

모두를 위한, 그리고 당신을 위한 온라인 정리 서비스.



BIUM 비움



여러분의 드라이브 혹시 이런 상태이신가요?

 저장용량(85% 사용 중)

400GB 중 342.73GB 사용

 스팸메일함 64

 휴지통 999+

 비즈니스용 WORKS Drive

24.3GB / 30GB 18% 남음

MYBOX+ 업그레이드

☐ ☆ > AliExpress

☐ ☆ > Vercel 4 Failed pre

☐ ☆ > AliExpress (광고) 단 U

☐ ☆ > AliExpress (광고) 지금

☐ ☆ > AliExpress (광고) 상품

☐ ☆ > AliExpress (광고) 당신

☐ ☆ > AliExpress (광고) 핫 세

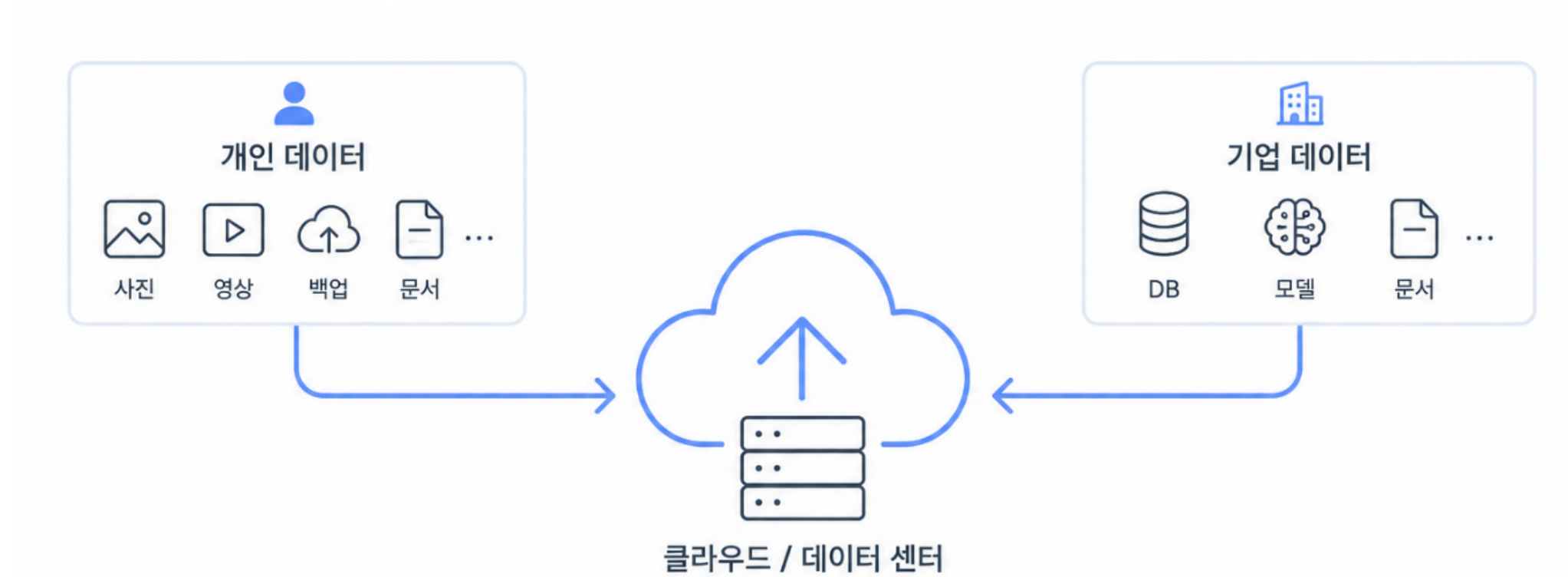
☐ ☆ > AliExpress (광고) 잠깐

400GB 중 85% 사용됨 

 저장용량(98% 사용 중)

15GB 중 14.79GB 사용

여러분만 그런것은 아닙니다. AI 도입 이후 클라우드 저장량은 더 늘어났습니다



개인과 기업이 생성·저장하는 데이터가 급격히 증가하면서,
사진·영상·문서·백업부터 기업의 서비스 데이터까지 클라우드에
축적되는 데이터의 규모가 지속적으로 커지고 있습니다.

하지만 이런 데이터를 저장하는 것도 비용이라는것 알고 계신가요?

2024년

415
TWh

데이터센터 전력 소비

연평균

+12%

매해 소비 전력 증가율

2030년

945
TWh

데이터센터 예상 전력 소비



탄소 526만 톤 !!

국제 에너지 보고서 (~2024)

전 세계 모든 국가가 수개월 동안 사용하는
전력량

데이터의 증가, 환경의 부담

DATA CENTERS IN TEXAS

Planned Texas data centers could emit more greenhouse gases than many countries

Texas regulators are allowing AI data centers across the state to start up thousands of new pollution sources with no public notice or environmental reviews.

종합적으로 볼 때, 플러드라이트가 텍사스에서 확인한 수천 개의 새로운 발전기는 매년 텍사스 지역 사회에 거의 2,500톤의 질소산화물을 배출하도록 허가받았는데, 이는 주에서 가장 최근에 건설된 석탄 화력 발전소의 배출량보다 세 배 이상 많은 양입니다. (질소산화물은 심각한 호흡기 질환과 조기 사망을 유발할 수 있는 매우 유독한 가스입니다.)

클린뷰에 따르면 스타게이트의 애빌린 캠퍼스를 포함해 텍사스주에는 데이터 센터와 연계된 가스 발전소가 최소 15곳 건설될 예정입니다. 플러드라이트가 검토한 허가 자료에 따르면 이 중 9곳은 연간 1억 3천만 톤 이상의 온실가스를 배출할 수 있습니다. 미국 환경보호청(EPA)의 계산기에 따르면 이는 석탄 화력 발전소 35곳의 연간 배출량과 맞먹는 양입니다.

우리가 저장하는 데이터, 정말 모두 필요할까요?

중복된 파일, 오래된 데이터, 잊힌 백업...
불필요한 데이터가 우리의 **저장 공간과 비용, 그리고 환경**을 차지하고 있습니다.



중복 파일

이름만 다를 뿐,
내용이 동일한 파일들

전체 데이터의
20~30%
중복 데이터



유사한 사진/영상

연속 촬영, 편집본, 캡처 등
거의 동일한 사진과 영상

전체 사진/영상의
30~40%
유사 이미지



오래된 데이터

1년 이상 열어보지 않은 파일,
더 이상 사용하지 않는 자료

기업 데이터의
60% 이상
비활성 데이터



불필요한 백업/첨부

자동 백업, 대용량 첨부파일 등
잊혀진 데이터들

메일함 용량의
10~20%
불필요 데이터

하지만, 사람들은 환경을 위해 파일을 지우지 않습니다

귀찮은게 먼저 다가옵니다

환경을 위해서라는 필요성도 와닿지 않지만, 정리의 번거로움이 행동을 가로막고 있습니다



결국, 방치

어피니티 다이어그램

인터뷰와 리서치를 통해 도출한 사용자 니즈를 그룹화했습니다.

