

엔터프라이즈 스포트라이트 클라우드 컴퓨팅에 대한 새로운 시각



Credit: puhha/Shutterstock

오늘날 클라우드 컴퓨팅은 기업의 핵심 인프라로 자리 잡았다. 클라우드 컴퓨팅을 바라보는 IT의 시각은 개념을 이해하는 데서 벗어나 기술과 비즈니스 과제를 해결하기 위한 최선의 사용 방법을 찾는 방향으로 바뀌었다. 지금은 잠시 숨을 고르며 클라우드에 대한 생각을 정리하기 좋은 시점이다.

CONTENTS

- 2 2025년, 클라우드 아키텍처의 르네상스
- 5 “비용·AI·규제...” CIO가 클라우드 전략을 재고하는 이유
- 13 송환과 지속가능성, 2025년 클라우드 지형을 다시 그리다

FOUNDRY EDITORIAL 제공

CIO **CSO** **COMPUTERWORLD** **InfoWorld** **NETWORKWORLD**

2025년, 클라우드 아키텍처의 르네상스

David Linthicum | InfoWorld.com

‘퍼펙트 스톰’이라 불리는 복합적인 경제 위기가 다가오면서 기업은 클라우드 전략을 재고해야 하는 상황에 놓여 있다. 2025년에는 클라우드 아키텍처가 인프라 선택지를 넘어 비즈니스 무대의 중심에 서게 될 전망이다. 이는 또 하나의 하이프 사이클로 치부해서는 안 되는 중요한 변화다.

먼저 가장 큰 주제인 생성형 AI에 대한 이야기부터 해보자. 생성형 AI 모델을 실행하기 위한 컴퓨팅 요구사항을 보면 전통적인 클라우드 환경은 소폭장난처럼 느껴진다. 가트너의 예측에 따르면, 전체 클라우드 인프라 용량에서 엔터프라이즈 AI 워크로드가 소비하는 비중은 2025년 30%를 넘어설 전망이다. AI로 인한 클라우드 비용 증가를 감안하면 이 전환은 이미 진행되고 있다.

퍼블릭 클라우드 사용 확대에 따른 비용 증가는 이제 기업 경영진의 주요 고민거리로 자리 잡고 있다. 필자는 이와 같은 상황에 대해 오래전부터 경고했다. 지난 10년간 이어진 ‘리프트 앤 시프트(Lift and Shift)’

전략은 막대한 기술 부채를 남겼다. 청구서 금액은 CFO의 말문을 막히게 할 정도이며, 이제 기업은 클라우드 서비스에 책정했던 처음 예산의 2~3배에 달하는 비용을 지출하고 있다. 게다가 지금의 비용에는 AI 워크로드가 아직 본격적으로 반영되지도 않았다.

최근 들어 필자가 가장 자주 받는 컨설팅 의뢰는 IT 부서가 퍼블릭 클라우드 자원에 왜 그렇게 큰 비용을 지출하는지 분석해달라는 요청이다. 의뢰하는 사람들은 다름 아닌 CEO, CFO다. 즉, 10년 전만 해도 IT 인프라에 거의 관심이 없었던 C 레벨 경영진이다.

흥미로운 부분은 이익에 밝은 투자자가 클라우드 비용 절감을 위한 전략적 기술에 집중적으로 자금을 투입하고 있다는 점이다. 여기서 말하는 이익에 밝은 투자자란 클라우드 아키텍처에 대해 더 정교한 접근 방식을 추구하는 기관 투자자를 의미한다. 이들은 ‘퍼블릭 클라우드 올인’에 더 이상 동의하지 않는다. 하이브리드 클라우드

아키텍처, 엣지 컴퓨팅 통합, 클라우드 네이티브 최적화 패턴, 클라우드 송환에 대해 적극적으로 질문하며 기업의 클라우드 전략 전반을 면밀히 들여다보고 있다.

이렇게 생각해 보자. LLM(Large Language Model)을 실행하면서 민감한 데이터를 처리해야 한다면, 굳이 비싼 요금을 내며 퍼블릭 클라우드 GPU 인스턴스를 사용할 이유가 있을까? 프라이빗 클라우드 아키텍처가 부활하고 있는 이유가 바로 여기에 있다. 예전과 다른 부분은 지금의 프라이빗 클라우드는 AI 워크로드를 지원하는 동시에 데이터 주권을 유지하도록 설계된다는 점이다.

업계는 이 흐름을 따라잡기 위해 고군분투하고 있다. AWS·MS·구글 같은 전통적인 클라우드 서비스 업체는 더 나은 하이브리드 솔루션을 제공하기 위해 경쟁하고 있으며, 델·HPE·IBM 등의 엔터프라이즈급 IT 업체는 마침내 사용할 만한 프라이빗 클라우드 플랫폼을 갖추기 시작했다. 컨설팅 기업의 메시지는 “클라우드 파트너를 활용하십시오”에서 “지난 10년간 해온 방식을 재고해 봅시다”로 바뀌고 있다.

결론적으로 클라우드의 발전에 있어

2025년은 단순히 하나의 단계로 끝나지는 않을 것이다. 지금 나타나는 흐름은 시스템 설계 방식의 근본적인 변화다. 혁신적인 기업은 이미 이 변화를 준비하고 있다. 그렇지 못한 기업은 기술 측면에서 뒤쳐질 수 있다.

2025년을 위한 게임 플랜

다가오는 클라우드 아키텍처 르네상스에 대비하려면 여러 조치를 취해야 한다. 다행히도 방법은 복잡하지 않다. 이미 검증된 기존 기술에 새로운 개념과 프로세스를 적용하고 도입함으로써 목표를 달성할 수 있다. 이런 접근 방식은 허황된 이야기가 아니라 당장 실행할 수 있는 매우 현실적인 전략이다.

