

Fortune 1000 대 기업이 선정한 데이터 클라우드 확산 전략

2022.5

Snowflake Professional Service

김용대

Snowflake 회사 소개



- **엔터프라이즈 데이터베이스 경험에서 출발**
 - Public Cloud 에 최적화된 데이터 플랫폼 필요
 - 엔터프라이즈 및 최신 데이터 클라우드 기능 추가
- **전세계 최초 클라우드 Native 데이터 분석 플랫폼 출시**
 - AWS, Azure, GCP 로 확대 / AWS 서울 Region 런치
- 직원 : 4,000명
- 고객사 : 5,000+

2012년
회사 설립



2015년
AWS GA



2020년
NYSE 상장



2021년
한국 지사 설립





수천의 고객이 입증하였습니다



데이터 공유를 통한 가치 극대화



데이터 공유 기회

연간 3조 달러의 가치

McKinsey
& Company

WORLD
ECONOMIC
FORUM

"전 세계 데이터분석가들의 70%가
외부 데이터 사용을 확대하고
있습니다"

FORRESTER

"Chief Data Officers: Invest In Your Data Sharing Programs Now", 2021

새로운 기회

새로운 "스마트 소비" 로 기술과 콘텐츠가 융합되고 제품 출시가 빨라지고, COVID-19 이후 온라인 가속화 되고 있습니다

데이터 인식변화

데이터 = 비즈니스 자산
데이터가 Silo에 갇혀 있거나 개별 부서만의 소유가 아님
소유주의 가공 데이터 접근
쉽고 안전하게, 데이터가 필요한 모든 직원에게 접근 및 활용 가능해야 함
데이터 활용도 높임
의사결정 개선과, 효율성 혁신을 위해 데이터가 분석과 머신러닝에 활용

데이터 중심 기업 "빠른 전환"
Digital Transformation

Data

IT 속도 개선

소비자의 진화

- 디지털 경험 확대
- 스마트 소비의 확산
- 정보 탐색과 가격 비교 증가
- 개인 취향과 밀접한 서비스에 대한 니즈 확대

판매형태의 변화

- 유통 및 제조 테크의 본격화
- 기존에 없던 새로운 형태의 서비스 등장 - 라이브 커머스
- 고객 세분화를 통한 개인화
- 가망 고객 Pool 확대 강구

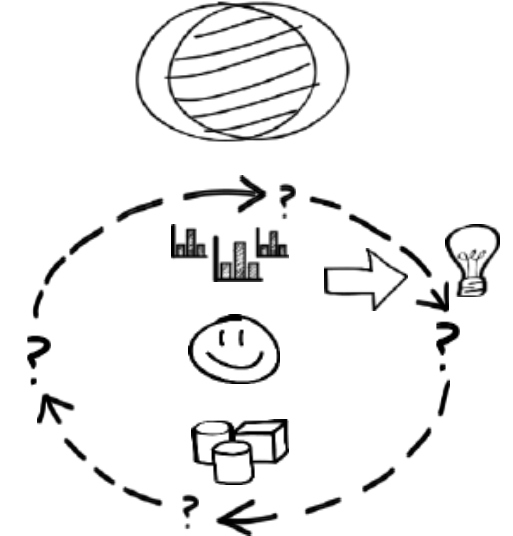
사업자의 진화

- 유통, 금융, 제조 등 사업영역의 확장 및 이동
- 비즈니스 차별화를 위한 다양한 접근과 시도
- 극대화된 편의성 혹은 내부 운영의 빠른 조치

디지털 경험 (데이터) 확대

자기 주도형 데이터 아키텍처 지향

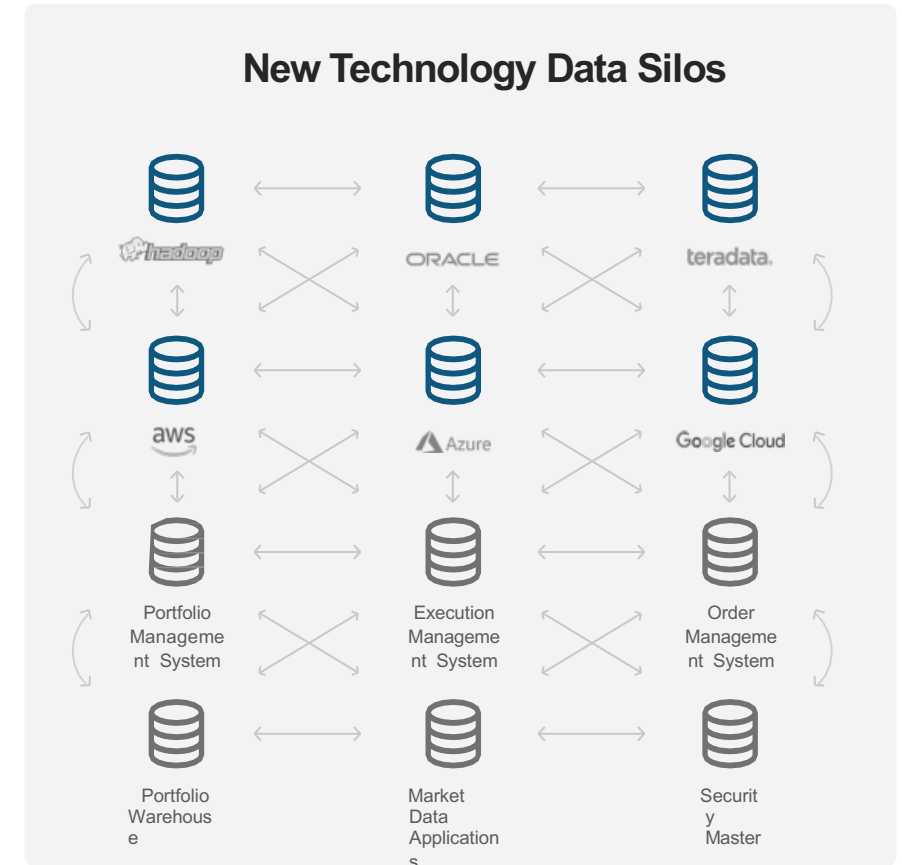
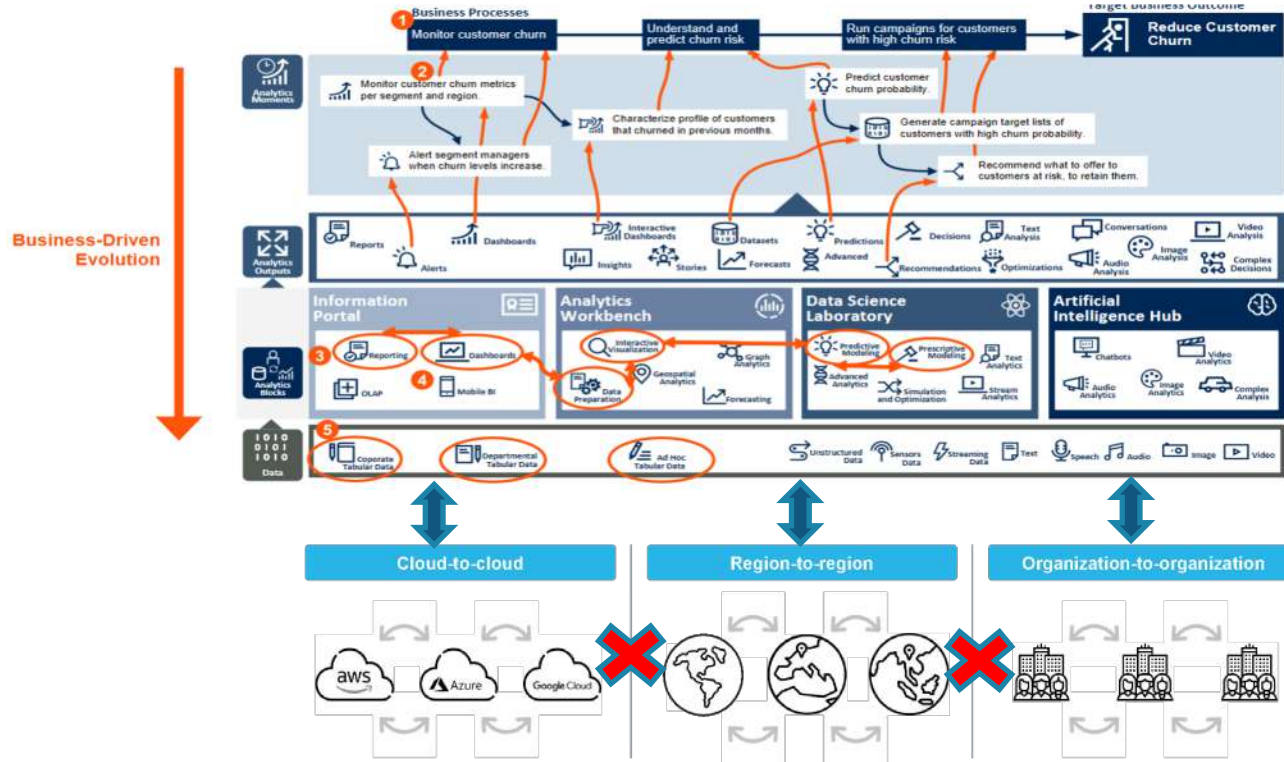
Want Can



Speed to Insight (Secs/Mins)

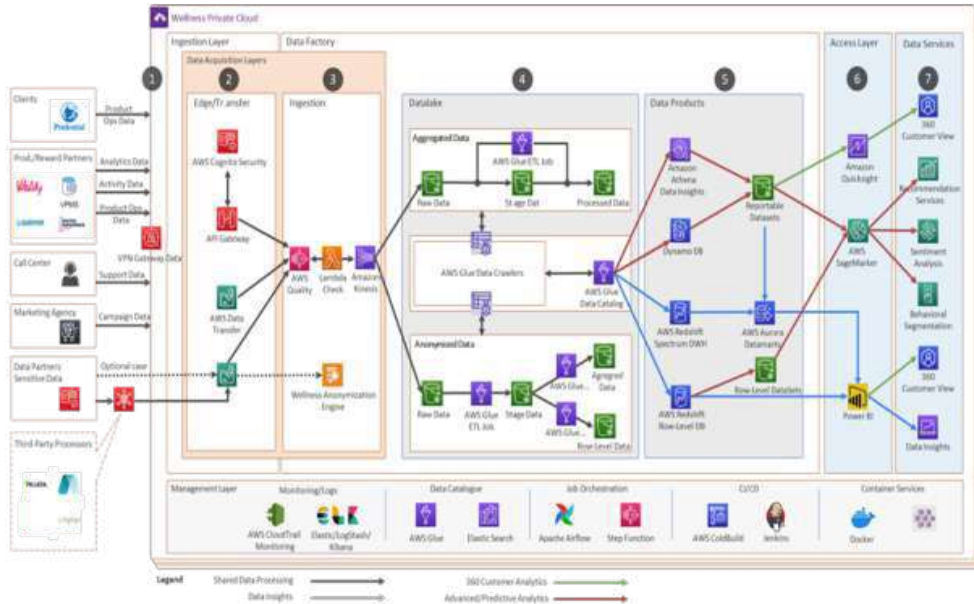
데이터 레이크 > 기술사일로

내부 부서 간 데이터 공유 확대, DMP(Data Management Platform) 강화 - 2차 데이터(구글 Analytics , 쿠키) 고객 데이터와 비식별 기반 고객데이터 매칭, ERP & POS 데이터 및 Supplier들간 재고보충 , 그룹사간 데이터 판매, 광고 분석 효과 측정 등 증가 > 고급분석과 별도로 다양한 데이터 결합 및 활용 요건 증가



