

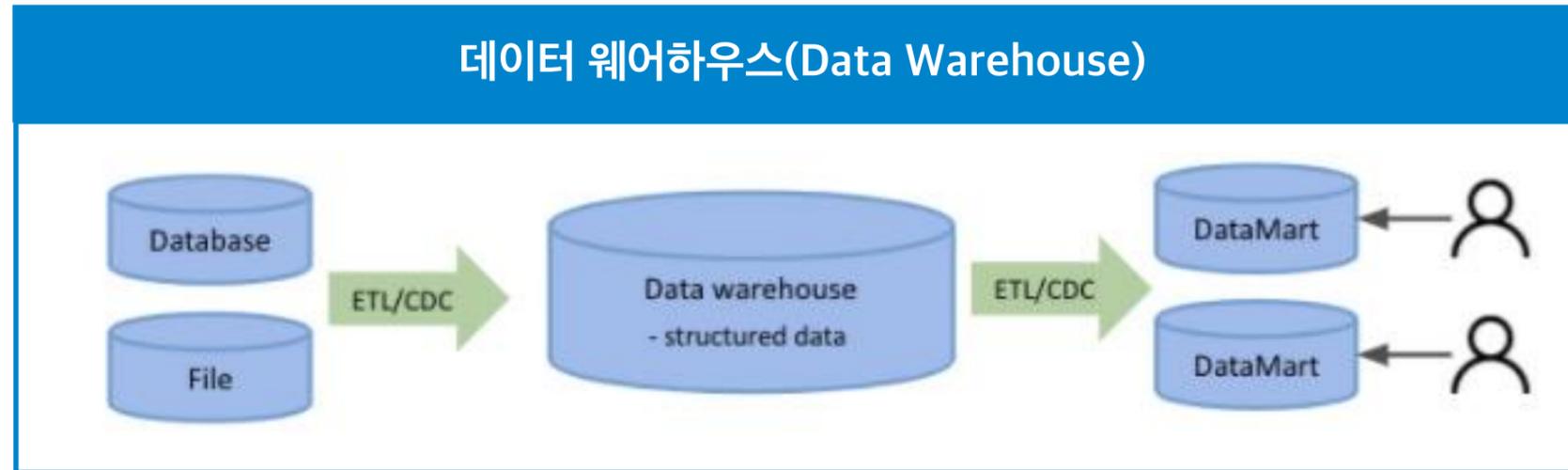
세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

기업IT 담당자 커뮤니티, 쉐어드IT

데이터 레이크를 지나서 요즘 키워드는
데이터 메쉬라고 하던데 데이터 메쉬의 핵심
기술은 뭐가 있을까요?

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 시스템 아키텍처 변천사



데이터 웨어하우스 개념

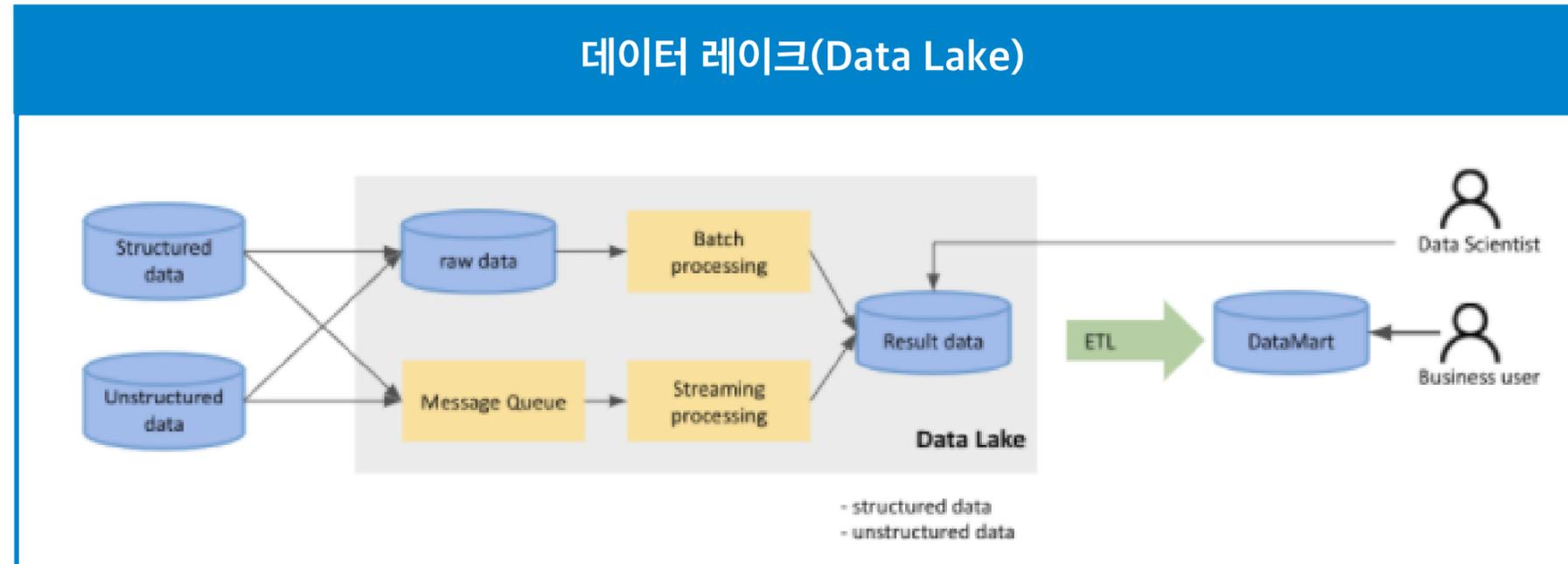
- RDBMS에서 데이터를 모아 분석하는 아키텍처
- OLTP에 저장된 데이터를 ETL/CDC 방식으로 수집 후 Data warehouse에 저장하고 분석
- 만약 현업에서 데이터 분석 요구가 많을 경우 그 부서에 필요한 데이터만 Data Mart라는 형태로 별도의 공간에 저장하며, 현업은 DM을 활용해 분석

