



# Tableau와 빅 데이터: 개요

# 목차

<b>빅 데이터 현황</b> .....	<b>3</b>
변화하는 데이터 및 분석 요구 .....	3
빅 데이터는 가능성인 동시에 위험.....	4
<b>Tableau가 빅 데이터를 활용하는 방법</b> .....	<b>5</b>
빅 데이터라는 큰 그림 .....	5
데이터 액세스 및 연결 .....	5
모든 규모의 데이터와 빠른 상호 작용 .....	6
<b>Tableau와 빅 데이터 분석 에코시스템</b> .....	<b>7</b>
클라우드 인프라 .....	8
수집 및 준비.....	8
저장 및 처리.....	9
쿼리 가속화.....	10
데이터 카탈로그 .....	10
<b>빅 데이터 분석 아키텍처</b> .....	<b>10</b>
주요 클라우드 제공업체 예.....	11
Tableau 고객 사례 .....	12
공통 패턴 .....	13
<b>Tableau 정보 및 추가 리소스</b> .....	<b>14</b>



# 빅 데이터 현황

## 변화하는 데이터 및 분석 요구

데이터는 어디에나 있으며 데이터 이용 및 분석 요구 또한 그러합니다. 유행어로서 '빅 데이터'의 인기는 진정되었을지 모르지만, 규모, 다양성 및 속도라는 V자로 시작되는 빅 데이터의 세 가지 요소(volume, variety, velocity)는 빅 데이터 분석 사용 사례에 그 어느 때보다도 많이 적용됩니다. 주관적이긴 해도, 이 세 요소와 V로 시작되는 다른 요소들(변동성, 유효성, 진실성 등)에 대한 업계의 논의 결과, 빅 데이터도 여전히 데이터라는 사실을 알게 되었습니다. 단지 너무 복잡해졌으므로 이를 효과적으로 수집, 정리, 이해 및 이용하려면 조직이 혁신해야 할 것입니다.

모든 산업 분야와 모든 규모의 조직에 걸쳐 다양한 형식과 원본의 엄청난 양의 데이터를 생성하는 수많은 '일'이 벌어지며 디지털 전환이 일어나고 있습니다. 조직은 과거 어느 때보다도 더 다양한 데이터를 수집, 처리 및 분석하고 있습니다. 스키마가 없는 JSON에서 중첩 유형의 다른 데이터베이스(관계형 및 NoSQL), 비플랫 데이터(Avro, Parquet, XML)에 이르기까지 데이터 형식이 매우 다양해지고 있으며 그러한 다양한 형식을 이용하기 위해서는 커넥터가 매우 중요합니다.

### 조직에는 보통 다음과 같은 다양한 데이터가 있습니다.

- **정형 데이터** - 특정 질문에 대해 사전 계산된 집계 형태, 혹은 인메모리 컴퓨팅용 추출의 형태로 끌어왔거나, 분석을 위해 집계된 형태입니다. 조직이 보유한, 가장 잘 정리되고 쉽게 이용 가능한 데이터입니다.
- **반정형 데이터(또는 개체 저장소)** - 관계형 데이터베이스, 데이터 웨어하우스 또는 데이터 마트에서 볼 수 있습니다. 흔히, 엔터티 분석용으로 정기적으로 새로 고쳐지는 비즈니스 개념, 예를 들어, 거래 수, 기회 수 또는 개별 영업 담당자가 기회에 대해 취한 조치 등 답은 알 수 없으나 잘 알려진 질문이 여기 속합니다.
- **원시, 비정형 데이터** - 데이터 레이크 또는 클라우드 저장소에 존재합니다. 여기에는 소셜 네트워크 피드, IoT 기기 등에서 생성된 스트림 데이터가 포함됩니다. 데이터 과학자가 이러한 데이터를 마이닝 및 변환할 수 있지만 개발할 잠재력이 더 있는 것으로 보입니다.