우선 내부를 정비해야 한다. 앞으로 3~6개월 동안은 현재의 클라우드 지출과 활용 패턴을 심층적으로 분석하는 데 주력한다. 중요한 것은 경영진에게 보여주는 다듬어진 수치가 아닌 실제 수치 데이터다. AI/ML 워크로드의 예상 설계도도 작성해야 한다. 이 워크로드는 현재의 추정치를 훨씬 뛰어넘어 급증할 것이다. 또한 퍼블릭 클라우드 환경에서 비용 낭비를 유발하는 워크로드도 파악한다. 분석을 마치면 많은 기업이 예상치 못한 결과를 마주하게 될 것이다.

다음으로 할 일은 합리적인 워크로드 배치 전략 수립이다. 이 과정에서는 데이터 중력, 성능 요구사항, 규제에 따른 제약을 고려한다. 최신 유행을 맹목적으로 따르지 말고 비즈니스 현실과 정렬되는 결정을 내리는 것이 핵심이다. 하이브리드 및 프라이빗 클라우드 투자에 대한 명확한 ROI 모델을 만들어야 한다.

이제 기술 아키텍처에 대한 논의가 필요하다. 초점을 맞춰야 할 영역은 데이터 파이프라인 최적화, 엣지 컴퓨팅 통합, AI/ML 인프라 요구사항 충족이다. 멀티클라우드 연결은 더 이상 선택 사항이 아니라 생존을 위한 필수 요건이다. 아키텍처를 설계하고 구성하는 과정에서 주의할 부분은 철저한 보안과 규정 준수 프레임워크를 유지해야 한다는 점이다.

기술뿐 아니라 조직적인 요소도 매우 중요하지만 대부분 기업이 이 부분에서 실수한다. 인프라 전문가, 데이터 과학자, 재무 분석가, 보안 전문가로 구성된 클라우드 비용 관리 전담팀을 구성하라. 이 조직은 또 하나의 IT팀이 아니라, 비즈니스 가치를 창출하는 핵심 기능으로 자리 잡아야 한다. 또한 투자 우선순위에도 재조정이 필요하다. 자동화된 오케스트레이션 툴, 클라우드 관리 플랫폼, 데이터 패브릭 솔루션에 초점을 맞춰야 한다.

비용의 중요성

재무 관리는 클라우드 전략에서 빼놓을 수 없는 요소다. 기업은 적절한 비용 청구 메커니즘을 구현하고 명확한 총소유비용(TCO) 모델을 개발해야 한다. 클라우드 비용 지출에 대한 책임 소재를 분명히 하는 것도 중요하다. 각 부서별로 클라우드 사용 비용이 투명하게 드러나면 책임 의식과 비용 관리 행태가 크게 달라질 것이다. 핀옵스(FinOps)는 조심해야 한다. 분명 가치가 있는 접근법이지만, 기업의 실제 상황과 장기적인 전략에 맞춰 신중하게 적용하지 않으면 잘못된 성과 지표에 의존할 수 있다.

이 같은 트랜스포메이션 여정은 평가와 계획부터 시작해 파일럿 프로젝트를 거쳐 전면적인 구현까지 대략 12~24개월 동안 진행된다. 기억해야 할 점은 이 프로젝트가 단순한 IT 프로젝트가 아니라 모든 이해관계자의 동의가 필요한 비즈니스 트랜스포메이션 이니셔티브라는 사실이다.

2025년의 승자는 클라우드 서비스에 가장 많은 비용을 지출한 기업이 아니다. 비즈니스 목표와 정렬되는 지능적이고 유연한 클라우드 아키텍처를 구축하는 기업이 될 것이다. 시장의 압력에 떠밀려 어쩔 수 없이 뒤따라가는 상황에 놓이지 않으려면 지금이 적기다. 후회하지 않으려면 지금 움직여야 한다.

“비용·AI·규제...” CIO가 클라우드 전략을 재고하는 이유

Robert Mitchell | CIO.com

기업은 지난 몇 년에 걸쳐 자본 지출 (CapEx)을 최소화하고 비용을 절감하기 위해 최대한 신속하게 퍼블릭 클라우드 환경으로 전환했다. 그러나 이제 CIO는 기존의 투자를 다시 검토하고 합리성을 따져보고 있다. 이 투자가 실제로 생산성 향상과 비용 절감으로 이어지는가?

포레스터의 수석 애널리스트 트레이시 우는 “퍼블릭 클라우드로 앞다퉈 전환할 때 많은 기업이 가격에 대해서는 깊이 생각하지 않았다”라고 지적했다. 그 결과 일부 기업에서는 연간 클라우드 지출이 큰 폭으로 증가했다. 우는 “클라우드 지출은 점점 늘어나고 예산은 줄어드는 상황에서 이제 기업은 도대체 어떤 상황인지, 어떻게 해야 상황을 바로잡을 수 있는지 묻고 있다”라고 말했다.

미국 재보험 그룹(Reinsurance Group of America, RGA)의 수석 부사장이자 CTO인 론 홀로웰은 “2025년에는 프라이빗 및

퍼블릭 클라우드 전반에 걸쳐 업무 수용 프로세스, 분배 기준, 구현 프랙티스의 성숙도를 높여 퍼블릭 클라우드 규모를 적절히 조정해야 한다. 퍼블릭 클라우드와 프라이빗 클라우드 간 워크로드 배치는 비용 최적화와 명확히 정의된 선택 기준에 의해 결정될 것”이라고 말했다.

소프트웨어 업체 엔다바(Endava)의 클라우드 역량 부문 부사장이자 대기업 CIO를 대상으로 컨설팅을 하는 라두 봄블레아는 “올해는 하이브리드 클라우드와 멀티클라우드, 온프레미스 송환에 대해 더 많은 이야기가 나올 것”이라고 말했다. 그 배경에는 예상보다 높은 비용뿐 아니라 성능과 지연 문제, 보안, 데이터 프라이버시, 규정 준수에 대한 우려가 자리한다. 데이터의 위치와 전송 경로, 처리 지점을 규정하는 지역별 디지털 주권 규정 역시 중요한 요인이다.