데이터 웨어하우스 특징

- 테이블과 같은 구조화된 정형 데이터 분석에 유리
- 데이터 소스로부터 배치 형태로 주기적으로 데이터를 공급받음
- 데이터 신뢰도 높음
- 특정 벤더에서 제공되는 솔루션 기반으로 구축되며, RDBMS의 특성 상 스케일 아웃 형태의 확장에 한계가 있음
- 상용 소프트웨어 + 이를 지원하기 위한 인증된 고성능 하드웨어 필요 -> 가격이 비쌈 / Oracle Exadata

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 시스템 아키텍처 변천사



데이터 레이크 개념

- 데이터의 양이 늘어나고 유형이 다양해짐(정형 + 비정형 + 반정형)
- 이런 현상을 해결하기 위해 탄생한 것이 데이터 레이크
- 데이터 웨어하우스와는 다르게 비정형 데이터 저장과 분석을 지원하며, JSON과 같은 반정형 데이터도 지원함

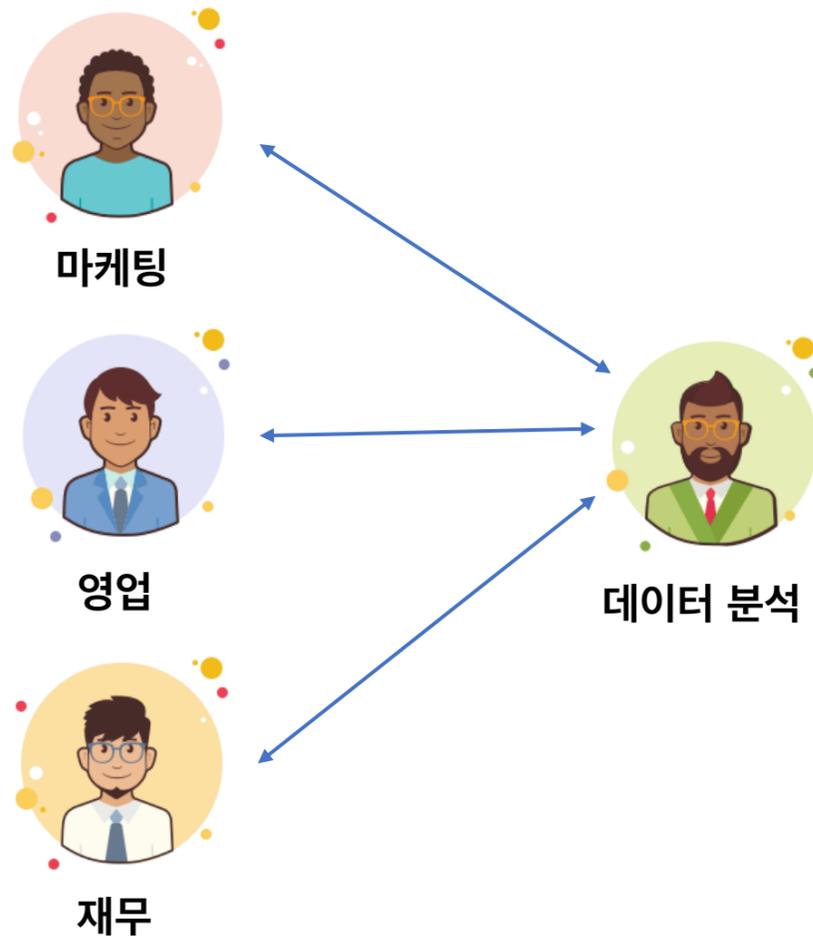
데이터 레이크 특징

- 테이블과 같은 구조화된 정형 데이터 분석에 유리
- 스트림이나 이벤트 로그와 같은 실시간 스트리밍 데이터에 대한 실시간 처리가 가능
- 이렇게 처리된 데이터를 데이터 저장소에 저장해서 분석하거나 DM으로 넘겨 사용자에게 제공
- 데이터 신뢰도 낮음
- Hadoop/Spark 기반으로 구축해 HDFS에 저장 후 Hadoop/Spark로 분석 / 실시간 스트리밍은 Kafka 활용