일부 데이터는 가장 가치 있게 사용될 사례를 아직 찾아야 하지만, 이 모든 데이터는 지식 근로자가 의사 결정을 위해 데이터에 액세스하여 분석해야 한다는 더 커진 요구 앞에 놓여 있습니다. 데이터 분석 및 비주얼리제이션에 사용되는 응용 프로그램은 데이터 자체에 끌려갑니다. 이는 곧 견고한 저장소와 더 유연하고 더 큰 규모의 데이터 처리 서비스를 동시에 제공하며 분석할 수 있는 클라우드로의 대규모 전환을 의미합니다. 조직이 광범위한 클라우드 기반의 빅 데이터 실행을 운영하든, 현재 데이터를 아주 조금밖에 분석하지 못하고 있든 간에, 비즈니스 및 IT 부서에 걸쳐 패턴을 시각화하고 거기에 담겨 있는 인사이트를 얻기 위해 분석하는 기능을 부여함으로써 상당한 이익을 얻을 수 있습니다.

최신 분석을 통해 모든 기술 수준의 더 많은 비즈니스 사용자에게 더 다양한 기능을 제공하게 되었음에도 불구하고 이러한 모든 데이터를 전체 조직에 유용한 리소스로 만드는 방법을 찾는 데에는 많은 복잡한 과제가 있습니다. 비즈니스 요구 사항은 데이터 그 자체와 마찬가지로 자주 변화하므로 민첩하고 조정 가능한 빅 데이터 전략 및 아키텍처가 필요합니다. 조직은 데이터 연결에 초점을 맞춘 단일형 플랫폼을 구축하기보다는 빅 데이터 기회의 범위를 확장하고 진화하는 분석 사용 사례에 대해 검토하는 것이 현명합니다. 그렇지 않을 경우 더 큰 그림을 놓치게 되는 위험이 따를 수 있습니다.

## 빅 데이터는 가능성인 동시에 위험

데이터 자산은 수익성이 좋은 비즈니스와 어려움을 겪는 비즈니스 간의 핵심 차별화 영역이 되고 있습니다. 하지만 데이터는 규모가 너무 방대하고, 계속 증가하고 있으며, 아주 다양하여 관계형 데이터베이스 관리 시스템으로 처리하기에는 너무 복잡하고 비용이 많이 듭니다. 사전 컴퓨팅 및 공유 컴퓨팅으로 인한 하드웨어 비용 절감 외에 고객은 또한 데이터 이동을 최소화하고자 합니다. 가장 민첩하게 데이터를 이동할 수 있는 인프라는 원시, 비정형 데이터와 사용자가 분석할 수 있게 준비된 데이터 간의 격차를 해소하는 데 도움이 됩니다.

조직은 또한 연결 및 성능 문제에 당면합니다. 라이브 연결 또는 인메모리 분석 옵션에도 불구하고 거대한 데이터 레이크는 추출 생성 작업이나 다른 데이터와 통합 작업으로 분주할 수 있습니다. 최신의 셀프 서비스 분석 접근 방식이 민첩성을 약속하지만, 이러한 데이터 집합에서 대량의 조인을 생성하면 시스템이 중단될 수 있습니다.

IT와 비즈니스 부서는 협력해야 하지만 해당 분야 전문가가 메타데이터, 비즈니스 규칙 및 보고 모델을 작성하는 상향식 방법론을 통해서입니다. 이러한 프로세스는 끊임없이 반복되어야 하며 변화하는 비즈니스 요구에 맞도록 개선되어야 합니다. 현재의 디지털 전환 시대에서 비즈니스는 멈춰 있지 않으므로 빅 데이터 분석 프레임워크 또한 멈춰서는 안 됩니다.

# Tableau가 빅 데이터를 활용하는 방법

## 빅 데이터라는 큰 그림

Tableau에서 하는 모든 일은 사람들이 데이터를 보고 이해할 수 있도록 지원한다는 사명과 관련이 있습니다. Tableau는 기본적으로 데이터 대중화를 믿으므로 디지털 경제를 위한 최신 분석 플랫폼입니다. 데이터를 아는 사람에게 데이터에 질문할 권한이 있어야 합니다. 즉, 모든 기술 수준의 지식 근로자가 데이터가 어디에 있는지 데이터에 액세스하고 그것을 분석하여 인사이트를 발견할 수 있는 능력을 가져야 한다는 뜻입니다.

