

고객 사례

# 빠르게 진화하는 이커머스 데이터 관리를 위한 통합 플랫폼

**4x** 기존 하둡 맵리듀스 코드 대비 4배 성능 향상      **25%** 기존 하둡 코드 대비 1/4의 자원 사용

업종: 유통 및 소비재  
 솔루션: 주문 픽업 최적화  
 플랫폼 사용 사례: [Delta Lake, Databricks SQL](#)  
 파트너: [Tableau](#)  
 클라우드: [Azure](#)



"Gmarket은 End to End 데이터 과학이 가능한 플랫폼이며, 데이터브릭스는 Gmarket의 도전에 함께 동행하고 있습니다."

— Gmarket 플랫폼 기술팀 서대홍 매니저

## 빠르게 변화하는 이커머스 환경과 데이터 요구 사항

전 세계 이커머스 플랫폼들은 빠른 속도로 변화하고 있는 이커머스 환경과 급증하는 디지털 쇼핑 수요에 발맞추어 빅데이터와 인공지능 기술을 활용한 데이터 솔루션을 모색하고 있습니다. 이러한 흐름에 따라 국내외 다양한 시장에서 수백 만 명의 고객을 보유하고 있는 Gmarket은 이커머스 분야의 선두주자로서 고객 만족도를 높이기 위해 데이터 분석에 기반 된 최적화된 상품 추천, 개인화된 서비스, 편리한 구매 경로 등을 제공하는 데에 진력하고 있습니다. Gmarket은 '상상하는 모든 거래'를 실현하기 위해 오픈마켓에서 생성되는 대용량 데이터를 신속하고 정확하게 분석하려는 것을 미션으로 삼아 성장의 발판을 마련하고자 했습니다. 하지만 시간이 지날수록 더욱 방대해지고 있는 데이터 요구 사항을 충족하기에는 레거시 시스템이었던 하둡 온프레미스 시스템은 여러 측면에서 부족함이 제기되었습니다. 이에 Gmarket은 차세대 빅데이터 플랫폼의 필요성을 인식하고 데이터브릭스를 도입하여 대용량 데이터 관리 문제를 해결하고 통합 레이크하우스(Lakehouse) 플랫폼을 통해 효율성의 증가와 함께 비용 절감의 효과를 경험할 수 있었습니다.

## 유연하고 안전하며 효율적인 시스템을 모색

Gmarket은 국내 시장에서 가장 큰 이커머스 플랫폼을 운영하고 있어 국내외 소비자 시장을 아우르며 다양한 카테고리 온라인 거래를 지원하고 있습니다. 이커머스 분야의 오픈마켓 강자로서 셀러와 바이어의 연결 구축에 초점을 두고, 데이터 기반 기술을 핵심적으로 활용하여 셀러에게는 최상의 비즈니스 환경을 제공하고, 바이어에게는 차별화된 고객 경험을 제공하고 있습니다.

하지만 기존의 하둡 온프레미스 시스템으로 빠르게 변화하는 이커머스 비즈니스 환경과 빅데이터 요구 사항을 충족하기에는 여러 어려움이 있었습니다. 하둡 시스템 사용은 점차 레거시화 되어서 새로운 기술의 도입에 적절히 대응할 수 없는 상황이 초래되었고, 이로 인해 전체 생산성에서 기술 부채의 비중이 점차 증가했습니다. Gmarket 플랫폼 기술팀의 서대홍 매니저는 이를 빚대어 데이터브릭스를 도입하기 전에 "기술 부채가 부채를 낳는 상황이 도래했다." 라고 말했습니다.