공통된 문제점

복잡하고, 비싼, 제한된, 여러 복제본



- 시스템 운영 / 비용 / 내/외부 데이터 활용 어려움
- 점점 복잡해 지는 아키텍처 및 관리 포인트 증가
- DW 을 위한 고정적인 모델링 / 튜닝
- BI 사용자 쿼리 증가에 따른 성능저하 및 비용 증가
- 현업 데이터 요구사항 증가 (운영 비용 증가)
- 비정형 데이터 처리를 위한 별도 서비스 설계 (S3 + Athena etc)



다양한 Data
및 사일로
이슈



확장성 및 성
능 이슈



복잡하고, 인프라
에 고 비용 투
자



느리고, 제한
된 의사결
정

- 다양한 데이터 소스 및 유형
- 비-정형 데이터 활용 증가
- 데이터 동기화 어려움 증가

- 아키텍처 유연성 결여: 환경 및 워크로드 변경에 따른 확장성 이슈
- 빈번한 ETL 및 쿼리 성능 이슈
- 자원 경합으로 인한 SLA 미준수

- 고 비용의 인프라 서비스 투자 필요
- 오랜 플랫폼 구축 기간 필요
- 플랫폼 관리 및 튜닝에 많은 투자 필요

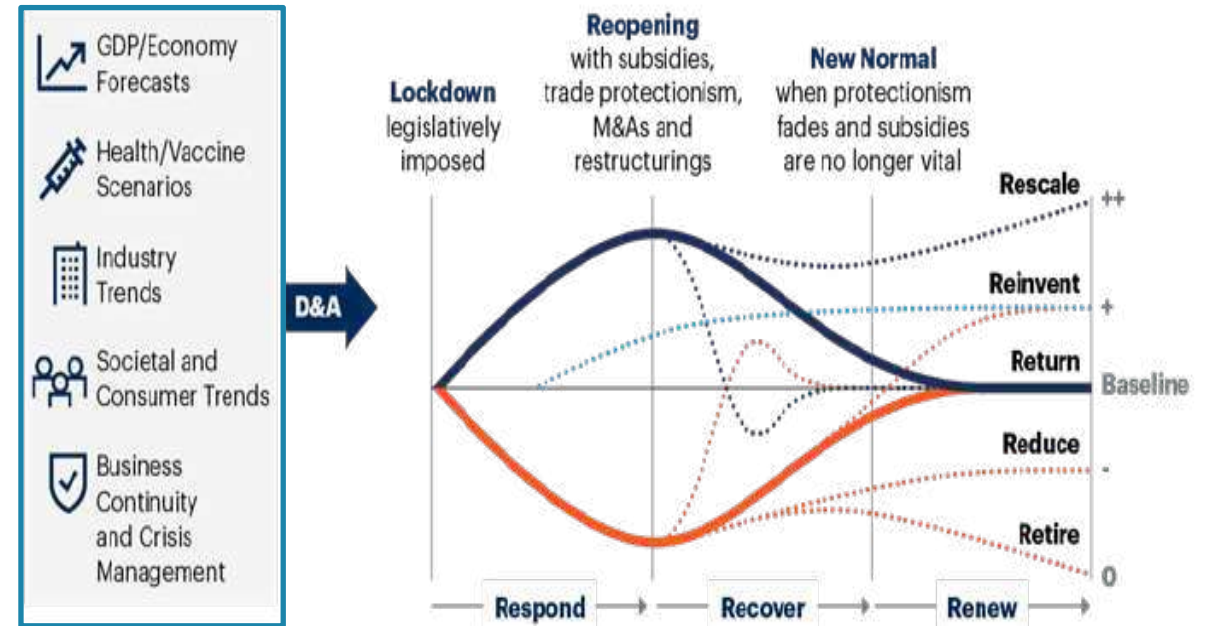
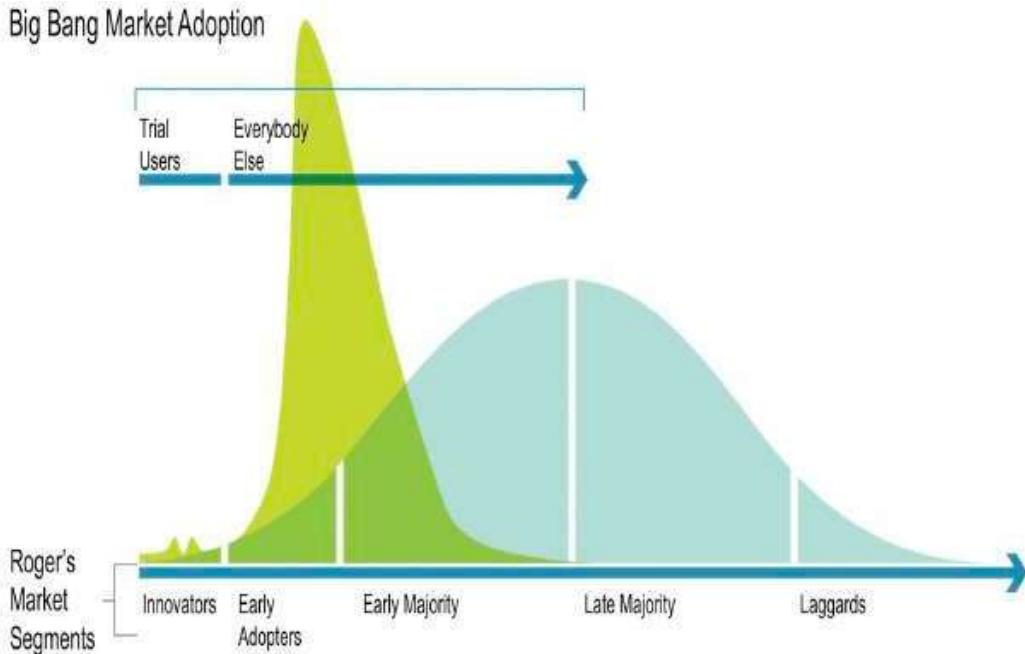
- 제한된 사용자 및 워크로드 수용
- 사용자 불만족 증가

보다 빠른 소통을 위한 유연한 데이터 활용 전략

기업 경쟁중심에서 고객 중심 신규 비즈니스 창출 및 확대를 위한 다양한 데이터 활용 고민
 실질적으로 실행 가능한 아키텍처 구축은 시간이 오래 걸림

빠른 시장 변화 반응 감지 통한
 매출 확대 + 비용절감 위한 해법의 주기가 빨라짐

가망 고객 Pool 확대 위한 CDP 적용? -
 외부데이터 / 내부데이터 / 그룹사 데이터 연계 통한 데이터 활용 극대화



데이터 공유 > 확대

안전하고, 즉각적인 데이터 접근

복잡도 제거, 비용절감

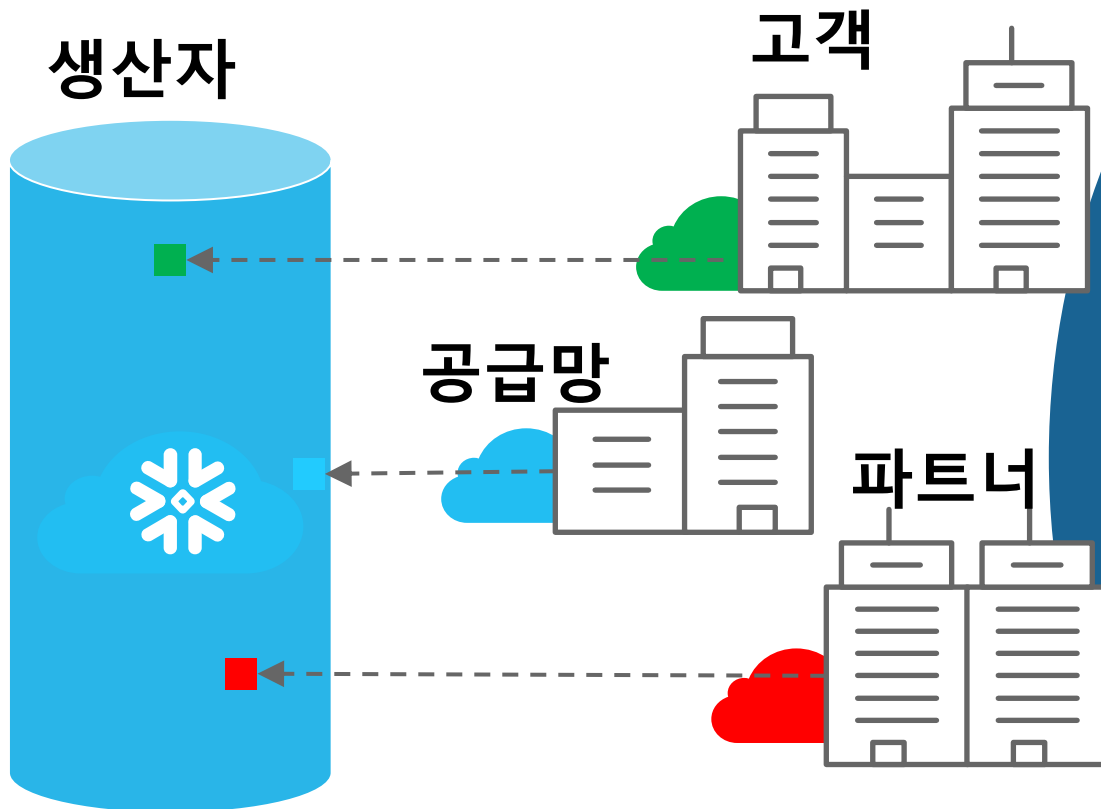
- > 기존 저장소 활용
- > 별도의 복잡한 데이터 파이프 라인 없음 (E, T)
- > 특정 소비자를 위한 스테이징 필요 없음

데이터 경험치 확대

- > 즉각적인 부서 & 기업 외부 데이터 활용 증가
- > 실시간 외부 원천 데이터 접근
- > 외부 데이터, 동시 사용

데이터 보호

- > End to end 암호화 내장
- > 데이터 공유 보호 기능 내장 : 상세 권한, 비 식별 조치, 데이터 기능



검증된 데이터 클라우드의 효과



“Snowflake
gave me back
2 Days
of my week”