많은 고객이 다양한 빅 데이터 기술을 다루고 있으므로 엔지니어링 투자, 에코시스템 내 파트너십, 그리고 전체 비전을 데이터 환경의 진화에 맞춰 조정해 왔습니다. Tableau는 빅 데이터 시대에 앞서 오랫동안 여러모로 투자해 왔습니다. 이러한 투자에는 Hadoop 및 NoSQL 플랫폼 모두에 대한 데이터 연결뿐만 아니라 대규모의 온프레미스 및 클라우드 데이터 웨어하우스로의 연결도 포함됩니다.

“매우 한정된 비즈니스 사용 사례로 시작했는데 그 후 빠르게 확산되었습니다. 누구나 빅 데이터 분석에 대해 이야기하려 하지만 Tableau는 이를 단순화합니다.

—ASHISH BRAGANZA, 글로벌 비즈니스 인텔리전스 담당 이사, LENOVO

[Lenovo가 28개 국가에서 보고 효율성을 95% 향상한 방법 알아보기](#)

## 데이터 액세스 및 연결

모든 규모 및 형식의 데이터 분석이 가능하도록 데이터 위치와 관계없이 데이터에 대한 광범위한 액세스를 지원합니다. Tableau는 현재 75개 이상의 기본 데이터 커넥터뿐만 아니라 확장 옵션을 통해 다른 수많은 데이터 커넥터를 지원하고 있습니다. 새로운 데이터 원본이 출현하고 사용자에게 중요하게 됨에 따라 지속적으로 공급업체의 커넥터를 Tableau와 통합 및 인증하여 Tableau 제품 내에 통합함으로써 새로운 데이터에 액세스할 때 따르는 마찰을 줄입니다. 웹 트래픽이든, 데이터베이스의 레코드든, 로그 파일이든 사람들이 사용하려는 데이터 원본은 다양하며, 항상 다양할 것이라고 믿습니다.

- **SQL 기반 연결** — Tableau는 SQL을 사용해 Hadoop, NoSQL 데이터베이스 및 Spark와 연결합니다. Tableau가 생성하는 SQL은 ANSI SQL-92 표준을 따릅니다. SQL은 매우 압축적(하나의 표현식)이고 오픈 소스이며 표준화되어 있고, 라이브러리에 의존하지 않으며 표현력이 매우 풍부하므로 SQL을 통해 매우 강력한 기능을 활용할 수 있게 됩니다. 예를 들어 SQL을 사용하면 조인 작업, 함수, 기준, 요약, 그룹화 및 중첩 작업을 표현할 수 있습니다.

- **NoSQL 인터페이스** — 이름이 암시하는 바대로 NoSQL('not only SQL') 데이터베이스는 관계형 형식 외에도 비관계형으로 모델링된 데이터를 보유할 수 있어 열, 문서, 키-값 및 그래프를 포함한 추가 저장 형식을 지원할 수 있습니다. 또한 SQL과 유사한 인터페이스도 지원할 수 있습니다.
- **ODBC** — Tableau는 SQL 및 빅 데이터 플랫폼에서 제공되는 SQL과 유사한 데이터 인터페이스 간의 변환 계층으로서 오픈 데이터베이스 연결(ODBC) 프로그래밍 표준을 활용하는 드라이버를 사용합니다. ODBC를 사용하면 SQL 표준을 지원하고 ODBC API를 구현하는 모든 데이터 원본에 액세스할 수 있습니다. Hadoop 사용을 위해서는 Hive Query Language(HiveQL), Impala SQL, BigSQL, Spark SQL 등의 인터페이스가 포함됩니다. 최고의 성능을 달성하기 위해 Tableau는 생성된 SQL을 사용자 지정 튜닝하고 집계, 필터 및 기타 SQL 작업을 빅 데이터 플랫폼으로 푸시합니다.
- **웹 데이터 커넥터** — Tableau 웹 데이터 커넥터 SDK를 사용하여 기존 커넥터 범위를 벗어나는 데이터에 연결할 수 있습니다. 셀프 서비스 분석 사용자는 내부 웹 서비스, JSON 데이터, REST API를 포함한 HTTP를 통해 액세스 가능한 거의 모든 데이터에 연결하여 외부의 데이터를 사용해 빅 데이터 분석을 강화할 수 있습니다.

