

02_Data analytics in the cloud

Cloud Data Analytics 과정 다음 섹션 소개

- 학습 목표: 다음 섹션에서는 온프레미스 시스템에서 클라우드로 데이터를 마이그레이션하는 방법과 온프레미스, 하이브리드, 클라우드 데이터 시스템 아키텍처의 차이점을 탐구함.
 - **Google Cloud** 아키텍처 프레임워크: Google Cloud Architecture Framework에 대해 자세히 학습함.
 - 클라우드와 데이터 분석의 영향: 클라우드가 데이터 분석 및 기타 여러 산업에 미치는 영향을 확인함.
 - 비용 최적화: 클라우드 비용 최적화 전략과 그 이점에 대해 알아보고, 스토리지, 쿼리 실행, 리소스 프로비저닝 비용 및 다양한 결제 모델에 대해 살펴봄.
 - 활용: 이러한 지식을 바탕으로 미래에 고용주에게 가장 비용 효율적인 클라우드 솔루션을 제안할 수 있음.
-

Cloud Data Migration의 개요

- 정의:
 - 온프레미스 컴퓨팅 네트워크 인프라를 클라우드 플랫폼으로 이전하는 과정.
 - 신중한 계획과 노력이 필요함.
- 마이그레이션 전 고려 사항:
 - 적합한 클라우드 환경 선택: 조직에 맞는 클라우드 환경을 선택하는 것이 중요함.
 - 데이터 양: 이전할 데이터의 양이 많을 경우, 시간이 오래 걸려 운영에 지연이 발생할 수 있음.
 - 운영 중단 시간(**Downtime**): 마이그레이션 중 시스템 중단 시간을 모든 이해관계자와 합의하는 것이 필요함.

Cloud Migration 전략

- **Rehosting (Lift and Shift):**
 - 기존 온프레미스 시스템을 변경 없이 클라우드로 통째로 옮기는 전략.
 - 조직이 클라우드의 향상된 효율성, 안정성, 혁신 기술을 빠르게 활용할 수 있음.
- **Replatforming:** 클라우드로 마이그레이션 후 온프레미스 시스템에 작은 변경을 가하여 성능을 개선하는 전략.
- **Repurchasing:**
 - 기존 애플리케이션을 새로운 클라우드 기반 서비스(주로 **SaaS**)로 이전하는 전략.
 - 새로운 경험에 대한 팀원 교육이 필요함.

- **Refactoring:** 기존 애플리케이션을 버리고 서버리스 컴퓨팅과 같은 새로운 기능을 가진 완전히 새로운 애플리케이션을 처음부터 구축하는 전략.
- **Retiring:** 더 이상 유용하지 않은 애플리케이션을 중단시키는 전략.

Google Cloud를 활용한 마이그레이션

- 마이그레이션 파트너 선택: 비즈니스에 적합한 인프라와 서비스를 제공하고, 지속적인 개발에 투자하며, 신뢰할 수 있는 고객 지원을 제공하는 클라우드 서비스 제공업체(CSP)를 선택하는 것이 중요함.
 - **Google Cloud Adoption Framework:** 조직의 클라우드 기술 채택 준비 상태를 평가하는 데 도움이 되는 프레임워크. 다음 4가지 테마를 평가함.
 - **Learn (평가):** 조직의 학습 프로그램 품질 및 규모.
 - **Lead (리더십):** 리더십이 IT 부서에 제공하는 지원 수준.
 - **Scale:** 클라우드 기반 서비스 사용 범위 및 시스템 관리 자동화 필요성.
 - **Secure:** 무단 접근으로부터 클라우드 환경을 보호하는 능력.
 - **Google Cloud 마이그레이션 경로:**
 - **Assess (평가):** 기존 네트워크 인프라를 철저히 검토함.
 - **Plan (계획):** 워크로드가 존재할 기본 클라우드 인프라를 설정함.
 - **Deploy (배포):** 워크로드를 실제로 Google Cloud로 옮김.
 - **Optimize (최적화):** 클라우드 기반 기술 및 기능을 사용하기 시작하며, 개선된 접근성, 확장성, 비용 절감, 보안, 효율성을 누림.
 - 결론: 클라우드 마이그레이션은 신중한 계획이 필요한 작업이며, 클라우드 데이터 분석가로서 이러한 단계를 이해하고 조직을 도울 수 있는 능력을 갖추는 것이 중요함.
-

Cloud Deployment Models 개요

- 배경: 조직의 고유한 비즈니스 요구에 맞는 적절한 환경을 선택하는 것이 중요함.
- 3가지 주요 모델: Public cloud, Private cloud, Hybrid cloud.
- 클라우드 전문가의 역할: 유연하고, 적응력이 뛰어나며, 변화하는 조건에 신속하게 대응할 수 있는 모델을 선택하는 데 도움을 줄 수 있음.
- 비유: 건물을 난방하는 여러 방식에 비유할 수 있음.
 - **Public cloud:**
 - 전기 회사로부터 난방에 필요한 전기를 공급받는 것.
 - 사용한 만큼만 지불하고 유지보수 걱정이 없음.
 - **Private cloud:**
 - 자체 태양광 패널을 설치하여 전기를 생산하는 것.
 - 패널을 직접 구매하고 관리해야 함.
 - **Hybrid cloud:**
 - 전기 회사 서비스와 자체 태양광 패널을 모두 사용하는 것.
 - 더 많은 옵션과 유연성을 가짐.