1. 파일이 계속 쌓인다

과제, 프로젝트
파일이 많아
계속 쌓인다

게임, 앱 설치
파일이 쌓여
공간을 차지한다

팀/부서별로
데이터가 중복
저장된다

예전 파일을
지우기
망설여진다

다운로드, 캡처
파일이 너무 많다

임시 파일까지
모두 저장된다

2. 정리하기 귀찮고 어렵다

뭘 지워도 되는지
판단이 어렵다

수작업 확인에
시간이 너무
많이 든다

중복 파일을
일일이 찾기
힘들다

정리보다
새 작업이
우선이다

폴더 구조가
복잡해서
찾기 어렵다

버전 구분이
어려워
혼란스럽다

3. 비용과 리스크가 크다

클라우드 저장
비용이 부담된다

스토리지 비용이
계속 증가한다

불필요한 데이터로
예산 낭비가
발생한다

데이터 관리
정책이
필요하다

필요한 파일을
못 찾는 경우가
많다

중요 데이터가
중복으로
백업된다

4. 자동화 & 통합 관리 필요

자동으로 중복을
찾아줬으면
좋겠다

여러 기기/드라이브
를 한 번에
관리하고 싶다

안전하게 정리하고
복원할 수
있었으면 좋겠다

정책/규칙 기반
자동 정리가
필요하다

직관적인 UI로
쉽게 관리하고
싶다

권한 및 보안
관리 기능이
필요하다

사용자 관점 PERSONA 1



Persona1 — Digital-Native Student

이름: 기찬아

나이: 24세

직업: 컴퓨터공학부 3학년

특징: 개발 프로젝트·공모전·대외활동이 많음

디지털 사용량: 높음

“ 정리해야 한다는 건 알지만,
하나하나 확인하는 것이 너무 귀찮아
일단 저장해두고 나중에 미룸. ”

과제, 개발 프로젝트, 공모전, 대외활동을 하면서
중복 파일이 계속 쌓이지만, 제출 후
정리하는 과정이 귀찮아
매월 불필요한 cloud 비용을 지불함

사용자 관점 PERSONA 2



Persona2 — Game Data / Cloud Optimization Manager

이름: 김구름

나이: 38세

직책: Game Data & Cloud Infrastructure Manager

회사: Nexon

경력: IT 인프라 14년 · Cloud 8년

게임 Asset은 파일 하나가 수십 GB까지 커질 수 있고, 같은 Texture와 Mesh가 여러 버전으로 반복 저장된다. 데이터를 지우기는 어렵지만, 불필요한 저장 비용은 줄여야 한다.

사용자 관점 PERSONA 3



Persona — Cloud Storage Optimization Manager

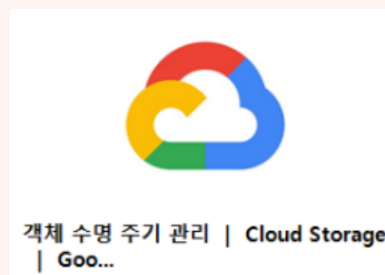
이름: Steve Jobs

나이: 39세

직책: Cloud Storage Manager

회사: Google

경력: IT 인프라 12년, Cloud 6년



“고객 데이터를 안전하게 보관하면서도,
필요 이상의 비싼 스토리지 비용이
발생하지 않도록 만들어야 하는 사람.”

How Might We?

사용자가 데이터의 가치를 일일이 판단하지 않아도,
필요한 데이터는 **안전하게 보존**하면서
불필요한 저장 공간은 줄일 수 있을까?



Solution



원터치로,
모든 기기의 데이터는
필요한 만큼만, 필요한 곳에.



