

## 데이터의 진정한 가치를 만나다

데이터옵스(DataOps) 세 가지 활용 분야

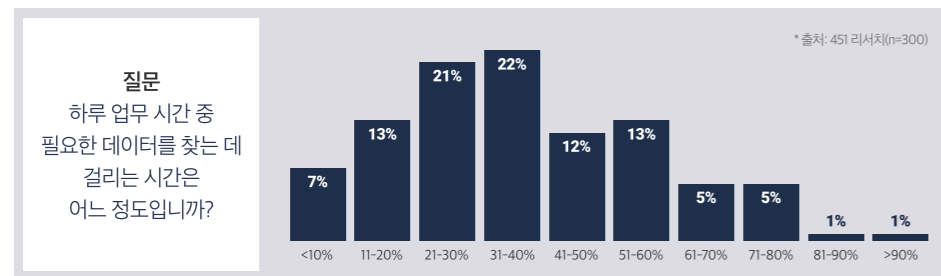
비즈니스 민첩성 향상 및 수익 극대화, 비즈니스 통찰력과 의사결정 개선을 위해 데이터 관리 기술과 비즈니스 프로세스를 개선하려는 기업이 늘고 있다. 이 가운데 특히 주목해야 할 것이 데이터옵스(DataOps)다. 데이터옵스는 데이터 보안 및 개인정보 보호, 데이터 사일로 문제, 분산된 데이터 증가 등 핵심 데이터 관리의 어려움을 극복할 수 있는 최적의 방법이기 때문이다. 451 리서치가 최근 실시한 데이터옵스 현황 설문조사 결과를 통해 데이터옵스를 활용할 수 있는 세 가지 분야에 대해 알아보자.

### 데이터옵스, 이젠 선택이 아닌 필수

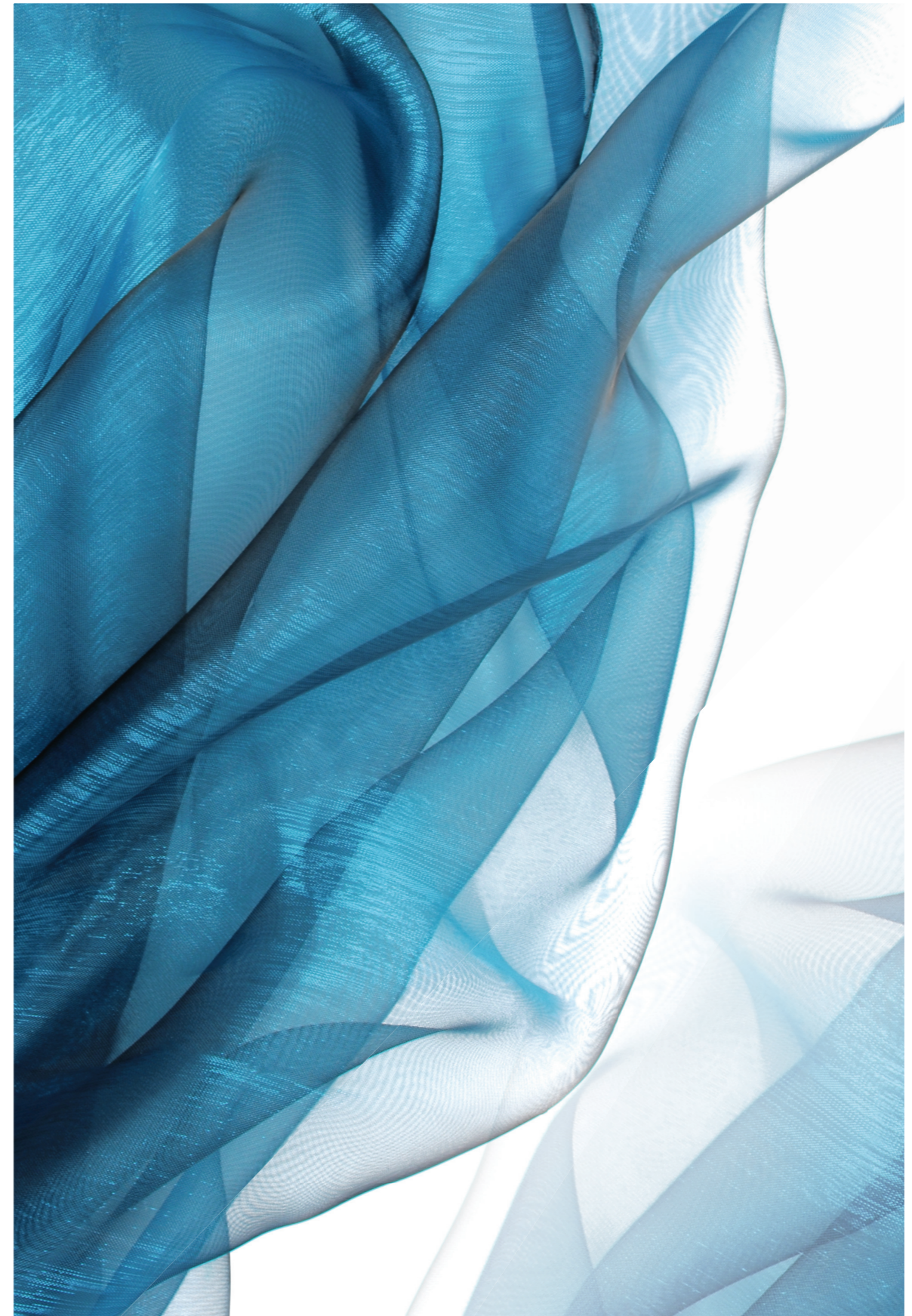
451 리서치의 설문조사에 참여한 응답자의 80%는 기업 내 데이터 소스가 100개 이상인 것으로 조사됐다. 데이터 분석가나 비즈니스 분석가의 요청에 걸맞은 데이터 환경을 준비하기 위해 하루 이상이 소요된다고 답한 응답자 비율은 92%에 달했다. 응답자의 30%는 미가공 데이터로부터 통찰력을 얻기까지 일주일 이상 걸린다고 답했으며, 50% 이상은 3일이 넘는다고 응답했다. 급속도로 변화하는 비즈니스 환경에 대응하기에는 지나치게 느린 속도다.