각 모델의 특징 및 장단점

- **Public Cloud**
 - 특징:
 - 컴퓨팅, 스토리지, 네트워크 리소스를 인터넷을 통해 제공
 - 여러 사용자와 조직이 공유
 - 제3자 클라우드 서비스 제공업체가 인프라를 관리하고 유지보수함.
 - 장점: 사용한 만큼만 지불, 손쉬운 확장/축소, 유지보수 걱정 없음, 높은 신뢰성, 빠른 배포, AI/ML과 같은 최신 혁신 기술 및 업데이트 제공.
 - 단점: 없음.
- **Private Cloud**
 - 특징:
 - 모든 클라우드 리소스를 단일 사용자 또는 조직에 전용으로 제공
 - 온프레미스 데이터 센터 내에 구축되고 관리 및 소유
 - 장점:
 - 적절히 보호될 경우 사적이고 안전한 네트워크 제공, 데이터 저장 및 컴퓨팅 위치를 제어하여 규정 준수 용이,
 - 하드웨어를 공유하지 않아 일관된 성능 제공.
 - 단점: 높은 유지보수 및 관리 비용, 부적절한 보안 설정 시 해킹에 취약.
- **Hybrid Cloud**
 - 특징:
 - Public cloud와 Private cloud 모델의 조합.
 - 클라우드 서비스와 온프레미스 클라우드 모델의 제어 기능을 모두 활용 가능
 - 장점: 기존 온프레미스 인프라에 public cloud를 추가하여 데이터 센터 비용 없이 컴퓨팅 파워 증가, AI/ML과 같은 최신 혁신 기술에 접근, 보안 및 규정 준수 장점, 사용자에게 더 가까운 컴퓨팅으로 인한 빠른 성능, 특정 작업에 가장 적합한 클라우드 환경을 선택할 수 있는 유연성.
 - 단점: Public/Private 모델을 혼합하여 복잡성이 증가할 수 있음.
- 결론: 조직의 고유한 요구에 맞춰 Public, Private, Hybrid 클라우드 모델 중 가장 적합한 모델을 선택하는 데 클라우드 전문가의 조언이 중요함.

클라우드 배포 모델의 이점

- 배경: 이전 학습에서 public, private, hybrid 클라우드 배포 모델에 대해 학습. 이 모델들은 조직이 클라우드 환경을 활용하는 다양한 방법을 설명함.
- 클라우드 전문가의 역할: 클라우드 데이터 전문가로서 이러한 배포 모델을 이해하는 것은 조직의 필요에 따라 다양한 클라우드 환경에 적응하는 데 도움이 됨.

클라우드 배포 모델 요약

배포 모델	정의	이점	예시 도구

Public	인터넷을 통해 컴퓨팅, 스토리지, 네트워크 리소스를 제공하여 사용자가 특정 비즈니스 요구에 따라 온디맨드 리소스를 공유하도록 허용하는 클라우드 모델.	비용 효율적, 확장 가능함, 유지보수가 필요 없음, 신뢰성이 높음, 빠르고 쉽게 배포 가능, 혁신적.	Google Cloud Platform, Amazon Web Services, Microsoft Azure
Private	모든 클라우드 리소스가 단일 사용자 또는 조직에 전용으로 제공되며, 온프레미스 데이터 센터 내에서 생성, 관리, 소유되는 클라우드 모델.	안전함, 내부 규정을 준수함, 일관되고 예측 가능한 성능.	Amazon Virtual Private Cloud, VMWare, HPE Greenlake
Hybrid	Public 및 Private 모델을 결합하여 조직이 클라우드 서비스와 온프레미스 클라우드 모델의 제어 기능을 모두 활용할 수 있는 클라우드 모델.	강력한 컴퓨팅 성능, 최신 혁신 기술 활용, 안전함, 내부 규정 준수, 빠른 성능, 유연한 운영.	Google Anthos, Microsoft Azure Arc, AWS Outposts

관리형 클라우드 서비스 (Managed Cloud Services)

- 정의: 조직이 클라우드 시스템 관리를 외부 업체에 위탁하는 경우
- 역할:
 - 제3자 업체가 조직의 클라우드 리소스 및 인프라의 일부 또는 전부를 관리하여 클라우드 시스템이 원활하게 운영되도록 함.
 - 예를 들어, 소매업체가 온라인 상점을 관리하는 데 사용하는 SaaS 도구인 Salesforce가 관리형 클라우드 서비스에 해당함.
 - 이러한 서비스는 회사의 더 큰 클라우드 생태계의 일부로 간주됨.

핵심 내용 요약

- 클라우드 데이터 전문가로서 조직의 현재 클라우드 시스템과 해당 도구를 관리하는 방법을 이해하는 것이 업무의 일부임.
- 조직의 요구 사항을 해결하기 위한 새로운 클라우드 환경을 설정하는 방법을 알아야 할 수도 있음.
- 다양한 클라우드 배포 시스템과 이것이 회사의 전반적인 데이터 생태계에 어떻게 기여하는지 파악하는 것이 중요한 일상적인 역할이 될 것임.