Annual \$25M
→ \$2.5M
“We can query
3 years of data
instead of **13 Weeks**”

logitech

7 x
Improved
performance
of reports



5 x
Scalable
Enterprise BI
Infrastructure



95%
Fewer
Downtime
Events

SIEMENS

Global Data Mesh
Improve plant
efficiency &
WW SCM
Control Tower



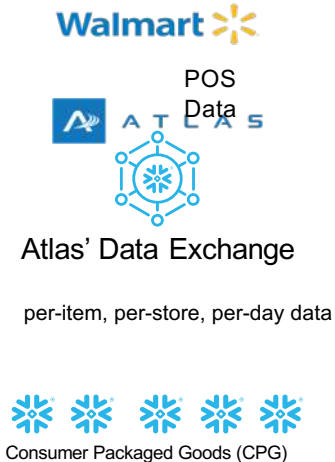
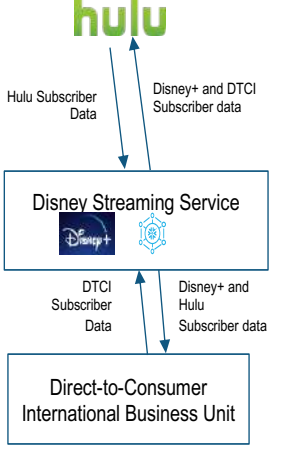
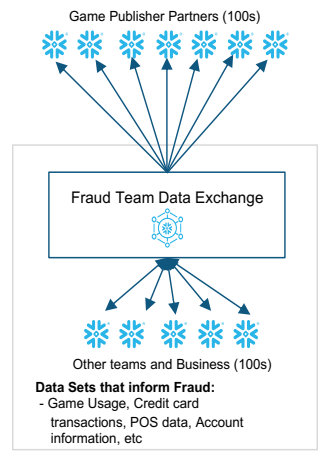
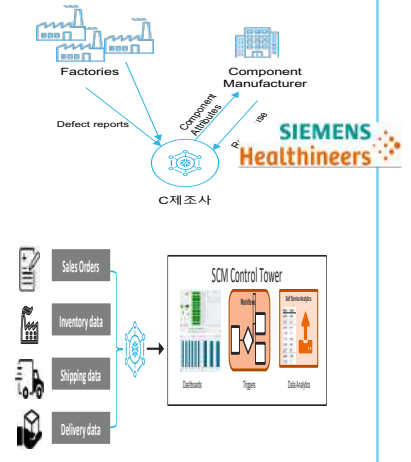
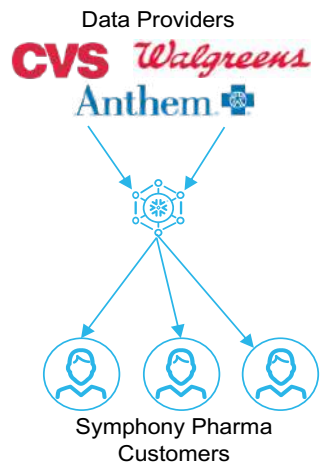
150x
Improvement
(Oracle BI Query)

Sainsbury's

6 hour → 3sec
Query Improvement
& Eliminate Group Silo
for Cost Reduction



Data Sharing 기반 비즈니스의 진화 및 차별화

산업	Retail/CPG – ATLAS	Media – Hulu & Disney	Game – Top S 사	Manufacturing – Global C사 & SIEMENS	Healthcare - SYMPHONY HEALTH
비즈니스 문제점	<ul style="list-style-type: none"> 100s CPG – Walmart내 자사 제품 매출 현황에 대한 빠른 이해 필요 대용량 데이터 실시간 기반 ETL 비용절감 필요 	<ul style="list-style-type: none"> 고객 분석 확대를 통한 최적화된 마케팅 전략 수립 CDP 기반 고객 행동 데이터 확보 방안 필요 (개인정보 식별) 	<ul style="list-style-type: none"> 빠른 게임 Fraud Detection을 위해 Speed가 관건이었으며, 100+ 유관 부서와의 빠른 데이터 Share 및 game Publisher 들과의 빠른 데이터 소통이 필요 	<ul style="list-style-type: none"> 보다 빠른 Supply chain partners들과의 협업을 통한 운영 효율성 향상 – 기존 매뉴얼 중심의 데이터 receive & send 비효율적 운영 방식 	<ul style="list-style-type: none"> 고객에게 Ftp보다 빠르게 데이터 제공 필요 –대 고객 서비스 만족향상
데이터 해결방안	Share with Customers - 실시간 데이터 access/secure View확보	Share with Partners / Across Business Unit - Data Clean Room 기반 외부 고객 정보 없이 식별 가능 구조 구현	Share with Partners / Across Business Unit - Fraud data 통합 기반 실시간 운영 구조 확립	Share with Suppliers - Quality 데이터 중심의 협업 극대화 가능 / 내부 운영 효율성 개선 가능	Share with Across Business Unit - Data Clean Room을 이용 고객 정보 안정적으로 빠르게 전달
시사점	<ul style="list-style-type: none"> 비용절감/빠른 판매 현황 파악 통한 생산/마케팅 의사결정 가능 	<ul style="list-style-type: none"> 고객의 다양한 VIEW 확보 / 마케팅 최적화 기반 매출 증대 	<ul style="list-style-type: none"> Reduce fraud: \$150k per day of fraud costs & Reduce ETL Cost of 100 different DB 	Faster time-to-insights for operation (SCM Control Tower)	1주일에서 1일 기준 데이터 전달 가능 / 비용절감 / 대고객서비스 만족 상승
적용 모델	 <p>per-item, per-store, per-day data</p> <p>Consumer Packaged Goods (CPG) Companies</p>		 <p>Data Sets that inform Fraud: - Game Usage, Credit card transactions, POS data, Account information, etc</p>		

Snowflake 주요 기능



데이터 플랫폼 구축 관점

전통적인 데이터 플랫폼



분석 리드타임 증가

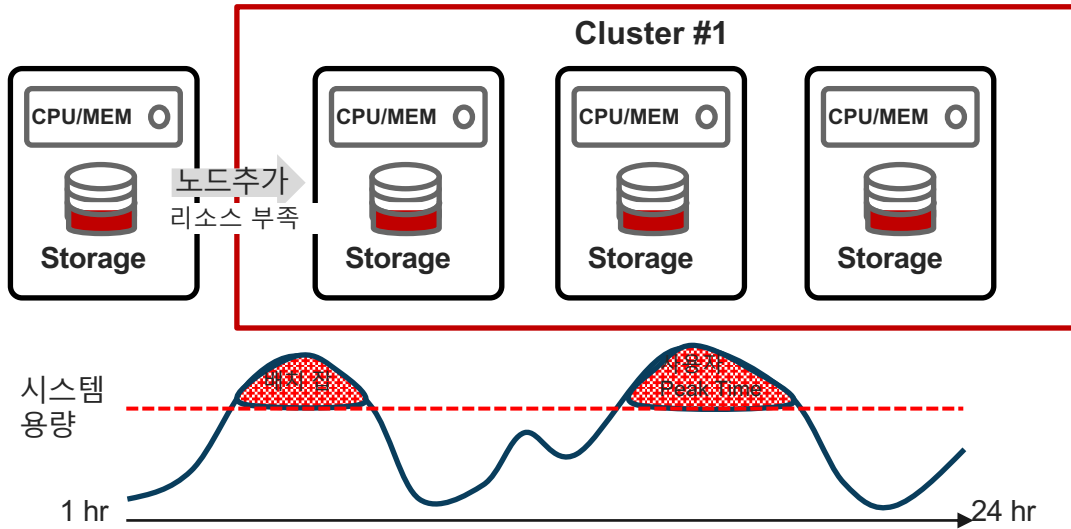
Snowflake



빠른 분석 환경 제공

사용자 관점

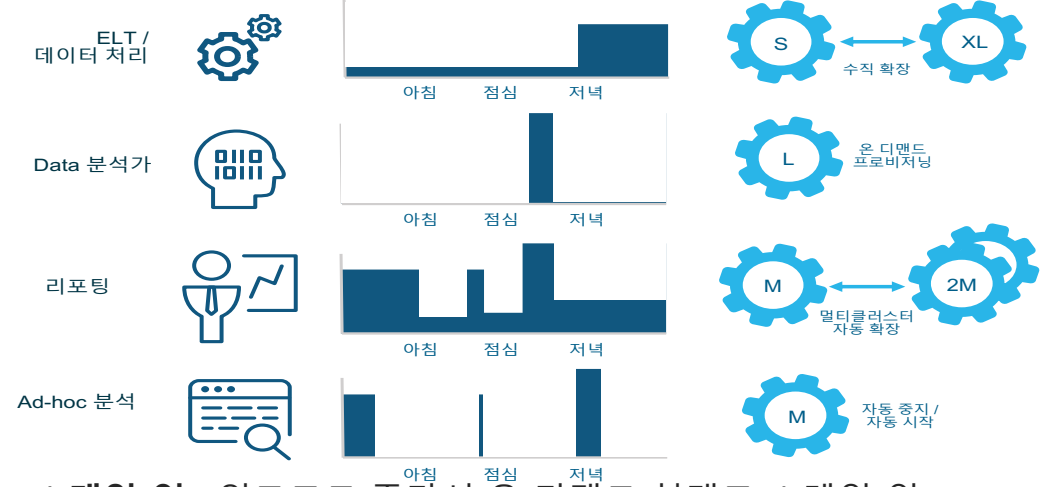
전통적인 데이터 플랫폼



- 신규 워크로드 유입으로 인한 성능 저하
- 시간대별 작업 **스케줄링 필요**
- 분석 워크로드를 제한
- 리소스 부족 시 컴퓨팅 직접 추가

Silo 분석 환경 유발

Snowflake

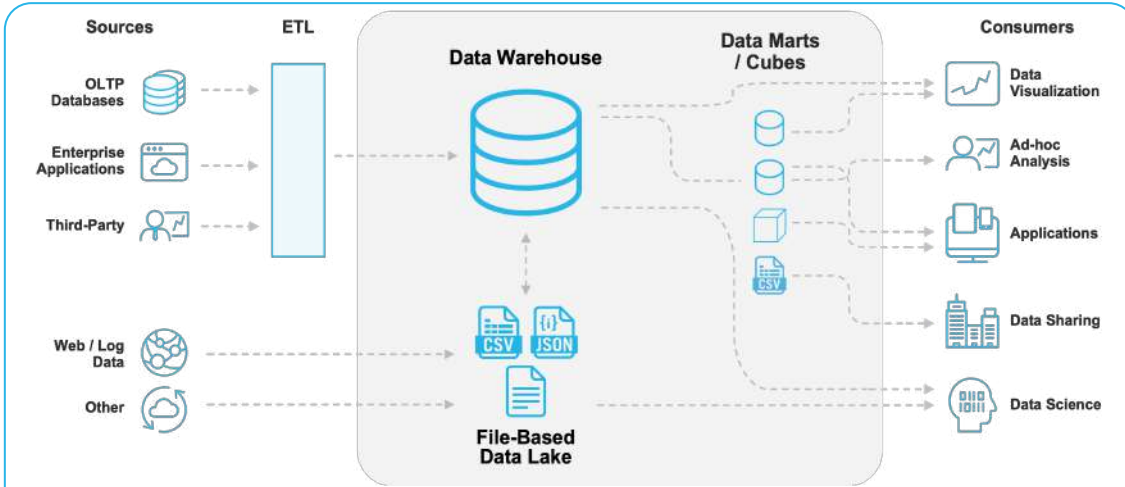


- **스케일-업** : 워크로드 증가시 온 디맨드 형태로 스케일-업
- **온-디맨드 프로비저닝**: 요청시 컴퓨팅 클러스터 즉시 프로비저닝
- **탄력성** : 동시 사용자 증가/감소시 Auto-Resume & Suspend 로 유연한 비용 통제
- **격리된 작업 환경**: 워크로드 유형에 따른 가상의 별도의 웨어하우스 제공