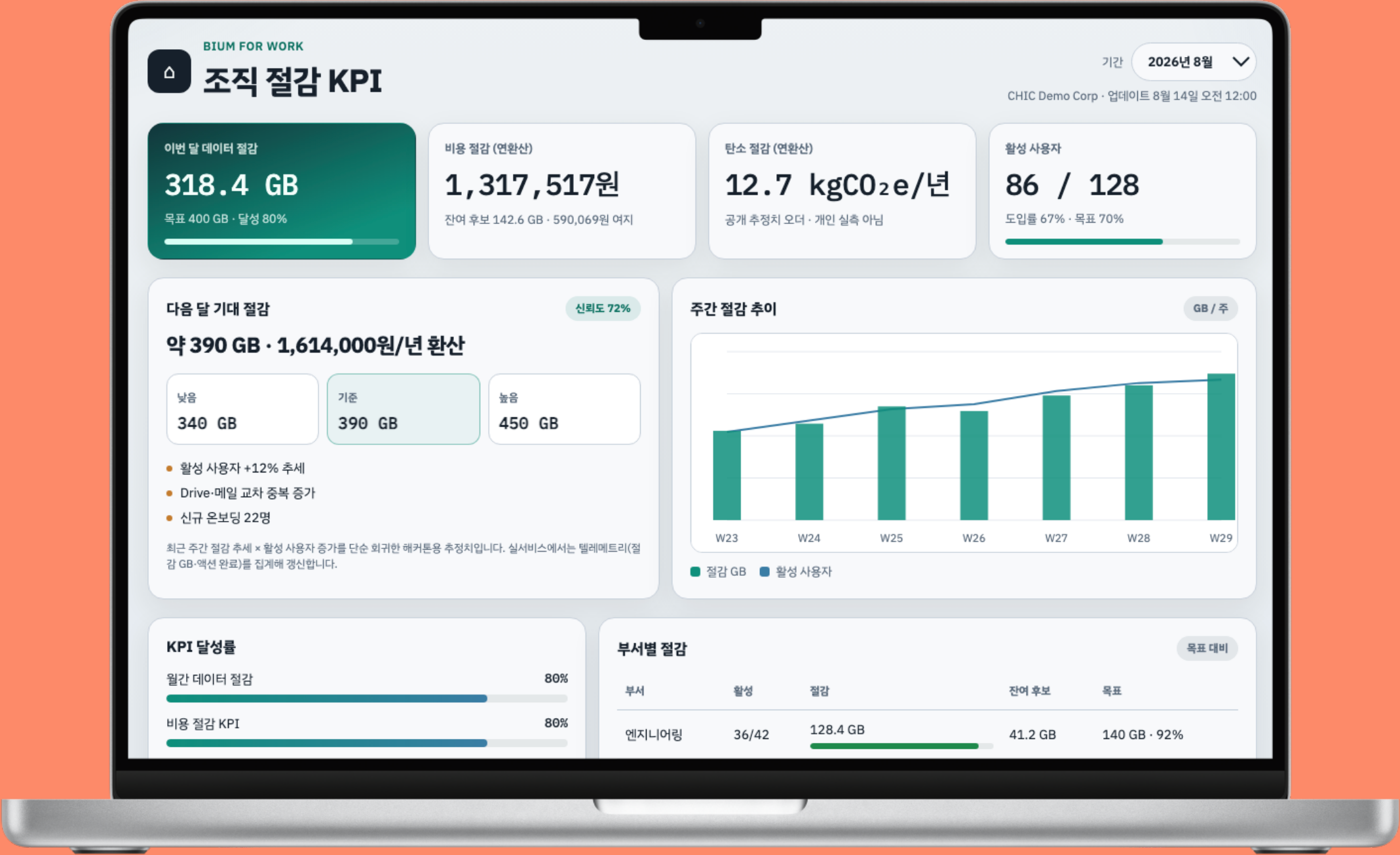


Solution



BIUM 비움

조직절감 KPI로, 절감 효과를 한눈에 파악





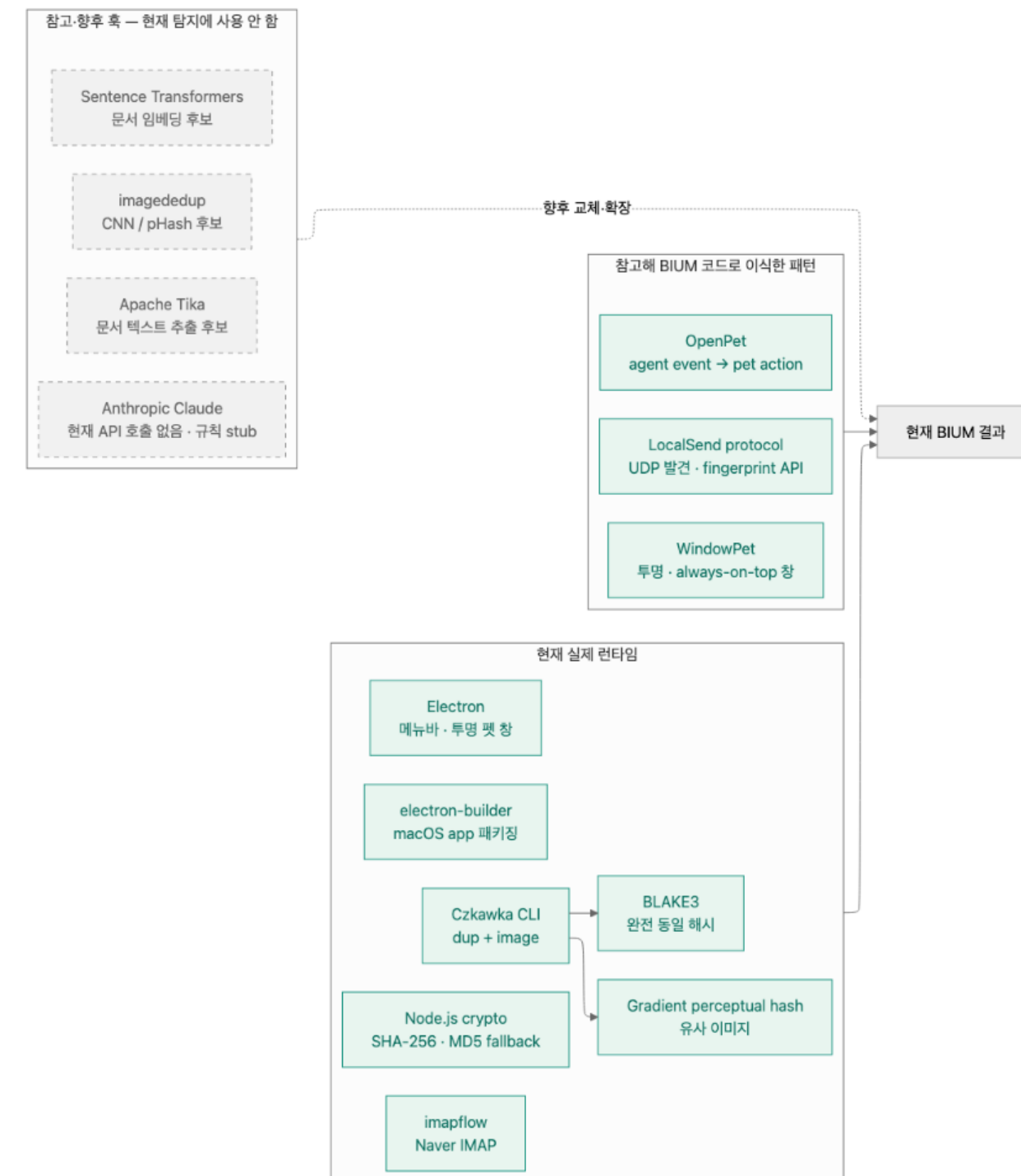
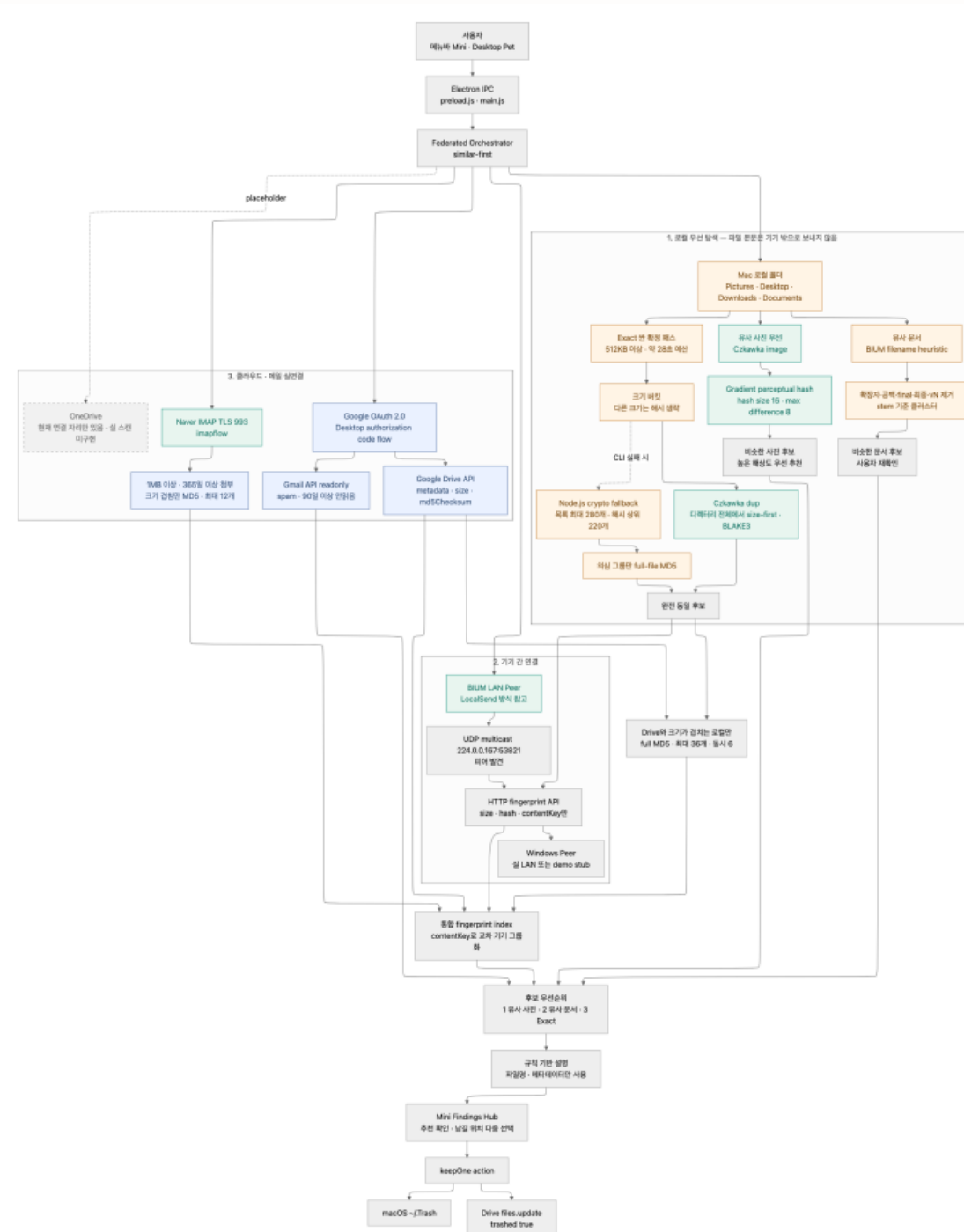
사용한 오픈 소스