Cloud Data Analytics의 역할

- 전통적 데이터 관리의 한계: 과거에는 데이터 관리가 수작업으로 이루어져 노동 집약적이고, 리소스가 많이 소모되며, 오류 발생 가능성이 높았음.
- 클라우드의 역할: 클라우드 데이터 분석은 이러한 과정을 자동화하고 향상하여 데이터 관리를 훨씬 효율적으로 만들고, 인적 오류를 줄임.
- 데이터 통합: 클라우드를 통해 다양한 소스의 데이터를 원활하게 통합하여 사용자가 실시간으로 접근하고 분석할 수 있는 단일 소스를 생성함

데이터 관리 방법

- 데이터 통합 (**Data Integration**): 다양한 소스의 데이터를 단일의 사용 가능한 데이터 소스로 결합함.
 - **ETL (Extract, Transform, and Load)**: 데이터를 웨어하우스에 로드하기 전에 변환함.
 - **ELT (Extract, Load, and Transform)**: 데이터를 웨어하우스에 로드한 후에 변환함.
- 데이터 수집 (**Data Ingestion**): 나중에 사용하거나 저장하기 위해 데이터를 얻고, 가져오고, 처리함.
 - 스트림 수집 (**Stream ingestion**): 다양한 소스에서 수집된 데이터를 실시간으로 지속적으로 처리함.
 - 배치 수집 (**Batch ingestion**): 미리 정해진 간격 또는 더 큰 덩어리로 데이터를 처리함.

데이터 유형 및 클라우드 도구

- 데이터 유형: 클라우드 데이터 분석을 통해 파일(files), 객체(objects), 블록(blocks)과 같은 다양한 유형의 데이터를 저장할 수 있음.
 - 파일 데이터: 컴퓨터나 다른 저장 장치의 파일에 저장된 정보.
 - 객체 데이터: 고유 식별자가 있어 저장 위치에 관계없이 찾을 수 있는 정보 조각.
 - 블록 데이터: 더 큰 정보에서 잘라내어 자체 파일 경로를 가진 정보 조각.
- 데이터 접근 도구: 웹 인터페이스, **APIs**, **SQL**, **Pub/Sub**와 같은 수집 도구, **Looker** 및 **Jupyter Notebooks**와 같은 비즈니스 인텔리전스 솔루션 등이 있음. 이러한 도구들은 사용자가 언제 어디서든 클라우드에 저장된 데이터에 접근하도록 도움.

Cloud Data Analytics의 이점

- 향상된 프로세스: **Big-data** 분석, 여러 데이터 소스의 비동기적 시각화, **AI** 및 **ML**, 맞춤형 보고서 분석, 데이터 마이닝, 데이터 과학 등 많은 데이터 분석 활동이 클라우드 데이터 분석 프로세스를 통해 크게 향상됨.

- 혁신의 원동력: 데이터는 많은 기업의 혁신을 이끄는 원동력이며, 클라우드 분석은 데이터 분석 분야를 발전시켜 모든 종류의 조직이 강력한 분석 도구와 프로세스를 사용할 수 있게 함.
- 비용 효율성: 분석 프로세스를 향상하여 사용자가 데이터에서 가치 있는 인사이트를 더 쉽고, 빠르고, 비용 효율적으로 발견하도록 함.

Cloud Data Analytics의 이점

- **Cloud vs. On-premises:** Cloud data analytics는 여러 소스에서 다양한 유형의 데이터를 더 효율적으로 수집 및 저장하고, 사용자가 언제든지 필요한 데이터에 쉽게 접근할 수 있도록 하는 이점이 있음.
- **Cloud data analytics의 주요 이점:** Cloud data analytics는 기존 온프레미스 데이터 분석이 제공할 수 없는 확장성, 낮은 비용, 유연성, 접근성 및 낮은 유지보수를 포함한 다양한 이점을 제공함.

Cloud vs. On-premises 데이터 분석 비교

기능	Cloud data analytics	On-premises data analytics
확장성 (Scalability)	조직의 변화하는 요구에 맞춰 Cloud 스토리지 및 리소스를 쉽게 확장 가능함. 물리적 공간에 리소스를 구축할 필요가 없어 업데이트가 용이함.	온프레미스 리소스 확장 과정은 매우 복잡함. 더 많은 스토리지가 더 나은 처리 하드웨어를 추가하려면 물리적 컴퓨터 시스템을 업데이트해야 함.
비용 (Cost)	물리적 공간, 하드웨어 및 소프트웨어에 투자할 필요가 없어 다양한 규모의 조직에 비용 효율적인 솔루션임.	온프레미스 데이터 분석은 물리적 공간, 공과금, 현장 보안 및 조직의 분석 요구를 지원하는 컴퓨터를 구축하기 위한 초기 투자가 필요함.
유연성 (Flexibility)	조직에 필요한 서비스를 언제든지 선택하고 조정할 수 있어 유연성이 높음. 필요에 따라 전략을 수정할 수 있음.	온프레미스 시스템을 설계할 때 서비스 및 기능을 미리 결정해야 함. 새로운 요구 사항이 발견되면 변경하는 데 많은 시간이 소요됨.