데이터옵스 현황에 대한 설문조사는 기업이 데이터 탐색 및 접근을 위해 상당한 시간을 투자하고 있음을 시사한다.

↓ 데이터를 탐색하고 접근하는 데 보내는 시간



데이터옵스는 해마다 급증하는 데이터 관리의 어려움을 극복할 방법으로 꼽힌다. 응답자의 81%는 자사의 성공에 데이터옵스가 긍정적 영향을 줄 수 있다는 데 동의했다.





## 데이터옵스 적용 분야

데이터옵스의 가장 큰 이점으로 분석 프로젝트의 효율성 개선을 들 수 있지만 이는 데이터옵스의 여러 이점 중 일부에 불과하다. 데이터옵스가 필요한 주요 분야로 아래의 세 가지를 꼽을 수 있다.

**1** 민첩한 운영을 위한 데이터옵스

**2** 분석 이니셔티브를 위한 데이터옵스

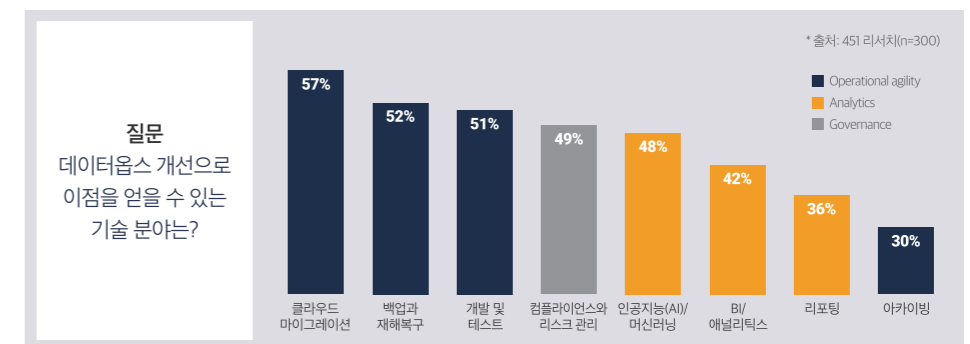
**3** 거버넌스를 위한 데이터옵스

설문 대상자의 53%가 데이터옵스의 영향력이 가장 높은 분야로 운영의 민첩성을 꼽았으며, 32%는 분석, 14%는 거버넌스 측면의 이점을 언급했다.