RGA의 홀로웰이 퍼블릭 클라우드보다

프라이빗 클라우드를 활용하는 주된 이유는 비용이다. 홀로웰은 계절적이거나 특정 기간에 수요가 집중되는 온디맨드 워크로드에는 퍼블릭 클라우드가 가장 비용 효율적이라면서 “용량 수요가 비교적 일관적인 워크로드의 경우 프라이빗 클라우드와 고정 용량 솔루션이 더 매력적인 경제성을 제공할 수 있다”라고 설명했다.

봄블레아 역시 다수의 CIO가 같은 이유를 들고 있다고 전했다. 엔다바가 관리하는 엔터프라이즈급 시스템의 80%는 퍼블릭 클라우드를 일부 또는 전면적으로 사용 중이지만, 이런 기업의 약 60%는 하나 이상의 시스템을 다시 온프레미스로 마이그레이션하고 있다. 봄블레아는 “이 흐름은 일시적 현상이 아니라 뚜렷한 추세로 자리 잡고 있다”라고 설명했다.

그렇다면 퍼블릭 클라우드에서 송환된 워크로드는 어디로 향하고 있을까? 가트너의 우는 “퍼블릭 클라우드보다 온프레미스(온프레미스 프라이빗 클라우드 혹은 호스팅형 프라이빗 클라우드)에 대한 관심이 다시 높아지고 있다. 생성형 AI 같은 데이터 집약적인 워크로드로 인해 클라우드 비용이 천문학적으로 증가하기 시작한 것이 이런 변화를 촉발했다. CIO는 애플리케이션을 온프레미스로 다시

가져오거나 온프레미스 또는 호스팅형 프라이빗 클라우드 서비스를 사용하는 방식을 통해 데이터 프라이버시를 보장하는 동시에 멀티테넌시를 피할 수 있다”라고 말했다. 이는 클라우드 리더로 분류되는 기업 5곳 중 4곳이 프라이빗 클라우드에 대한 투자를 20% 늘릴 것으로 예측되는 이유이기도 하다.

물론 2025년이 단순히 송환의 해가 된다는 의미는 아니다. 우는 “생성형 AI, 비용, 데이터 주권 문제, 성능 요구사항 등으로 인해 프라이빗 클라우드에 대한 투자가 늘고 있지만 퍼블릭 클라우드 역시 채택 확산과 생성형 AI 서비스, 인프라 사용 면적 감소, 새로운 인프라에 대한 접근성에 힘입어 투자가 증가하고 있다”라고 말했다.

퍼블릭 클라우드의 숨은 비용

세인트 주드 어린이 연구 병원(St. Jude Children's Research Hospital)의 SVP 겸 CIO인 키스 페리는 퍼블릭 클라우드가 병원 생태계에 속하지 않은 연구원에게 지식을 제공하는 좋은 방법이라고 말했다. 이 병원은 온프레미스 슈퍼컴퓨터로 연구 데이터를 대부분을 생성한다. 하지만 생성된 데이터를 퍼블릭 클라우드로 전송하고 다시 가져오는 과정에서 상당한 비용이 발생할 수 있다. 페리는 “학계는 데이터를 고성능 컴퓨팅

자원 가까이 두는 편을 선호하지만, 이 같은 이그레스 비용으로 인해 어려움을 겪는 경우가 많다”라고 말했다.

데이터 집약적인 워크로드는 지속적인 고성능 컴퓨팅이 필요한 상황에서 상당한 비용을 초래한다. 홀로웰은 “비용뿐 아니라 성능 측면에서 또 다른 요인은 데이터 이동이다. 그래서 RGA는 클라우드 간 데이터 전송을 최소화하기 위해 데이터 라이프사이클을 신중하게 관리한다”라고 설명했다.

우는 데이터 집약적인 워크로드를 퍼블릭 클라우드에서 실행할 경우 비용이 많이 드는 이유는 저장된 데이터에 대한 요금과 AZ(Availability Zone), 지역, 그리고 클라우드 간에 전송되는 데이터에 대한 요금, 두 가지가 부과되기 때문이라고 덧붙였다. 또한 클라우드 업체는 특정 AZ에서 나가는 데이터와 들어오는 데이터에 대해 모두 이그레스 요금을 부과한다. 우는 “따라서 AZ 간 전송에서는 사실상 두 번 요금이 부과된다. 이런 숨겨진 전송 요금이 상당히 커질 수 있다”라고 지적했다.

불불레아는 페타바이트 급으로 가면 데이터 전송 비용이 커지고 데이터 전송과 동기화가 복잡할 수 있다면서 “AI

프로젝트에서 클라우드 비용의 약 45%는 퍼블릭 클라우드에서 다른 위치로 데이터를 전송하는 과정에서 발생한다. 서비스를 위해 필요한 모든 요소가 포함된 시스템을 갖출 경우 솔루션 비용이 초기 추정치보다 3~4배 더 높아질 수 있다”라고 경고했다.

예를 들어 오픈AI를 사용해 AI 솔루션을 구축하는 기업은 AI 서비스 외의 다른 요소도 고려해야 한다. 기밀 정보를 보호하기 위한 보안 저장소가 필요하고, 보안 어플라이언스와 정책을 정의하고 구성해서 자격을 갖춘 사람과 서비스에만 액세스를 허용해야 한다. 안전한 스토리지와 데이터 변환, 모니터링, 감사, 규정 준수 계층은 시스템의 복잡성을 높인다. 결국 AI 서비스를 중심으로 그 주변에 엔터프라이즈 시스템의 요구사항을 충족시키기 위한 10~12개의 부가적인 클라우드 서비스를 더한 솔루션을 구축해야 한다.

