

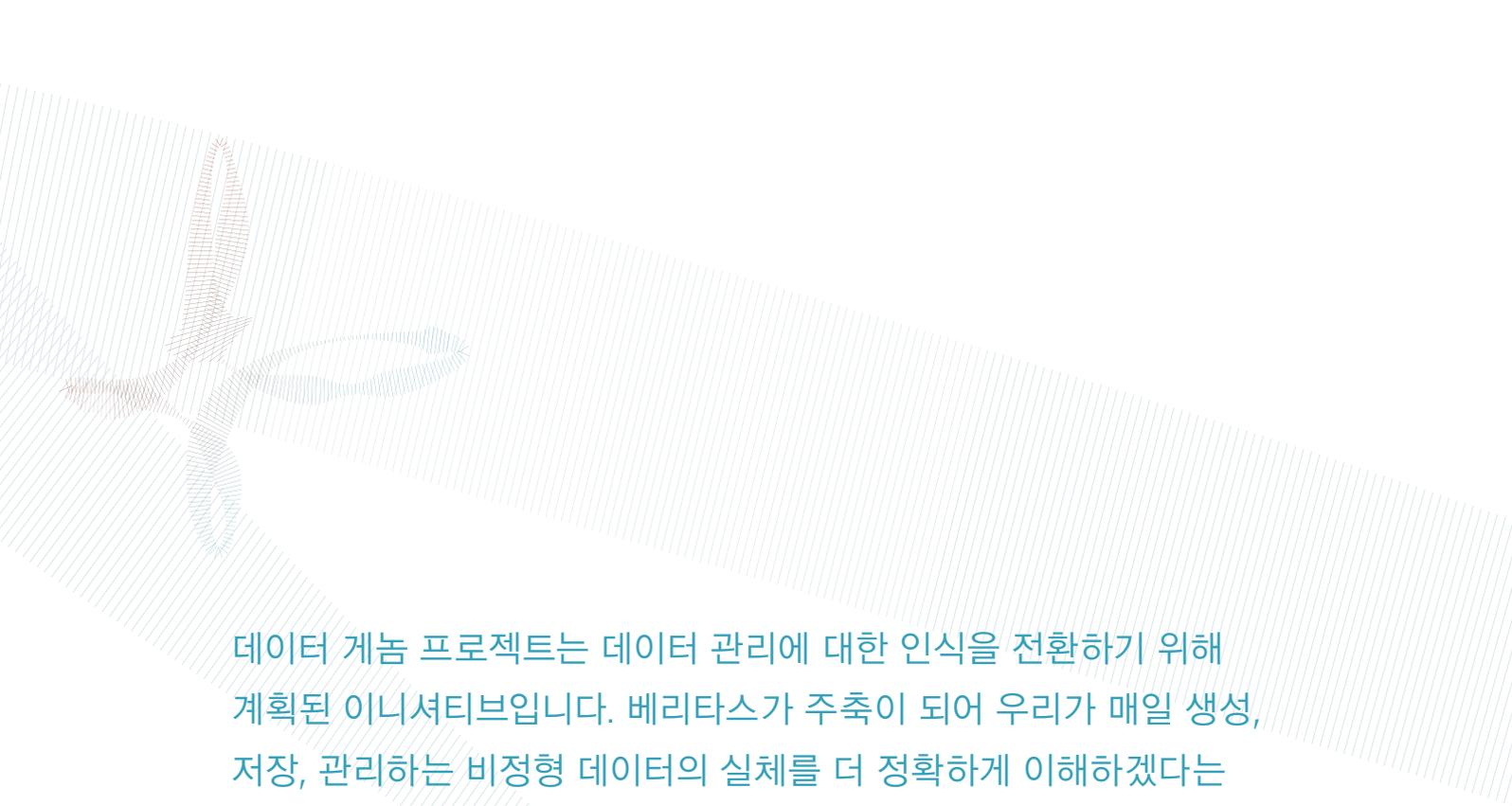
VERITAS™

# 데이터 게놈 지수

스토리지 환경의 비정형 데이터 분석 리포트

2016

3월



데이터 계층 프로젝트는 데이터 관리에 대한 인식을 전환하기 위해 계획된 이니셔티브입니다. 베리타스가 주축이 되어 우리가 매일 생성, 저장, 관리하는 비정형 데이터의 실체를 더 정확하게 이해하겠다는 목표 아래 이와 뜻을 같이 하는 데이터 과학자, 산업 전문가와 앞선 사고를 가진 리더들이 한데 모였습니다. 그리고 그 첫 결과물로 실제 스토리지 환경의 구성에 관한 벤치마크 리포트인 데이터 계층 지수를 발표했습니다.