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 시스템 아키텍처 변천사

데이터 웨어하우스, 데이터 레이크의 문제점



문제점 1. 분석가의 데이터 도메인 지식 부족

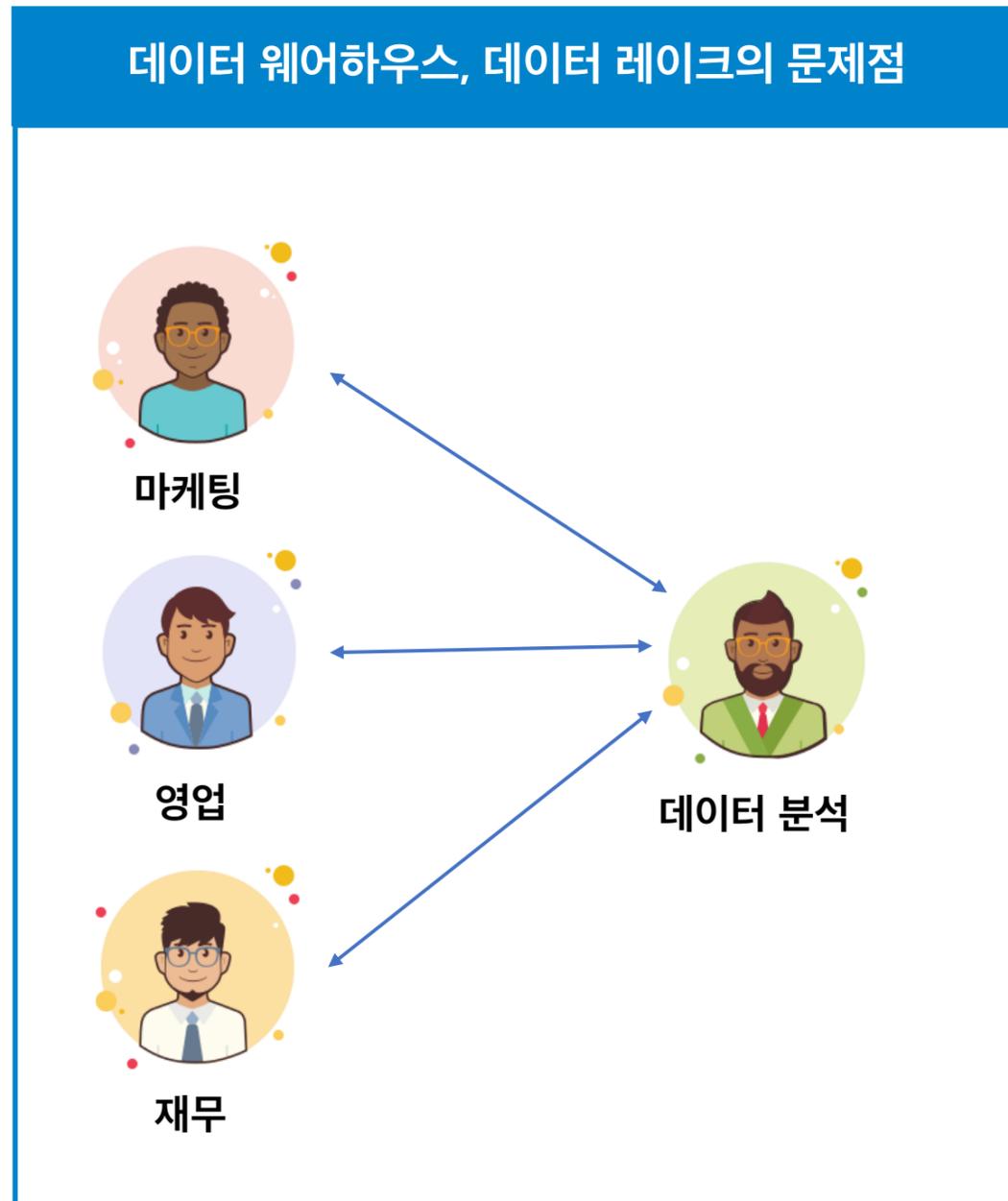
- 분석하려는 영역의 지식은 현업 담당자가 가장 잘 알고 있음
- 하지만 데이터 분석가는 주어진 데이터를 잘 분석하는 데에 집중, 이 데이터가 어떤 의미를 가진 데이터인지는 잘 모름
- 그래서 현업과 커뮤니케이션 시도하지만 분석가 한명이 여러 부서의 현업 담당자를 상대하다 보니 커뮤니케이션 지속성이 떨어짐
- 신속하고 깊이있는 분석이 어려움

