

팀프레시 데이터를 효율적으로 정제하고 손쉽게 시각화 스노우플레이크-테블로 연동, 데이터 기반 의사결정 수행

TimF

콜드체인 물류 플랫폼을 시작으로 폭발적인 성장세를 보이며 사업을 확장하고 있는 팀프레시는 늘어나는 데이터를 정제해 효과적으로 활용하기 위한 방안으로 스노우플레이크 플랫폼을 도입했다. 그 결과 데이터 정제와 가공, 분석을 위한 전 처리 과정에 많은 시간과 노력을 들여오던 비효율성을 걷어내고, 손쉽게 데이터에 접근해 시각화된 데이터를 확인할 수 있게 됐다.

팀프레시는 3자 물류(3PL)와 유통을 중심으로 사업을 펼치고 있는 기업이다. 콜드체인 물류센터나 차량을 기반으로 새벽배송과 당일·택배배송 서비스를 제공하고 있고, 상품 재고관리와 출고업무 전반을 대행하는 풀필먼트 사업도 진행하고 있다. 주요 고객사는 자체 물류망 구축에 어려움을 느끼고 있는 중소 이커머스 기업들이다. 팀프레시는 신선식품, 의류, 화장품에서부터 식자재 유통 서비스, 기업 전문 패키지 보험 컨설팅, 화물 주선 플랫폼으로 사업을 빠르게 확장해나가고 있다. 팀프레시는 2018년 5월 설립 이후 현재까지 2000억 원 규모의 투자를 유치한 예비 유니콘 기업이다.

팀프레시의 기존 데이터 아키텍처

팀프레시는 윈드라이브, 구글 시트, IDC 서버(DB) 등 다양한 사내 시스템과 아마존웹서비스(AWS) 클라우드 인프라에 새벽배송과 창고관리와 관련된 수많은 데이터를 보유하고 있었다. 이같은 데이터를 아파치 에어플로우(Airflow)를 이용해 가져와 IDC 데이터웨어하우스에 로우(Raw) 데이터를 저장해왔다. 이렇게 저장된 로우 데이터를 지표와 통계 데이터로 정제해 데이터마트를 구축했다. 데이터 소

비자들은 플라스크(Flask) 서버에서 데이터를 확인하거나 스케줄러를 이용해 매일 전송된 슬랙 보고를 확인하고, 테블로(tableau)에서 시각화된 데이터를 확인했다.

문제점

이같은 데이터 아키텍처에는 여러 문제점이 있었다. 우선 데이터가 실시간으로 변화하는 상황에서 데이터를 정제하는 데 많은 시간이 소요됐다. 데이터 담당자들이 데이터 분석보다는 데이터를 정제하고 조합해 가공하는 전처리에 더 많이 매달리는 상황이 이어졌다.

데이터 소비자들이 원하는 데이터를 찾아보는 것도 쉽지 않았다. 슬랙에서는 꽤 많은 시간을 들여 필요한 보고 데이터를 일일이 검색해야 했고, 필요한 데이터를 제대로 시각화할 수 없었다. 그러다보니 휴먼 에러가 발생하기도 했다.

“스노우플레이크는 데이터 조력자입니다. 원하는 부분들이 이미 준비돼 있어서 우리는 잘 쓰기만 하면 됩니다. 우리가 어떻게 잘 쓸 것인지 도와주는 조력자 역할을 하고 있습니다.” - 엄성렬 팀프레시 솔루션 아키텍처팀장

해결책

팀프레시는 빠르게 돌아가는 새벽배송·물류 환경에서 비즈니스에 효과적인 인텔리전스를 확보하기 위해서는 데이터를 빠르게 정제하고 가공할 방법을 찾아야 했다. 한정된 인력 상황에서 데이터 소비자들이 매우 쉽게 필요한 데이터에 접근할 수 있는 방안을 찾은 끝에, 스노우플레이크를 도입했다.

스노우플레이크를 도입하면서 팀프레시는 기존 아키텍처를 재구성했다. 데이터를 아마존웹서비스(AWS)의 글루(glue)를 이용해 RDS(Relational Database Service) 시스템에 S3(Simple Storage Service) 증분 백업을 진행한 뒤, 스노우플레이크에 저장하는 절차를 진행했다. 또 스노우플레이크에서 제공하는 대시보드를 보고서 자료로 활용하고, 위치 데이터를 기반으로 지도가 포함된 보고형 데이터들은 스노우플레이크와 태블로를 연동해 제공하도록 구현했다.

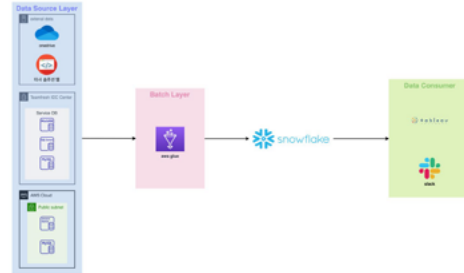
결과

스노우플레이크 도입 효과는 예상보다 더 컸다. 우선 아키텍처를 단순화해 데이터 자원 관리 포인트가 크게 감소했다. 기존 아키텍처상에 존재하던 IDC 에어플로우 애플리케이션, 각종 DB 리소스, 플라스크 웹서버 등 데이터 정제와 가공을 위한 작업을 더 이상 할 필요가 없어졌다. 스노우플레이크의 서버리스 환경 특성으로 인해 서버 관련 리스크도 사라졌다. 이전 데이터웨어하우스에서는 오랜 시간이 소요됐거나 불가능했던 조회도 빠르게 진행할 수 있게 됐다.

스노우플레이크 도입 후 쿼리 튜닝이나 인덱스 등과 같은 데이터 가공 작업도 크게 줄어 데이터 담당자들은 분석 처리(OLAP) 업무에 전념할 수 있게 됐다. 필터링 기능도 적시에 제공해 각 부서에 필요한 데이터를 빠르게 제공, 공유함으로써 데이터 사일로도 줄었다. 실제로 쿼리 데이터 조회 수치는 기존 대비 평균 20% 상승했다는 결과를 얻었다. 이전에는 조회할 수 없었던 쿼리는 제외할 수 치다.

무엇보다 기존에는 쿼리 조회가 불가능해 수행하지 못했던 데이터

TO-BE 환경에 대한 구성진행



를 모두 대시보드 화면에 표출해 시각화할 수 있게 됐다. 팀프레시 시스템 안에서 고객사가 받는 배송지역별 주문 건수와 라스트마일 배송 현황, 고객사별 현황을 한 눈에 볼 수 있게 됐다.

태블로와의 연동 효과도 크다. 태블로에서 지원되는 지도 데이터 기능으로 배송 현황을 파악할 수 있게 됐을 뿐만 아니라 배송 기사들의 배송구역 지정, 물류센터 선정 등 다양한 의사결정과 비즈니스 전략 수립에 활용하고 있다.

향후 계획

팀프레시는 앞으로 고객사에게 스노우플레이크와 태블로를 활용해 다양한 주문 관련 데이터를 제공할 계획이다. 아울러 스노우플레이크의 스트림릿(Streamlit)을 도입해 데이터 예측 모델링을 바탕으로 한 지표를 개발할 예정이다. 데이터 예측으로 물류 흐름을 보다 정확하게 예상할 수 있게 될 것으로 전망된다. 이같은 계획은 고객사 이탈 방지, 도급 비용 절감과 생산 효율성 향상에 기여할 것으로 팀프레시는 기대하고 있다.

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.