오픈소스	라이선스	BIUM에서의 역할
Electron	MIT	macOS 메뉴바 앱 · Desktop Pet
electron-builder	MIT	.app 패키징
imapflow	MIT	네이버 메일 IMAP 연결
Czkawka	MIT	완전 중복 · 유사 이미지 탐지
imagededup	Apache-2.0	사진 Near-Duplicate 분석
Apache Tika	Apache-2.0	문서 텍스트 추출
Sentence Transformers	Apache-2.0	문서 의미 유사도 분석
LocalSend	MIT	LAN 기기 발견 · 연결 구조 참고
WindowPet	MIT	투명 · Always-on-top Pet 구조
OpenPet	-	Agent 이벤트 → Pet 행동 구조
crgimenes/neko	BSD-2-Clause	Neko 픽셀 스프라이트
oneko.js	-	픽셀 애니메이션 구조 참고
PawPal 계열	저장소별 상이	Golden Puppy 애니메이션





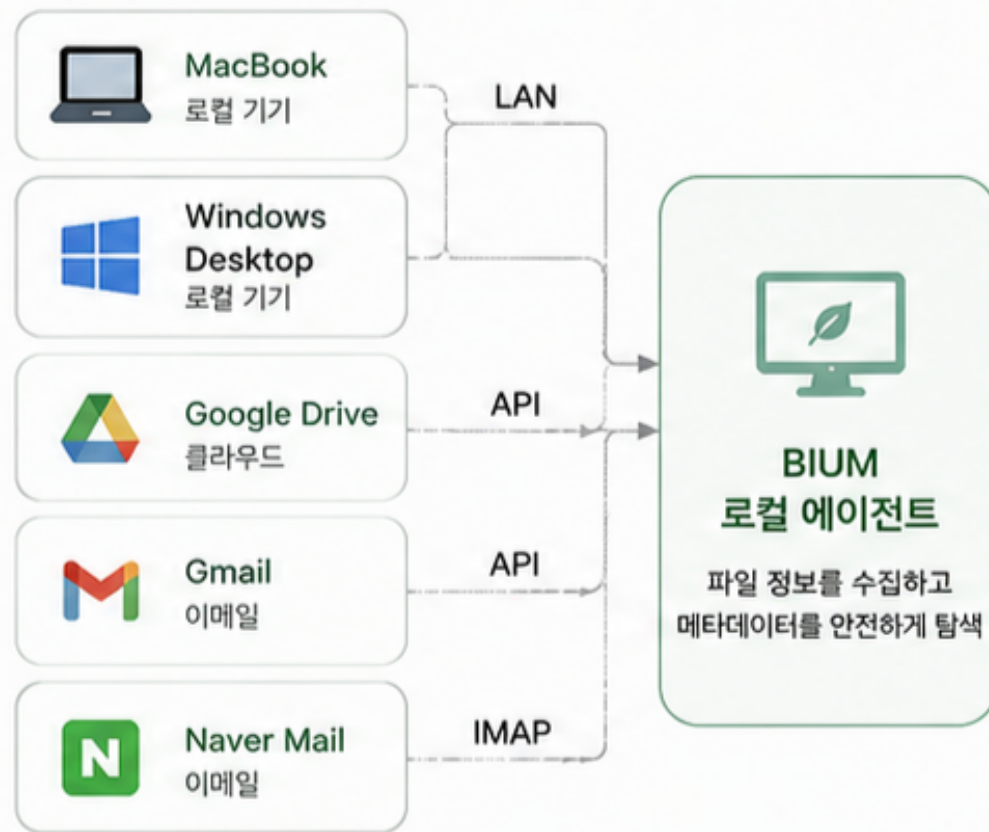
파이프라인



BIUM 비움 파이프라인

1 연결 (Connect)

여러 공간의 데이터를
안전하게 연결합니다.



2 탐색 (Scan)

파일 정보와 메타데이터를
탐색합니다.



3 발견 (Find)

중복·유사·불필요 데이터를
자동으로 발견합니다.



4 추천 (Recommend)

정리 후보를 모아
사용자에게 보여줍니다.