문제점 2. 타 부서의 관심도 부족

- 현업과 데이터 분석팀은 서로 다른 부서임, 그래서 각자의 업무를 이해하기 어려움
- 회사에서 차세대 데이터 분석 시스템을 만들기 위해 데이터 분석팀에서 프로젝트를 수행하려 할 경우, 분석팀은 데이터 수집 + 분석 요구사항 수집에 집중
- 하지만 현업은 일단 자기가 속한 부서의 우선순위에 집중함
- 때문에 데이터 분석 성과가 나오기 어려운 구조

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 시스템 아키텍처 변천사



문제점 3. 서로 다른 데이터 분석 기술 활용

- 데이터 분석 시스템은 기본적으로 중앙 집중화 되어있음
- 따라서 단일 시스템이기 때문에 단일화된 분석 기술 활용
- 그런데 데이터 도메인 특성 상 분석에 유용한 솔루션이 따로 있는 경우가 있음(마케팅 분석이나 트래픽 분석에 특화된 솔루션이 있는 것 처럼)
- 이 경우 기존의 단일화된 기술 체계를 사용하는 분석 시스템이 이러한 도메인 별 특화된 기술을 수용하기 어려움

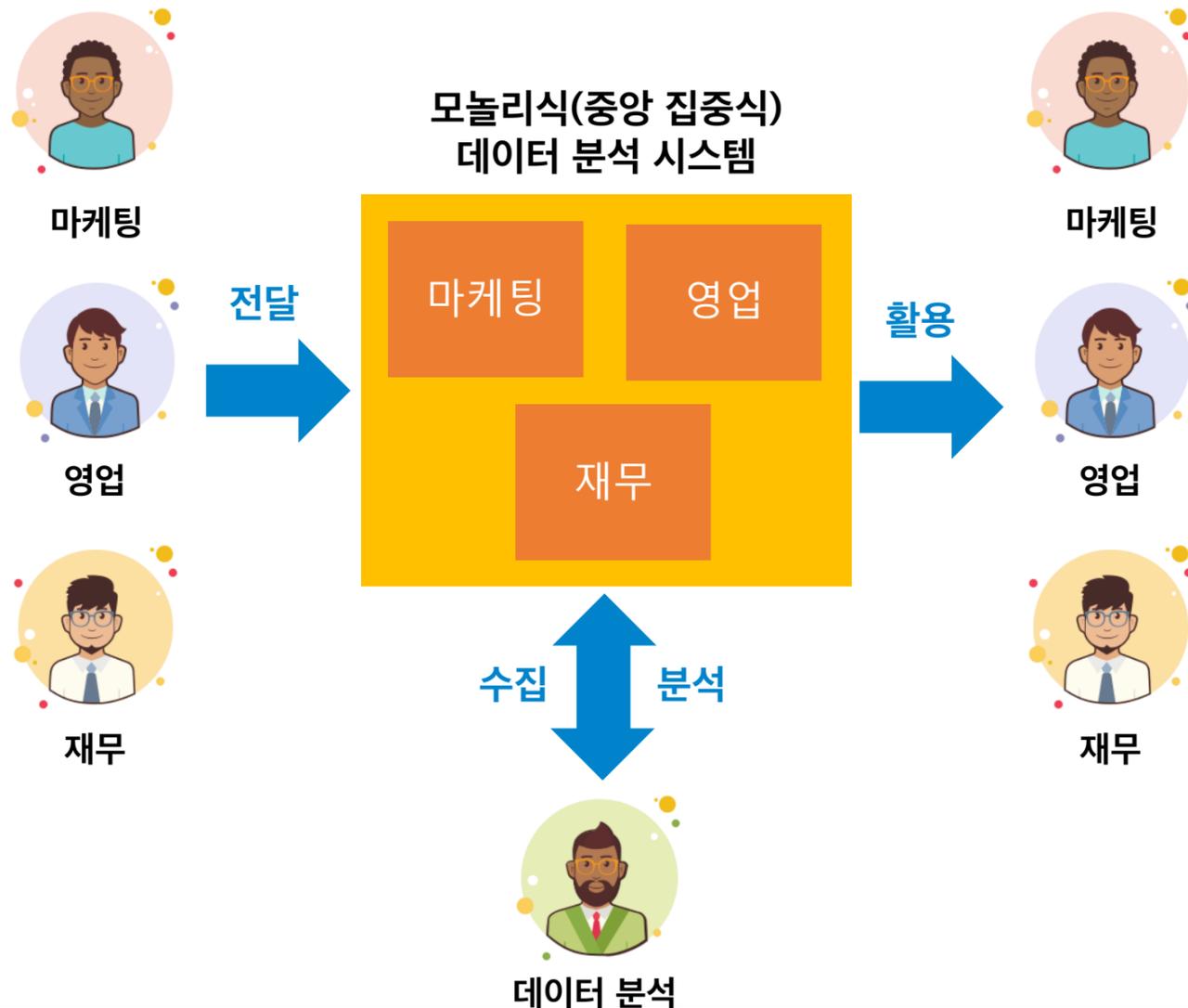
문제점 4. 예산 및 인력 부족

- 데이터 분석팀은 수익창출 부서가 아닌 비용집행 부서(연구조직에 가까움)
- 독자 수익을 내기 어려우며, 데이터 분석 시스템 구축을 위해 특정 데이터 도메인 영역의 현업 부서의 예산을 가져와 구축을 시도함 (외부 투자와 지원이 절실)