이질적인 워크로드가 가능한 데이터 플랫폼 제공

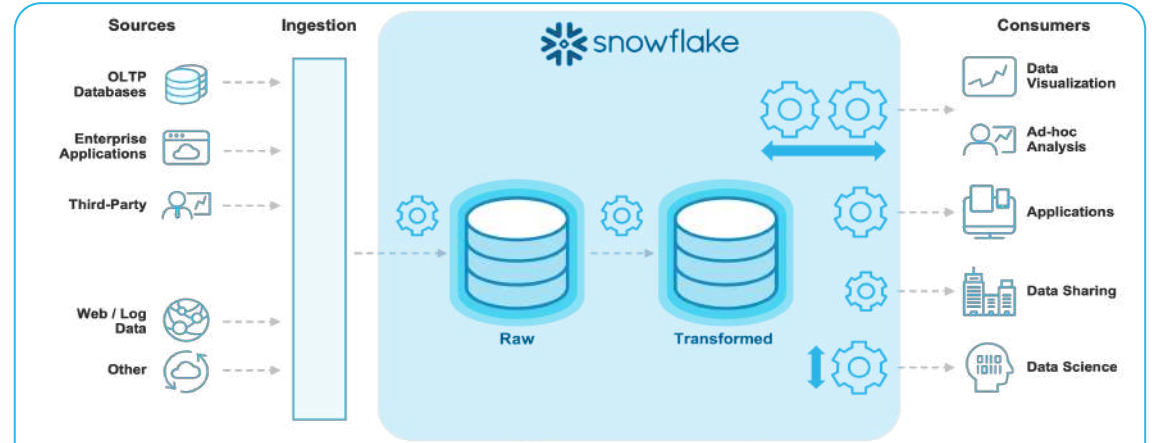
데이터 파이프라인 관점

전통적인 ETL 기반 데이터 파이프라인



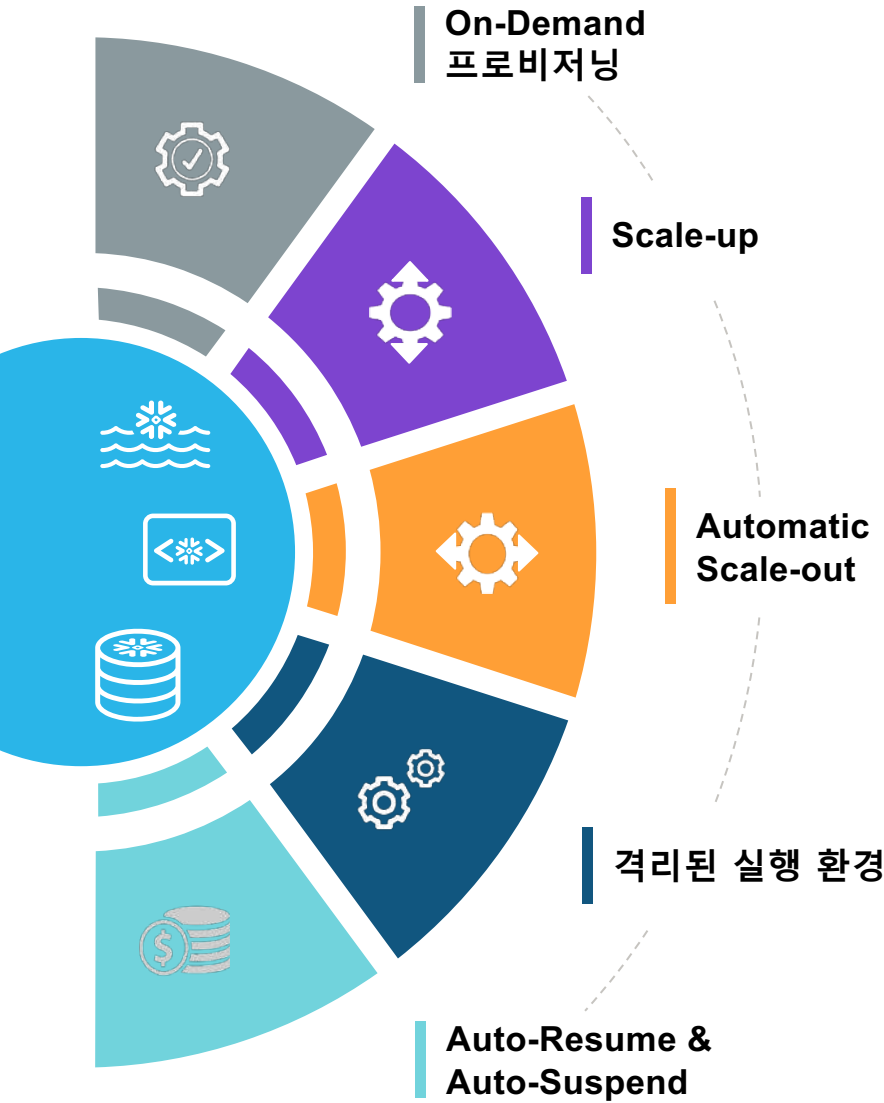
- 원천 데이터 추출, 변환 후 적재
- 적재 대상 및 리포팅 요구 증가로 인한 **ETL 부하 증가 및 속도 저하**
- **클라우드 환경**: 오브젝트 스토리지를 활용한 추출 및 변환, 이를 통한 ETL 부하 경감, 하지만 **데이터 관리 이원화 및 지속적인 변환 개발 비용 증가(Spark 및 Hadoop 기반)**

Snowflake: ELT 기반의 데이터 파이프라인



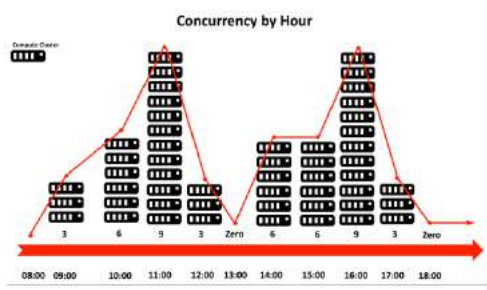
- **추출 후 바로 적재** : 원천데이터로부터, 빠른 원본 확보, 재작업 용이
- **데이터 처리(ETL) 비용 절감**
 - ✓ **범용 SQL** 이용한 데이터 변환, **고급 Spark 개발자 없이**
 - ✓ **파이프라인 간소화**
 - ✓ 정형/반 정형 데이터, 테이블로 바로 저장 & 자동 **Catalog (컬럼, 통계정보등)**. 작업 감소. 과거 Procedure 순차 처리 변환이 아닌 **ELT 도구(DBT 등)**등을 이용한 빠른 개발 및 동시처리 강화

SNOWFLAKE 아키텍처



Snowflake Virtual Warehouse - Auto-Scale out

- * 피크 시점의 성능 저하 없이 많은 Concurrent request를 처리하기 위한 멀티-클러스터 기능 제공
- * 부하에 따라 자동으로 Scale-out & down 기능 제공



```
CREATE OR REPLACE WAREHOUSE "WH_PROD_REPORTING"
WAREHOUSE_SIZE = 'SMALL'
AUTO_SUSPEND = 600
AUTO_RESUME = TRUE
MIN_CLUSTER_COUNT = 1
MAX_CLUSTER_COUNT = 10
SCALING_POLICY = 'STANDARD';
```

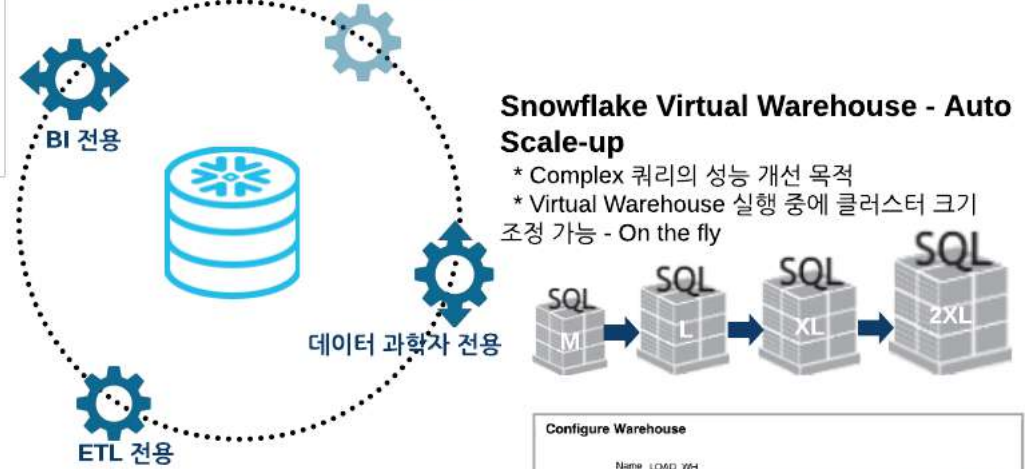
On Demand Provisioning

- * 새로운 워크로드를 위한 전용 실행 환경을 빠르고 쉽게 프로비저닝
- * Auto-Resume & Auto-Suspend 방식으로 사용한 시간 만큼 비용 발생
- * 새로운 Virtual Warehouse Provisioning 또는 Suspend된 Virtual Warehouse를 Resume하는데 수초 이내 완료

Ad-hoc 쿼리 전용

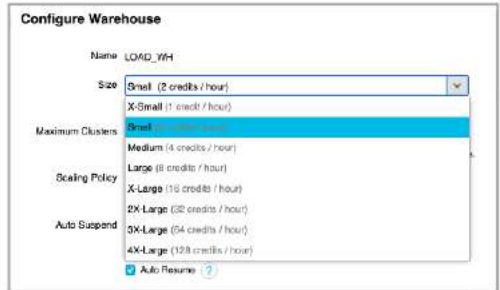
Snowflake Virtual Warehouse - Auto Scale-up