<p>접근성</p> <p>(Accessibility)</p>	<p>팀원이 현장에 없어도 필요한 데이터에 접근할 수 있게 해줌. 데이터는 이해관계자가 어디에 있던 제공될 수 있음.</p>	<p>접근이 더 어렵고 더 많은 관리가 필요함.</p>
<p>유지보수</p> <p>(Maintenance)</p>	<p>물리적 인프라가 Cloud 제공업체에 의해 관리되므로, 온프레미스 시스템이 요구하는 물리적 유지보수가 필요하지 않아 유지보수 비용이 일반적으로 낮음.</p>	<p>데이터베이스 유지보수와 함께 온프레미스 컴퓨터 시스템은 IT팀이 업데이트하고 기능을 유지해야 함.</p>

클라우드 비용에 영향을 미치는 주요 요인

- 배경: 클라우드 데이터 분석 비용 관리는 시장에 갈 때 쇼핑 목록을 만드는 것과 비슷함. 필요한 리소스와 양을 정확히 아는 것이 돈을 절약하고 낭비를 막는 핵심임.
- 주요 비용 요인: 클라우드 전문가들이 비용을 관리하기 위해 사용하는 세 가지 주요 방법은 리소스 프로비저닝, 스토리지, 쿼리 실행임.

리소스 프로비저닝 (Resource Provisioning)

- 정의: 사용자가 적절한 소프트웨어 및 하드웨어 리소스를 선택하면, 클라우드 서비스 제공자가 이를 설정하고 사용 중에 관리하는 과정.
- 세 가지 제공 모델:
 - 사전 프로비저닝 (**Advance provisioning**): 사용자가 클라우드 서비스 제공자와 공식 계약을 맺고 고정 가격을 지불하거나 월별 청구를 받음. 제공자는 합의된 리소스를 수집하여 사용자에게 전달함.
 - 동적 프로비저닝 (**Dynamic provisioning**): 사용자의 변화하는 요구에 따라 리소스가 조정되며, 사용한 만큼만 요금이 부과됨. 사용량 요구에 따라 쉽게 확장하거나 축소할 수 있음.
 - 자체 프로비저닝 (**Self-provisioning**): 사용자가 웹사이트나 온라인 포털을 통해 리소스를 구매하며, 리소스는 몇 시간 또는 몇 분 내에 신속하게 제공됨.
- 결제 방식: 고정 요금(fixed), 종량제(pay-as-you-go), 즉시 구매(instant purchase)의 세 가지 요금제로 지불할 수 있음.

스토리지 비용 (Storage Costs)

- 주요 비용 요인: 스토리지 비용은 데이터 저장, 데이터 처리, 네트워크 사용량에 따라 달라짐.
 - 데이터 저장 (**Data storage**) = 저장되는 데이터의 양.
 - 클라우드 전문가들은 이를 '버킷(buckets)'이라고 부름.

- 스토리지 비용은 버킷의 위치와 저장되는 데이터의 유형(클래스)에 따라 달라질 수 있음.
- 예를 들어, 보관(archive) 스토리지는 백업 전용으로 가장 저렴함.
- 데이터 처리 (**Data processing**): 원시 데이터를 정리, 구성, 분석을 위한 형식으로 변환하는 단계. 처리하는 데이터가 많을수록 더 많은 저장 공간이 필요함.
- 네트워크 사용 (**Network use**): 스토리지 버킷 간에 읽거나 이동하는 데이터의 양.

쿼리 실행 비용 (Running Queries Costs)

- 과금 기준: 대부분의 클라우드 서비스 제공자는 쿼리 실행 후 반환되는 양이 아니라, 처리되는 데이터의 양을 기준으로 요금을 부과함.
- **Google BigQuery**의 두 가지 요금 모델:
 - 온디맨드 가격 책정 (**On-demand pricing**):
 - 각 쿼리가 처리하는 바이트 단위의 데이터 양을 기준으로 함.
 - 컴퓨팅 및 스토리지 요구 사항이 비즈니스 우선순위에 따라 변동하는 사용자에게 이상적임.
 - 용량 가격 책정 (**Capacity pricing**):
 - 시간에 걸쳐 쿼리 실행에 사용된 컴퓨팅 파워를 기준으로 함.
 - '슬롯(slots)'이라 불리는 가상 중앙 처리 장치로 측정됨.
 - 예측 가능하고 통제 가능한 비용을 원하는 사용자에게 이상적임.
- 클라우드 전문가의 역할: 이러한 지식을 활용하여 비용 효율적인 클라우드 리소스를 사용하여 최상의 성능과 가치를 얻고, 리소스, 프로세스 및 수익을 최적화하는 데 기여할 수 있음.

Cloud Data Analytics의 산업별 영향

- 배경: 과거에는 의료 기록, 재고 관리 등이 수작업으로 이루어져 정보 접근에 한계가 있었음. 오늘날 **Cloud data analytics**는 여러 산업을 혁신하고 있음.
- 클라우드 전문가의 역할: 클라우드 전문가는 모든 종류의 비즈니스가 다음 큰 트렌드를 예측하고, 혁신을 이끄는 패턴을 발견하며, 운영, 시스템, 고객 만족도를 개선하는 신속한 결정을 내리도록 도울 수 있음.