- 만약 해당 데이터 도메인 영역의 현업 부서 사정이 어려워질 경우 데이터 분석 장비와 인력에 투자하기 어려워짐 -> 데이터 분석 성과 저조

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 시스템 아키텍처 변천사

기존의 데이터 분석 시스템은 모놀리식 아키텍처



부서 별로 각기 다른 상황

- 현업에서는 이런 상황을 잘 인식하고 있기 때문에 중앙 집중식 데이터 분석 플랫폼이 있다 하더라도 부서 별로 분석 시스템을 따로 구축하려 함
- 예산이 충분하고 데이터 도메인 지식도 많을 경우 자신의 부서를 위한 분석 시스템 구축 시도
- 이러다 보니 현업 별로 개별적인 데이터 분석 플랫폼이 만들어지게 됨

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 시스템 아키텍처 변천사

데이터 메쉬(Data Mesh) : 분산형 데이터 분석 아키텍처



마이크로서비스(분산형)
데이터 분석 시스템

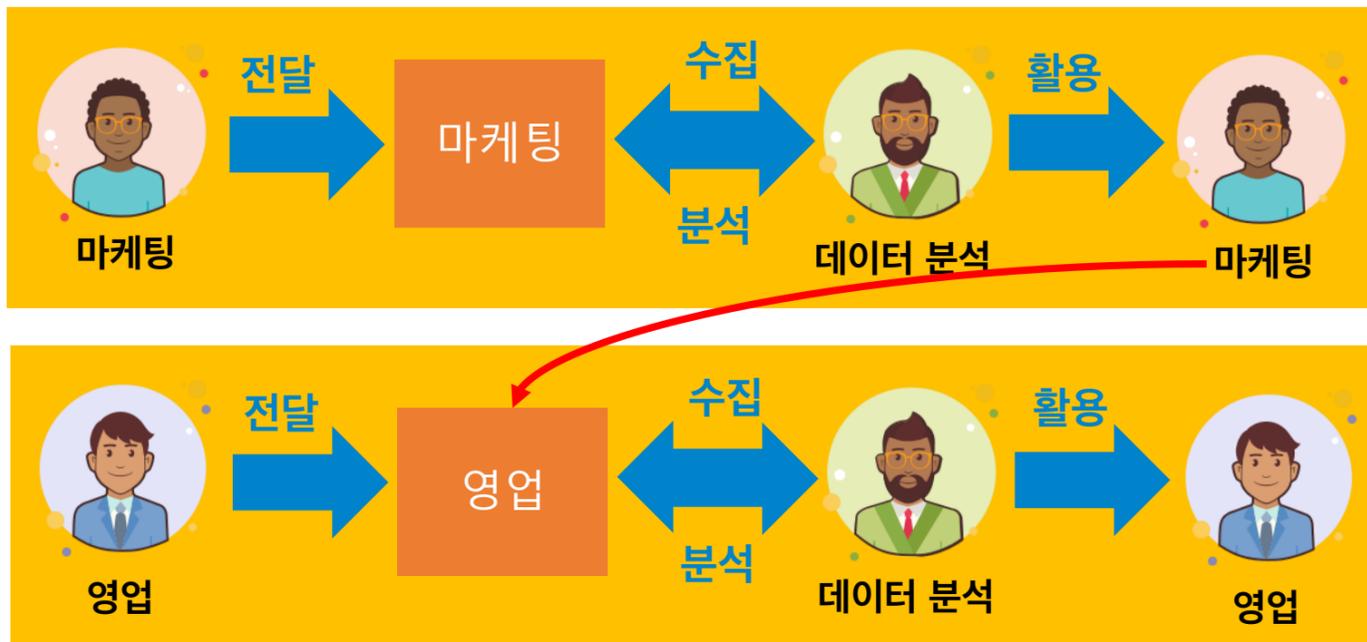
데이터 메쉬 개념 및 특징

- 부서별로 독립된 데이터 분석 시스템과 팀 + 현업 사용자를 하나로 묶어 요구 사항을 최대한 빠르게 반영하고 독립된 예산을 활용해 필요한 데이터 분석 시스템을 빠르게 구축하는 아키텍처
- 해당 데이터 도메인에 적합한 기술을 활용해서 기술적인 최적화도 가능
- 데이터 도메인별로 팀과 시스템을 나누는 방식을 통해 데이터에 대한 이해도와 분석 속도를 높이고자 하는 것을 목적으로 함

