

GC녹십자지놈 GC녹십자지놈



IBM Elastic Storage System 3000

고객 사례

GC녹십자지놈

스토리지 인프라 고도화로 유전체 분석 시스템 혁신

2013년 설립된 GC녹십자지놈은 GC녹십자의 가족사로 임상 유전체 분석에 특화된 기업입니다. GC녹십자지놈은 질병 진단, 예측, 맞춤형 치료를 위해 유전체 분석 정보를 활용하고 있으며, 연구 목적 기관과 달리 병원의 임상 데이터 분석에 집중하면서 임상 유전체 분석 분야를 선도하고 있습니다. 또한 GC녹십자 지놈은 차세대 염기서열 분석기(NGS) 등 첨단 장비를 활용해 다양한 프로젝트에 참여하고 있는데, K-DNA 프로젝트의 일환으로 추진되는 1차 시범 사업 수주를 계기로 첨단 분석을 위한 인프라 고도화에 나섰습니다. 그 첫 시작으로 IBM ESS를 중심으로 올플래시, 디스크, 테이프 스토리지 계층화 기반을 마련했습니다.

고객 상황

GC녹십자지놈은 시장의 요구에 맞는 유전체 분석 서비스를 제공합니다. 이를 위해 실제 협력 병원에서 수집한 임상 데이터를 사용해 분석을 수행하고, 이를 바탕으로 시장이 요구하는 다양한 유전체 분석 검사 항목을 출시하고 있습니다. 더불어 중요 국가 과제에도 적극적으로 참여하고 있습니다. 관련해 GC녹십자지놈은 K-DNA 사업에 앞서 시행되는 한국인 5천 명 유전체 분석 시범 사업을 수주하였습니다. 이 사업은 수주 여부를 떠나 여러모로 GC녹십자지놈에게 의미가 큼니다. 본격적인 바이오 시장 성장이 곧 시작된다는 신호탄이기 때문입니다.

유전체 분석 분야는 거대한 시장 기회를 맞이하고 있습니다. 빅데이터3법 시행으로 규제 수위가 낮아져 바이오 빅데이터의 비즈니스 가치가 커질 것이라 기대와 함께 디지털 뉴딜 정책의 일환으로 바이오 빅데이터 구축이 본격화되는 시기가 성큼 다가오고 있습니다. 이런 분위기 속에서 K-DNA의 일환으로 추진되는 시범 사업 수주는 앞으로 더욱 커질 시장 기회에 앞서 무엇을 개선하고 강화해야 하는지를 알려주는 일종의 나침반과 같았습니다.

GC녹십자지놈 GC녹십자지놈



IBM Elastic Storage System 3000
고객 사례

GC녹십자지놈

GC녹십자지놈의 눈에 들어온 개선 대상은 바로 IT 인프라 운영 전략이었습니다. GC녹십자지놈의 IT 환경은 다른 유전체분석 기관과 크게 다르지 않았습니다. 일루미나 시퀀싱 장비에서 나온 데이터를 HPC 인프라를 통해 분석하는 환경이었습니다. 고성능 HPC와 스토리지를 핵심 인프라 요소로 관리하던 것이 기존 전략이었습

니다. GC녹십자지놈은 유전체 분석 시장 성장세에 맞춰 가려면 IT 인프라가 아니라 데이터 센터라는 더 큰 개념을 전제로 전략을 수립해야 한다고 판단하고 자체 데이터센터를 구축하였습니다. 이 과정에서 기존 사업과 과제를 병행하다 보니 확장 요구가 있었고, 이에 대응하는 가운데 예상치 못한 문제가 불거졌습니다. 바로 스토리지 I/O 병목이 데이터센터 확장의 발목을 잡았던 것입니다.

솔루션

GC녹십자지놈은 차세대 데이터센터 전략에 맞는 스토리지 인프라의 중요성을 체감하고 곧 새로운 솔루션 도입 검토에 나섰습니다. 이에 고려하기 시작한 솔루션이 분산 파일 시스템이었습니다. 초기에는 Lustre와 같은 오픈소스를 포함해 여러 솔루션을 살펴보았지만, 핵심 역량이 필요한 주요 업무에 고성능과 안정성이 절대적으로 요구되었던 GC녹십자지놈은 IBM ESS 스토리지 도입을 결정했습니다.