광산 기업 모자이크 컴퍼니(The Mosaic Company)의 CIO 제프 와이소키는 예산 초과 가능성을 인정하면서도 퍼블릭 클라우드 서비스 업체와 협력하면 비용을 적정 수준으로 유지할 수 있다고 말했다. 실제로 최근 모자이크는 마이크로소프트의 Bing 플랫폼을 기반으로 광산 운영을 위한 데이터 집약적 안전 모델인 ‘모자이크

GPT(Mosaic GPT)’를 개발해 파일럿 운영을 앞두고 있다. 이 모델에는 모자이크가 수년간 축적한 안전 정보가 통합돼 있어 광산 현장 작업자가 특정 환경에서 안전에 대한 질문을 입력하면 적절한 처리 방법을 확인할 수 있다.

이 모자이크 GPT에 대해 와이소키는 “비용을 합리적인 수준으로 낮추기 위해 마이크로소프트와 함께 몇 가지를 변경했다”라고 설명했다. 모자이크 팀이 모델을 구축한 방식이나 마이크로소프트가 솔루션을 설계한 방식 모두 프로젝트를 예산 범위 내에 유지하는 것이 목표였다.

와이소키는 모자이크의 ERP 시스템이 초기에는 온프레미스에 위치했지만, 지금은 SAP 프라이빗 클라우드에서 운영되고 있다고 설명했다. 그러나 와이소키는 일부 서버는 여전히 온프레미스에 남아있으며, 이는 앞으로도 변함없을 것이라고 덧붙였다. 다만 엣지 서버 솔루션이 클라우드와 동기화하는 경우는 발생할 수 있다. 와이소키는 “현재 회사 IT 운영의 80~85%가 클라우드에서 이뤄지는데, 앞으로도 이 수준에서 큰 변화는 없을 전망”이라고 말했다.

예산을 초과할 수 있는 AI 프로젝트

AI/ML 워크로드에는 데이터 집약적 특성 때문에 클라우드 비용을 급격히 끌어올릴 수 있다. 가트너의 클라우드·엣지·AI 인프라 서비스 및 기술 부문 부사장 시드 내그는 기업이 GPT-3.5 혹은 4.0과 같은 대규모 기반 모델을 활용할 때 선택의 여지가 거의 없다고 설명했다. 필요한 규모의 컴퓨팅 성능을 내부적으로 구현하기에는 비용이 지나치게 크기 때문이다.

한편 가트너는 2027년까지 기업이 활용하는 LLM의 절반 이상이 산업별 모델이 될 것이라고 내다봤다. 산업별 모델은 대규모 범용 기반 모델보다 훨씬 작아 다양한 환경에서 실행할 수 있다. 레드햇의 인스트럭트랩(InstructLab) 같은 툴을 활용해 자사 데이터로 모델을 보완하더라도, 규모 면에서는 여전히 기반 모델보다 작다. 가트너의 내그는 “산업별 모델은 학습에 필요한 자원이 상대적으로 적기 때문에 온프레미스, 프라이빗 클라우드, 또는 호스팅형 프라이빗 클라우드 인프라에서 실행할 수 있을 것”이라고 설명했다.

그러나 봄블레아는 이런 모델을 학습시키거나 최적화하는 데 필요한 컴퓨팅 성능과 인프라를 온프레미스에서 직접 확보하거나 구매하는 것은 쉽지 않다면서

“컴퓨팅 요구사항은 가장 중요한 요소 중 하나”라고 말했다. 다행히 클라우드 서비스 업체는 기업이 자체 데이터를 활용해 모델을 학습하도록 지원하는 AI 플랫폼을 제공한다. 붐블레아는 “따라서 AI 모델을 온프레미스에서 실행하더라도 굳이 온프레미스 인프라를 새로 구성할 필요는 없다”라고 설명했다.

하지만 반드시 그렇게 해야 할까? 가트너의 내그는 “프라이빗 클라우드 호스팅이나 온프레미스 선택은 신중해야 한다. 법적·재무적 책임을 지는 의사결정권자라면 설득력 있는 이유가 없는 한 CapEx 중심의 시절로 다시 돌아가기를 주저할 것”이라고 말했다.

클라우드 서비스 업체는 서비스형 플랫폼의 일부로 제공하는 AI/ML 서비스를 지속적으로 확대하고 있다. 기업은 사전 학습된 모델을 기반으로 시작해 자체 데이터를 추가해도 무리 없이 서비스를 활용할 수 있다. 붐블레아는 “퍼블릭 클라우드 서비스 업체가 제공하는 모델이 머지않아 대부분 기업 표준 요구사항의 90%를 충족할 만큼 성숙해질 것”이라고 전망했다. 다만 이런 서비스를 활용할지는 결국 비용 문제로 귀결된다. 즉, 비즈니스 모델 관점에서 충분한 설득력이 있어야

한다는 것이다.

비용은 저렴하지만 낮은 성능

포레스터의 우는 CIO가 처음에는 비용 절감에 초점을 맞췄지만 이 초점이 성능 고려사항 또는 최종 목표와는 맞지 않을 수 있다고 말했다. 퍼블릭 클라우드가 비용이 덜 드는 옵션이라 해도 잠재적인 지연 또는 기타 성능 문제까지 고려할 때 최적의 옵션은 아닐 수 있다. 결제 처리, 금융 서비스와 같이 지연이 용인되지 않는 업종에서는 더욱 그렇다.