- * Complex 쿼리의 성능 개선 목적
- * Virtual Warehouse 실행 중에 클러스터 크기 조정 가능 - On the fly

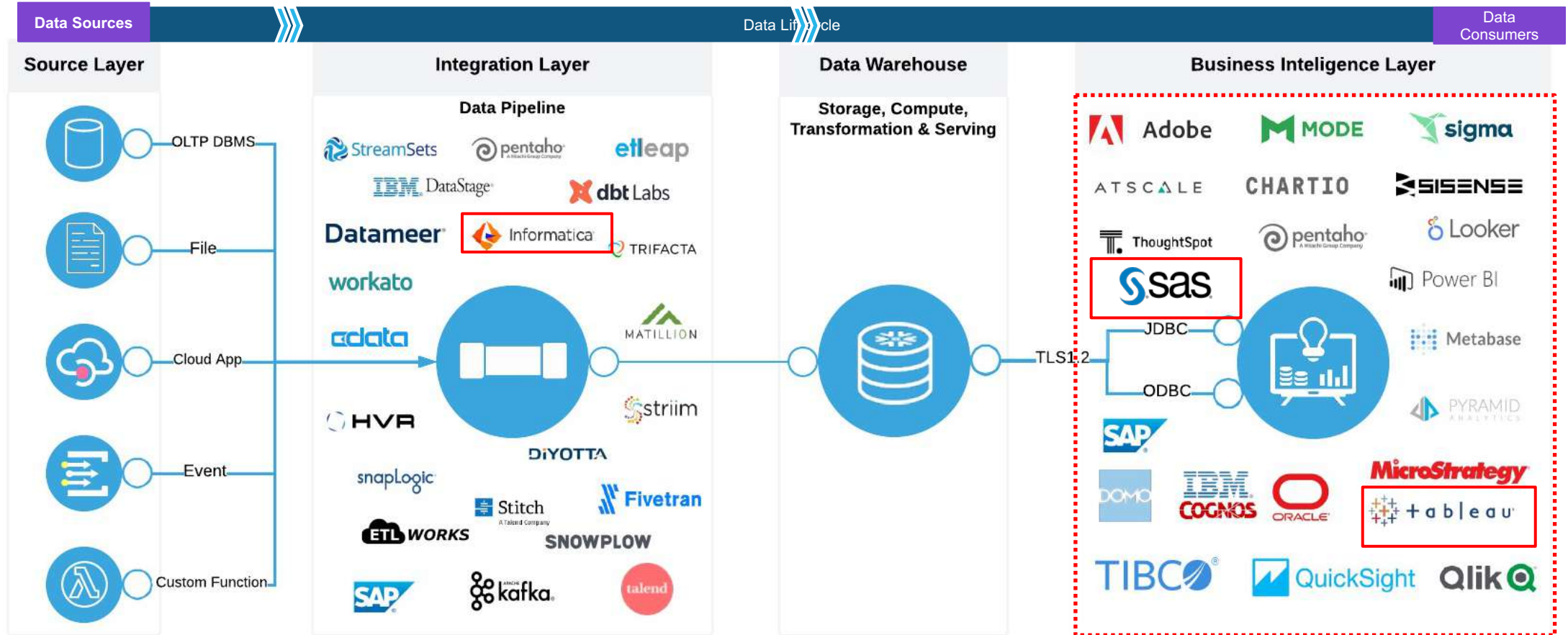


Workload 유형에 맞는 물리적으로 분리된 전용 실행 환경

- * 동일한 데이터에 대해 격리된 실행 환경 제공 - Sandbox
- * 여러 유형의 워크로드가 중복 실행되더라도 리소스 경합 현상 제거



Snowflake Data Cloud 파이프라인

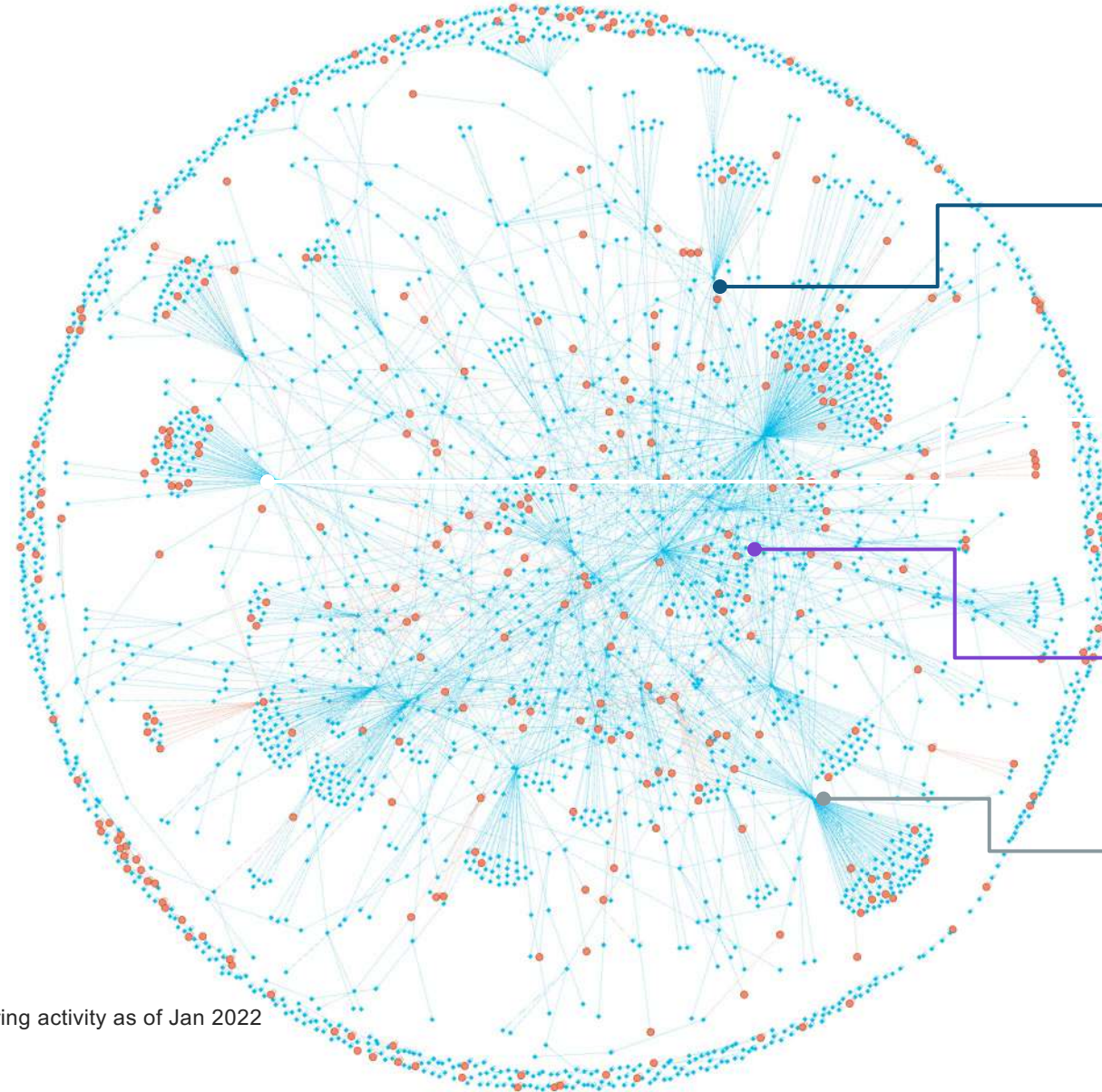


Retail & CPG



DATA CLOUD IN RETAIL AND CPG

수천 개의 조직이
서로 간의 데이터
공유를 통하여
비즈니스 강화를
수행하고 있습니다.



Retailers



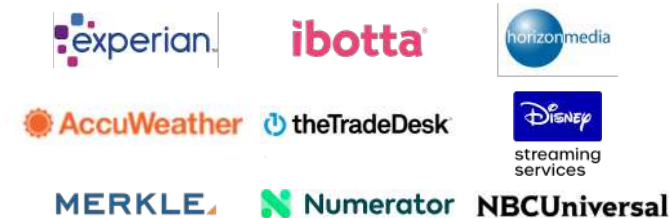
Manufacturers



Fulfillment & Distribution



Data Providers & Media Partners



* Visualization based on actual Data Cloud sharing activity as of Jan 2022

DATA CLOUD for CPG & retail

❄️ 디지털 환경에서 소비자와 브랜드의 상호 작용이 확대되며, CPG & RETAIL 산업에도 변화의 바람이 불고 있습니다. 데이터를 기반으로 고객 경험을 개선하고 공급망을 혁신하여 새로운 가치창출 및 이익중심 경영을 실현합니다.

고객으로부터
시작하는
비즈니스의 새로운
원칙/ 변화 이해



- Customer expectation
- Supply chain & OPS
- Digital & eCommerce

❄️ Snowflake Retail Data Cloud를 통해 CPG & Retail 데이터에 대한 접근, 관리 및 공유를 보다 원활하게 하며 개인화된 고객 경험 제공, 공급망 최적화, 그리고 데이터 기반 상품화를 가속화할 수 있습니다.

❄️ 다양한 업무 영역에서 검증된 비즈니스 ROI

- Increase Personalization
- Optimize Supply Chain Operations
- Enable Data-Driven Merchandising
- New Biz. Model - Data Monetization

❄️ Snowflake Value Enablers

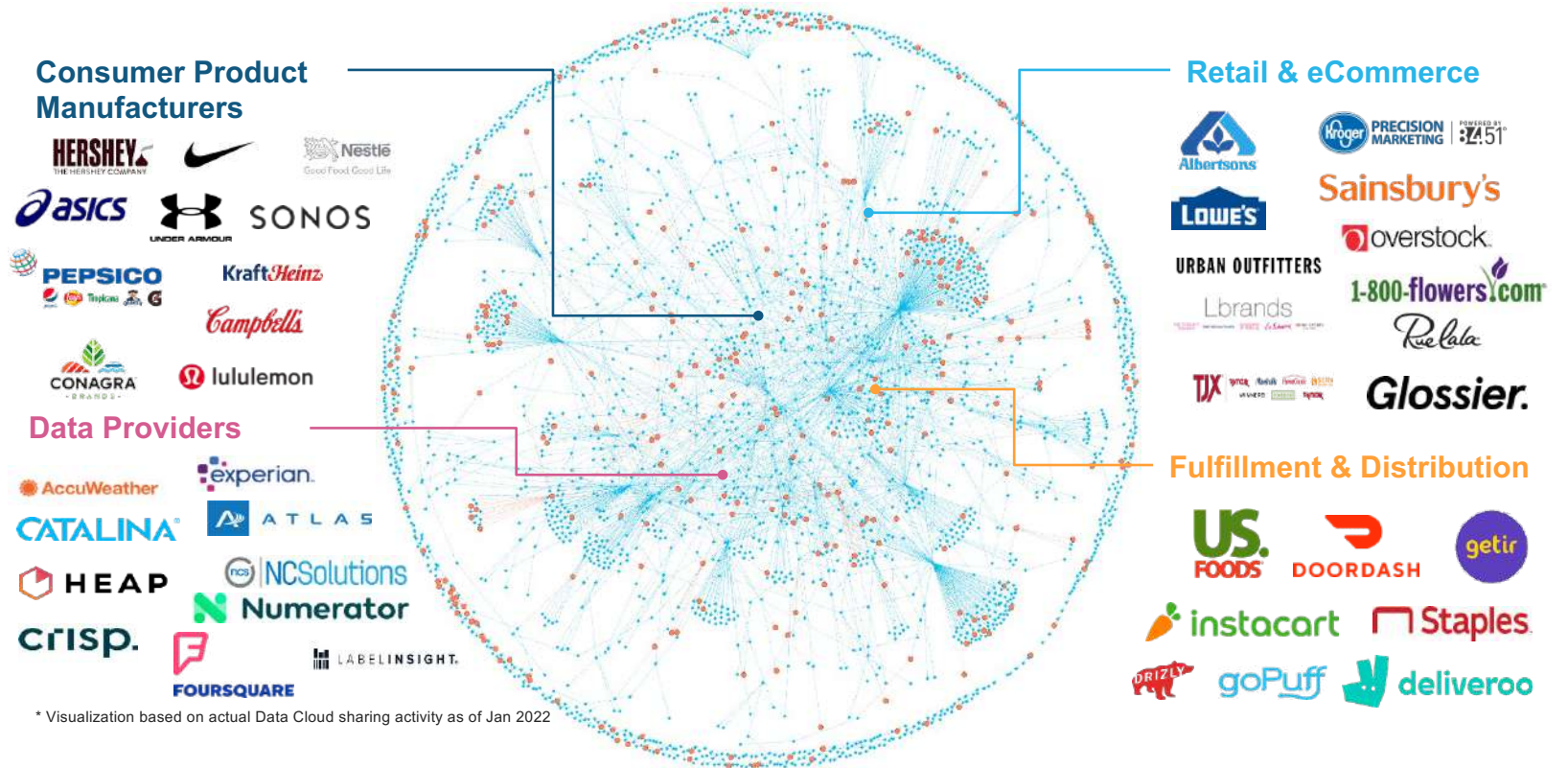
- 더 높은 성능 및 확장성 To 데이터
- 운영 효율성 증대 & 적은 IT 리소스
- OP 데이터 및 클라우드 플랫폼에 비해 50-80% TCO 절감



DATA CLOUD for CPG & retail

새로운 비즈니스 모델 발굴 / 실시간 수요 최적화 / 데이터 활용 극대화 위한 새로운 접근
 A ONE-STOP-SHOP TO MEET NEW CONSUMER DEMANDS & BEHAVIORS

❄️ Snowflake Retail Data Cloud를 통해 CPG & Retail 데이터에 대한 접근, 관리 및 공유를 보다 원활하게 하며 개인화된 고객 경험 제공, 공급망 최적화, 그리고 데이터 기반 상품화를 가속화할 수 있습니다.