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 시스템 아키텍처 변천사

데이터 메쉬(Data Mesh) : 분산형 데이터 분석 아키텍처



마이크로서비스(분산형)
데이터 분석 시스템

데이터 메쉬에서는 타 부서 시스템의 데이터 조회가 어려움

- 데이터 도메인 별로 별도로 구축된, 전혀 다른 시스템이기 때문
- 타 부서 데이터 조회를 위해서는 해당 데이터 접근 권한 필요
- 타 부서 시스템의 경우 데이터 활용을 위한 분석 도구가 다를 수 있고, 도구의 UI도 달라서 타 부서 데이터에 대한 접근성이 떨어짐

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 시스템 아키텍처 변천사

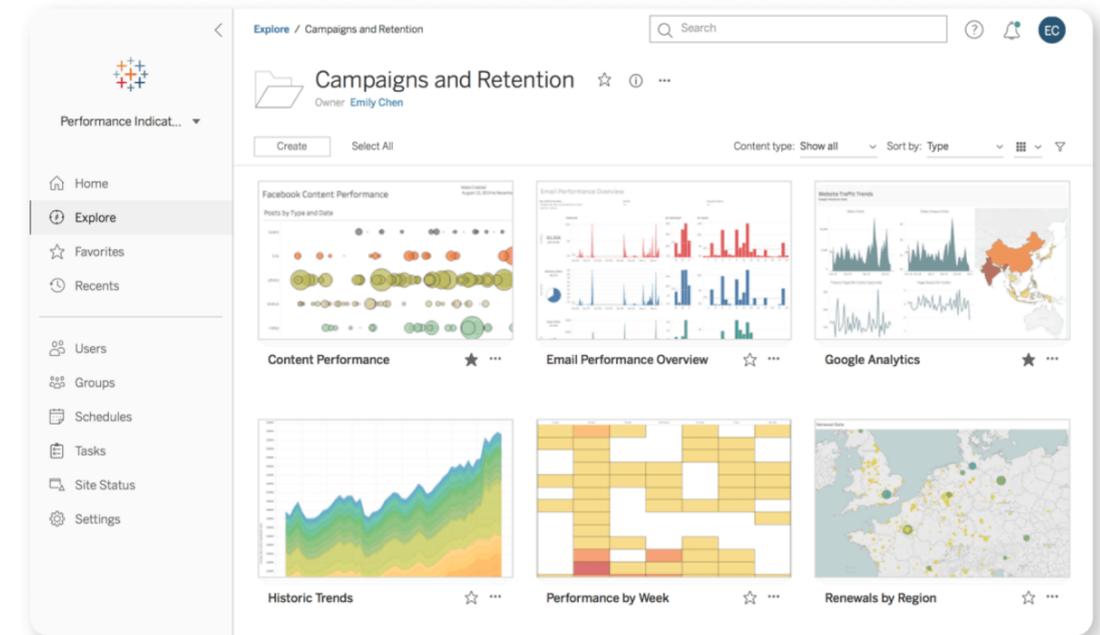
데이터 메쉬(Data Mesh) : 분산형 데이터 분석 아키텍처



마이크로서비스(분산형)
데이터 분석 시스템

타 부서 데이터 활용 방안 1. 통합 데이터 분석 도구 활용

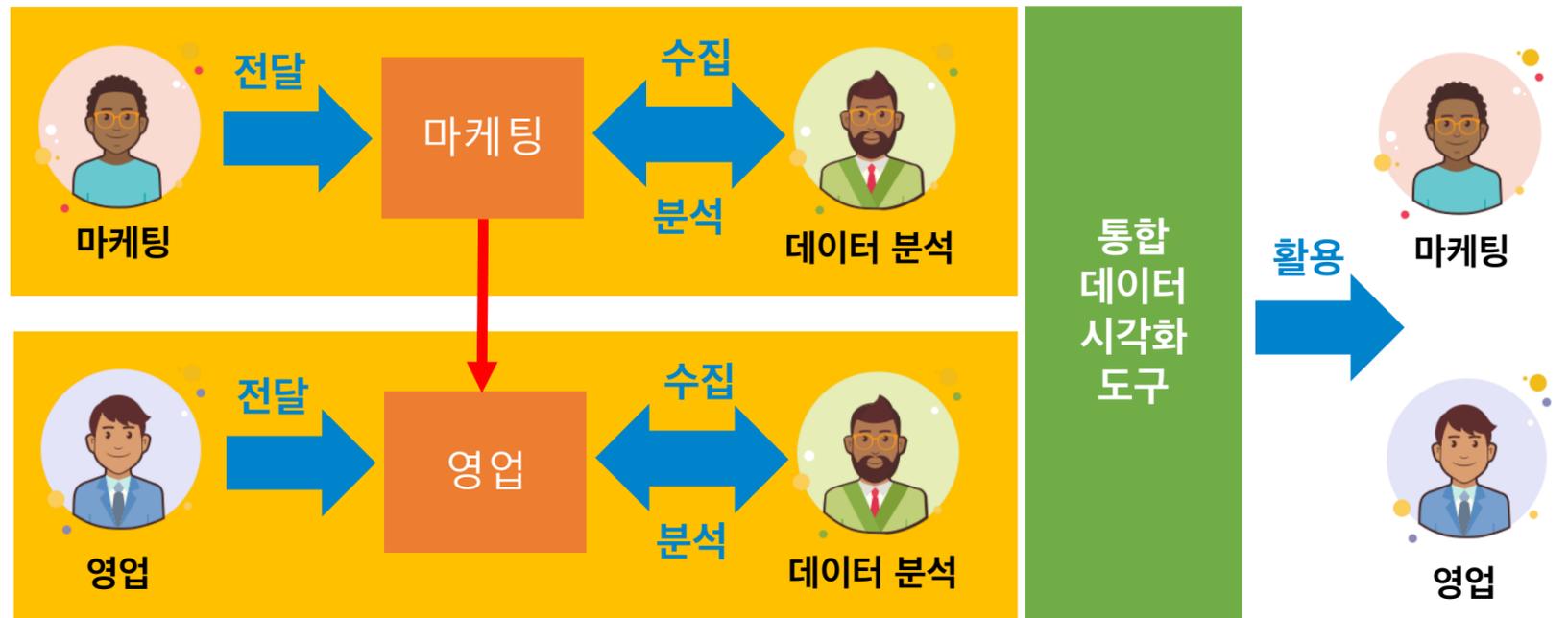
- 데이터 도메인 별로 구축된 분석 시스템은 각각 다르나, 이 데이터를 실제 분석해서 활용하기 위한 시각화 도구나 쿼리 도구는 통합된 하나의 도구를 사용해서 해결
- Tableau가 많이 사용됨