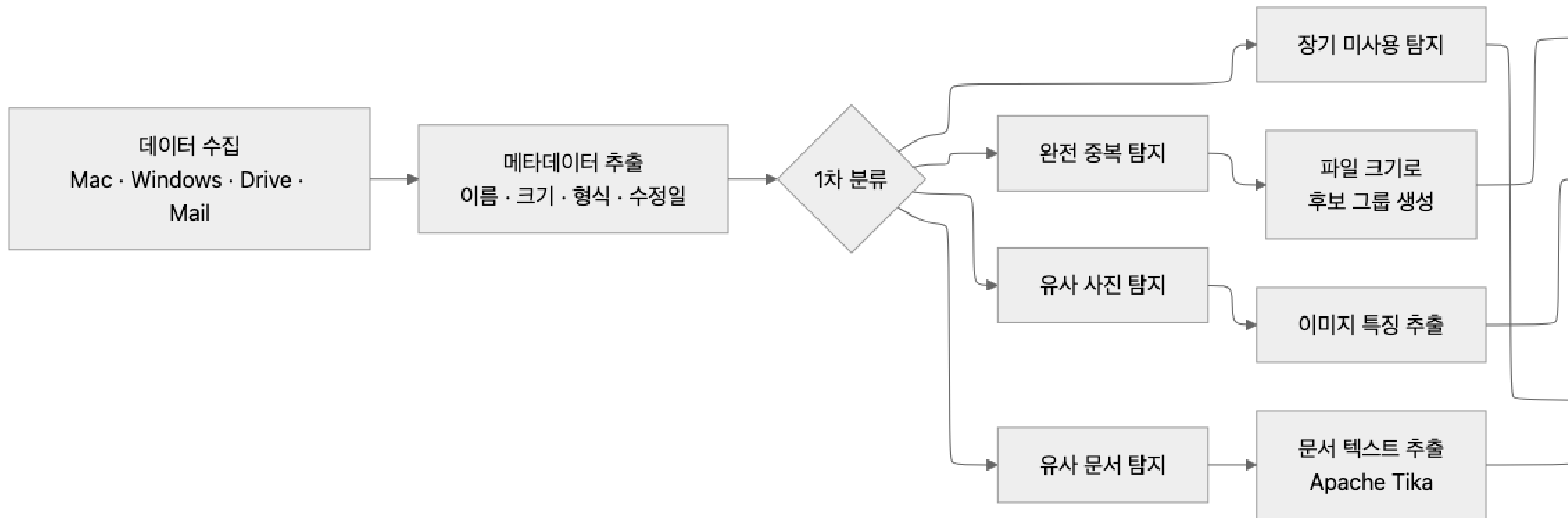


5 결정 (Action)

사용자가 최종 결정을 내리고
정리가 완료됩니다.

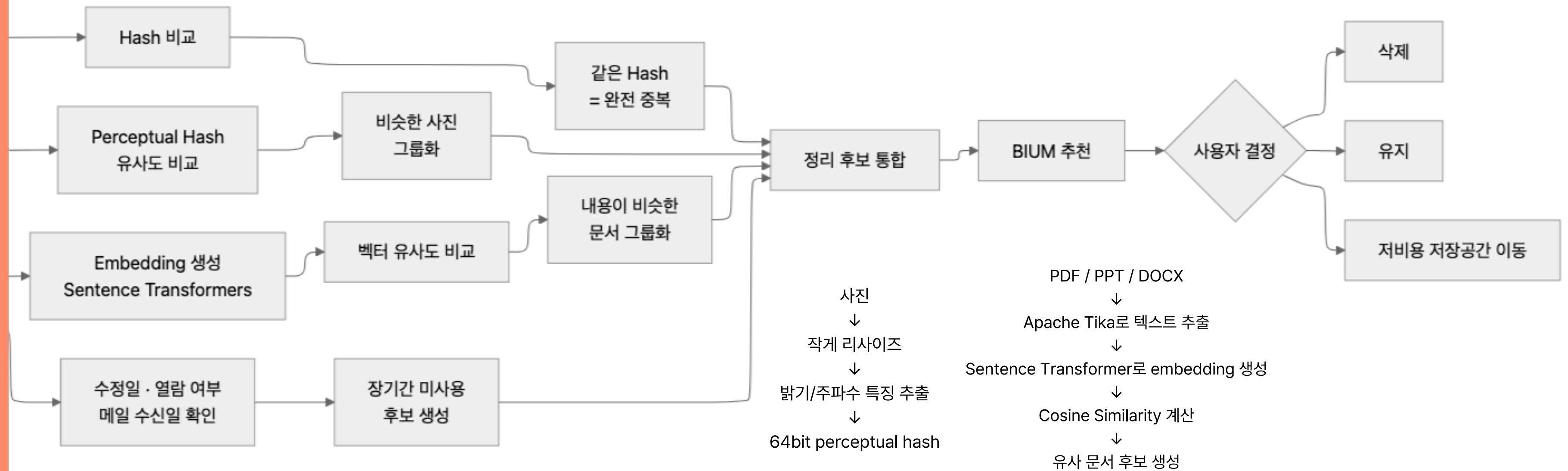


파이프라인





파이프라인



파일 전체 내용을 함수에 넣어서 고정 길이 지문(hash)을 만든다 → 완전 중복 = 파일의 디지털 지문(Hash)을 비교.
비슷한 사진 — perceptual hash 또는 이미지 임베딩 → 사진 유사도 = 사진의 시각적 지문을 만들고, 지문 간 거리를 비교
문서는 파일 자체가 아니라 내용의 의미를 숫자로 바꾼다 → 그다음 cosine similarity를 계산한다.