주요 산업별 영향

- 의료 (**Healthcare**):
 - 개인 맞춤형 의학 및 예측 분석: 환자의 결과를 크게 개선하고 더 건강한 삶을 돕는 데 기여함.
 - 대규모 데이터 분석: 여러 환자 데이터에 대한 분석을 통해 의료 제품의 장기적인 효과나 처방 빈도 추세 등을 파악할 수 있음.
- 제조업 (**Manufacturing**):
 - 도전 과제 해결: 예측 불가능한 수요 변동 및 공급망 중단과 같은 과제를 해결함.

- 실시간 분석: 제조 공정 및 고객 상호 작용에서 수집된 대량의 데이터를 실시간으로 분석하여 운영의 안정성과 효율성을 최적화함.
 - 스마트 기술 활용: 스마트 기술을 통해 품질을 검사하고, 수리 필요성을 식별하며, 공급망 중단을 추적하고 예방할 수 있음.
 - 고객 중심 제품: 고객 데이터를 기반으로 제품을 생산하여 효율성을 높이고 매출을 증대함.
 - **교육 (Education):**
 - 학생 역량 강화: 학습자가 직업 준비 및 실제 업무 시나리오에 필요한 지식과 기술을 갖추도록 함.
 - 과정 최적화: 학생 선호도, 성과 데이터, 등록 패턴을 분석하여 과정 제공을 최적화함.
 - 개인화된 학습 경험: 학생이 콘텐츠와 상호 작용한 데이터를 기반으로 맞춤형 학습 경험을 설계할 수 있도록 도움.
 - **운송 (Transportation):**
 - 효율적인 경로 설계: 실시간 클라우드 데이터 분석을 통해 시간, 돈, 리소스를 절약하는 효율적인 경로를 설계함.
 - 운영 최적화: 지연 가능성을 예측하고, 운영을 최적화하며, 신뢰성을 높이고, 고객 서비스를 개선함.
 - 차량 유지보수: 제조와 마찬가지로, 차량에 설치된 기술이 운영 상태를 모니터링하고 유지보수 알림을 제공함.
 - **AI 및 ML 활용:** AI 및 ML 기능을 사용하여 과거 여행 데이터를 분석하고 미래 여행 동향을 예측하는 데 사용함.
 - 물류 계획: 최신 가상 모델링 소프트웨어를 활용하여 지능형 물류 및 효율적인 경로 계획을 세울 수 있도록 함.
 - **결론: Cloud data analytics**는 언급된 네 가지 산업 외에도 거의 모든 분야에 큰 영향을 미치고 있으며, 그 영향력은 계속 커질 것임. 이로 인해 클라우드 전문가에 대한 수요가 매우 높음.
-

Cloud Architecture 개요

- **비유:** 다양한 상점들이 있는 거대한 쇼핑몰에 비유할 수 있음. 쇼핑객은 건물 유지보수나 보안에 대해 걱정할 필요 없이 필요한 서비스를 이용할 수 있음.
- **정의:** 클라우드를 구축하는 청사진. 다양한 구성 요소와 기술이 어떻게 배열되어 가상 네트워크를 통해 리소스를 공유하고 확장할 수 있는지 정의함.
- **주요 구성 요소:** 프론트엔드 플랫폼, 백엔드 플랫폼, 클라우드 기반 배포 모델, 네트워크.

주요 구성 요소 상세 설명

- **프론트엔드 플랫폼 (Frontend Platform)**
 - **역할:** 사용자가 상호 작용하는 아키텍처의 부분. 화면 디자인, 앱, 인터넷 네트워크 등을 포함.

- 예시: 모바일 폰에서 웹 브라우저를 열어 **Google Doc**을 편집할 때, 브라우저, 폰, **Google Docs** 앱이 모두 프론트엔드 구성 요소임.
- **백엔드 플랫폼 (Backend Platform)**
 - 역할: 클라우드 자체를 구성하는 구성 요소. 컴퓨팅 리소스, 스토리지, 보안 메커니즘 및 관리를 포함.
 - 주요 구성 요소:
 - 애플리케이션 (**Application**): 사용자가 프론트엔드에서 접근하고 백엔드에서 작업을 완료하는 소프트웨어.
 - 서비스 (**Service**): 클라우드 아키텍처의 핵심. 컴퓨팅 시스템에서 발생하는 모든 작업을 처리하고 사용자가 접근하는 리소스를 관리함.
 - 런타임 클라우드 (**Runtime cloud**): 모든 클라우드 서비스가 효율적으로 작동할 수 있는 공간을 제공. 노트북이나 휴대폰의 운영 체제와 유사하며, 클라우드 서비스가 원활하게 실행되도록 보장함.
 - 스토리지 (**Storage**): 클라우드 서비스 제공업체가 시스템을 실행하기 위해 데이터를 보관하는 곳. 방대한 양의 데이터를 보관하고 정리하기 위해 유연하고 확장 가능한 스토리지 제공함.
 - 인프라 (**Infrastructure**): 클라우드가 작동하도록 하는 중요한 하드웨어(CPU, GPU, 네트워크 장치 등)로 구성됨.
 - 보안 (**Security**): 데이터 및 네트워크에 대한 보안 계획 및 설계. 시스템에 대한 중요한 감독을 제공하고, 데이터 손실을 방지하며, 가동 중단 시간을 피함.

