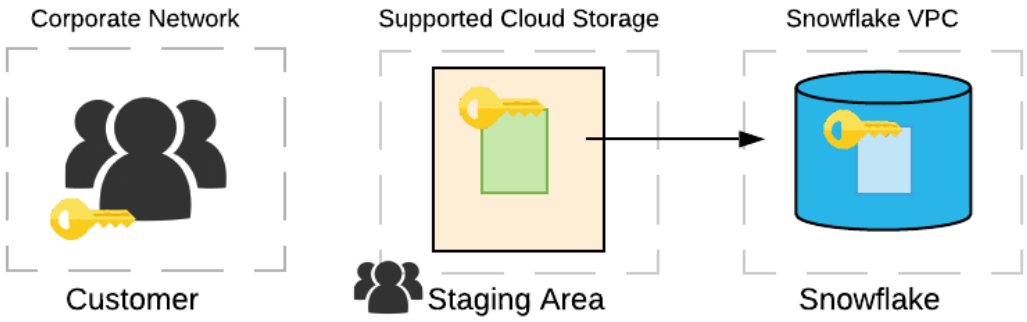
내부 스테이지 사용 (Snowflake-provided Staging Area)

- E2EE 흐름 상세
  1. Customer 는 Staging Area 로 데이터 파일을 업로드
     a. 외부 스테이지를 사용할 경우 클라이언트 측 암호화를 사용하여 선택적 암호화
     b. 내부 스테이지를 사용할 경우 전송 전 고객 시스템의 Snowflake 클라이언트에 의해 자동으로 암호화 & 스테이지로 로딩 된 후 암호화
  2. Staging Area 에서 Snowflake로 데이터를 로드
     a. 데이터는 Snowflake 독점 파일 형식으로 변환되어 클라우드 저장소 컨테이너에 저장, 모든 미사용 데이터는 항상 암호화 되며 전송중에도 TLS 암호화
     b. 데이터가 테이블에서 변환되거나 연산 시 데이터의 암호를 해독 하여 사용 후 다시 암호화
  3. 쿼리 결과를 Staging Area 로 언로딩
     a. 외부 스테이지의 경우 클라이언트측 암호화, 내부 스테이지의 경우 자동 암호화
  4. 스테이지에서 데이터 파일을 다운로드하고 클라이언트 측에서 데이터 해독

- **클라이언트 측 암호화**

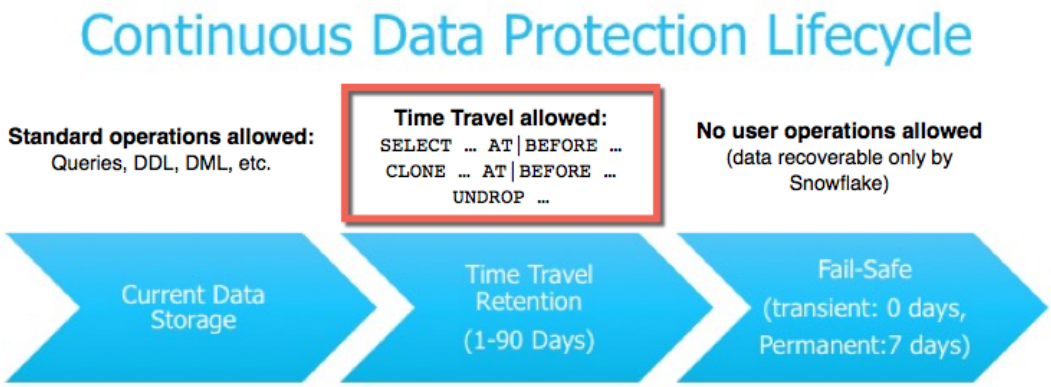


  - 고객이 Snowflake와 공유하는 시크릿 마스터 키를 생성
  - 클라우드 저장소 서비스(Staging Area)에서 임의의 암호화 키를 생성하여 파일을 업로드, 임의의 암호화 키는 공유된 마스터키를 사용하여 암호화
  - 암호화 된 파일 및 암호화 된 임의 키 모두 클라우드 저장소 서비스에 업로드 되며 암호화된 임의키는 파일의 메타데이터와 함께 저장