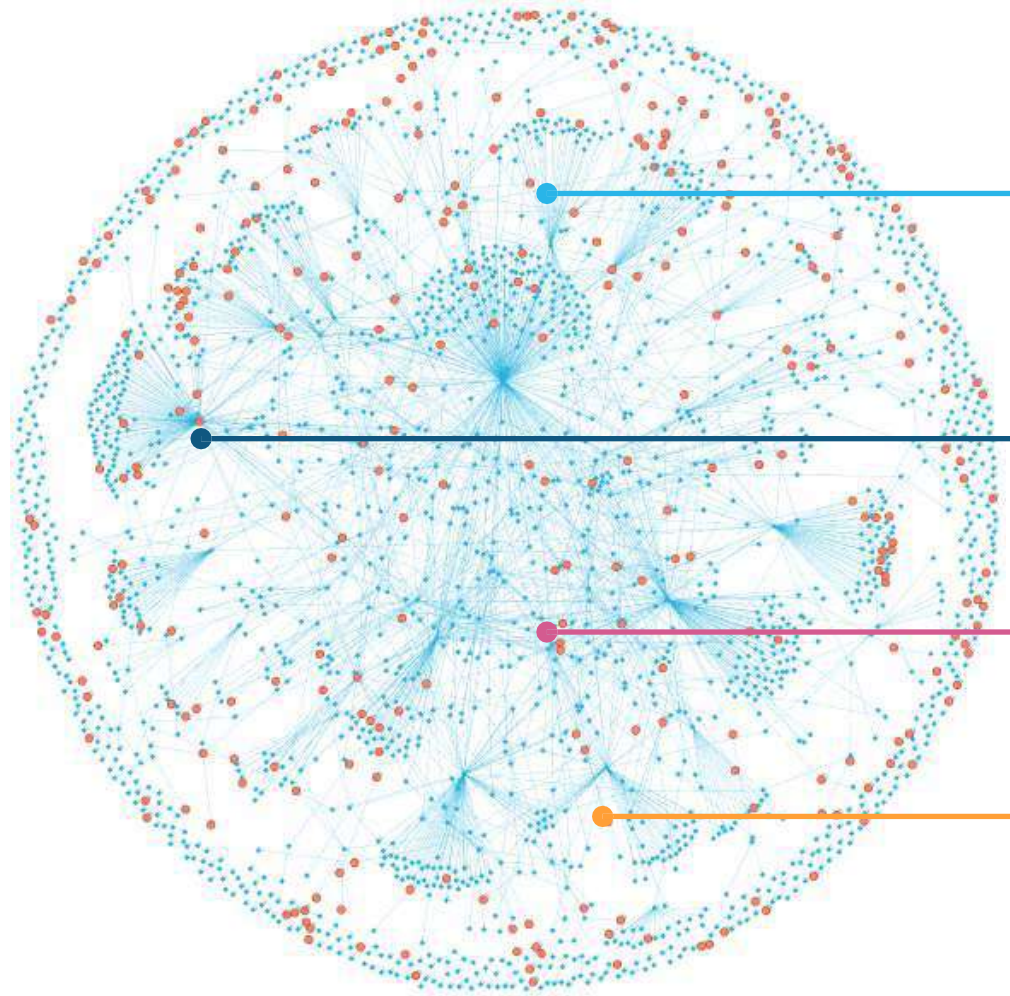


Media



DATA CLOUD for Media

개인 정보 및 보안을
유지하면서 미디어 및
광고 생태계 전반에
데이터 민주화를
지향한다



Agencies / Brand & Advertisers
Enabling the 'buy side' to leverage their 1PD to enhance their marketing approach and effectiveness

Data Providers & Services
Adopting to a modern, more secure method to license and distribute their data solutions and services

Media Companies & Publishers
Developing the content that creates the supply for the entire ecosystem

Platforms (AdTech / MarTech)
Leveraging Powered By Snowflake to evolve and build new products and capabilities

Visualization based on actual Data Cloud sharing activity as of Oct 28 2021.



SNOWFLAKE MEDIA MARKET

5 out of 6

Big Media Companies*



WarnerMedia

VIACOMCBS

SONY

9 out of 15

Top AdTech & MarTech
Platforms (SSP & DSP)**



DEMANDBASE

Magnite



OneView

INNOVLD

PubMatic

choozle

9 out of 20

Top Gaming Companies***



Sony
Interactive
Entertainment



Electronic Arts



zynga



UBISOFT



Playtika

*Motely Fool; **G2 DSP & G2 SSP; ***New Zoo



데이터 활용 관점에서 단일 뷰가 필요합니다



유치

유지

Multi-channel marketing programs

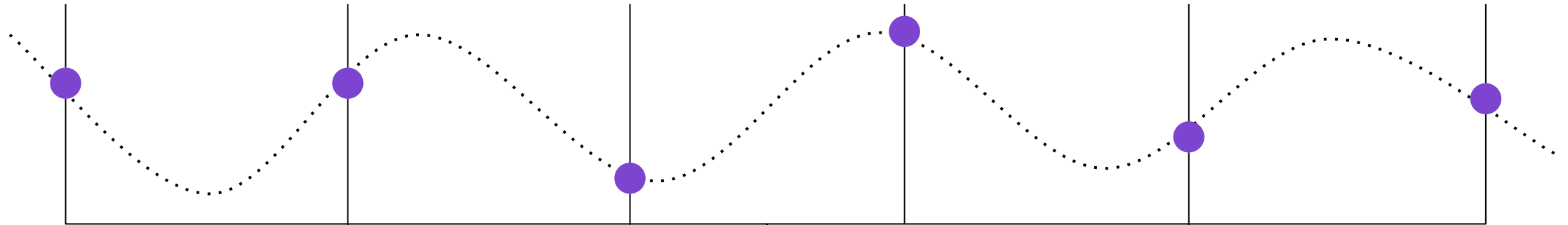
Identity & enrichment

Offers, bundles, promotions

Engagement across platforms & devices

Engagement with content and advertising

Add'tl touchpoints (e.g., customer service)



구독자 분석

Segmentation & Personalization

Portfolio and elasticity analysis

Media-mix modeling

ARPU, Look-alike segmentation

Recommendation engines, CX

Propensity to churn/renew

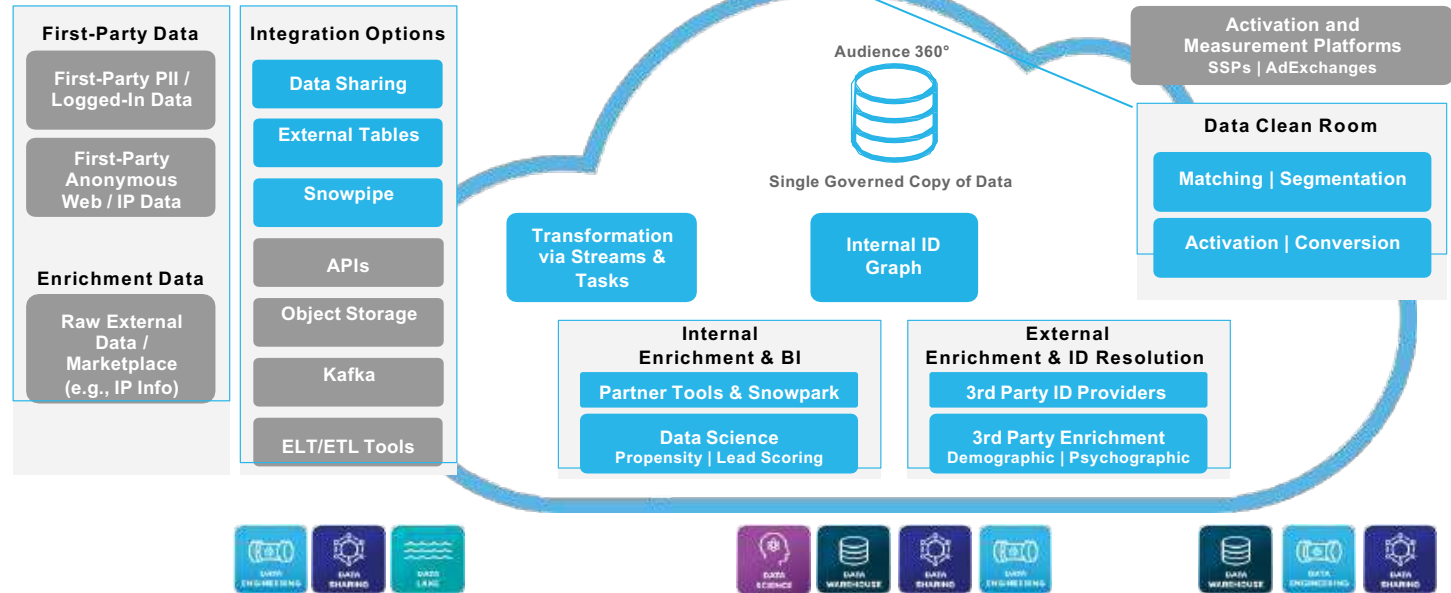
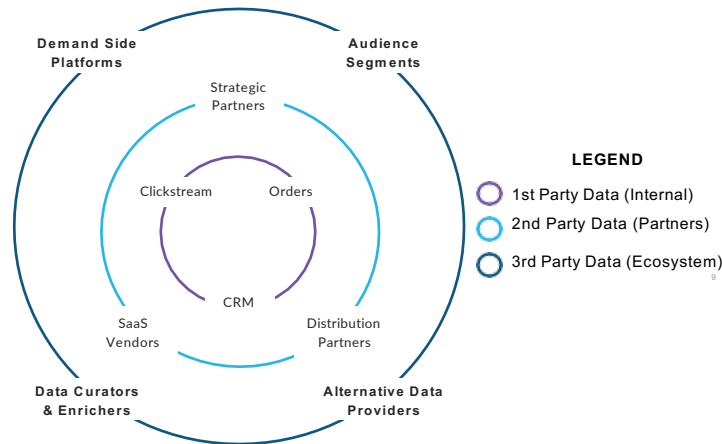


Media & Entertainment

Create 360° audience views / Resolve identity & enrich profiles / Plan & execute campaigns



Data Sources > Data Access > Quality & Hygiene > Customer ID & 360 > Planning | Analysis | Activation | Attribution