포춘 500대 기업의 86%를 고객으로 둔 베리타스는 고객들을 위해 엑사바이트 단위의 데이터를 백업, 아카이빙, 분석하면서 기업 환경의 특성을 연구하고 있습니다. 현재 베리타스가 주목하는 특성은 바로 메타데이터입니다. 베리타스는 파일 분석 제품을 통해 고객으로부터 수집한 메타데이터를 활용하여 정확하고 면밀한 관점에서 실제 환경의 구성을 조명할 수 있습니다.

## 현재의 데이터 환경



# 41%

전체 환경 중 41%가 지난 3년 동안 변경된 적이 없습니다.

# 3년



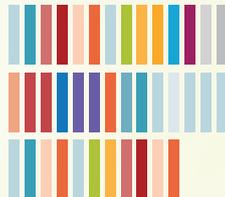
가장 큰 비중을 차지하는 파일 형식은 예상치 못한 결과였습니다.



Office 파일이 업무 환경에 과중한 부담이 되고 있습니다.



주인 없는 데이터



'주인 없는 데이터'가 오랜 기간 방치되어 기업에 큰 부담이 되고 있습니다.

시스템 아카이빙으로



스토리지 비용 절감

# → 50%

스토리지 공간을 재확보하려면 삭제 또는 아카이브 전략을 우선 순위에 두어야 합니다.



## 첫 번째 베리타스 데이터 게놈 지수

베리타스는 고객의 데이터 환경이 실제로 어떻게 구성되어 있는지 정확히 알아보기 위해 2015년에 수많은 베리타스 고객의 비정형 데이터 환경을 대상으로 수십억 개의 파일과 그 특성을 분석했습니다. 대표적인 파일 형식 확장자 8,000여 개가 분석에 포함되었습니다. 일반적으로 이 데이터는 각 고객의 전체 파일 시스템 환경을 대표하는 하위 집합입니다.