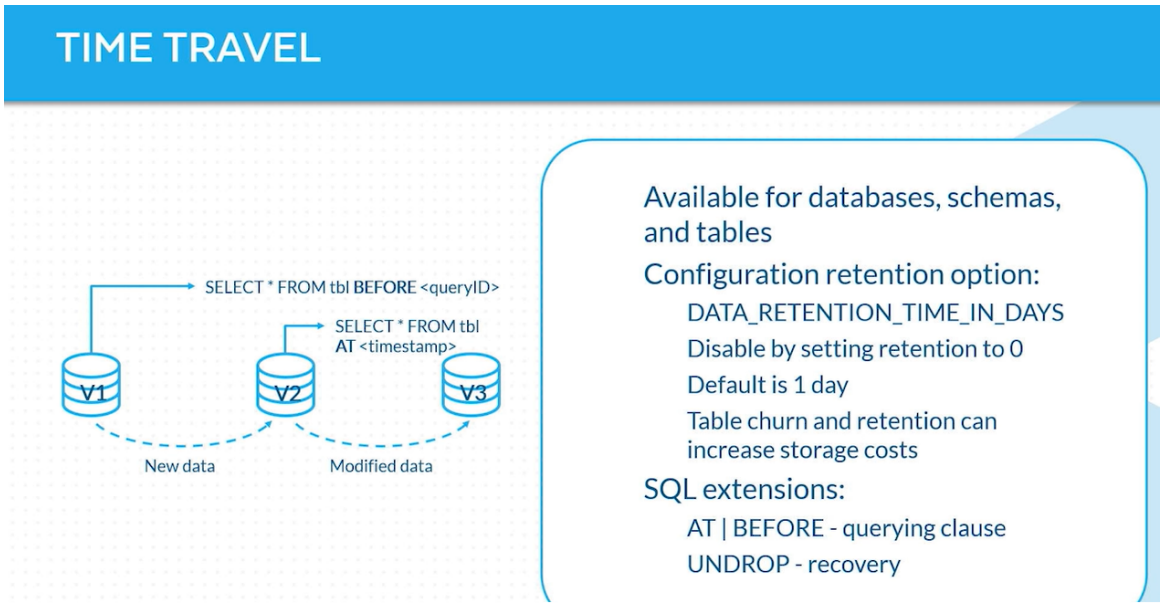


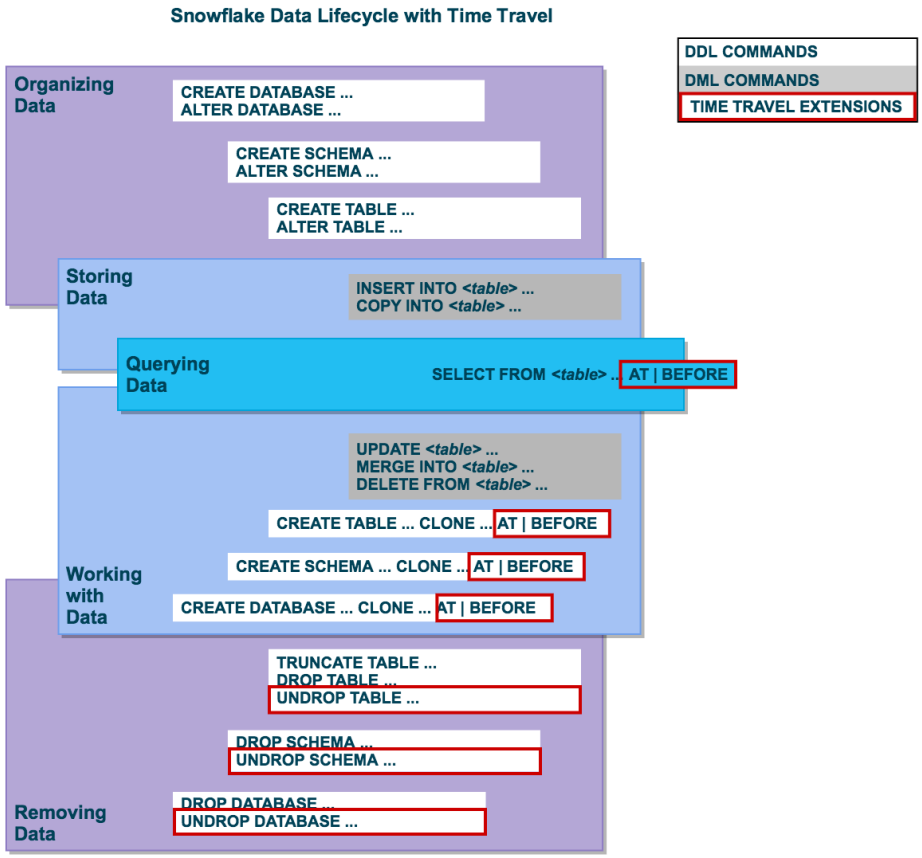
  - 암호화 된 데이터를 Snowflake로 로드할 때 CREATE STAGE 명령을 사용하여 추가 MASTER_KEY 매개변수가 포함된 스테이지 오브젝트를 생성하여 데이터 로드
  - MASTER_KEY 매개변수는 Base64로 인코딩 된 128, 256비트의 AES 암호화 키 필요 = 고객과 공유한 마스터 키

- **Time Travel**