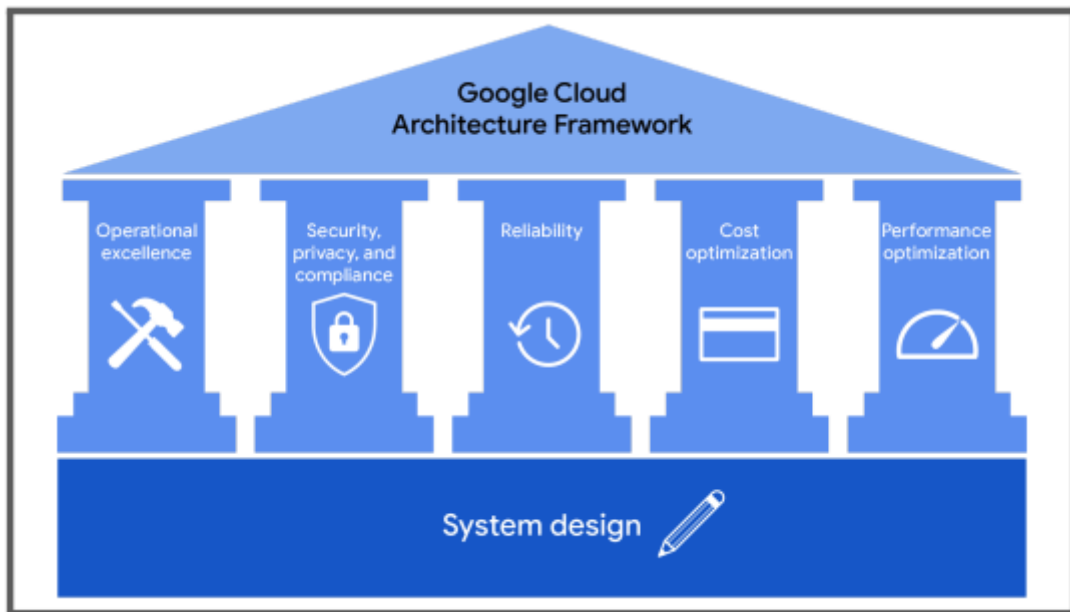
결론

- 클라우드 아키텍처는 이 모든 구성 요소가 어떻게 고도로 민첩하고 확장 가능한 솔루션을 제공할 수 있는지를 보여주는 청사진임.
 - 클라우드 아키텍처를 이해하는 것은 숙련된 클라우드 전문가가 되기 위한 중요한 단계임.
-

Google Cloud Architecture Framework

- 배경:
 - 클라우드 아키텍처는 다양한 구성 요소와 기술이 가상 네트워크를 통해 리소스를 모으고, 공유하며, 확장할 수 있도록 배열하는 방식
 - 클라우드 데이터 분석가로서 클라우드 시스템을 생성하고 운영하는 방법을 이해하는 것은 매우 중요한 업무의 일부
- 프레임워크의 역할:
 - **Google Cloud Architecture Framework**는 안전하고, 효율적이며, 탄력적이고, 고성능이며, 비용 효율적인 클라우드 시스템을 개발하고 사용하는 데 도움
 - 이 프레임워크는 클라우드 설계 및 배포 과정을 안내하는 역할

Google Cloud Architecture Framework의 6가지 기둥 (Six Pillars)



프레임워크의 6가지 기둥은 시스템 설계(System design)를 기반으로 하여 클라우드 작업에 대한 설계 고려 사항을 안내합니다.

- 시스템 설계 (**System design**):
 - Google Cloud Architecture Framework의 토대
 - 시스템을 설계할 때는 클라우드 시스템 요구사항을 충족하는 데 필요한 아키텍처, 구성 요소, 모듈, 인터페이스 및 데이터를 정의
- 운영 우수성 (**Operational excellence**): 클라우드 워크로드를 얼마나 효율적으로 배포, 운영, 모니터링 및 관리하는지를 결정
- 보안, 개인 정보 보호 및 규정 준수 (**Security, privacy, and compliance**): 클라우드 내 데이터의 보안을 보장하고, 정보의 프라이버시를 보호하며, 설계가 조직 표준에 부합하는 것을 포함
- 신뢰성 (**Reliability**): 시스템이 클라우드 내의 워크로드를 처리하도록 설계되었음을 의미
- 비용 최적화 (**Cost optimization**): 조직을 위한 클라우드 아키텍처에 대한 비즈니스 투자를 극대화하는 것을 의미
- 성능 최적화 (**Performance optimization**): 최적의 성능을 촉진하기 위해 클라우드 리소스를 지속적으로 개선하는 것을 의미

핵심 요약

- 클라우드 데이터 분석가로서 클라우드 솔루션을 설계, 구현, 사용하는 것이 업무의 큰 부분을 차지
- 설계 프레임워크를 갖는 것은 시스템을 이해하고 작업을 관리 가능한 우선순위로 나누는 데 도움
- 다양한 클라우드 아키텍처에 대한 이해를 구조화하는 데 유용하게 사용