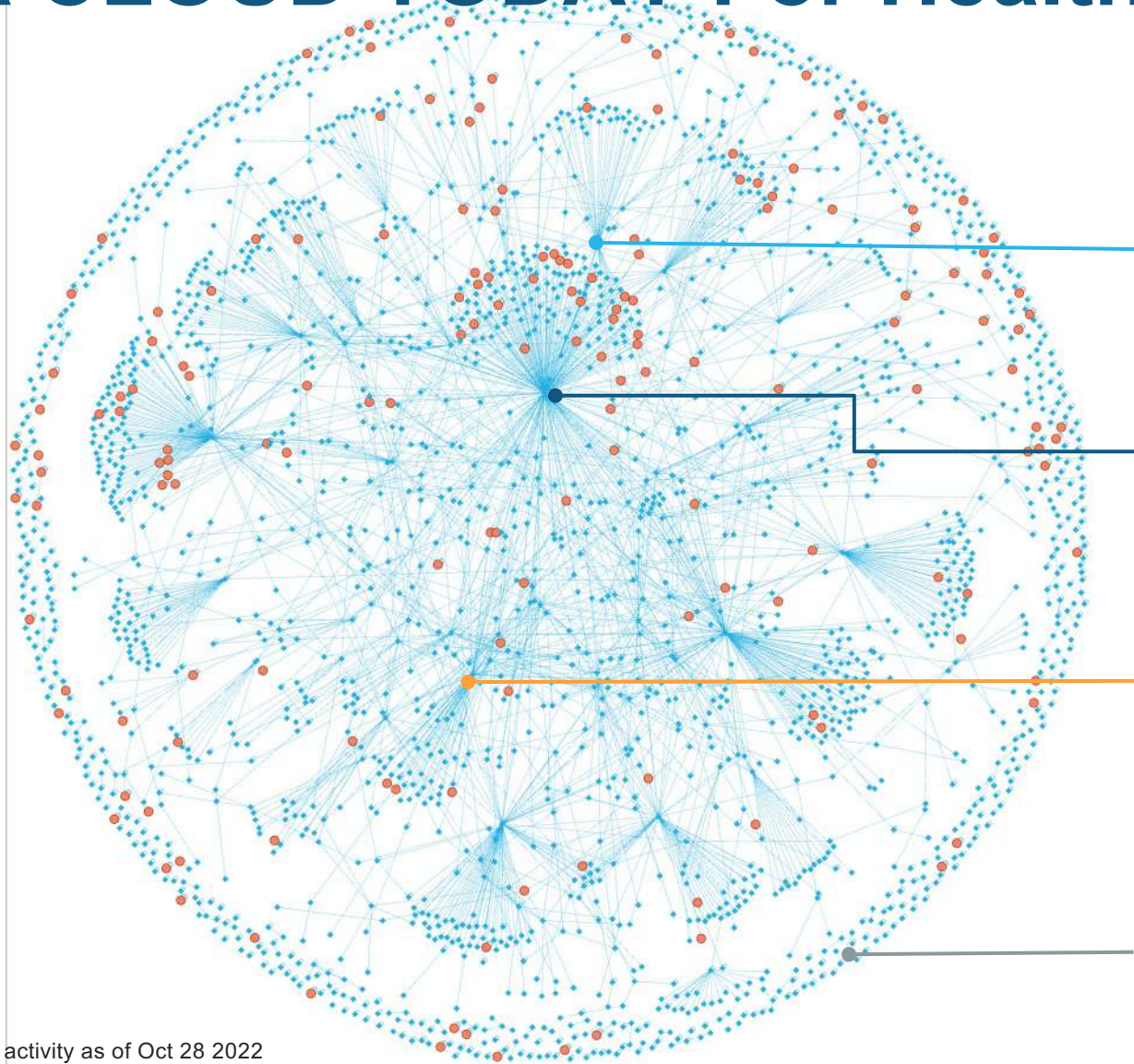


Life Sciences & Healthcare



DATA CLOUD TODAY For Health Care

수천 개의 조직이
서로 간의 데이터
공유를 통하여
비즈니스 강화를
수행하고 있습니다.



Leading Payers & Providers

Anthem

PRISMA
HEALTH

Center for Medicare
& Medicaid Services

NYC
HEALTH+
HOSPITALS

COVID-19 & Vaccine Insights

COVID-19
RESEARCH DATABASE

STARSHEMA

CA.GOV

MCKESSON

Data Providers

IQVIA

experian.

plunge digital

INVITAE

PRECISIONxtract

knoema

Health Tech

HealthCatalyst

komodo
HEALTH

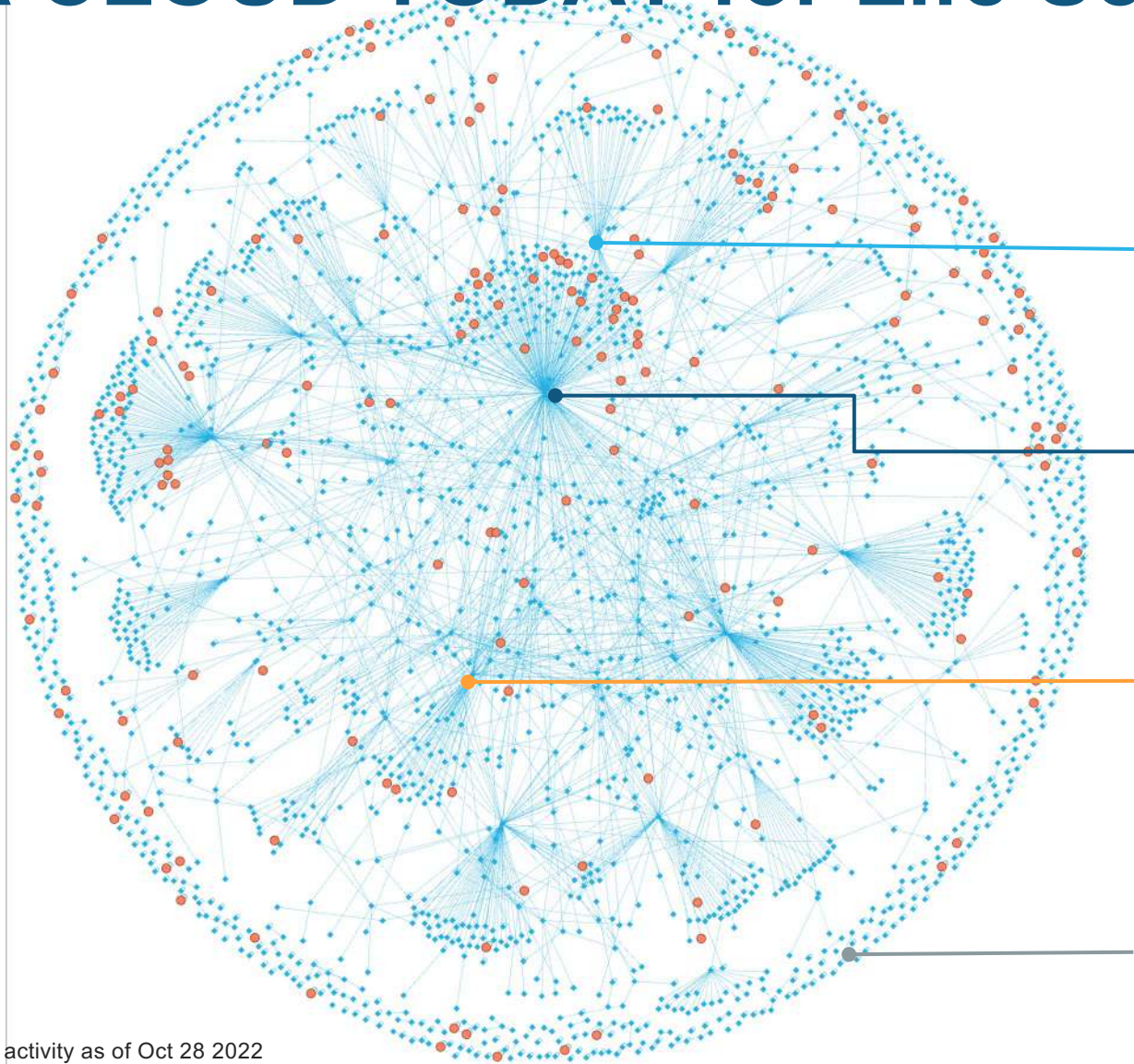
athenahealth

fireflyhealth

* Visualization based on actual Data Cloud sharing activity as of Oct 28 2022

DATA CLOUD TODAY for Life Science

수천 개의 조직이
서로 간의 데이터
공유를 통하여
비즈니스 강화를
수행하고 있습니다.



Fortune 500 pharmaceutical,
medical device & genomics
companies

COVID-19 & Vaccination Insights



Data Providers



* Visualization based on actual Data Cloud sharing activity as of Oct 28 2022

주요 사례 In Life Sciences & Healthcare

신약개발, 바이오, 게놈 및 CRO 관련 다양한 생명공학 영역에서 데이터 활용 극대화를 위한 데이터 클라우드 활용 사례 보유

Pharmaceuticals & Biotechnology



Medical Devices



Genomics



Analytics Service Providers

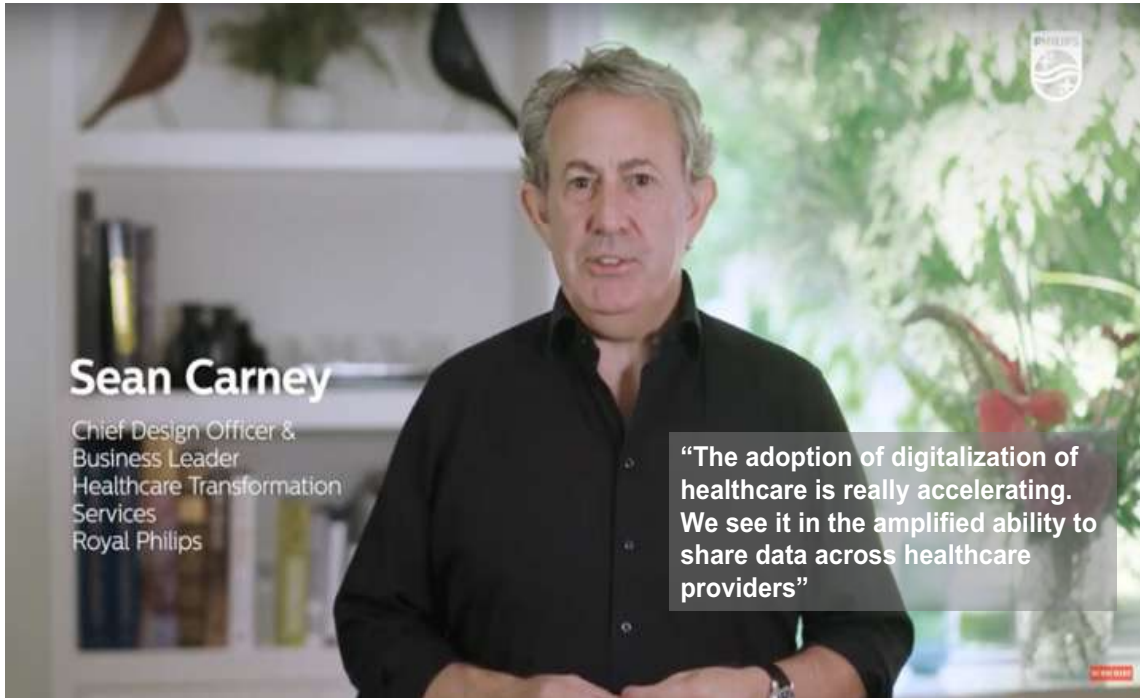


Data Partner (Analytics / CRO 등)