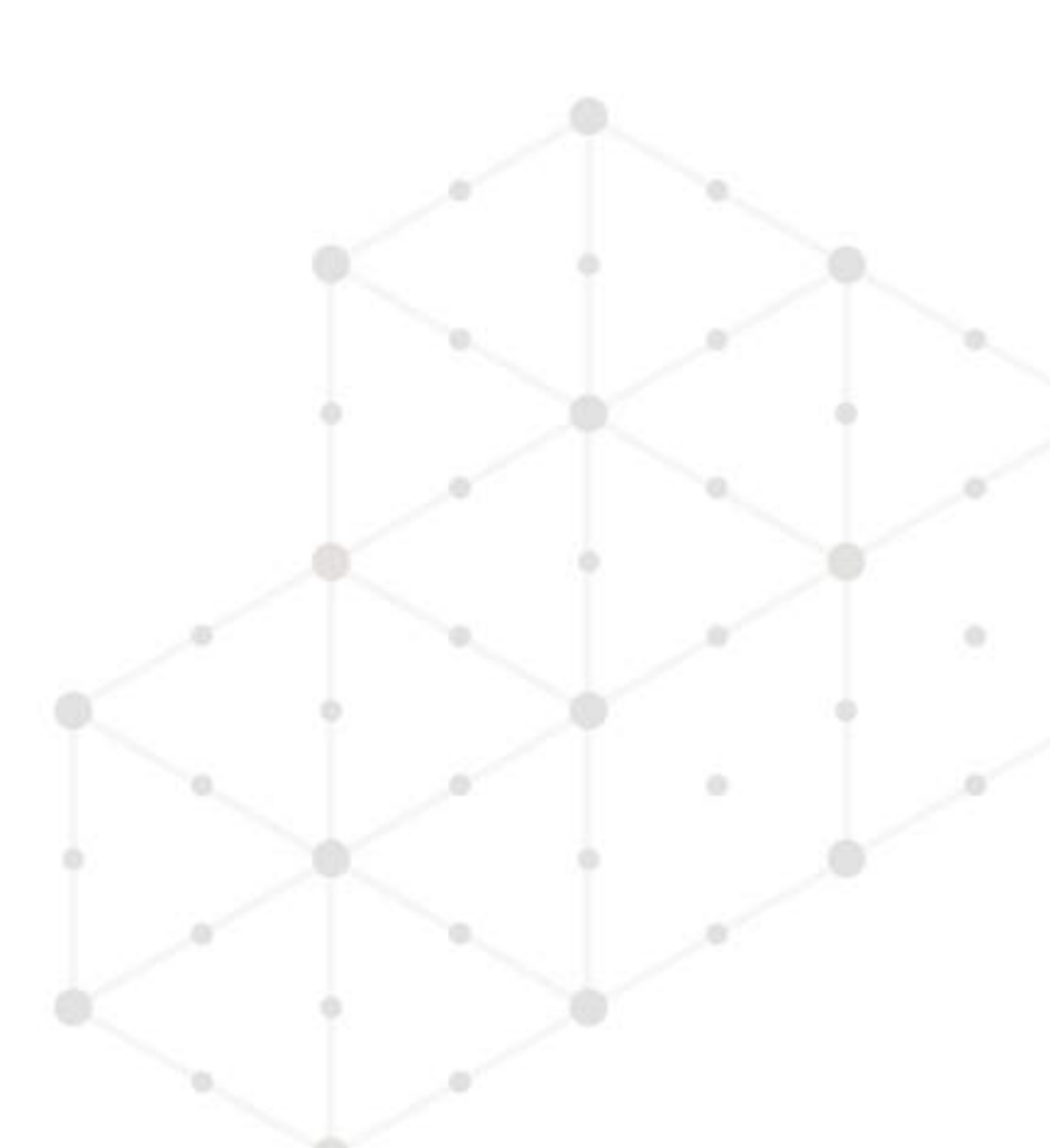
데이터브릭스를 사용하기 전에는 대용량 데이터 관리 문제가 Gmarket 기술 팀의 가장 큰 과제였습니다. 레거시 시스템에서는 계속 증가하는 대용량 데이터 분석 및 저장 요구를 충족하기 위해 장비의 추가와 인프라의 구축에 많은 시간과 비용이 요구되었습니다. 또 하둡 에코 시스템은 여러 개의 전용 솔루션으로 구성되어 있어 데이터 파이프라인의 작성에 상당한 학습 곡선이 요구되었기 때문에 신입 직원이나 직원 교체 시에도 기존의 여러 가지 솔루션을 조합하여 파이프라인을 구축하는 작업은 많은 시간과 리소스 소비가 소요되었습니다. 이에 따른 시스템의 복잡성과 유지 보수 비용도 증가했습니다. 이는 결국 이커머스 비즈니스 성공의 핵심 요소인 시장 진입 시간을 축소하는 데에 제약 요인으로 작용했습니다.