Cloud Cost Optimization 개요

- 정의: 비용 절감 전략을 구현하여 클라우드 관련 비용을 줄이는 과정.
 - 단순히 돈을 절약하는 것을 넘어, 워크로드 효율성을 높임
 - 클라우드 리소스가 사용자에게 최대 가치를 제공하도록 보장함.
- 주요 이점:
 - 비용 가시성 (**Cost visibility**): 조직이 돈을 어디에 쓰는지, 특정 클라우드 서비스에 대한 비용이 어떻게 청구되는지 정확히 알 수 있음.
 - 애플리케이션 성능 향상: 클라우드 리소스를 최적화하여 앱이 원활하게 실행되도록 보장함으로써 사용자 경험을 개선하고 클라우드 비용을 절감함.
 - 탄소 배출량 감소: 클라우드 리소스를 최적화함으로써 조직의 탄소 발자국을 줄이는 데 중요한 역할을 함.

Cloud Cost Optimization 전략

- **Rightsizing**: 애플리케이션이나 워크로드의 정확한 요구사항에 맞게 처리 능력 및 스토리지와 같은 컴퓨팅 리소스를 조정하는 과정. 이를 통해 사용량을 최적화함.
- **Auto-scaling**: 애플리케이션을 모니터링하고, 사용자 요구를 충족하는 데 필요한 컴퓨팅 리소스에 따라 자동으로 확장하거나 축소하는 클라우드 서비스.
- **Reserved instances**: 조직이 특정 기간 동안 특정 양의 리소스를 구매하고, 이 약정에 대한 대가로 할인을 받는 클라우드 결제 모델. 장기적으로 사용량 변동이 크지 않을 것으로 예상될 때 유용함.

결론

- 이러한 비용 최적화 전략들은 비용을 절감하고, 앱 효율성을 개선하며, 낭비를 줄이고, 환경적 책임을 높이며, 클라우드 투자에 대한 더 나은 수익을 제공함.
- 이러한 지식은 유능한 클라우드 데이터 분석가로서 조직에 가치를 제공하는 훌륭한 방법임.

클라우드 비용 최적화 팁

- 클라우드 비용 최적화의 의미: 비용 절감 전략을 실행하여 클라우드 비용을 줄이는 과정.
 - 돈을 절약하는 것 외에도, 비용 가시성, 애플리케이션 성능 향상, 탄소 배출량 감소와 같은 중요한 이점들을 제공함.
 - 클라우드 데이터 전문가로서 비용 최적화는 조직에 가장 큰 이점을 제공하는 데 도움이 될 것임.

비용 최적화를 위한 5가지 모범 사례

- 사용하지 않는 리소스 삭제 및 유휴 시스템 통합
: 더 이상 필요하지 않은 임시 기능이나 데이터 스토리지를 계속해서 비용 지불하는 경우가 흔함.
 - 따라서 사용하지 않는 리소스를 확인하고 중지 및 삭제하는 것이 좋음.
 - 일부 클라우드 서비스는 자동 종료 기능을 제공하여 사용하지 않는 리소스에 대한 비용 지불을 피할 수 있음.

- 시스템 **Rightsizing**:
 - 처리 능력 및 스토리지와 같은 컴퓨팅 리소스를 애플리케이션 또는 워크로드의 정확한 요구 사항에 맞게 조정하는 과정.
 - 이를 통해 사용량을 최적화하고 불필요한 컴퓨팅 리소스에 대한 비용 지출을 피할 수 있음.
- 컴퓨팅 요구 사항 **Autoscaling**:
 - 사용자 수요에 맞춰 필요한 컴퓨팅 리소스에 따라 자동으로 확장하거나 축소하는 클라우드 서비스.
 - 비용이 많이 드는 애플리케이션을 지속적으로 실행하는 것보다 필요에 따라 확장하여 비용을 최적화하는 것이 더 나은 방법.
- 히트맵 사용:
 - 시스템 사용량의 최고점과 최저점을 시각화하여 보여주는 도구임.
 - 히트맵을 사용하여 어떤 시스템이 가장 자주 사용되는지 파악하고 이를 최적화하는 데 활용할 수 있음.
- 단일 클라우드 vs. 다중 클라우드 환경:
 - 다중 클라우드 시스템은 가용성과 가동 시간을 높일 수 있지만, 비용 최적화 측면에서 단점이 있을 수 있음.
 - 다중 클라우드 환경은 사용자에게 플랫폼 간 전환을 강요하여 리소스를 소모하고, 클라우드 간 네트워크 트래픽 지원 비용을 증가시킴.
 - 따라서 조직의 필요에 따라 단일 클라우드 환경이 더 나은 선택일 수 있음.

핵심 정리

- 클라우드 환경은 빅데이터 요구가 있는 조직에게 비용 효율적인 솔루션이 될 수 있음.
- 클라우드 데이터 전문가로서, 조직이 클라우드 비용을 최적화하고 시스템으로부터 가능한 한 최고의 가치를 얻도록 보장하는 것이 업무의 일부임.