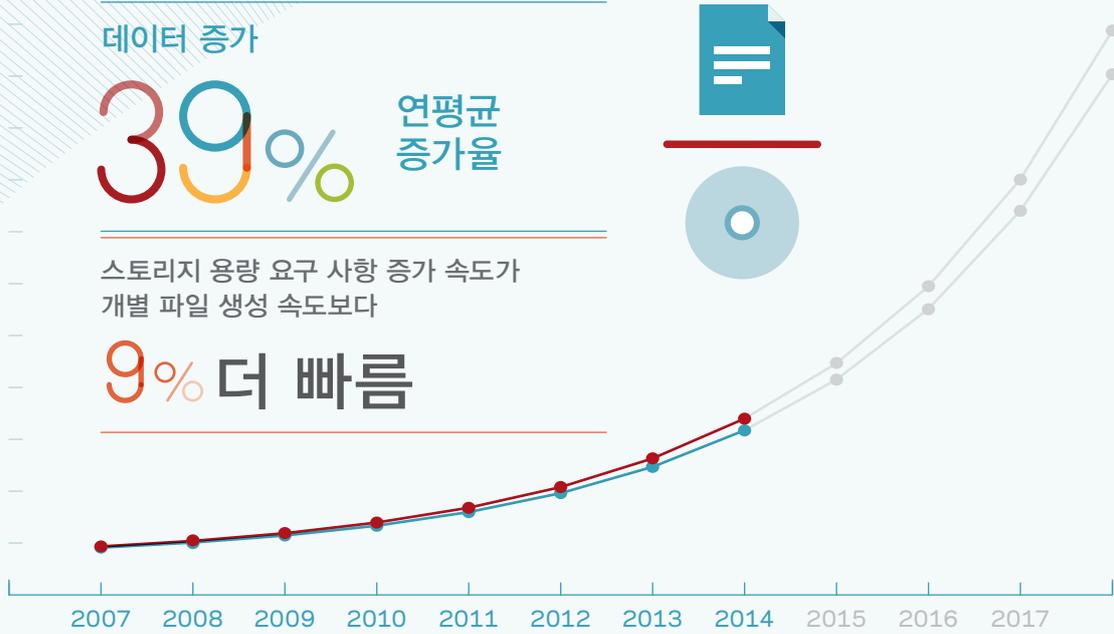
데이터 증가

39%

연평균  
증가율

스토리지 용량 요구 사항 증가 속도가  
개별 파일 생성 속도보다

9% 더 빠름



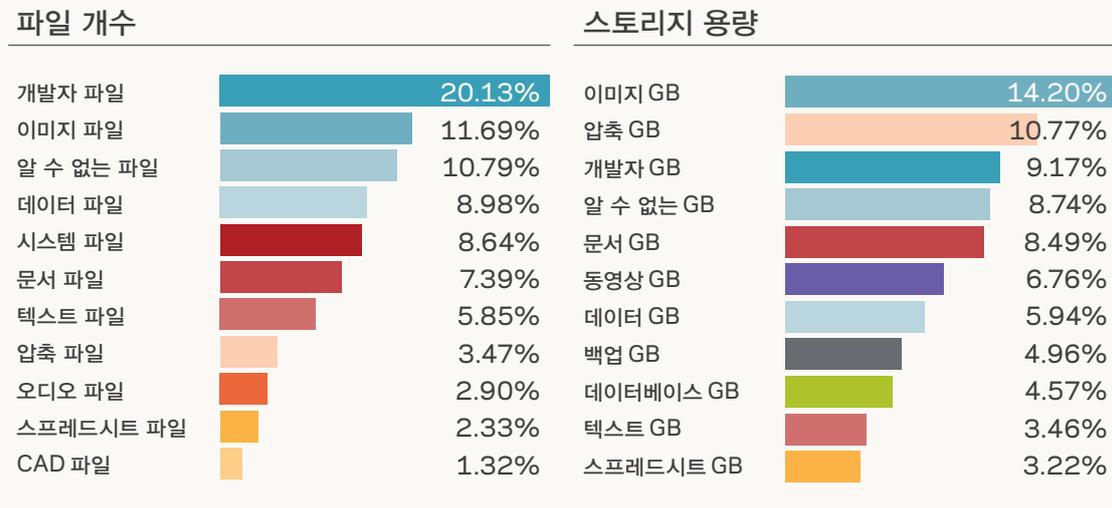
## 데이터가 급증하고 있습니다

지난 7년간 파일 레벨에서 실제 데이터 증가 속도를 보면 전년 대비 39.2481189% 증가했습니다. 이렇듯 스토리지 용량 요구 사항의 증가 속도는 개별 파일 생성 속도보다 9% 빠릅니다. 따라서 업무 방식을 개선하여 증가 속도를 다소 늦춘다 해도 결국 스토리지 관리 문제로 귀결됩니다.

스토리지 용량 제한은 단지 스토리지 문제로 끝나지 않습니다. 스토리지 환경은 매우 복잡합니다. 정보 1PB당 평균 2,312,000,000개의 파일이 있습니다.

## 주범은 이미지 파일과 개발자 파일

전체 데이터를 살펴보면 파일 개수가 가장 많은 파일 형식과 공간 사용량이 가장 큰 파일 형식이 전혀 다릅니다. 이러한 차이는 환경의 복잡성과 비용에서 비롯된 것으로 볼 수 있습니다.



## 가장 왕성한 활동을 보인 데이터

지난 10년간 다른 파일 형식보다 두드러진 활동을 보인 대표적인 파일 형식을 꼽자면 아래와 같습니다.



스토리지 용량 면에서는 아래와 같은 파일 형식이 선두를 차지합니다.