프로토타입 / 시연

메뉴 바 아이콘 클릭



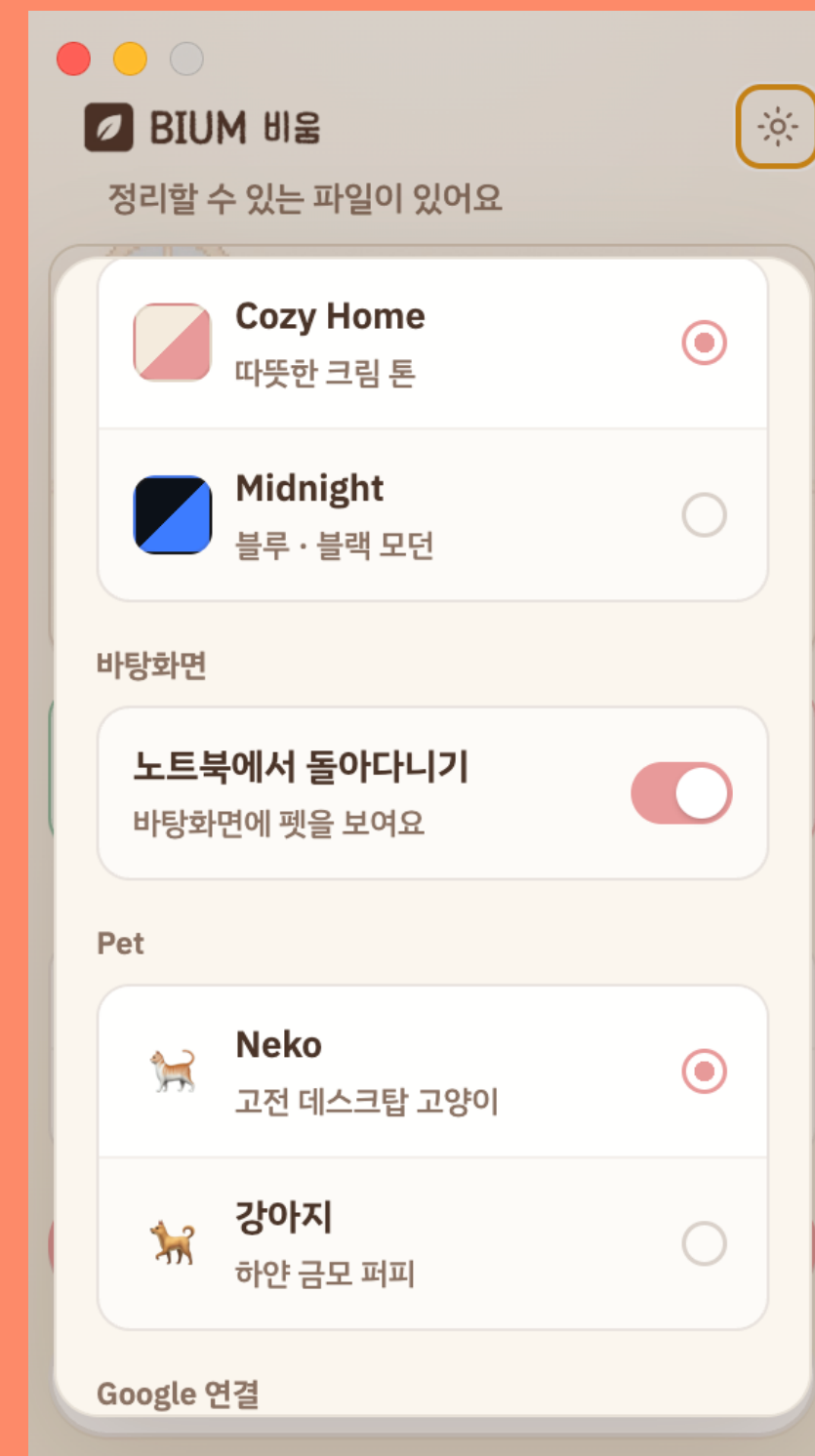
데이터 관리 펫 애니메이션



홈 화면

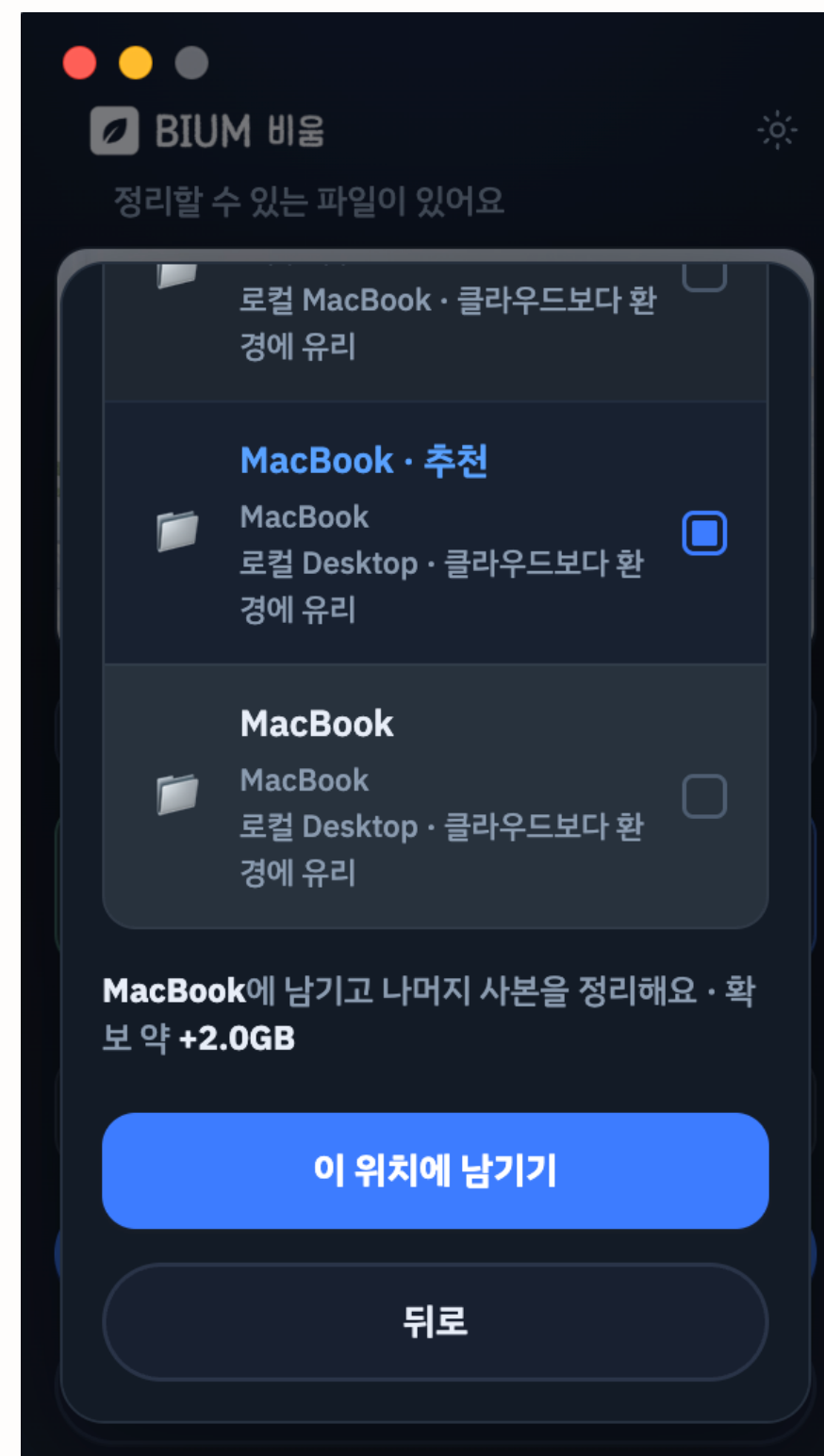
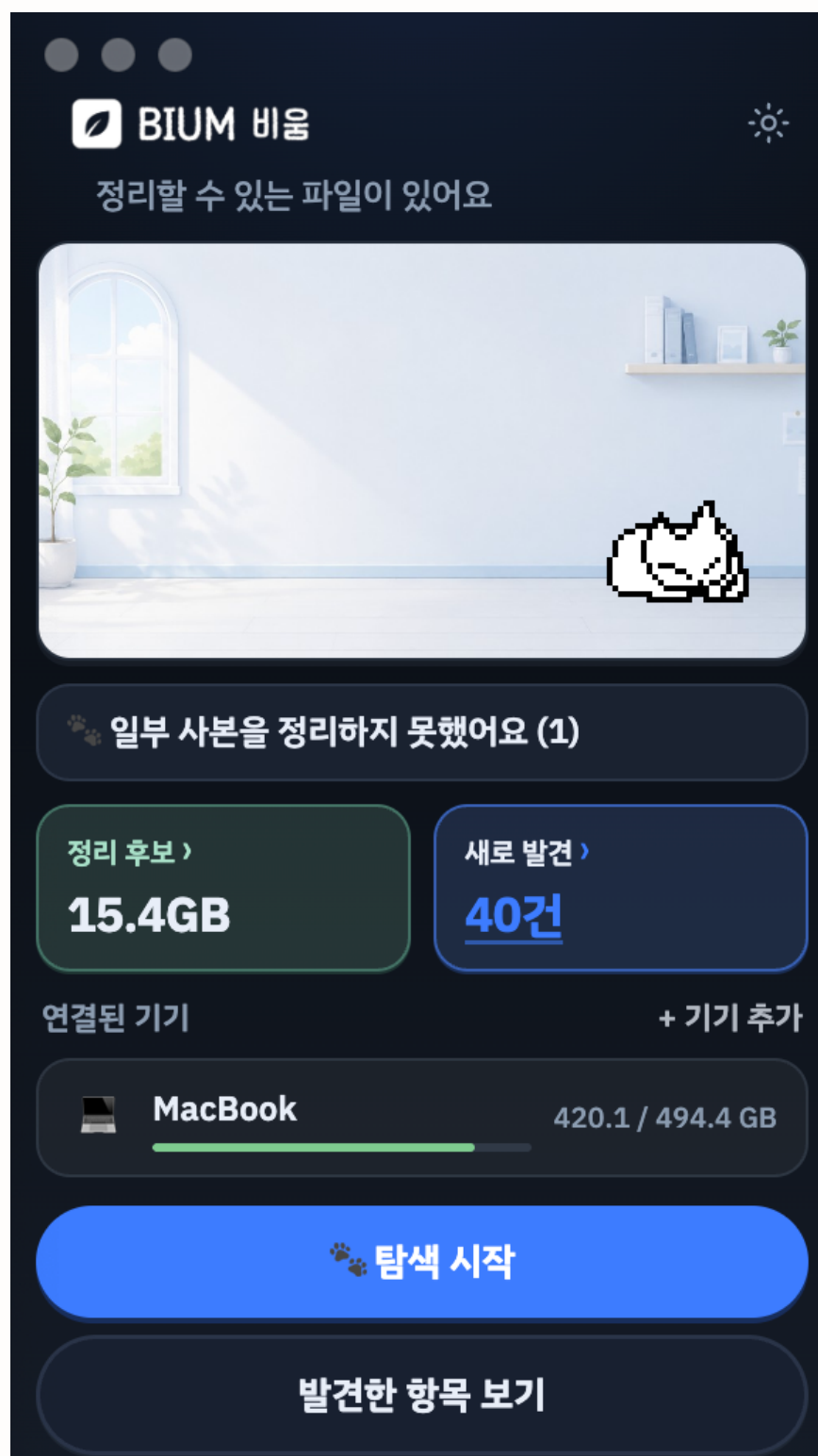