Cloud Data Analytics 학습 내용 요약

- 클라우드 마이그레이션 및 시스템 환경 선택: 클라우드로 이전하는 과정과 올바른 데이터 시스템 환경을 선택하는 것이 중요함을 학습함.
- 클라우드의 영향: 클라우드가 데이터 분석 분야 및 다양한 산업에 미치는 영향에 대해 알아봄.
- 클라우드 비용 및 모델: 비용 및 청구 모델을 검토하고, 클라우드 비용 최적화와 그 이점에 대해 탐구함. 스토리지, 쿼리 실행, 리소스 프로비저닝, 다양한 청구 모델의 비용을 이해하게 됨.
- 클라우드 아키텍처: 클라우드 아키텍처가 클라우드에서 가능한 모든 것을 위한 청사진임을 개괄적으로 학습함.

이러한 지식은 고용주가 가치 있는 비용 절감을 달성하도록 돕는 데 기여할 것임.

Module 2 용어집

- **Advanced provisioning (고급 프로비저닝):** 사용자가 클라우드 서비스 제공업체와 공식 계약을 체결하고, 고정 가격을 지불하거나 매월 청구되는 클라우드 제공 모델.
- **Autoscaling (자동 확장):** 애플리케이션을 모니터링하여 사용자 수요에 따라 필요한 컴퓨팅 리소스에 맞게 자동으로 확장하거나 축소하는 클라우드 서비스.
- **Backend platform (백엔드 플랫폼):** 컴퓨팅 리소스, 스토리지, 보안 메커니즘 및 관리 등 클라우드를 구성하는 구성 요소들.
- **Cloud cost optimization (클라우드 비용 최적화):** 비용 절감 전략을 실행하여 클라우드 비용을 줄이는 과정.
- **Data ingestion (데이터 수집):** 추후 사용 또는 저장을 위해 데이터를 획득, 가져오기 및 처리하는 과정.
- **Data integration (데이터 통합):** 여러 출처의 데이터를 단일하고 사용 가능한 데이터 소스로 결합하는 과정.
- **Data processing (데이터 처리):** 원시 데이터를 정리, 정리 및 분석하기 쉬운 형식으로 변경하는 과정.
- **Data storage (데이터 스토리지):** '버킷(buckets)'이라고 불리는 저장 공간에 보관되는 데이터의 양.
- **Dynamic provisioning (동적 프로비저닝):** 사용자의 변화하는 요구에 따라 리소스를 조정하고, 사용한 만큼만 요금을 청구하는 클라우드 제공 모델.
- **Frontend platform (프론트엔드 플랫폼):** 사용자가 상호 작용하는 클라우드 아키텍처의 구성 요소.
- **Hybrid cloud (하이브리드 클라우드):** 공용 클라우드와 개인 클라우드 모델을 결합하여, 조직이 클라우드 서비스와 온프레미스 클라우드 모델의 제어 기능을 모두 활용할 수 있는 모델.
- **Network use (네트워크 사용):** 클라우드 제공업체가 사용자의 컴퓨팅 및 스토리지 리소스를 연결하는 네트워크에 대해 부과하는 요금.
- **Private cloud (개인 클라우드):** 모든 클라우드 리소스가 단일 사용자 또는 조직에 전용으로 제공되는 클라우드 모델.
- **Public cloud (공용 클라우드):** 인터넷을 통해 컴퓨팅, 스토리지 및 네트워크 리소스를 제공하여, 사용자가 특정 비즈니스 요구에 따라 온디맨드 리소스를 공유할 수 있게 하는 클라우드 모델.
- **Refactoring (리팩터링):** 기존 애플리케이션을 버리고 완전히 새로운 애플리케이션을 처음부터 구축하는 클라우드 마이그레이션 전략.
- **Rehosting or "lift and shift" (재호스팅 또는 "리프트 앤 시프트"):** 전체 온프레미스 시스템을 클라우드로 옮기는 클라우드 마이그레이션 전략.
- **Replatforming (재플랫폼화):** 온프레미스 시스템을 클라우드로 마이그레이션한 후 약간의 변경을 가하는 클라우드 마이그레이션 전략.
- **Repurchasing (재구매):** 애플리케이션을 새로운 클라우드 기반 서비스 플랫폼(주로 SaaS)으로 옮기는 클라우드 마이그레이션 전략.
- **Reserved instances (예약 인스턴스):** 조직이 특정 기간 동안 특정 양의 리소스를 구매하고, 그 대가로 할인을 받는 클라우드 결제 모델.
- **Resource provisioning (리소스 프로비저닝):** 사용자가 적절한 소프트웨어 및 하드웨어 리소스를 선택하면, 클라우드 서비스 제공업체가 이를 설정하고 사용하는 동안 관리하는 과정.
- **Retiring (폐기):** 더 이상 유용하지 않은 애플리케이션을 끄는 것.
- **Rightsizing (라이트사이징):** 애플리케이션 또는 워크로드의 정확한 요구사항에 맞게 컴퓨팅 리소스를 조정하는 과정.

- **Self-provisioning** (자체 프로비저닝): 사용자가 웹사이트나 온라인 포털을 통해 클라우드 제공업체로부터 리소스를 구매하면, 리소스가 빠르게 제공되는 클라우드 제공 모델.
-