  - Snowflake Time Travel 은 정의된 기간 내 모든 시점의 과거 데이터에 엑세스할 수 있는 기능

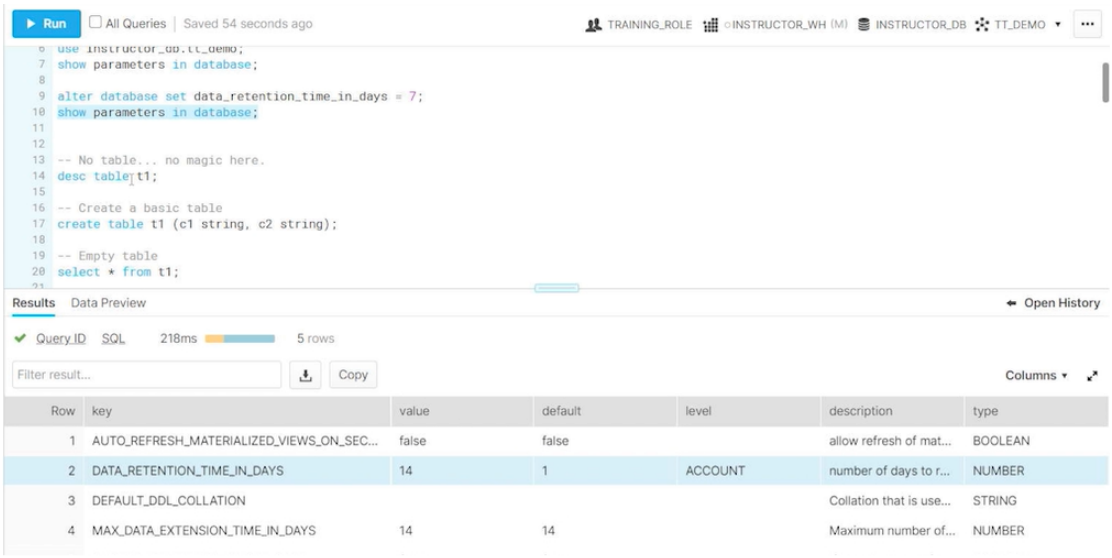


- Time Travel을 사용하면 정의된 기간내 과거 데이터 관련 작업이 가능
  - 실수로 또는 의도적으로 삭제한 데이터 관련 오브젝트 복원
  - 과거의 주요 시점의 데이터 복제 및 백업
  - 지정된 기간 동안 데이터 사용, 조작 분석
- 정의된 기간이 경과되면, 데이터는 Snowflake Fail-Safe로 이동되며, Time Travel 작업을 수행할 수 없음