서대홍 매니저는 "하둡보다 유연하고 안전하며 효율적인 솔루션이 필요하다."라고 강조했습니다. 특히, 빠른 확장이 가능한 유연한 인프라, 독립 확장이 가능한 처리 및 저장 자원, 빠른 범용 분산 처리 등 다양한 요구 사항을 소화할 수 있는 통합 플랫폼의 도입이 시급함을 체감했습니다. 그리하여 Gmarket은 시대에 맞는 기술 요구 사항의 변화를 인식했고 결국 레거시 시스템의 전환을 결심했으며, 이 문제를 해결하기 위한 데이터 클라우드 시스템의 도입 방안을 모색하기 시작했습니다.

## 차세대 빅데이터 플랫폼, 데이터브릭스 레이크하우스 도입

Gmarket은 차세대 빅데이터 플랫폼의 필요 기술을 파악한 다음에 이상적인 플랫폼을 '클라우드 네이티브 데이터 레이크하우스 플랫폼'이라고 명명하고 이에 부합하는 플랫폼을 찾기 시작했습니다. Gmarket이 가장 중요하게 여겼던 조건은 다음과 같습니다. 먼저, 클라우드의 장점을 최대한 활용하고 자동 스케일 인/아웃을 통한 신속한 확장 능력, 분리 아키텍처를 통해 독립적으로 확장이 가능한 처리 성능과 저장 능력, 인-메모리 기반의 분산 데이터 파이프라인 처리 엔진을 사용하여 하둡보다 더 빠른 처리 속도 능력이 필요했습니다. 그리고 데이터 레이크(Data Lake)과 데이터 웨어하우스(Data Warehouse)를 결합한 데이터 레이크하우스(Data Lakehouse)를 통해 하나의 플랫폼에서의 데이터 처리 능력 등이 필요했습니다.

Gmarket은 이와 같은 다양한 조건에 부합하는 차세대 빅데이터 플랫폼으로 데이터브릭스 레이크하우스를 선택했습니다. 분산 처리 오픈 소스인 아파치 스파크(Apache Spark)의 커미터들이 데이터브릭스 창업에 참여했기 때문에 데이터브릭스가 지니고 있는 강력한 인력, 이에 따라 데이터브릭스가 오픈 소스 생태계에 큰 영향력을 미치고 있음을 높이 샀습니다. 이와 함께 데이터브릭스 플랫폼의 뛰어난 확장성과 탄력성을 고려했습니다. 수용에 따른 스케일 업/다운 작업을 통해 적절한 하드웨어를 선택할 수 있으며, 부하에 따라 자동으로 자원을 조정하는 자동 스케일 인/아웃 기능을 높게 평가했습니다. 또 클러스터 미사용 시에 자동 종료와 분리 아키텍처를 통해 필요에 따라 각각의 자원을 확장 및 축소할 수 있는 효율성을 높이 샀습니다. 그리고 데이터브릭스 플랫폼의 우수한 편의성을 고려했습니다. 기존의 시스템과는 달리 데이터브릭스의 레이크하우스는 잡 수행 시에 노트북 소스를 바로 사용할 수 있는 장점을 지니고 있었으며, 하나의 통합된 플랫폼에서 더욱 편리하게 개발할 수 있다는 장점 또한 도입 결정의 중요 요소로 손꼽았습니다.