테마선택





프로토타입 / 시연





프로토타입 / 시연

깃허브를 통해 어플리케이션을 깔고 시연 해 보실 수도 있습니다

v0.1.0 Latest

Compare ▾



github-actions released this 3 hours ago · 2 commits to main since this release v0.1.0 40f4a2b

Full Changelog: <https://github.com/dlwldn4824/BIUM/commits/v0.1.0>

Assets 7

BIUM-0.1.0-mac-arm64.dmg	sha256:3bd7c55a1628e67295...		119 MB	3 hours ago
BIUM-0.1.0-mac-arm64.zip	sha256:c778b5daf1501f04cf...		119 MB	3 hours ago
BIUM-0.1.0-mac-x64.dmg	sha256:b0f9efcff9367ffc21...		124 MB	3 hours ago
BIUM-0.1.0-mac-x64.zip	sha256:96c8017ae9d2354d1e...		123 MB	3 hours ago
BIUM-0.1.0-win-x64.exe	sha256:9c24dd57f3f286989b...		103 MB	3 hours ago
Source code (zip)				3 hours ago
Source code (tar.gz)				3 hours ago



문제 해결

클라우드 사용자 관점에서 BIUM이 제공하는 솔루션

문제	CS 전공 3학년 : Mr. 기찬아	게임회사 클라우드 담당자 : Mr. 김구름	BIUM (우리 솔루션)
무엇이 필요한지 알기 어렵다	사진·문서를 하나씩 열어봐야 삭제 여부를 알 수 있음	오래된 DB·모델·문서가 현재 업무에 필요한지 한눈에 파악하기 어려움	판단에 필요한 정보 제공, 이에 따른 정리 서비스
데이터가 여러 곳에 흩어진다	스마트폰·PC·외장하드 ·iCloud·Google Drive 등에 같은 파일이 존재	개발 서버·클라우드·백업·프로젝트 저장소 등에 데이터가 분산	Local + Cloud를 함께 분석해 동일·유사 데이터를 통합해서 보여줌
삭제하면 안 되는 데이터가 섞여 있다	추억 사진·중요 문서까지 실수로 삭제할까 걱정	업무 데이터·원본·백업 데이터를 잘못 삭제하면 복구 비용과 업무 리스크 발생	삭제 대상과 보호 대상을 구분하고, 근거를 제시한 뒤 사용자가 결정
결국 계속 쌓인다	정리를 미루다가 클라우드 용량을 추가 구매	데이터 증가에 따라 스토리지 비용과 관리 부담이 지속 증가	중복·불필요 데이터를 선별해 저장공간을 효율적으로 사용

문제 해결

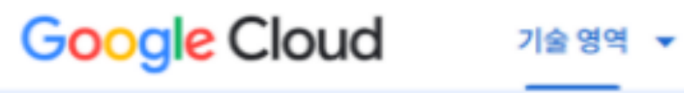
클라우드 운영 기업 관점에서 BIUM이 제공하는 솔루션

클라우드 운영 기업에서는 살짝 다른 역할



Cloud Storage Optimization Layer

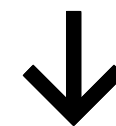
How?



- 객체를 가장 저렴한 저장 데이터 가격이 책정된 스토리지 클래스로 전환하는 `SetStorageClass` 작업이 우선합니다.

구글 `Lifecycle Policy` 공식 문서 중 내용: 미사용 부분을 더 저렴한 스토리지 클래스로 이동하여 비용 최적화

구글 클라우드 방식: 사용 빈도·경과 시간 기반으로 정리 우선순위 추정



BIUM: 실제 파일을 스윕해 상태를 파악 → 사용자 판단 기반 우선순위·storage 최적화

환경과 관련한 긍정적 영향

- 데이터센터 전력 사용 부담 감소

저장·백업·복제해야 할 데이터가 줄어 서버와 냉각에 필요한 에너지 수요를 낮춘다.

