CUSTOMER CASE STUDY

LG화학 디지털 트랜스포메이션 강력 추진 '시민 데이터 과학자(CDS)' 플랫폼 구축

LG화학은 디지털 트랜스포메이션을 조직의 핵심 전략으로 추진하면서 그 일환으로 'CDS(Citizen Data Scientist)' 플랫폼을 구축했다. CDS는 기술에 대한 특별한 지식이 없어도 누구나 쉽게 필요한 데이터를 찾고 분석할 수 있는 플랫폼이다. LG화학은 스노우플레이크와 데이터이쿠의 조합으로 CDS 플랫폼을 완성했다.

석유화학 분야의 글로벌 리더인 LG화학은 석유화학뿐 아니라 2차 전지 등 첨단소재와 생명과학 등의 강력한 성장동력을 보유한 사이언스 기업이다. 2030년까지 매출 60조 원 달성, 명실상부한 글로벌 톱(Top) 기업이 되겠다는 포부를 안고 있다.

LG화학의 디지털 트랜스포메이션

LG화학은 디지털 트랜스포메이션(DX)을 강력하게 추진하고 있는 기업이다. 지속성장을 위해 가장 필수적인 조건으로 DX를 바라보고 있다. 이를 위해서는 조직 전체가 DX에 적극적으로 참여하고, 성과를 거둬야 한다. LG화학은 DX를 성공적으로 진행하기 위해 다양한 내부 과제를 설정, 추진했다.

DX의 핵심과제 중 하나는 모든 조직원이 데이터를 자유롭게 활용하고, 그 속에서 통찰력을 얻어 비즈니스에 적극 활용하는 것이었다. 이를 위해 AWS S3와 같은 오브젝트 스토리지에 데이터레이크를 구성하고 조직 내의 데이터를 모았다.

그러나 모든 조직원이 디지털 기술에 익숙한 것은 아니기 때문에 이 데이터를 활용하는 것이 어려운 과제였다. IT 지식이 부족한 현업 담당자들은 이 데이터를 어떻게 가져와야 하는지, 질의(Query)를 어떻게 해야 하는지 알지 못했다. 데이터를 만든다고 해도 동료들에게 공유하는 방법도 알지 못했다.

이 때문에 현업 담당자들에게 데이터를 활용할 수 있는 파이썬 프로그래밍을 교육하기도 했지만, 현장에서 호응은 크지 않았다. 본연의 업무만도 바쁜 조직원들에게 부가적인 학습을 요구하는 것은 무리였다. 힘들게 DX 과제를 추진해도 일회성이 되는 경우가 많았고 과제 유지관리가 안 되는 모습이 많이 보였다.

CDS 플랫폼 구축

LG화학은 이런 문제를 해결하기 위해 CDS 플랫폼 구축에 나섰다. CDS는 Citizen Data Scientist의 약자로, IT에 전문지식이 없는 현업 담당자라고 하더라도 데이터 과학자처럼 자유자재로 데이터를 분석할 수 있는 사람을 말한다. CDS라는 용어는 가트너가 처음 사용한 말로 가트너는 "주요 업무는 통계, 분석 분야가 아니지만 고급 진단 분석이나 예측, 처방 기능을 사용하는 머신러닝 모델을 만들거나 생성하는 사람"이라고 정의하고 있다.

LG화학은 스노우플레이크와 데이터이쿠라는 솔루션을 조합해 CDS

> **"스노우플레이크는 AI나 머신러닝 기술을 계속해서 업데이트 하고 있고, 타 회사를 인수하는 등 새로운 기능을 위한 투자를 계속하고 있기 때문에 미래가 밝은 기업이라고 생각한다.** - 김낙회 LG화학 DX담당 AI 추진팀 책임

플랫폼을 구축했다. 데이터 관리 플랫폼으로 스노우플레이크를 활용하고, 데이터 분석 도구로 데이터이쿠를 사용하는 구조다.

여러 부서나 시스템에서 발생하는 데이터를 스노우플레이크에 통합하고, 이 데이터를 데이터이쿠에 가져가서 모델을 개발하거나 분석을 하는 방식이다. 데이터이쿠에서 생성된 데이터는 다시 스노우플레이크에 푸시다운 형태로 저장된다.