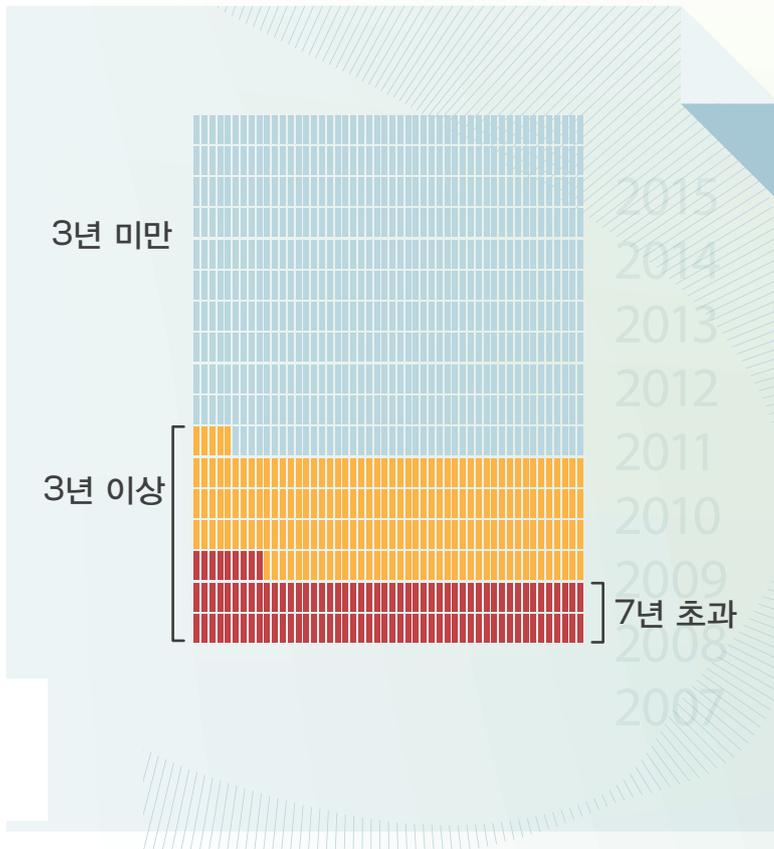




## 지속적인 데이터 증가

파일 생성량만 보면 가을철이 최대 성수기입니다. 가을에는 다른 계절보다 91% 더 많은 텍스트 파일이, 48% 더 많은 스프레드시트가, 89% 더 많은 지리 및 정보 시스템 파일이 생성됩니다.

가을에서 겨울로 넘어갈 때 파일 수가 증가하는 유형은 백업과 문서뿐입니다. 모든 연간 백업 프랙티스를 고려할 때 백업의 크기는 756% 증가합니다. 연간 총 동영상의 68%가 여름과 가을에 촬영되며, 겨울이 되면 이미지는 63% 줄어듭니다. 한편 이메일(pst) 패턴은 제법 예측 가능합니다. 실제로 이메일의 계절간 표준 편차는 0.7%에 불과합니다.