세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 시스템 아키텍처 변천사

데이터 메쉬(Data Mesh) : 분산형 데이터 분석 아키텍처



마이크로서비스(분산형)
데이터 분석 시스템

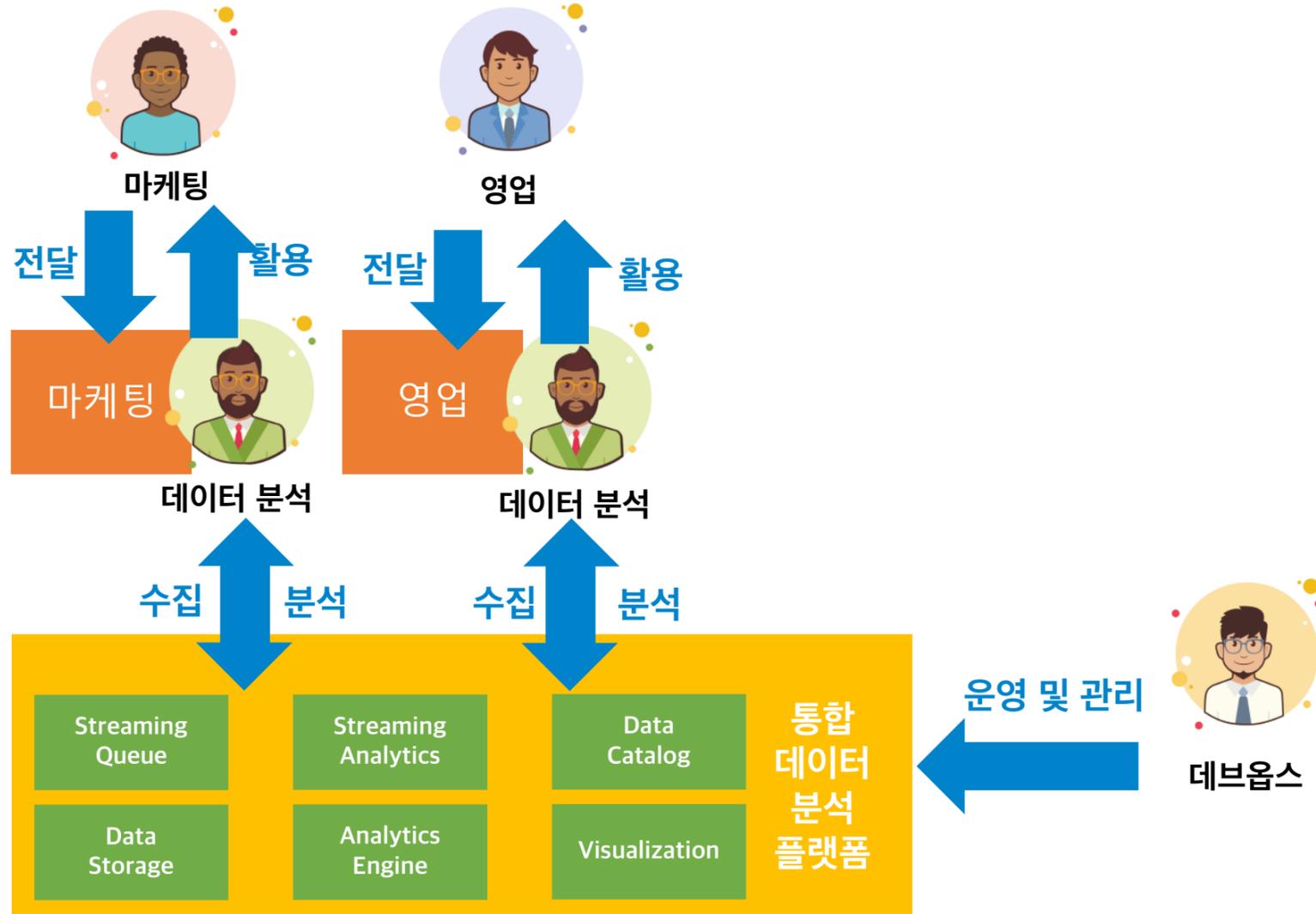
타 부서 데이터 활용 방안 2. 데이터 페더레이션

- 데이터 페더레이션 : 보유한 데이터셋을 타 부서의 데이터 분석 프로젝트로 공유해주어 서로 다른 시스템 간의 데이터를 서로 조회할 수 있는 기능
- 이것이 가능해야 서로 다른 도메인 데이터를 서로 연결해 보다 다각도로 분석할 수 있음
- 분석 사례 : 마케팅 + 영업 데이터를 연결해 마케팅 캠페인이 실적에 미치는 영향 분석 등

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 시스템 아키텍처 변천사

전사 통합 데이터 분석 플랫폼



데이터 메쉬를 위한 데브옵스의 필요성

- 데이터 메시 아키텍처를 활용해 부서 별로 각기 다른 데이터 분석 시스템을 빠르게 구축해야 함
- 하지만 데이터 분석 시스템 구축에 필요한 기반 기술, Hadoop, Kafka같은 소프트웨어를 설치하고 데이터 카탈로그도 구축하고 시각화 도구도 준비해야 하고, 이를 운영할 하드웨어 인프라도 준비해야 함
- 부서 별로 데이터 분석 시스템을 구축하기 위해 위와 같은 작업은 반드시 필요, 회사 전체로 보면 부서 별로 같은 작업을 반복하는 셈
- 따라서 데브옵스 팀에서 이러한 인프라와 솔루션은 플랫폼 형태로 구축해서 부서에 제공
- 각 부서는 이 플랫폼 위에서 셀프 서비스로 필요한 시스템을 구축하고 분석 로직을 만들어 분석 업무를 수행하는 형태로 발전해야 함

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터는 축적하고 있습니다. 보유하고 있는 데이터를 어떻게 분석해야 할지 모르겠습니다.