## 데이터브릭스로 데이터 파이프라인의 성능 및 비용 효율성 증가

Gmarket은 데이터브릭스의 도입을 통해 혁신적인 성과를 얻었습니다. 기존에 사용하던 하둡에서 개발된 맵리듀스(Map Reduce) 코드를 데이터브릭스를 이용해 재개발하여 성능을 4배나 향상하고 자원 사용을 1/4로 획기적으로 절감하는 성과를 올렸습니다. 이로 인해 데이터 파이프라인의 성능과 비용 효율성이 크게 향상되었습니다.

Gmarket은 기존 레거시 시스템에서의 복잡한 데이터 파이프라인의 요구로 스트림 데이터 처리에 어려움을 겪었지만 데이터브릭스의 하나로 통합된 플랫폼을 사용하여 간결하게 데이터 파이프라인을 개발하고 운영할 수 있게 되었습니다. 이와 같은 도입의 결정으로 프로젝트 리딩 시간은 50% 이상 개선되었으며, 이에 따라 개발 생산성과 속도 및 편의성이 크게 향상되었습니다.

그리고 Gmarket은 데이터브릭스를 통해 데이터를 한 곳에 통합하여 종합적인 데이터 분석을 수행할 수 있게 되었습니다. 이를 통해 통합적인 인사이트를 얻을 수 있었으며, 이는 데이터의 가치를 극대화하는 데에 큰 도움이 되었습니다. 또 데이터브릭스는 클라우드 네이티브 환경에서 리소스를 분리하여 사용할 수 있는 기능을 제공하여 피크 타임(peak time)에도 데이터 분석 작업이 원활하게 수행될 수 있게 지원했습니다. 이는 리소스 관리의 유연성을 높여 효율적인 데이터 처리와 분석을 실현하고 비즈니스 성과를 크게 향상하는 데에 도움을 주었습니다.

특히, 데이터브릭스의 작업 조정(Job Orchestration) 기능과 노트북을 통해 빠르게 파이프라인을 배포하고 관리할 수 있는 기능으로 파이프라인의 개발과 관리의 효율성을 높였습니다. 이는 데이터 처리와 분석 과정에서 높은 생산성을 달성하고, 비즈니스 목표를 실현하는 데 큰 도움이 되었습니다. 동시에 인프라 준비와 플랫폼 학습에 소요되는 리드 타임의 감소와 데이터베이스를 기반으로 통합적인 개발을 통해 개발 속도를 높이고 업무 효율성이 향상될 수 있었습니다.

## 데이터 과학의 실현

Gmarket은 데이터브릭스의 빠르고 효율적인 확장 및 처리 능력을 활용하여 효율성을 증가하였고, 시간과 비용 측면에서도 크게 절감하는 데에 성공했습니다. 레이크하우스의 통합적인 플랫폼은 데이터 팀뿐만 아니라 다양한 부서 간의 협업을 개선하여 혁신적인 비즈니스 성과를 거두었습니다. 이와 같이 Gmarket은 기존 시스템상에서 겪었던 어려움을 데이터브릭스를 통해 극복하고 더욱 고도화된 데이터 성숙도를 달성해 나가고 있습니다. Gmarket 플랫폼 기술팀의 서대홍 매니저는 "Gmarket은 End to End 데이터 과학(Data Science)이 가능한 플랫폼이며, 데이터브릭스는 Gmarket의 도전에 함께 동행하고 있다."라고 말하며, 현재와 그리고 향후 Gmarket의 데이터 레이크하우스 로드맵에서의 데이터브릭스와의 협업을 강조했습니다.

# 시작할 준비가 되셨나요?

[Databricks 무료로 시작하기](#)      [Databricks 제품 더 알아보기](#)

전문가 상담