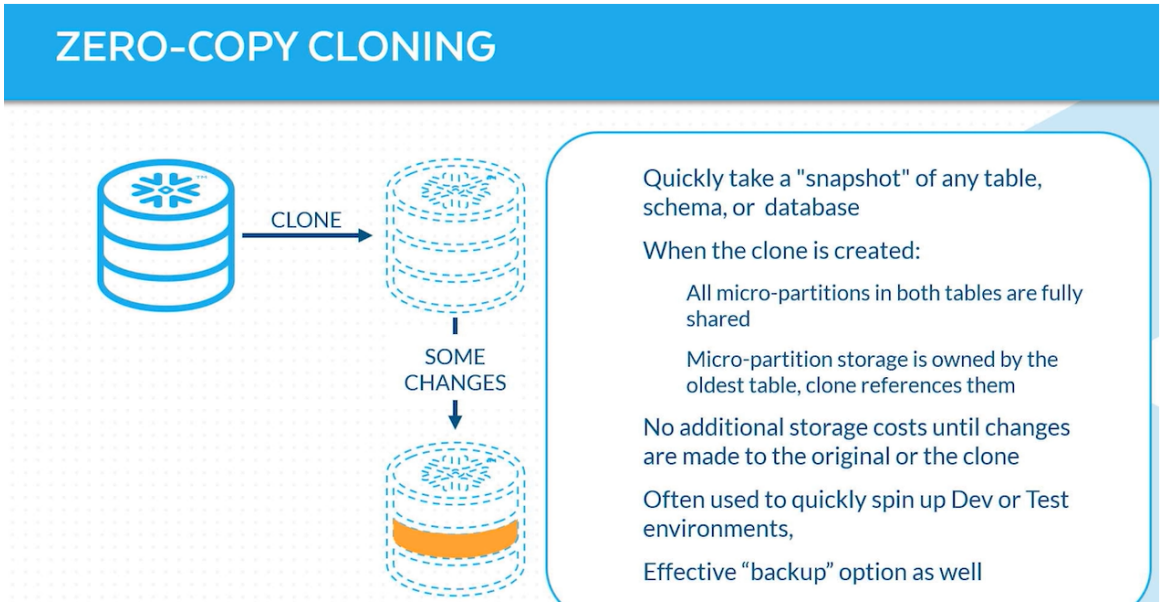
- Time Travel 의 SQL 확장
  - AT, BEFORE 절을 사용하여 정확한 과거 데이터를 찾으며, 이 절은 다음 매개변수를 사용
    - TIMESTAMP
    - OFFSET → 현재 시간과 초 단위 시간 차이
    - STATEMENT → 식별자 (ex. 쿼리 ID)
  - 테이블, 스키마, 데이터베이스에 대한 UNDROP 명령(복원 명령)
- Time Travel 의 데이터 보존 기간
  - 표준 보존 기간은 Default 값 = 1일(24시간)
  - Standard Edition 의 경우 0-1 설정 가능
  - Enterprise Edition 의 경우 0~90일 사이의 값으로 설정 가능
  - 보존기간이 끝난 경우 과거 데이터는 Snowflake Fail-safe로 이동