41%

전체 환경 중 41%가 지난 3년 동안 변경된 적이 없는 영역

3년

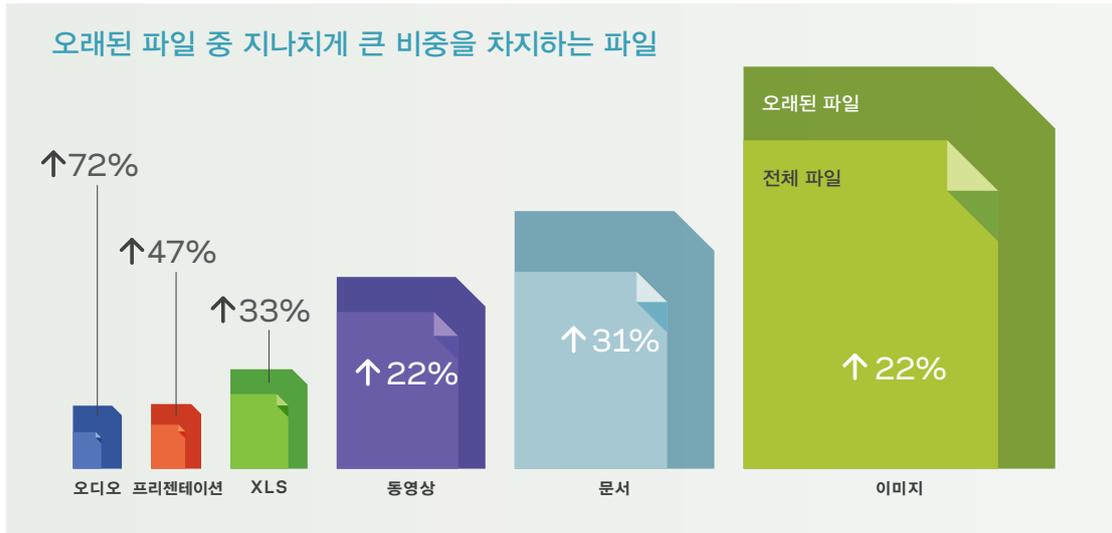
12% = 아주 오래됨 (7년간 변경 없음)

## 왜 보관하고 계십니까?

정보는 오늘날 기업의 최대 자산이지만, 지금처럼 엄청난 속도로 생성되는 상황에서는 개별 정보의 효용성이 순식간에 사라지고 맙니다.

## 조치: 파일 형식에 대한 우선 순위 설정

오래된 데이터가 너무 많고 그 처리 방법을 결정하는 것도 만만치 않을 경우, 정보 관리 '결정 비용'이 가장 유용하게 사용되는 곳에 우선 순위를 두는 것이 좋습니다. 오래된 데이터 및 전체 데이터 중 각각 지나치게 큰 비중을 차지하는 파일 형식을 살펴보니 "Office" 파일이 큰 부담으로 작용하고 있습니다.



개별 파일을 처리하여 스토리지 공간 재확보 효과를 극대화할 수 있는 영역을 찾는다면 파일당 재확보 GB가 가장 큰 5개 형식에 주목하십시오.

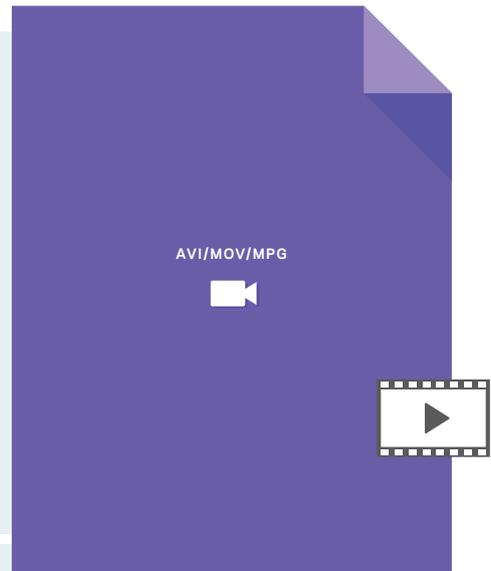
1. 가상 시스템 파일 형식
2. 보안 파일 형식
3. 게임 파일 형식
4. 과학 파일 형식
5. 지리 정보 시스템 파일 형식



동영상의 경우

15.8x

오래된 스토리지의 전체 용량 기준 15.8배 더 높음  
(오래된 총 파일 개수 기준 비율 대비)



7.3x



6.4x



2.2x



## 파일 불균형

특정 파일 형식에 따라 우선 순위를 정하려는 경우 파일 개수와 전체 대비 비율이 불균형을 나타내는 곳에 주목하십시오. 예를 들어, 동영상은 오래된 스토리지의 전체 용량 기준 비율이 오래된 총 파일 개수 기준 비율의 15.8배입니다. 가상 시스템 파일이 7.3배, 프레젠테이션이 6.4배, 이메일은 2.2배 더 많은 공간을 차지하면서 최상위 우선 순위 후보가 됩니다.