데이터옵스가 운영, 분석, 거버넌스 측면에서 제공하는 이점에 대해 구체적으로 살펴보기 전에 전반적인 기술적 이점도 알아볼 필요가 있다.

다음 그림을 보면, 예상되는 기술적 이점 8가지 중 4가지는 운영상의 민첩성과 관련된 것이다. 클라우드 마이그레이션(57%), 백업과 재해복구(52%), 그리고 개발 및 테스트(51%)가 가장 중요한 세 가지로 꼽힌다. 컴플라이언스와 리스크 관리의 거버넌스와 관련한 8가지 기술 혜택 중 하나에 불과하지만, 49%의 응답자는 인공지능/머신러닝(48%), BI/분석(42%), 그리고 리포팅(36%)의 세 가지 분석 관련 이점보다 더 중요하다고 답했다. 아카이빙(30%)은 운영상의 효율성 뿐만 아니라 거버넌스와도 관련이 있다.

↓ 데이터옵스와 관련한 잠재적 기술 혜택



## 1 운영의 민첩성을 위한 데이터옵스

운영의 민첩성을 위한 데이터옵스는 기업이 변화하는 비즈니스 니즈에 더욱 잘 대응하고 운영 효율성을 개선하는 데 초점을 맞춘다. 세 가지 주요 부문 중 운영상의 민첩성이 가장 중요하다고 답한 것은 '데이터옵스(DataOps)'에서 옵스(Ops), 즉 운영이 그만큼 중요하다는 것을 말한다. 데이터옵스는 데이터의 활용도를 높이는 데 초점을 두고 있으며, 이는 운영 효율성 및 민첩성 향상을 통해 얻을 수 있다.

데이터옵스를 채택하면 비즈니스 면에서 어떤 이득이 있을까. 응답자들은 인프라 활용을 통한 효율성 증진(48%), 데이터 장애 감소(36%), 그리고 자동화를 통한 데이터/IT 인력의 가용 시간 증가(16%) 순으로 꼽았다.

설문 조사 결과는 클라우드 트렌드가 운영의 민첩성을 저해하는 요인 중 하나라는 것을 보여준다. 가령, 응답자의 86%는 2개 이상의 클라우드 제공업체에 자사의 데이터가 분산되어 있다고 답했으며, 55%는 3개 이상의 클라우드를, 22%는 4개 이상의 클라우드를, 그리고 11%는 무려 5개 이상의 클라우드를 활용하는 것으로 조사됐다.

데이터 중력(Data gravity)이란 특정 환경에 많은 양의 데이터가 저장되어 있는 경우, 다른 곳으로 이동시킬 가능성이 작다는 것을 말한다. 이로 인해 다른 분석 툴이나 플랫폼에서 해당 데이터를 사용하는 것이 불가능해진다. 따라서, 복수의 클라우드 제공업체에 분산되어 있는 데이터를 제대로 관리하는 것이 중요하다.

그렇지 않으면 데이터 중력으로 인해 데이터가 통합되지 못하고 분리된 상태로 존재하게 된다. 한 부서에서 통제하는 데이터가 사내 다른 부서와 공유될 수 없기 때문에 분석 및 거버넌스와 관련된 어려움도 발생할 수 있다.

개별적으로 분리된 데이터는 다크 데이터, 즉 분석에 사용할 수 없는 데이터로 남게 된다. 다크 데이터가 존재하고, 어떤 종류의 데이터가 여러 곳에서 어떤 식으로 저장되고 처리되는지 알지 못하면 컴플라이언스 차원에서도 리스크가 발생할 수 있다. 이 문제는 기업 내에서 존재하는 데이터 레이크와 여러 클라우드에 분산된 데이터의 저장이 동시에 이루어지기 때문에 특히 중요하다.