show parameters in database

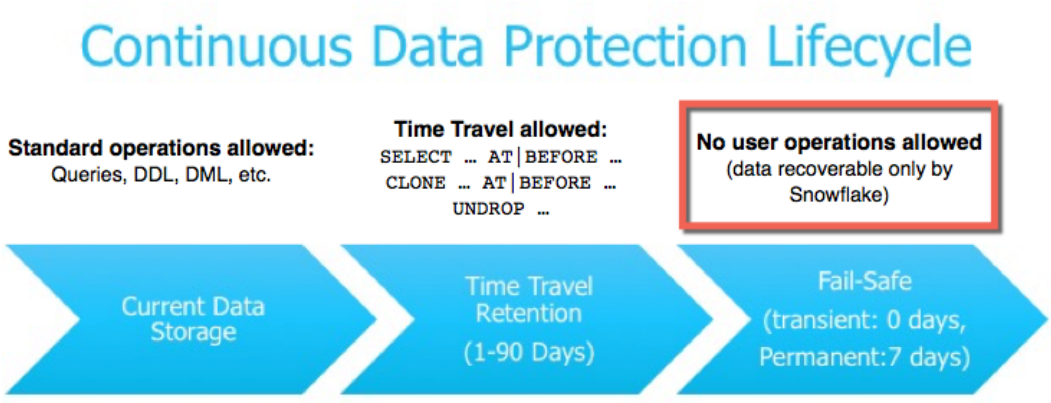
- ACCOUNTADMIN 권한을 가진 사용자는 DATA_RETENTION_TIME_IN_DAYS 오브젝트 매개변수를 사용하여 기본 보존 기간을 지정
- MIN_DATA_RETENTION_TIME_IN_DAYS 계정 매개변수를 사용해 계정의 최소 보존기간을 설정 가능
- MAX_DATA_EXTENSION_TIME_IN_DAYS 매개변수는 데이터 보존 기간을 연장할 수 있는 최대 일수