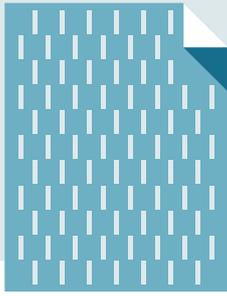
## 그저 책상 위를 정리했을 뿐입니다

주인 없는 데이터는 소유자가 없는 데이터입니다. 직원의 이직, 인사 이동, 일반적인 Active Directory의 혼잡성 등으로 인해 이전의 환경을 추적하는 것은 당연히 매우 어려운 일이며 경제적 부담이 될 수 있습니다.

이러한 비용을 입증하는 방법 중 하나는 지나치게 많은 스토리지 용량을 차지하는 주인 없는 데이터에서 찾을 수 있습니다. 실제로 주인 없는 데이터는 전체 파일의 1.6%에 불과하지만 전체 스토리지 용량의 5.1%를 사용합니다. 주인 없는 데이터는 콘텐츠 위주의 데이터 유형이 대부분을 차지하는데, 여기서 이미지는 일반 파일보다 88% 더 많은 공간을 차지하며 동영상 및 프레젠테이션도 각각 165% 및 229% 더 많습니다.

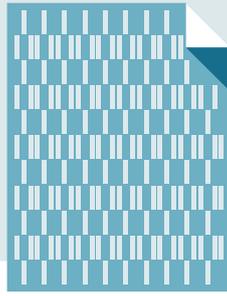
직원의 성향도 스토리지 환경에 영향을 미칠 가능성이 있습니다. 주인 없는 파일은 일반 파일보다 222% 더 큼니다. 아마도 관리자는 파일이 클수록 더 중요한 내용이 들어 있을 것으로 간주하고 직원이 떠난 후에도 남겨진 파일들을 차곡차곡 쌓아두었을 것입니다. 따라서 스토리지 공간을 재확보하려는 경우 먼저 소유자가 없는 콘텐츠에 중점을 두는 것이 좋습니다.

## 평균 파일 크기



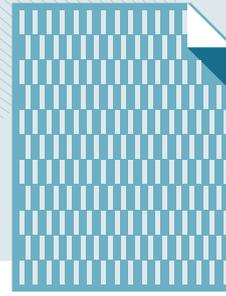
0.24 MB

지난 10여 년 간  
마지막으로 사용된 파일 크기



0.40 MB

지난 5년 간  
사용된 파일 크기



0.53 MB

지난해에 변경된  
파일 크기

## 밀도는 효용성 지표로 삼기에 부족합니다

오늘날에는 이전보다 밀도가 높은 콘텐츠가 생성되고 있지만, 이러한 콘텐츠는 지난 7년간 10.3% 늘어났을 뿐입니다. 평균 파일 크기는 아래와 같습니다.

지난 10여 년간 마지막으로 사용된 파일 크기: 0.24MB

지난 5년간 사용된 파일 크기: 0.40MB

지난해에 변경된 파일 크기: 0.53MB

오래된 것으로 분류된 파일이 지난해에 변경된 파일보다 33% 적습니다.

## 그렇다면 이제 어떻게 해야 할까요?

귀사의 스토리지 환경이 본 문서에서 분석한 환경과 비슷하다면 지금이 절호의 기회입니다. 귀사 환경이 평균 수치인 10PB라고 가정해 보십시오.



그중 41%가 오래된 환경이라면 3년간 손도 대지 않은 데이터를 관리하는 데 연간 최대 2,050만 달러를 소비하고 있을 가능성이 있습니다. 그렇다고 이러한 데이터를 정리하는 일도 쉽지 않습니다. 4.1PB의 데이터를 정리하려면 9,479,200,000개의 파일을 분류, 삭제 또는 보관할지 결정해야 합니다.



우선 순위를 정해야 합니다.



대체로 프레젠테이션, 스프레드시트, 문서, 텍스트 파일과 같은 콘텐츠 위주의 파일이 오래된 환경의 20%를 차지하므로 파일 시스템 아카이빙 프로젝트를 통해 스토리지 비용을 50% 이상 줄인다면 2백만 달러 이상의 투자 회수 효과를 거둘 수 있습니다. 오디오 및 동영상만 정리해도 11%의 공간을 재확보할 수 있습니다. 실제로 이미지는 7년 이상의 오래된 스토리지 공간의 18%를 차지합니다.