사용자가 뭘 해야할지 몰라도 데이터만 있으면 가치 있는 데이터를 만들 수 있을까요?

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 분석 인재를 뽑는게 좋은가요? 아니면 기존 IT인력을 교육시켜 진행하는게 좋은가요?

데이터 분석 관련해서 전문가를 고용해야할까요?
데이터사이언티스트라는 분들의 이력서가 꽤 오는데 고민입니다.

데이터 관리가 잘 되고 있다는 측정은 어떤 관점으로 할 수 있을까요?

경영진에게 어필 할 수 있는 활용 방법을 알고 싶습니다.

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

분석에서 딜레마는 향후 분석을 염두에 두고 당장 불필요해보여도 데이터를 미리 축적할 것인가 아니면 그때가서 축적을 시작할 것인가입니다.

분석관점이나 틀이 어떻게 바뀔지 모르다 보니, 지금까지 잘 분석하고 있었어도 관점이 바뀌어서 새 데이터가 필요하게 되면 미리 수집을 안하다 보니 충분한 데이터가 없는 경우가 많았습니다.

이에 대한 전문가의 의견이 궁금합니다.

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

다양한 데이터 관련 솔루션 중 가장 효과를 본 솔루션은 무엇인지?

해당 솔루션을 선택한 이유와 도입 이후 시행착오는 없었는지...

IT 담당자 관점에서 후회했던 case 살펴보기

대표적 성공 사례나 실패 사례(나 이거 했다가 3년 동안 죽을 뻔 했다)를 좀 길게 이야기 하는 것도 좋을듯 합니다.

세션 4. 데이터 분석 및 활용, IT 담당자에게 물어봤습니다!

데이터 관리를 잘 활용하려면 팁이나 노하우를 알려주시면 좋을 것 같습니다!

시간 관계 상 모든 질문을 다루지 못한 점
양해 부탁드립니다.

참여해 주셔서 감사합니다.

기업IT 담당자 커뮤니티, 웨어드IT