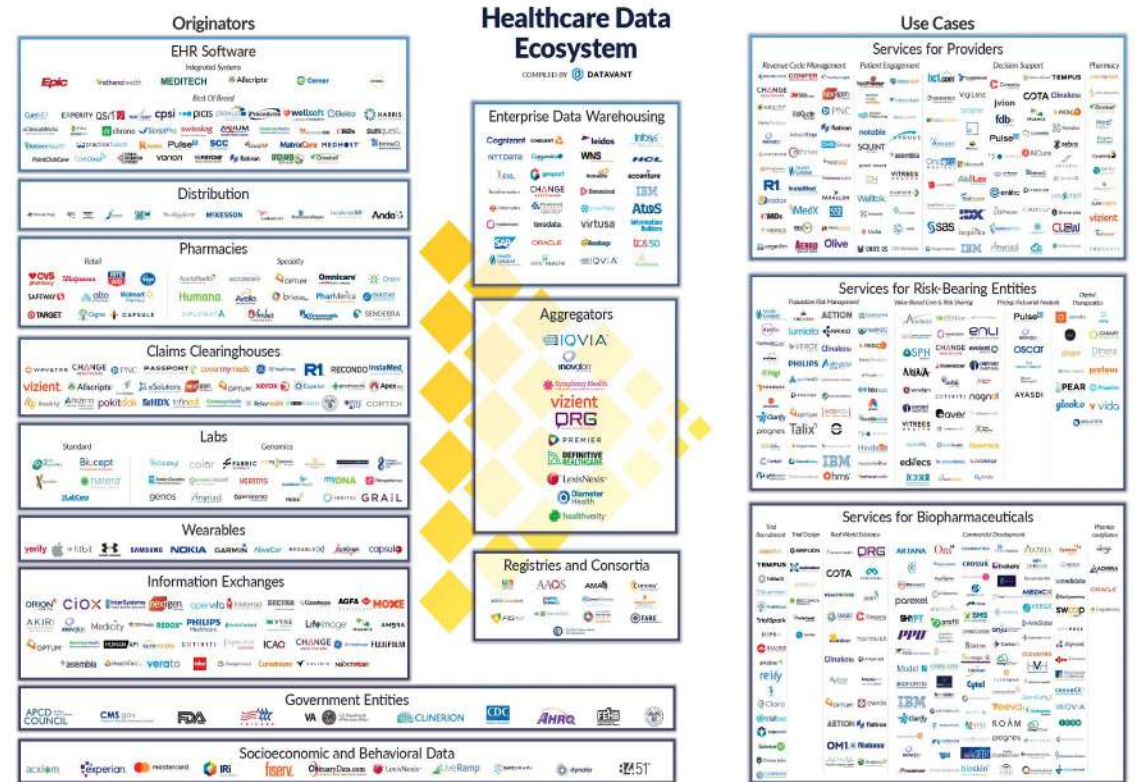
Journey to a Strategic Engagement

사업 비전 & 목적



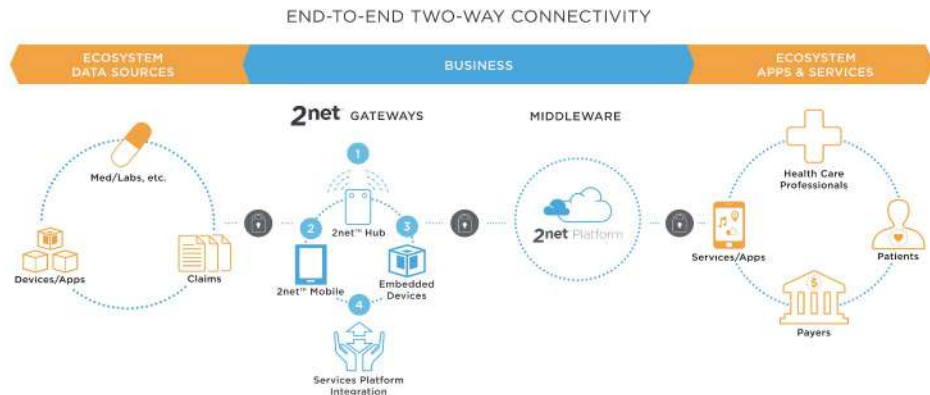
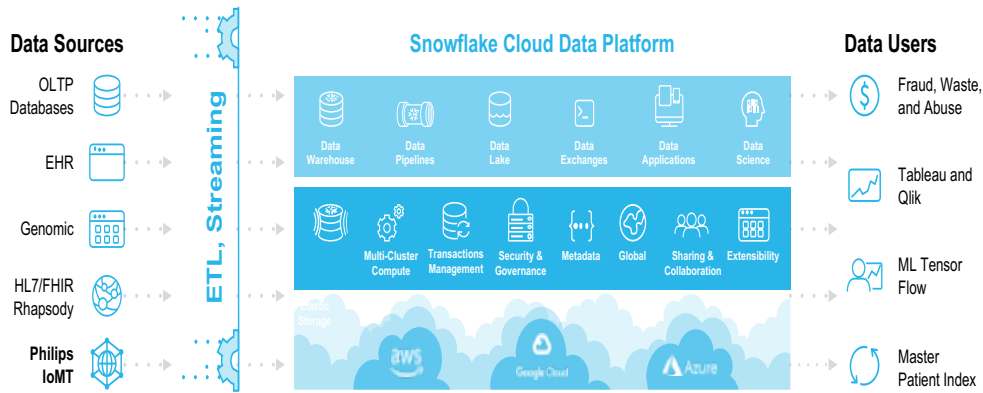
Source: <https://youtu.be/g-2somAVI30>

디지털 헬스케어 에코 시스템을 위한 혁신적인 데이터 클라우드 플랫폼 파트너로 Snowflake 선택

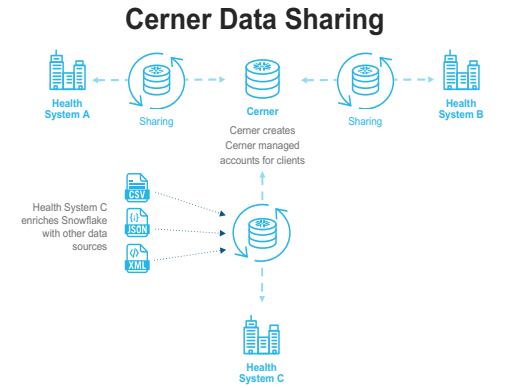
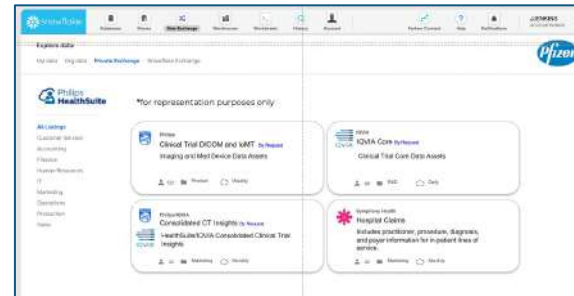


Journey to a Strategic Engagement

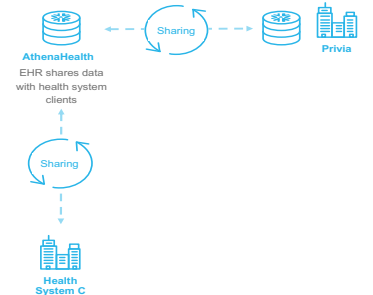
SNOWFLAKE HEALTHCARE CLOUD DATA PLATFORM & HEALTHSUITE



Snowflake Data Exchange



Active Secure Data Sharing: EHR



Quick-Start with Snowflake PS&T



Snowflake 를 빠르게 사용하기 위한 4가지 방법



사전 학습

데이터가 유효한지
탐색, 모델링 해보자

QuickStart Lab

Reverse Engineering
ETL Pipeline
Visualization
Snowflake Core



최선의 선택

구축/운영의
최적을 찾아보자

옵티마이제이션

Billing Metrics
Performance
Usage Monitoring
Configuration



점검

시행착오를
줄여보자

유즈 케이스 탐색 및 진단

마이그레이션 진단



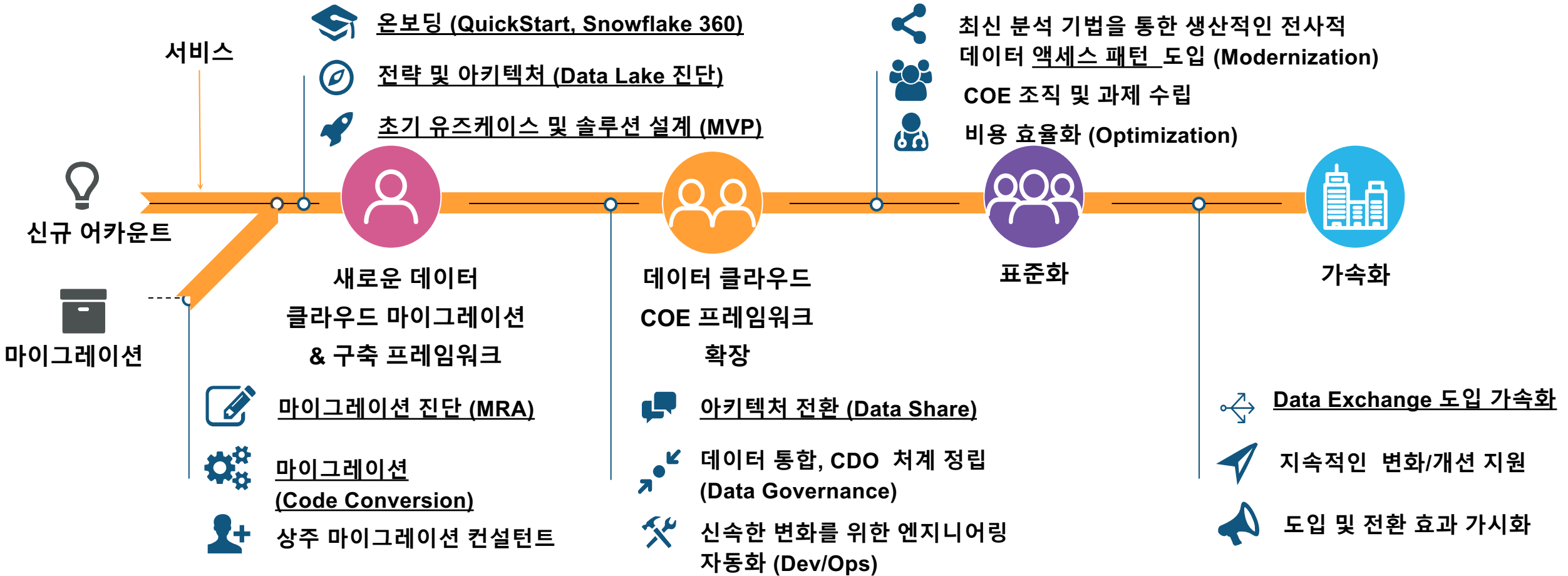
성장

App 개발
공유하자

App 개발

Data 공유(Share)
Data Clean Room

Snowflake Professional Service



Snowflake Education Service

Snowflake Fundamentals

대상:

- 데이터 분석가
- 데이터엔지니어
- 데이터 사이언티스트
- 데이터 아키텍트
- 데이터베이스 관리자

주요 개념:

- Snowflake 아키텍처 설명
- 데이터 로드 및 변환 과정
- 빠른 데이터 분석을 위한 Snowflake SQL 기능

가격:

- 2일(실습 없음):
 - 프라이빗 과정 - \$7,000(과정당), 최대 15명까지
- 4일(실습 있음):
 - 퍼블릭 과정 - \$1,500 (인당)
 - 프라이빗 과정 - \$14,000 (과정당), 최대 15명까지

Snowflake Advanced

대상:

- 데이터 분석가
- 데이터 엔지니어
- 데이터 사이언티스트
- 데이터 아키텍트
- Snowflake 관리자 등 기본 사항에 대해 이미 잘 알고 있는 분

주요 개념:

- 고급 데이터 이관
- 성능, 보안, 신속한 애자일 개발
- 데이터 공유 기능 및 스케줄링

가격:

- 3일:
 - 퍼블릭과정 - \$1,200 (인당)
 - 프라이빗과정 - \$10,500 (과정당), 최대 15명까지



Inside the Data Cloud

Snowflake의 워크로드와 다양한 리소스를 확인해보세요.

<https://www.snowflake.com/?lang=ko>

30일 무료 트라이얼을 시작해 보세요!

<https://signup.snowflake.com/>

Snowflake의 Industry별 솔루션을 만나보세요.

<https://www.snowflake.com/solutions/>

**Snowflake Blog 에서 더 다양하고 많은 소식을
전해드립니다.**

<https://www.snowflake.com/blog/?lang=ko>

질문이 있거나 추가 정보가 필요하신가요?
아래 QR Code를 스캔하고 정보를 남겨 주시면
최대한 빨리 연락 드리겠습니다.



[SNOWFLAKE에 연락하기](#)





감사합니다