데이터 분산으로 인한 어려움은 엣지 장치에서 생성되는 데이터가 늘면서 더욱 커지고 있다. 응답자의 66%는 엣지에서 생성되는 데이터가 전체 데이터의 20% 이상이라고 답했다. 금융 서비스와 제조업/건설업의 경우 이 수치는 70% 이상으로 더욱 높아진다. 사물인터넷(IoT)에서 생성되는 데이터양이 증가하고 생성 속도가 빨라지면서 달라진 점이 있다. 엣지에서 어떤 데이터를 처리할지 결정하는 것보다는 중앙 인프라에 어떤 데이터를 저장하고 분석할지 결정하기 위해 엣지에서의 데이터 처리가 필요하다는 점이다.

## 2 분석 이니셔티브를 위한 데이터옵스

분석을 위한 데이터옵스는 각 기관이 보유한 데이터로부터 효율적으로 통찰력을 얻는 데 초점을 맞추고 있다. 이는 전통적인 분석 기법과 머신러닝, 딥러닝 등 최첨단의 접근법을 사용해 이루어진다.

응답자들은 분석과 관련된 데이터옵스의 일차적 이점으로 다양한 데이터 소스에 액세스할 수 있다는 점과 분석 엔진과의 통합이 가능한 유연성 향상을 꼽았다(40%). 분석 이니셔티브를 위해 오브젝트 스토리지를 사용하는 추세를 고려하면 이는 특히 중요하다. 하둡(Hadoop) 환경과 달리 데이터 레이크를 기반으로 오브젝트 스토리지를 사용하면 여러 분석 엔진을 사용할 수 있다.

다음으로 가장 많이 언급되는 비즈니스 이점으로는 대내외 고객으로부터 정보 요청 시 응답 시간 단축(31%)과 생산 공정에 데이터 파이프라인 및 인공지능(AI) 모델을 더욱 민첩하게 업데이트할 수 있다는 것이다(29%).

또한 데이터 엔지니어들이 운영 데이터 요청에 부응하는 시간을 줄임으로써 얻을 수 있는 이점도 있다. 비즈니스 크리티컬한 데이터 요청 시 신속 대응(70%), 기획 및 데이터 모델링 시간 단축(59%), 분기 말이나 연말 등 실적 보고 시기의 스트레스 감소(47%) 등이 그것이다.

또한 머신러닝 모델 개발에 투자하는 기업들이 점점 증가하고 있다. 응답자의 42%는 현재 머신러닝을 사용하고 있으며, 34%는 개념증명(Proof of Concept)<sup>1)</sup> 단계에,

12%는 2020년에 머신러닝을 도입할 계획이 있다고 답했다. 머신러닝 도입을 저해하는 요소는 여러 가지다. 데이터 액세스 및 미흡한 준비가 가장 큰 이유로 꼽혔고(39%), 그 뒤를 이어 운영 시스템에 결과를 반영할 수 있는 역량 부족(35%), 제한적 예산(34%), 그리고 모델 구축 및 유지의 어려움(31%)이 이유로 꼽혔다.

머신러닝의 배포 혹은 개념증명 중이라 답한 응답자의 경우, 불과 19%만이 내부적으로 개발한 머신러닝 모델의 절반을 생산에 투입하고 있다고 답했다. 이는 시스템이 완전하지 못하기 때문이기도 하지만 머신러닝 개발과 배포가 복잡하다는 것을 의미한다.

설문 결과는 데이터옵스 채택 여부가 성숙한 단계에 있을수록, 머신러닝을 생산에 적용해 성공할 가능성 또한 더 높다는 것을 보여준다. 기업 문화에 데이터옵스가 완전히 통합되었다고 생각하는 응답자 중 42%는 사내에 설치된 머신러닝 모델이 생산에 투입되고 있다고 답했다.

1) 개념증명 : PoC, Proof of Concept. 기존 시장에 없었던 신기술을 도입하기 전 이를 검증하기 위해 사용하는 것으로, 특정 방식이나 아이디어를 실현하여 타당성을 증명하는 것을 뜻한다.

