snowflake®

CUSTOMER CASE STUDY

아모레퍼시픽 스노우플레이크 도입으로 태블로 성능 개선 데이터 파이프라인 운영 효율화로 업무 거버넌스 향상

AMOREPACIFIC

한국의 대표 코스메틱 기업 아모레퍼시픽이 글로벌 데이터 클라우드 기업 스노우플레이크의 솔루션을 사용해 자사가 보유한 빅데이터 분석과 시장 확장에 활용하고 있다. 데이터 최적화를 위한 업무 파이프라인 개선에 스노우플레이크 플랫폼을 활용했다.
아모레퍼시픽은 2019년부터 빅데이터 플랫폼팀을 꾸려 운영하고 있다. 빅데이터 플랫폼팀은 약 4700개 테이블의 350테라바이트 용량의 데이터를 기반으로 ▲시각화 개발·운영 ▲분석계 데이터 통합·관리 ▲통합 데이터 플랫폼 구축·운영 ▲데이터서비스 기획·운영을 진행하고 있다.

아모레퍼시픽은 우리나라뿐 아니라 해외에서도 각광받는 세계적인 기업이다. 화장품을 비롯해 향수, 헤어, 바디케어 제품까지 다양한 브랜드를 전개하며 아름다움을 전파하고 있다. 올해 7월 한국기업평판연구소의 '화장품 상장사 브랜드 빅데이터 분석' 결과, 1위에 자리할 만큼 뷰티 업계에서 정상의 자리를 지키고 있다.

빅데이터 활용 방식 개선

아모레퍼시픽 빅데이터 플랫폼팀은 기존에 타사 솔루션을 활용하면서 느꼈던 불편을 개선하고, 데이터 파이프라인 효율화를 위해 스노우플레이크 도입을 검토했다. 아모레퍼시픽은 본래 아마존웹서비스(AWS) 클라우드를 기반으로 시스템을 구축한 상태였다. 데이터 수집이나 가공 처리는 아마존 EMR 기반 아파치 스파크(Aparch Spark)와 S3를 사용하고, 데이터 제공은 레드시프트(RedShift), 아테나(Athena), 글루(Glue) 등 여러 서비스를 조합해 사용했다.

이러한 틀을 바탕으로 2~3년간 업무를 진행한 결과 여러 운영 상의 어려움이 발생했다. 데이터양과 사용자가 늘어나면서 성능은 떨어지고 운영 비용이 증가했다. 데이터 시각화에 활용하던 태블로(Tableau)의 경우 데이터 양이 증가하며 서버 또한 증설해야 했다. 시스템 운영 인력이 부족해졌고, 데이터 정합성 이슈도 발생했다.

특히 아테나를 통한 태블로 데이터 추출 작업에서는 ▲지속적인 스케줄링 시간 증가 ▲변경 적재 프로세스의 필요 ▲선별된 기능만 적재 가능한 이슈 ▲추출 데이터와 실제 데이터 간 오차 발생 등의 문제가 있었다. 태블로 추출 과정에서 오랜 시간이 소요되고 변경 적재 프로세스를 추가 개발해야 하는 불편함이 따랐다. 그 결과 추출된 데이터와 실제 데이터값이 달라지는 문제가 발생했고, 메모리에 업로드되어 있지 않은 오래 전 데이터는 별도의 헬프데스크 요청에 따라 추출해야 하는 애로사항이 있었다.

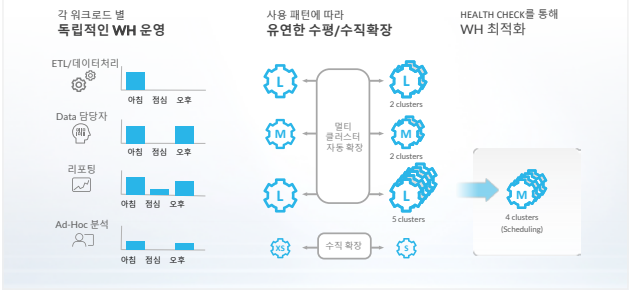
> **"스노우플레이크로 두 마리 토끼를 잡을 수 있었다. 현업은 더 빠른 서비스가 가능해졌고, 저희 또한 운영 상의 이슈를 해결할 수 있었다.** - 홍한선 아모레퍼시픽 빅데이터 플랫폼팀 담당