세인트 주드 어린이 연구 병원의 페리에 따르면, 데이터를 생산하는 장비와 이를 처리하는 컴퓨팅 자원 사이에서 발생하는 지연은 데이터 위치를 결정하는 핵심 변수다. 일부 경우에는 장비와 고성능 컴퓨팅 자원 간에 거의 즉각적인 연결이 필요하기 때문이다. 페리는 “병원 연구 장비와 온프레미스 고성능 컴퓨터, 그리고 퍼블릭 클라우드 간의 지연 문제로 인해 퍼블릭 클라우드를 실시간 검사에 활용하는 것은 적절하지 않다. 점점 더 많은 대형 클라우드 서비스 업체가 고성능 컴퓨팅을 처리할 수 있는 대규모 GPU 클러스터를 구축하는 만큼 비용도 고려해야 한다”라고 덧붙였다.

예를 들어 유전체 염기서열 분석은 비용이

적절하다는 전제 하에 로컬 슈퍼컴퓨터에서 수행하는 처리 작업 일부는 퍼블릭 클라우드로 떼어 내는 편이 합리적일 수 있다. 페리는 유전체 염기서열 분석과 관련된 일부 워크플로우는 시간이 지나면 어느 정도 표준화된다고 말했다. 이 경우 파이프라인의 규모를 최적화하고 클라우드에서 실행하는 것이 비용 측면에서 더 합리적일 수 있다. 페리는 “온프레미스 고성능 컴퓨팅의 사이클을 확보하기 위해 일부 유전체 염기서열 분석 파이프라인을 클라우드로 이전하는 작업을 진행하고 있다”라고 말했다.

모자이크 컴퍼니의 와이소키는 성능은 중요하지만 광산 운영 현장의 엣지 서버에서 실행되는 일부 경우를 제외하면 퍼블릭 클라우드 호스팅 여부를 결정할 정도의 요소는 아니라면서 “모자이크의 경우 효과적인 작업을 위해 디바이스 자체 또는 디바이스와 가까운 곳에 엣지 컴퓨팅을 배치해야 할 필요성이 앞으로도 항상 있을 것”이라고 말했다.

위치의 문제

와이소키는 “모자이크에 가장 중요한 요소 3가지는 보안, 개인정보 보호, 비용”이라고 강조하며, 지금까지 퍼블릭 클라우드 서비스에서 보안이나 개인정보

보호와 관련해 큰 문제를 겪은 적은 없다고 설명했다.

RGA 역시 사용 중인 퍼블릭 클라우드 서비스의 보안에 만족하고 있다. 홀로웰은 “RGA는 AWS의 베드록(Bedrock) 서비스를 통해 엔트로픽, 미스트랄 등의 파운데이션 모델을 활용한다. 데이터 격리와 보안이 제공되기 때문에 안전한 환경에서 챗GPT와 유사한 기능을 제공할 수 있다”라고 말했다.

그러나 포레스터의 우는 디지털 주권 문제는 별개라고 지적했다. 엄격한 현지화 규제가 존재하는 국가에서는 퍼블릭 클라우드가 선택지에서 아예 제외될 수 있다. 우는 “직접 관리하거나 관리 위탁 방식으로 운영하는 온프레미스 프라이빗 클라우드, 혹은 호스팅형 프라이빗 클라우드를 선택할 수 있다. 둘 중 어느 방식을 사용하든 데이터의 위치를 직접 통제할 수 있다는 점이 핵심”이라고 강조했다.

규제 환경 외에도 고려할 요소가 있다. 홀로웰은 “일부 지역에서는 데이터 현지화와 개인정보 보호 요구사항이 고객 계약서에 직접 포함된다”라며, 이런 경우에는 프라이빗 클라우드가 더 유연한 해법이 될 수 있다고 설명했다. 북불레아는 여러 국가에서 사업을 운영하는 기업이라면 온프레미스와

클라우드를 병행하는 하이브리드 접근 방식이 최선의 선택이라고 조언했다.

지역별 규제와 관련해 어떤 퍼블릭 클라우드 서비스 업체를 선택할지도 중요한 변수다.

봄블레아는 “각 국가마다 데이터 처리에 관한 규정이 다른 중동의 경우 특정 위치에서 워크로드를 실행하는 용도로 오라클 클라우드가 최선의 선택지”라고 말했다.

모든 중동 국가에 진출한 클라우드 서비스 업체는 없다. 하지만 오라클은 중동에서 상당한 입지를 갖고 있으므로 오라클과 다른 클라우드 서비스 업체를 통해 온프레미스 워크로드를 실행할 수 있다.

그러나 홀로웰은 하이브리드 클라우드에도 단점이 있다면서 “퍼블릭 클라우드와 하이브리드 클라우드 환경 전반에서 대규모 데이터 집합의 상호운용성과 성능을 관리하는 것은 여전히 큰 과제”라고 지적했다.

언제든 조정 가능한 유연함 유지

앞으로의 방향에 대해 홀로웰은 “전략적 목표는 단순히 모든 것을 퍼블릭 클라우드로 옮기는 것이 아니다. 비즈니스 요구사항과 애플리케이션의 자연스러운 수명 주기 관리 방식을 결합해 호스팅 결정을 평가하는 것”이라고 말했다. 즉, 용량 요구사항이 일정해 기존 통합 인프라로도 충분히

대응할 수 있는 애플리케이션은 프라이빗 클라우드에서 운영하고, 고성능 컴퓨팅이 지속적으로 필요하지 않은 애플리케이션은 퍼블릭 클라우드로 옮겨야 할 대상이다.

페리는 엔터프라이즈 애플리케이션을 위한 적절한 IT 인프라를 구축하기 위해서는 적절한 재료를 사용하는 것이 중요하다면서 “퍼블릭 클라우드는 솔루션을 구축하기 위해 필요한 재료 중 하나일 뿐이며, 적절한 균형을 맞추는 것이 중요하다”라고 말했다.