이렇게 구축된 CDS 플랫폼은 조직 내에서 환영받고 있다. 과거에는 데이터 분석이 필요하면 엑셀 파일을 서로 주고 받았다. 하지만 이제는 스노우플레이크에 피처 스토어(Feature Store) 개념을 도입해서 매일, 매분마다 업데이트 되는 분석용 데이터를 생성하고 있다.

또 기존에는 R이나 파이썬 등 코딩을 통해서 분석을 진행한 반면, 이제는 현업 담당자들이 프로그래밍 언어를 몰라도 사용자 인터페이스(UI) 기반으로 데이터 분석이나 시각화를 할 수 있게 됐다.

LG화학은 스노우플레이크를 엔터프라이즈 데이터 플랫폼으로 확장할 계획이다. 정형·비정형·반정형 데이터, 오브젝트 스토리지에 담겨 있던 데이터를 모두 스노우플레이크에 담아두고 BI 분석이나 머신러닝 분석 생성AI 등에 활용할 방침이다.

LG화학이 스노우플레이크를 선택한 이유

· **운영의 편리성 :** LG화학의 기존 분석 시스템은 온프레미스 시스템이었다. 시스템을 운영하기 위해 별도의 운영인력이 필요했고, 신경써야 할 일들이 많았다. 반면 SaaS 형태의 스노우플레이크는 계정을 한 번 세팅하고 레거시 데이터와 연결만 하면 추가적인 관리 비용 없이 편리하게 이용할 수 있다.

· **보안 :** 스노우플레이크는 엔터프라이즈급의 보안을 제공하는 플랫폼이다. LG화학의 CDS 플랫폼 구축 담당자는 스노우플레이크를 도입하는 과정에서 수시로 보안팀과 논의를 거쳤다. 혹시 모를 보안 위협이 없는지 체크하기 위해서였다. LG화학 데이터가 있는 VPC와 스노우플레이크의 VPC가 독립적으로 구성되어 있고, VPC 간에는 AWS 프라이빗 링크를 통해 안전하게 통신이 되기 때문에 보안 문제는 없다는 판단을 했다.

· **RBAC 설계 :** RBAC(Role Based Access Control)은 사용자의 권한에 따라 접속을 관리하는 정책을 말한다. 사용자나 사업부 별로 사용할 버추얼웨어하우스를 구성하거나, 권한에 따라 DB 테이블 접속을 통제할 수 있다. LG화학에는 국가 차원에서 보호해야 할 데이터를 보유한 부서도 있기 때문에 이런 기능이 매우 중요한 선택 요건이었다. RBAC은 각 사업부별로 사용량을 측정할 수 있다는 효과도 있다. 사용량을 기반으로 해당부서에 비용을 청구하는 것도 가능하다.

· **빠른 속도 :** LG화학의 기존 분석 시스템은 성능이 좋지 못해 이용자들의 불만이 적지 않았다. 스노우플레이크로 변경한 이후에는 훨씬 빠른 속도로 분석 결과를 받아볼 수 있게 됐다. 특히 많은 양의 데이터를 함께 조인했을 때 가격대비 성능이 우수했다. 이번 프로젝트 결과, 기존보다 성능이 10배 이상 개선된 것으로 나타났다.

· **푸시다운 :** 데이터이쿠의 데이터 저장, 처리, 트랜스포메이션 같은 작업을 푸시다운해서 스노우플레이크에서 할 수 있다. 데이터셋을 엑셀이나 스토리지에 저장을 해서 불러와서 쓰는 것보다 스노우플레이크에 저장을 한 다음에 처리를 하면 훨씬 빠른 성능을 경험할 수 있고, DB간 이동 비용을 줄일 수도 있다. 스노우플레이크는 무한정으로 확장 가능한 스토리지를 제공하고 있기 때문에 많은 유저들이 하나의 리소스를 공유하지 않아도 된다.

· **비용과 시간 :** 데이터 트랜스포메이션을 할 때 데이터가 이동하지 않고 스노우플레이크 안에 존재하기 때문에 비용 및 시간에서 많은 이득을 볼 수 있다. LG화학은 피처스토어 개념을 스노우플레이크 안에 구현했기 때문에 이용자들이 사용하는 모든 데이터가 스노우플레이크 안에 저장, 처리된다.

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.