데이터 최적화 실현

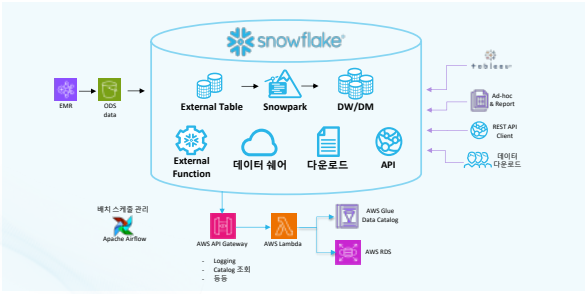
아모레퍼시픽은 스노우플레이크를 통해 문제 해결을 시도했다. 데이터 최적화에 초점을 맞췄다. AWS S3 상의 데이터를 익스터널(External) 또는 인터널(Internal) 테이블로 적재하는 과정에 스노우플레이크를 활용했다. 또한 기존 태블로 데이터는 SQL 추출 후 아테나를 통해 메모리 적재 작업을 진행하던 방식에서, UI에 각각 태블로를 띄운 뒤 원하는 테이블만 조인(Join)하는 전환 작업을 진행했다. 이러한 작업들을 통해 스케줄링 시간을 줄이고 변경 적재 과정의 어려움을 해소했으며 데이터 오차 문제도 해결했다.

데이터웨어하우스(DW) 작업 스케줄링도 스노우플레이크를 통해 용이하게 설정할 수 있었다. 각 워크로드별로 독립적인 업무 시간(WH)을 배분했다. 데이터 추출·변환·적재(ETL)는 아침에 처리하고, 데이터 담당자는 아침과 오후에 WH를 운영하는 방식으로 업무 효율을 향상시켰다. 클러스터 3~4개 가량을 사용하는 아모레퍼시픽은 멀티 클러스터 자동 확장으로 현재도 WH를 최적화하고 있다.

아모레퍼시픽의 Snowflake 사용 현황



Data Platform To-Be



적용 효과

아모레퍼시픽은 코스메틱 기업 특성상 수집한 빅데이터를 마케팅 캠페인이나 머신러닝에 활용한다. 하지만 현업 부서가 빅데이터 플랫폼팀으로부터 데이터를 공유 받을 때 애드혹(Ad-Hoc)이나 응용 프로그래밍인터페이스(API), 에어바이트(Airbite) 활용 등 부서별로 방식이 다르고 데이터 정합성이 떨어지는 문제가 있었다.

이러한 문제 또한 스노우플레이크를 통해 해결했다. SQL 명령어 한 줄로 손쉽게 데이터를 공유할 수 있다. 데이터 복사와 이동이 필요하지 않고 단일 플랫폼으로 데이터를 통합 관리해 업무에 드는 시간을 줄였다. 데이터를 일대일로 복제하는 게 아니라 단순한 공유 기능만으로 가져올 수 있어 업무 거버넌스가 향상됐다.

스노우플레이크 도입으로 태블로의 성능 이슈를 해결하고, API 제공과 데이터 공유를 통한 높은 사용성, 단일화된 스토리지 내 데이터 제공을 통한 정합성 확보 효과도 얻었다. 인력 부족 문제 또한 스노우플레이크가 제공하는 다양한 기능을 통해 별도의 개발이나 시스템 구축 절차 없이 여러 요청 사항에 대응할 수 있었다. 확장 기능을 통해 기존 AWS가 제공하던 기능 일부는 사용하되 스노우플레이크를 통한 데이터 공유와 분석 기능을 지속적으로 고도화할 계획이다.

스노우플레이크에 대하여

스노우플레이크는 데이터 클라우드를 통해 데이터를 가치있게 활용할 수 있도록 합니다. 데이터 사일로 없이 통합된 데이터를 활용할 수 있고, 데이터를 안전하게 공유하며, AI/ML 및 다양한 분석 워크로드를 지원합니다. 2023년 7월 30일 기준, Forbes Global 2000 기업 중 639개 기업을 비롯한 다양한 산업 분야의 8,537개사의 고객이 스노우플레이크 데이터 클라우드를 사용하여 비즈니스를 강화하고 있습니다. 자세한 내용은 www.snowflake.com/ko에서 확인할 수 있습니다.

© 2023 Snowflake Inc. All rights reserved.