### 3 거버넌스를 위한 데이터옵스

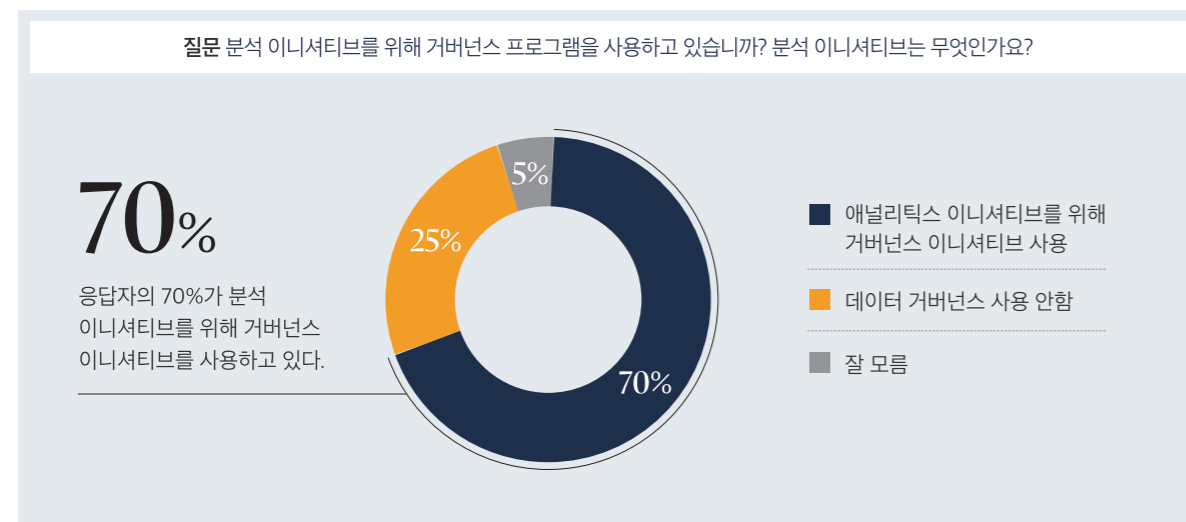
정부의 규제 요건에 대응할 수 있는 역량을 강화하는 것은 중요하다. 이것이 분석 이니셔티브를 가속할 수 있는 거버넌스 규칙을 수립하는 동시에 거버넌스를 위한 데이터옵스를 구현하는 이유다. 거버넌스를 위한 데이터옵스 측면에서 응답자들이 주로 기대하는 효과는 컴플라이언스 위반과 관련한 리스크 감소(44%)가 상위를 차지했으며, 뒤이어 법적 요건에 대한 신속한 대응(35%), 그리고 컴플라이언스 부적합 통지로 인해 생기는 변경 사항 감소(21%) 등이 언급되었다.

데이터 거버넌스와 관련한 어려움은 다양하다. 교육, 변화 관리, 그리고 거버넌스 기술/프로세스의 채택이 가장 높은 응답률을 보였으며(32%), 데이터 거버넌스 이해 당사자들의 이해 부족 및 기술 부족(31%)이 뒤를 이었다. 결과 전달의 속도를 늦추고 품질과 일관성에 영향을 미치는 수동 프로세스(28%)도 언급되었다.