- **데이터베이스 복제 (ZERO-COPY CLONING) - 추가 참고**

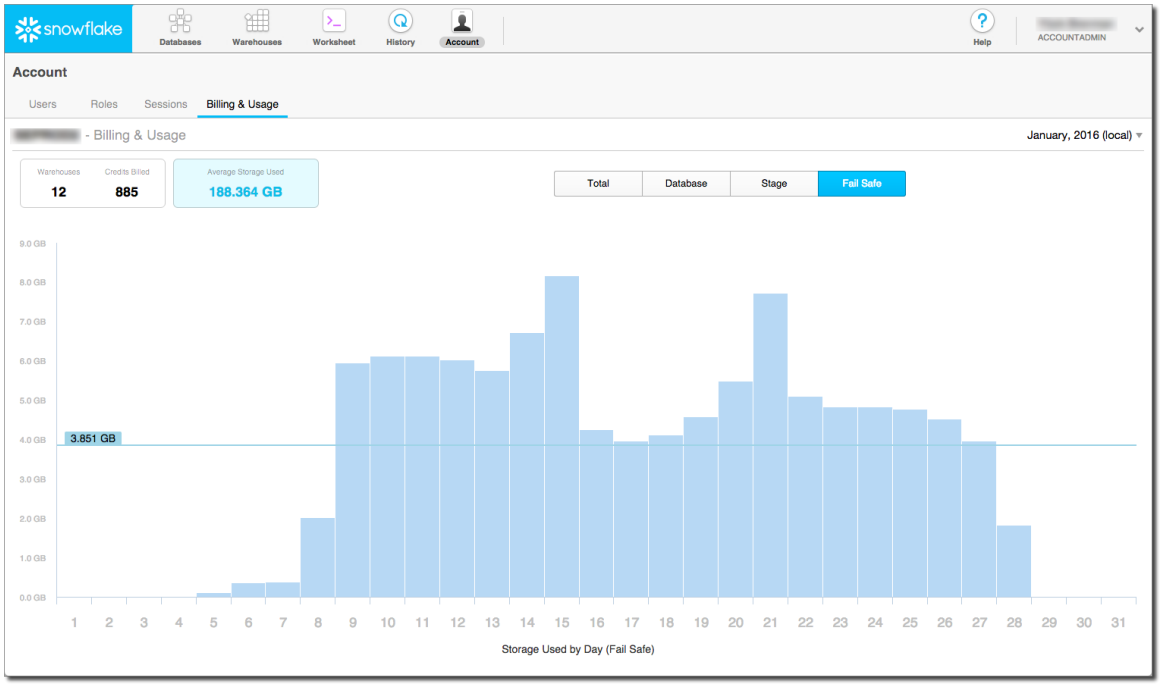


- Snowflake 의 복제는 테이블 스키마 또는 전체 데이터베이스 스냅샷 또는 복사 하며, 스토리지의 데이터는 복제하지 않는다.
- 복제된 데이터베이스는 복제 시점부터 독립적으로 작동되며 동기화 되지 않음.
- SQL 쿼리문 CLONE 구문을 사용하여 데이터베이스 복제
- 복제를 사용하여 스토리지 복제 없이 개발 → QA 환경으로 신속하게 스핀업 가능
- 복제된 데이터베이스를 일종의 백업 옵션으로 사용 가능

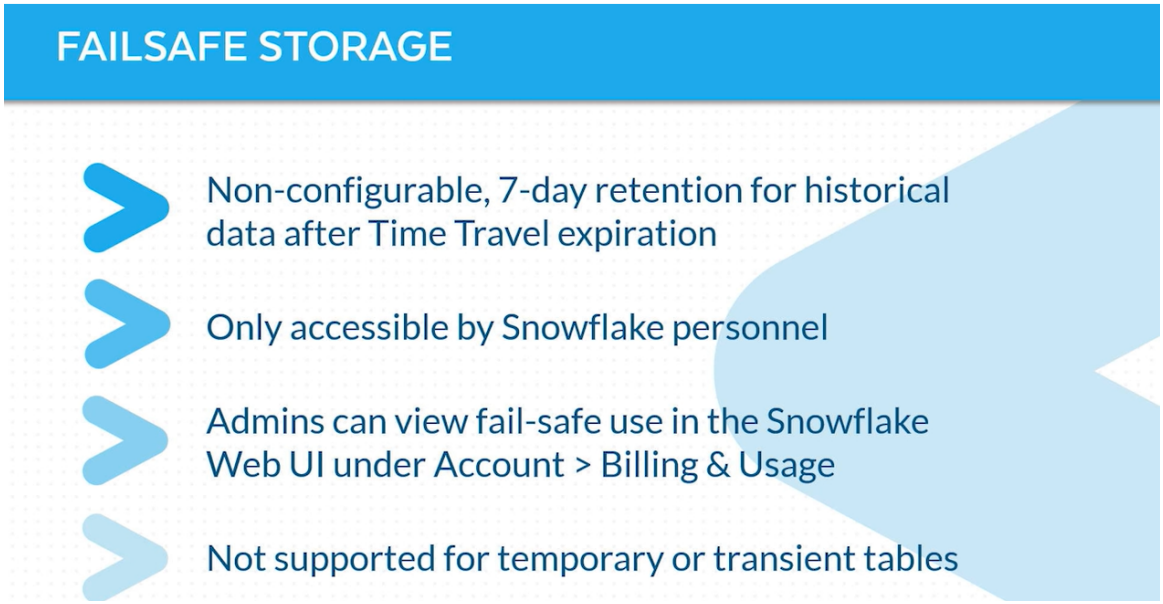
- **Snowflake Fail-Safe**



- Time Travel 이 만료 된 데이터를 보호하는 기능
- 시스템 오류 또는 기타 이벤트 발생 시 과거 데이터 보호
- 과거 데이터를 복원할 수 있는 7일의 기간을 제공 (설정 불가능)



계정에 대한 Fail-Safe 저장소 확인 (Snowsight)



- 영구 테이블을 사용하는 경우만 데이터 보호 가능 (임시, 일시, 외부 테이블은 Fail-Safe 기능이 없음)
- Fail-Safe 된 데이터는 사용자가 조작 불가능, Snowflake 담당자, 관리자, 기술지원과 협력하여 데이터 복구 가능
- Fail-Safe 용 별도의 스토리지가 있으며, 추가 비용 지불 필요