- 저장 인프라 자원 소비 완화

불필요한 데이터 축적을 줄여 추가 저장장치·서버 증설에 필요한 자원 수요를 완화한다.

- 디지털 탄소배출 저감에 기여

데이터 저장과 관리에 필요한 에너지 사용을 줄여 장기적으로 탄소배출 감소에 기여한다.

쓸데없는 저장 ↓ → 서버 부담 ↓ → 환경 부담 ↓



정량적 기대효과

10만명 기준

연간

200TB
/year

중복 삭제 기대치

연간

\$48,000
/year

저장 비용 절감

연간

최대 25tCO₂
/year

탄소 감축

* 1인당 저장 100GB × 정리 후보 3.5% 중 일부 실행
가정 → 1인당 2GB
총 10만 명 이용 가정 시 2GB * 10만 = 20만GB

출처:-자체 데모 실측(494.4GB 중 정리 후보
17.4GB) / Avast Global Photo Research, 2019

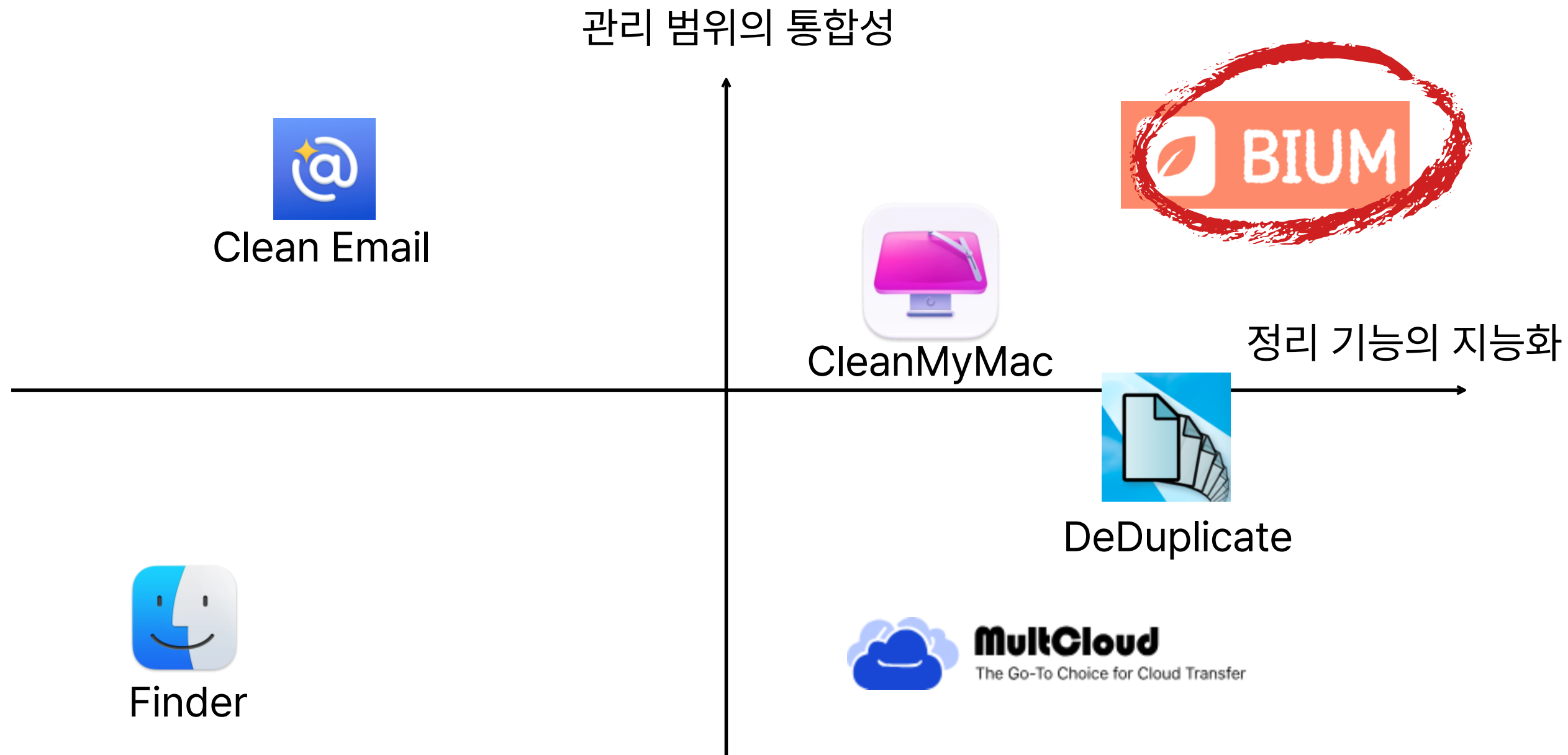
* 200TB × Google Cloud Storage Standard
\$0.020/GB·월 — 클라우드 저장분 기준 상한

출처:-Google Cloud Storage 공식 가격표
(cloud.google.com/storage/pricing)

* 46.33 kWh/TB·년(Corbett, POM 2018) × 545
gCO₂/kWh(arXiv, 2026) 기준

출처-Corbett, "How Sustainable Is Big Data?",
Production and Operations Management
27(9), 2018

경쟁사 분석



차별점



멀티 디바이스·멀티클라우드 관리가 어려움



장기 미사용·분산 저장 데이터까지
통합 정리하진 못함



무엇을 지워야 할지 판단해주는
정리 기능 부족



로컬·클라우드 파일 통합 관리 능력 부족



중복·유사 파일 판단과 클라우드 통합이 약함

프로젝트 의의

데이터를 비우는 작은 행동이 더 나은 환경으로 이어집니다

01 데이터 관리 문제를 환경 문제로 재정의

보이지 않던 데이터의 환경 비용을
사용자 경험의 문제로 연결



최근 데이터·클라우드 산업의 저장 효율화 흐름에서 출발해,
사용자가 환경 문제로 인식하기 어려웠던
중복·유사·장기 미사용 데이터의 관리 문제를
환경적 가치와 연결했습니다.

02 아이디어에서 실제 서비스까지, 하루 만에

기획에 그치지 않고
실제 사용 가능한 프로토타입으로 구현



문제 정의부터 UI/UX 설계, Desktop App 개발,
GitHub 공개, Google Drive 실연동까지
하루 안에 구현하며 아이디어의
기술적 실현 가능성을 검증했습니다.

03 오픈소스를 조합한 확장 가능한 탐색 파이프라인

기기와 클라우드를 넘나드는
데이터 탐색 구조 설계



오픈소스를 단순 활용하는 데 그치지 않고
해싱 기반 완전 중복 탐지 + 임베딩 기반 유사 탐지 +
로컬 기기·클라우드 연동을 하나의 파이프라인으로 설계했습니다.



BIUM 비움