오픈 소스 기술인 Lustre를 범용 스토리지에 적용하는 것은 아무래도 안정적인 기술 지원에 대한 우려가 있었습니다. 이와 달리 IBM ESS 스토리지는 시장에서 충분히 검증된 장비에 IBM Spectrum Scale(이전 General Parallel File System(GPFS))의 이점까지 누릴 수

있었습니다. 여기에 성능도 GC녹십자지놈이 원하는 수준 이상을 보여 주었습니다. 관련해 GC녹십자지놈 이태현 박사는 “Lustre 같은 환경이 필요했는데, 검토한 결과 Spectrum Scale이 우리 기준에 잘 맞았습니다.”라며 “성능의 경우 40~50Gbps 수준의 I/O가 나오길 원했는데, IBM ESS는 기대를 넘어서는 성능을 보여 주었습니다.”라고 말했습니다.

한편 GC녹십자지놈은 IBM ESS GL2S와 ESS 3000 스토리지를 들여오면서 스토리지 계층화 인프라를 완성하였습니다. 바로 IBM ESS 스토리지, TS4500 테이프 라이브러리, Spectrum Archive 솔루션을 통하여 핫 데이터, 콜드 데이터에 대한 장기 아카이빙을 위한 체계를 정립한 것입니다. 이에 따라 GC녹십자지놈은

GC녹십자지놈 GC 녹십자지놈



IBM Elastic Storage System 3000
고객 사례

GC녹십자지놈

높은 성능이 요구되는 유전체 데이터 분석부터 안정적으로 장기 보관이 필요한 아카이빙 데이터까지 일관성을 갖고 ILM(Information Lifecycle Management)을 수행할 기초를 마련하였습니다. 관련해 GC녹십자지놈 이태현 박사는 “성능이 빠른 스토리지 도입이 원래 목표였지만, IBM ESS를 도입하고 나서 보니 성능을 넘어 스토리지 계층화를 통한 운영 효율성까지 높이는 효과를 거두었습니다.”라고 말했습니다.

그리고 GC녹십자지놈은 IBM ESS 스토리지 도입을 통해 단일 네임스페이스 상에서 초고속 데이터 분석 스토리지를 확장해 나갈 수 있는 유연성도 확보 하였습니다. 이로써 새로운 서비스나 공공 프로젝트 추진 등에 따른 성능 확보가 요구될 때 신속한 대응이 가능해졌습니다.

 IBM Spectrum Scale



도입 효과

GC녹십자지놈은 스토리지 인프라 고도화를 통해 유전체 분석 역량을 한층 더 강화하였습니다. 기존 스토리지와 비교할 때 단일 분석 노드당 I/O 성능은 3배 이상 높아졌습니다. 스토리지 I/O 성능 향상은 분석 서버의 자원 효율성도 높였습니다. 기존에는 스토리지 병목으로 서버 CPU 자원이 제 성능을 발휘하지 못했습니다. 평균 CPU 사용률이 40~50%에 그쳤을 정도입니다. 그러던 것이 IBM ESS 스토리지 도입 후부터는 CPU 사용률이 크게 높아졌습니다. 자연히 유전체 분석 워크플로우는 더욱더 빨라졌습니다. GC녹십자지놈은 워크플로우 가속을 위해 향후 유전체 분석에 최적화된 FPGA

기반 하드웨어 도입도 검토 중입니다.

한편, 이번 스토리지 인프라 고도화는 GC녹십자지놈이 중장기 목표 아래 데이터 자산화와 데이터 거버넌스 체계 구축의 기초를 다지는 효과까지 이어집니다. GC녹십자지놈 이태현 박사는 “우리가 다루는 유전체 데이터는 양이 큰 편입니다. 한 사람당 1TB 정도 데이터가 나오는데, 일년에 15만 건 정도를 분석합니다. 이들 데이터는 모두 보관을 해야 합니다. 따라서 분석 단계에서는 더 빠른 성능의 스토리지가 필요하고, 대용량 데이터를 경제적이고 안정적으로 장기 보관할 수 있는 장치

GC녹십자지놈 GC녹십자지놈



IBM Elastic Storage System 3000
고객 사례

GC녹십자지놈

도 있어야 합니다. IBM ESS 스토리지, TS4500 테이프 라이브러리, Spectrum Archive 솔루션을 통해 우리는 성능과 용량 모든 측면에서 스케일아웃이 유리한 최신 스토리지 인프라를 갖출 수 있게 되었습니다” 라고 말했습니다.