개별 파일 태그 수가 적을수록 더 많은 스토리지 공간을 재확보할 수 있습니다. 이를테면 동영상, 가상 시스템 파일, 이메일을 중심으로 분류 프로젝트를 진행한다면 더 신속하게, 거의 15배 빠르게 정리를 시작할 수 있습니다. 또한 직원의 퇴사 또는 직무 이동 이후에 남겨진 데이터에 대한 정책을 평가함으로써 5%의 공간을 재확보하여 상당한 경제적 효과를 거둘 수 있습니다.

어떠한 환경이든 이와 같이 통찰력을 바탕으로 한 전략으로 폭발적인 데이터 증가에 신속하게 대응함으로써 보다 중요한 비즈니스에 주력할 수 있습니다.

## VERITAS™

Veritas Technologies는 기업이 가장 복잡한 환경에서도 효과적으로 정보를 활용하여 비즈니스 성공을 촉진할 수 있도록 지원합니다. 베리타스는 포춘 500대 기업의 86%를 포함하여 모든 규모의 기업에서 사용하고 있습니다. 실제로 베리타스는 10년 넘게 Gartner Magic Quadrant for Enterprise Backup Software and Integrated Appliances<sup>1</sup> 부문과 Gartner Magic Quadrant for Enterprise Information Archiving<sup>2</sup>에서 리더 그룹에 선정되었습니다.

베리타스는 풍부한 경험을 가진 파트너 커뮤니티와 함께 고객이 데이터 가용성을 높이고 통찰력을 향상시켜 경쟁력을 강화할 수 있도록 지원합니다. 전 세계 58개국에서 7,800여 명의 직원이 근무 중인 베리타스는 매출 규모가 26억 달러에 달하며 Amazon, Cisco, Fujitsu, Google, Hitachi, HP, IBM, Microsoft, NetApp, OpenStack, Symantec 등 세계 유수의 기업과 협력하고 있습니다.

베리타스는 파트너 커뮤니티와 손잡고 기존 데이터 센터부터 프라이빗, 퍼블릭, 하이브리드 클라우드까지 모든 유형의 엔터프라이즈 환경에서 지능적인 정보 관리 솔루션으로 데이터를 보호, 식별, 관리할 수 있도록 지원합니다. 베리타스를 선택한 기업은 현재 무슨 정보가 어디에 있는지, 이 정보를 어떻게 지속적으로 보호하며 어떤 데이터를 삭제해야 하는지 이해하는 데 필요한 통찰력과 가용성을 갖추게 됩니다.

[veritas.com/kr](http://veritas.com/kr)에서 베리타스와 함께 이러한 혜택을 누릴 수 있는 방법을 알아보십시오.

법적 고지:

<sup>1</sup>출처: Gartner, Inc., Magic Quadrant for Enterprise Backup Software and Integrated Appliances, Dave Russell, Pushan Rinnen, Robert Rhame, 2015년 6월 15일

<sup>2</sup>출처: Gartner, Inc., Magic Quadrant for Enterprise Information Archiving, Alan Dayley, Garth Landers, Anthony Kros, 2015년 10월 29일. 참고: 2005년 ~ 2009년에는 리서치 제목이 Magic Quadrant for Email Active Archiving이었음

Gartner는 리서치 간행물에 명시된 어떤 벤더, 제품, 서비스도 보증하지 않으며 기술 사용자에게 최고 평점 또는 기타 명칭의 벤더를 선택하도록 조언하지 않습니다. Gartner 리서치 간행물은 Gartner 리서치 조직의 견해를 담은 것이므로 사실의 진술로 간주해서는 안 됩니다. Gartner는 상업성 및 특정 용도의 적합성에 대한 보증을 비롯하여 해당 연구 조사와 관련하여 어떠한 명시적 또는 묵시적 보증도 거부합니다.

© 2016 Veritas Technologies LLC. All rights reserved. Veritas 및 Veritas 로고는 미국 및 기타 국가에서 Veritas Technologies LLC 또는 그 자회사의 상표 또는 등록 상표입니다.