그러나 온프레미스, 프라이빗 클라우드, 퍼블릭 클라우드 서비스를 최적의 조합으로 운영하는 일은 결코 쉽지 않다. 페리는 “기술은 지속적으로 발전하기 때문에 모든 것이 완벽하다고 말할 수는 없다. 시대에 맞춰 대응할 수 있는 역량이 필요하다”라고 말했다. 오늘 보유한 톨이 내일도 유효하다는 보장은 없다. 따라서 변화에 대응하고 준비할 수 있는 적절한 도구를 갖추는 것이 무엇보다 중요하다.

봄블레아는 기술 발전에 따라 변화해야 할 필요성은 업체 종속을 피해야 하는 이유이기도 하다면서 “다만 클라우드 워크로드를 가장 최적화된 방식으로 실행하기 위해서는 보통 업체가 독점적으로 제공하는 가장 고급 기능을 사용해야

하는 현실을 감안하면 딜레마에 처하게 된다”라고 말했다. 그러나 온프레미스, 퍼블릭 클라우드, 프라이빗 클라우드 사이를 간편히 이동할 수 있는 유연성을 확보하기 위해서는 업체 종속을 피하는 것이 좋다.

기술 발전에 맞춰 변화를 수용해야 한다는 점이 곧 업체 종속을 피해야 하는 이유다. 뎀블레아는 “클라우드 워크로드를

최적화하려면 보통 특정 업체가 독점적으로 제공하는 고급 기능을 활용해야 한다.

하지만 이 경우 업체 종속이라는 딜레마에 빠질 수 있다. 온프레미스, 퍼블릭 클라우드, 프라이빗 클라우드 간 워크로드를 자유롭게 이동할 수 있는 유연성을 확보하기 위해서는 가능한 한 업체 종속을 피하는 것이 바람직하다”라고 조언했다.

송환과 지속가능성, 2025년 클라우드 지형을 다시 그린다

Brian Adler | InfoWorld.com

클라우드 컴퓨팅 활용과 관련해 기업이 최우선으로 고려하는 과제와 주요 도전 과제는 무엇일까? IT 자산 관리 업체 플렉세라(Flexera)는 2024년 말 전 세계 759명의 클라우드 의사결정권자 및 사용자를 대상으로 실시한 설문조사 내용을 바탕으로 [2025 클라우드 현황 보고서\(Flexera 2025 State of the Cloud Report\)](#)를 발표했다. 조사 결과를 통해 지난 몇 년 동안 이어진 트렌드의 진화와 함께, 클라우드 사용을 이끄는 새로운 동력이 등장했음을 확인할 수 있었다.

데이터센터로 되돌아가는 워크로드

클라우드에 보냈던 애플리케이션과 데이터를 데이터센터로 다시 불러들이는 것을 의미하는 이른바 송환(repatriation)으로 인해 지금까지 워크로드와 데이터의 약 1/5(21%)이 데이터센터로 복귀했다. 그러나 지속적인 클라우드로의 마이그레이션과 클라우드 이탈을 상회하는 신규 워크로드 유입으로

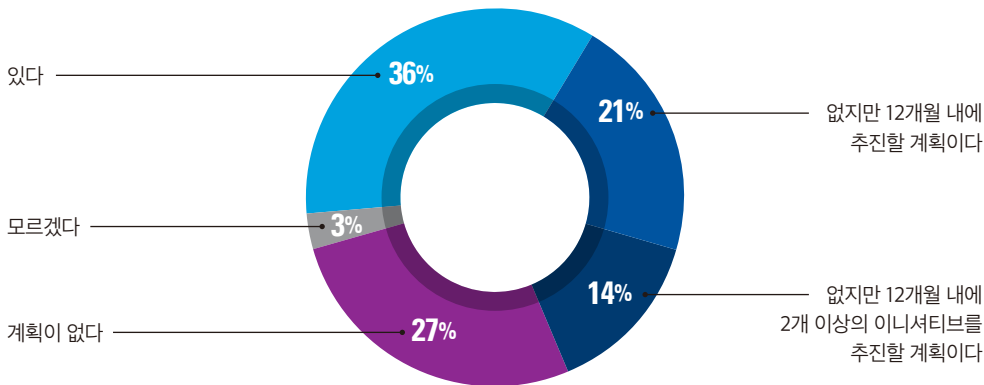
클라우드 성장세는 유지됐다.

많은 분석가와 전문가는 기업이 애플리케이션을 클라우드에 맞게 리팩터링하지 않아 발생하는 비효율성과 비용으로 인해 클라우드 워크로드를 자체 데이터센터로 되돌리는 상황을 몇 년 전부터 관측했다. 신규 클라우드 워크로드는 여전히 증가하고 있지만, 송환 빈도는 무시하지 못할 수준이다.

지속 가능성에 대한 관심 증대

설문 응답자 중 상당수는 클라우드 지속가능성 이니셔티브를 주요 관심사로 두는 것으로 나타났다. 응답자 57%는 클라우드 사용에 따른 탄소발자국 추적을 포함한 명확한 지속가능성 전략을 이미 수립했거나 향후 12개월 내에 도입할 계획이라고 답했다. 전체 응답자의 36%가 이미 클라우드 탄소발자국을 추적하고 있음을 고려하면 그 필요성에 대한 공감대는 확실해 보인다.

❖ 귀사에는 클라우드 사용 탄소발자국 추적을 포함한 지속가능성 이니셔티브가 있습니까?



N=759

자료 : 플렉세라 2025 클라우드 현황 보고서(그림 45)

전체 응답자의 36%가 이미 클라우드 탄소발자국을 추적하고 있다.

유럽 지역 응답자로 국한하면 클라우드 탄소발자국을 추적하는 비율은 43%로 전체 평균보다 더 높았다. 유럽 응답자와 전체 응답자 간의 격차는 점점 줄어드는 추세다. 글로벌 기업이 주요 지속가능성 기준을 채택하는 사례가 늘면서 이런 차이는 계속해서 좁혀질 것으로 예상된다.