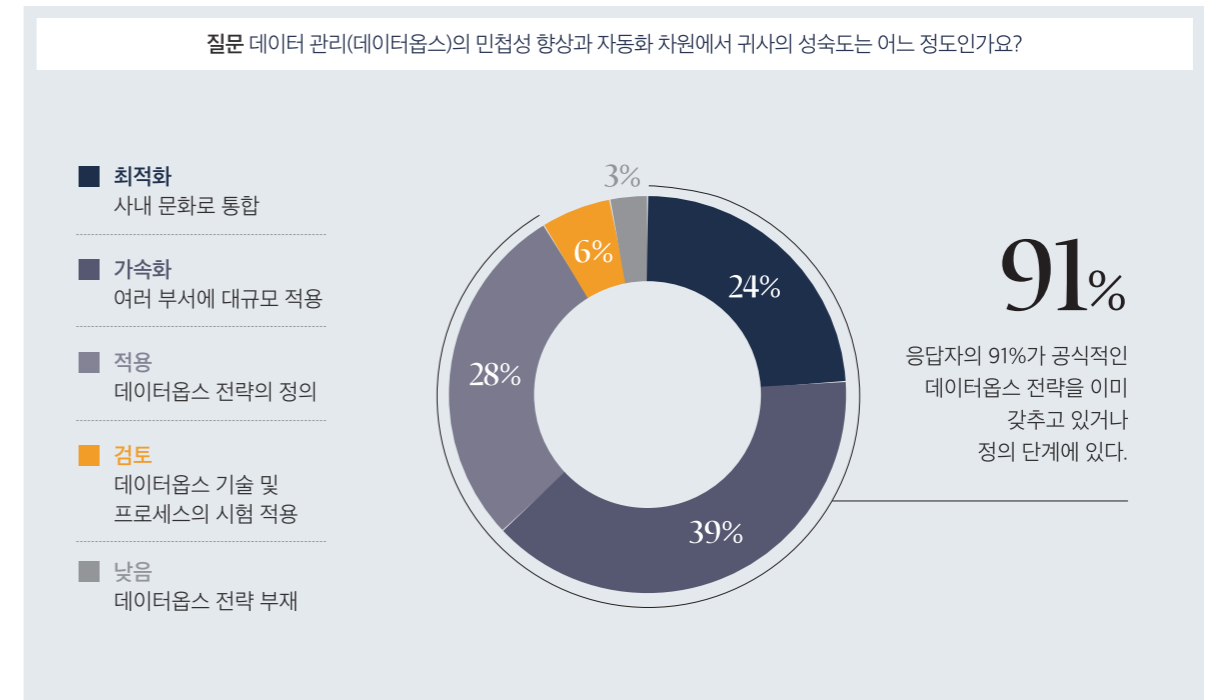
데이터 거버넌스에 대한 성숙도가 높다고 답한 응답자 일수록 어려움이 적은 것으로 나타났다. 실제로 데이터 거버넌스의 성숙도가 높은 기업의 25% 이상이 데이터 거버넌스와 관련한 어려움이 전혀 없다고 답했다. 하지만 이는 전체 응답자 중 12%에 불과했다. 데이터옵스 성숙도가 높은 기업의 21% 역시 데이터 거버넌스 관련 문제가 전혀 없다고 답했다.

데이터 거버넌스에 대해 성숙한 접근 방식을 채택할 경우 분석 이니셔티브를 가속할 수 있다는 이점이 있다. 451 리서치에서는 최근 몇 년 사이 기업들이 데이터 거버넌스를 위해 과거와 다른 방식을 채택하고 있다고 지적한 바 있다. 즉, 셀프 서비스와 애자일 분석을 통해 더욱 빠르게 이전할 수 있는 수단으로 데이터옵스를 고려하기 시작했다는 것이다.

↓ 분석에 힘을 실어주는 데이터 거버넌스



↓ 데이터옵스 이니셔티브의 성숙도



설문 조사 결과는 이러한 접근법이 더 폭넓게 채택되고 있으며, 채택한 경우 데이터 거버넌스 및 데이터옵스와 관련해서도 성숙도가 높다는 것을 보여준다. 전체적으로, 응답자의 70%는 자사의 분석 이니셔티브에 박차를 가하기 위해 거버넌스 이니셔티브를 사용하고 있다고 답했다. 데이터 거버넌스 성숙도가 가장 높다고 답한 이들의 경우 이 수치는 82%로 높아진다. 데이터옵스 성숙도가 가장 높다고 답한 이들의 경우도 81%에 달했다.

데이터 거버넌스에 대한 새로운 태도는 올바른 목적을 가진 사람들에게 데이터 액세스 권한을 제공해야 한다는 것이다. 이는 분석 이니셔티브를 위해 데이터 거버넌스를 사용한다고 답한 응답자들이 장점으로 꼽는 부분이기도 하다. 64%가 적절한 권한을 가진 이들에게 데이터가 배포되도록 보장하는 것을 데이터옵스의 장점으로 꼽았으며, 인공지능/머신러닝(61%), 그리고 전반적인 데이터 액세스(60%) 등이 뒤를 이었다.

출처: DataOps Unlocks the Value of Data, www.hitachivantara.com, 2020년 1월