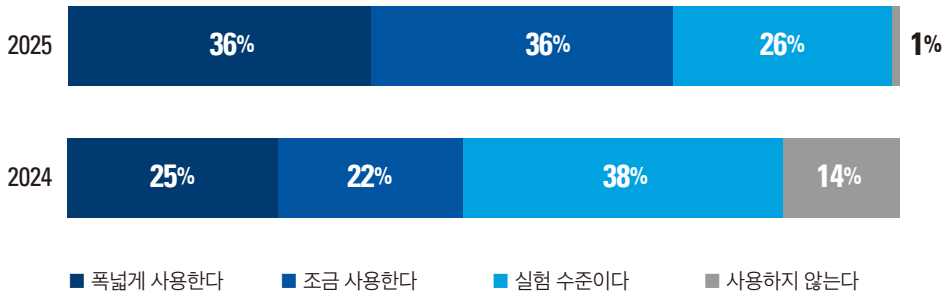
주류로 자리 잡는 생성형 AI

예상대로 AI 관련 퍼블릭 클라우드 서비스의 채택이 급증하고 있다. 설문 응답자 거의 절반이 AI/ML 기반의 서비스형 플랫폼(PaaS)을 이미 사용한다고 답했다.

또한 2025년 조사에서는 AI 모델에 데이터를 공급하는 데 핵심 역할을 하는 데이터 웨어하우스 서비스 채택도 빠르게 늘고 있다.

생성형 AI 역시 확산 속도가 가파르다. 전체 응답자의 72%가 이미 생성형 AI를 부분적으로 또는 광범위하게 사용하고 있으며, 26%는 실험 단계에 있다고 답했다. 생성형 AI는 단순한 유행을 넘어 주류 기술로 부상하고 있다.

❖ 귀사는 생성형 AI 퍼블릭 클라우드 서비스를 어느 정도 사용하십니까?



2024 : N=753, 2025 : N=759

자료 : 플렉세라 2025 클라우드 현황 보고서(그림 44)

72%의 기업이 이미 생성형 AI를 부분적으로 또는 광범위하게 사용하고 있다.

가장 큰 과제는 클라우드 비용과 보안

기업 규모와 관계없이 클라우드 전략에서 가장 큰 과제로 ‘지출 관리’를 꼽은 응답자는 84%에 달했다. 워크로드가 클라우드로 꾸준히 이동하고 있는 만큼 비용 통제와 최적화는 선택이 아닌 필수가 되고 있다. 클라우드 비용 평가에서 가장 중요한 지표로 ‘비용 효율성/절감’을 지목한 응답자는 87%로, 2024년(65%) 대비 큰 폭으로 증가했다. 특히 적절한 라이선스 관리를 통해 달성할 수 있는 ‘비용 회피’도 같은 기간 28%에서 64%로 상승했다. SaaS 비용이 클라우드 지출 증가에 상당한 영향을 미치는 만큼 SaaS 사용 증가에 따라 SaaS 라이선스에 대한 관심도 함께 높아지는 추세다.

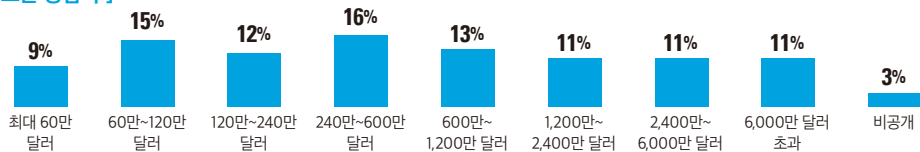
클라우드 지출 관리에 이어 두 번째 큰 과제로 ‘보안’을 꼽은 응답자는 77%였다. 디지털 시대에 항상 최우선 과제로 꼽히는 보안 문제는 클라우드 이니셔티브에서도 예외가 아니었다. 멀티클라우드 환경을 관리하는 데 가장 널리 사용되는 톨 역시 ‘보안 톨’이었다. 전체 응답자의 55%가 보안 톨을 가장 중요한 도구로 지목했다.

더 가속화되는 퍼블릭 클라우드 채택

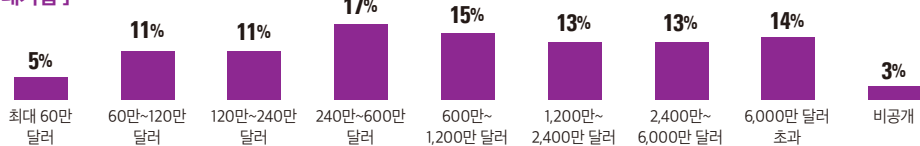
퍼블릭 클라우드 지출이 증가하는 가운데, 연간 1,200만 달러 이상을 지출한다고 답한 응답자는 33%로 지난해 29%보다 늘었다. 직원 수 1,000명 이상인 기업에서는 40%로 더 높았다. 클라우드 비용이 상승하고

❖ 귀사가 매년 퍼블릭 클라우드에 지출하는 비용은 얼마입니까?

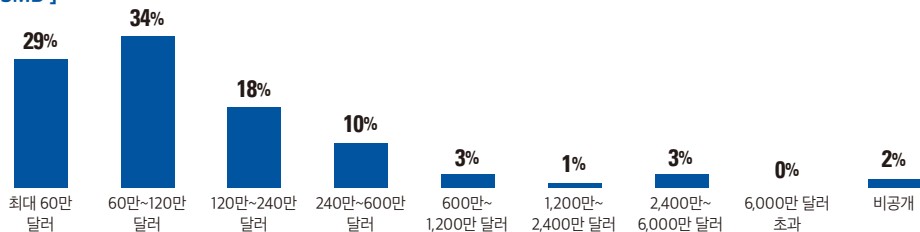
[모든 응답자]



[대기업]



[SMB]



전체 : N=759, 엔터프라이즈 : N=622, SMB : N=137
 자료 : 플렉세라 2025 클라우드 현황 보고서(그림 5)

매년 1,200만 달러 이상을 지출하는 응답자 비율은 33%로, 지난해의 29%에서 늘었다.

있음에도 워크로드가 클라우드로 몰려들고 있는 것이다. 한편 SaaS 지출은 전년과 비교해 큰 변동이 없는 것으로 나타났다.

기업이 여전히 신중한 태도를 보이는 분야는 민감 데이터 처리다. 많은 응답자가 민감 데이터를 클라우드로 옮기는 데

있어 조심스럽게 접근하고 있다고 답했다. 반면 민감하지 않은 데이터에 대해서는 응답자 1/3 이상이 전면적인 클라우드 마이그레이션을 계획하고 있었다.

클라우드 관리의 새 중심, CCoE와 핀옵스
 클라우드 및 SaaS 비용 관리의

❖ 귀사가 관심을 두는 주요 클라우드 과제는 무엇입니까?



N=759

자료 : 플렉세라 2025 클라우드 현황 보고서

클라우드 지출 관리(84%)와 보안(77%)이 가장 큰 클라우드 과제로 조사됐다.

주체는 기존의 업체 관리팀에서 CCoE(Cloud Centers of Excellence)와 핀옵스(FinOps)팀으로 옮겨가고 있다. 클라우드에 대한 중앙화된 접근 방식이 확산하고 있음을 보여주는 변화다. 실제로 응답자 69%가 현재 CCoE 또는 중앙 클라우드 전담팀을 보유하고 있다고 답했다.

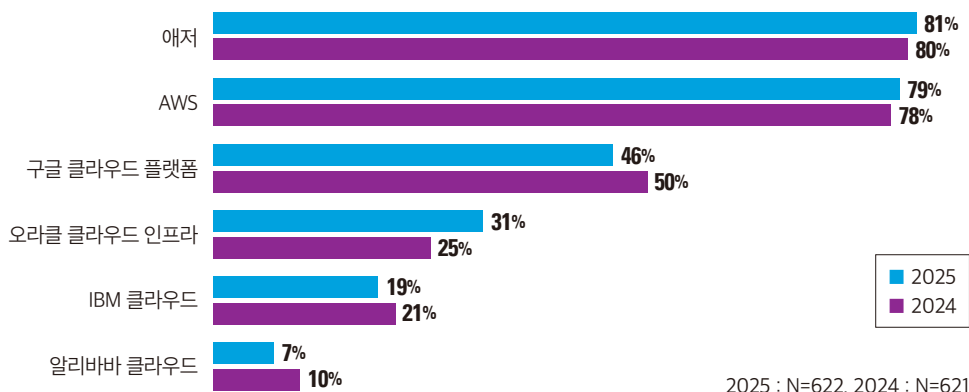
핀옵스팀이 클라우드 비용 최적화를 주도하는 경우도 늘고 있다. 2025년 조사에서 응답자 59%는 핀옵스팀이 비용 최적화 전략의 일부 또는 전체를 맡고 있다고 밝혀 2024년(51%)보다 소폭 증가했음을 알 수 있었다. 특히 SaaS와 데이터센터까지 핀옵스 관리 범위에 포함되면서 핀옵스팀은 클라우드 전략에서 점차 중심적인 역할을 하게 될 것으로 보인다.

AWS와 애저의 주도권 경쟁

2024년과 2025년 조사 결과를 비교하면 업계 리더 구도에 큰 변화는 없다. 많은 기업이 현재 사용 중인 클라우드 또는 클라우드 조합에 대해 어느 정도 안정화된 상태에 이른 것으로 보인다. 전체 응답자를 기준으로 보면 AWS와 마이크로소프트 애저가 퍼블릭 클라우드 서비스 업체 양강 구도를 형성하고 있다. 실행 중인 워크로드 수를 기준으로 두 업체는 최근 몇 년 동안 었치락뒤치락 치열하게 경쟁했다.

전통적으로 대기업은 직원 수 1,000명 미만 SMB보다 애저 사용 비중이 높다. 대기업만 놓고 보면 현재 ‘상당한 규모의 워크로드’를 실행하는 기업에서는 AWS(53%)가 애저(50%)보다 근소한 차이로 앞서고

❖ 귀사에서 사용 중인 퍼블릭 클라우드 서비스는 무엇입니까?



2025 : N=622, 2024 : N=621
자료 : 플렉세라 2025 클라우드 현황 보고서(그림 35)

AWS와 마이크로소프트 애저가 퍼블릭 클라우드 서비스 업체 양강 구도를 형성하고 있다.

있다. 반면, ‘일부 워크로드’까지 포함하면 애저(81%)가 AWS(79%)보다 높은 점유율을 보인다.

명 이상의 대규모 기업은 톨 유형과 관계없이 소규모 기업보다 멀티클라우드 톨을 더 많이 사용하는 것으로 조사됐다.

클라우드 전략의 일환으로 멀티클라우드를 채택하는 움직임은 계속 이어지고 있다. 응답자 70%는 하이브리드 클라우드 전략을 채택해 각각 하나 이상의 퍼블릭 클라우드와 프라이빗 클라우드를 사용 중이며, 나머지 30%는 퍼블릭 클라우드 또는 프라이빗 클라우드 중 하나만 사용한다. 직원 수 1만

앞으로의 전망
2025년 클라우드 시장의 핵심 키워드는 사용량 증가, 비용 최적화, 주요 클라우드 서비스 업체 간 경쟁, 그리고 AI 활용 확산이다. 여기에 더해 클라우드 송환과 지속가능성에 대한 관심이 높아지면서 클라우드 운영 방식 전반에 변화가 예상된다.

Brian Adler는 Flexera의 클라우드 시장 전략 담당 선임 디렉터로, 가트너의 선임 분석가와 핀옵스 재단의 이사회 회원으로 활동한 바 있다.