## 모든 규모의 데이터와 빠른 상호 작용

다른 데이터와 통합하여 빠르게 인사이트를 발견할 수 있도록 사용자가 규모에 맞게 모든 데이터에 액세스할 수 있기를 원합니다. 빅 데이터를 사용하는 셀프 서비스 및 시각적 분석이 가능해지도록 Tableau는 몇 가지 첨단 기술 개발에 노력해왔습니다.

- **Hyper 데이터 엔진** — **Hyper**는 고객이 규모가 크거나 복잡한 데이터 집합을 더 빠르게 분석할 수 있도록 지원하는 고성능 인메모리 데이터 엔진 기술입니다. 자체 개발한 동적 코드 생성 및 최신 병렬 처리 기술을 사용하는 Hyper는 최신 하드웨어 활용도를 높여서 이전 Tableau 데이터 엔진 대비 추출 생성이 최대 3배 더 빠르고 쿼리 속도 또한 5배 더 빠릅니다. Hyper는 또한 데이터를 추출하여 인메모리에 가져옴으로써 다소 속도가 느린 데이터 원본을 보강하고 속도를 개선할 수 있습니다.
- **하이브리드 데이터 아키텍처** — Tableau는 데이터 원본에 라이브 연결하거나 데이터(또는 하위 집합)를 인메모리로 가져올 수 있습니다. 필요에 따라 이 두 모드 간에 전환할 수 있습니다. Tableau의 하이브리드 데이터 액세스 접근 방식으로 사용자에게 더 많은 유용성을 제공하고 쿼리 성능을 최적화할 수 있습니다.

· **VizQL™** — Tableau의 핵심은 대화형 데이터 시각화를 데이터 이해 과정의 필수 요소로 만들어 주는 특허 기술입니다. 기존의 분석 도구를 사용하면 데이터를 행과 열로 분석하고, 표시할 하위 데이터 집합을 선택하고, 해당 데이터를 표로 구성한 다음, 해당 테이블에서 차트를 생성해야 했습니다. VizQL에서는 이러한 단계를 거치지 않고 데이터에 대한 시각적 자료를 즉시 만들어, 분석할 때 시각적 피드백을 제공합니다. VizQL을 사용하면 데이터를 무제한 탐색하여 가장 효율적으로 표현할 방법을 찾을 수 있으며 무제한으로 '실행 취소'가 가능하므로 실패를 두려워할 필요가 없습니다. 이러한 시각적 분석 주기에서 사용자는 작업해 나가며 배우고 필요에 따라 데이터를 추가하며 궁극적으로 더 깊은 인사이트를 확보할 수 있습니다. 이로써 더 풍부한 환경이 될 뿐 아니라 코드로 대시보드를 구축할 때에 비해 모든 기술 수준의 사용자가 더 많이 이용할 수 있습니다.

**“ Tableau를 사용하면 실시간으로 데이터 집합과 실제로 상호 작용할 수 있고 몇 분 만에 원하는 방식으로 분석한 다음 발표할 수 있습니다.**

—JAMIE FAN, 제품 분석 책임자, GRAB

**Grab이 수백만 개의 데이터 행을 분석해 고객 사용 환경(CX)을 개선한 방법 알아보기**

## Tableau와 빅 데이터 분석 에코시스템

Tableau와 같은 최신 분석 플랫폼은 인사이트 발견을 통해 빅 데이터의 잠재력을 끌어내는 열쇠가 될 수 있지만 아직은 전체 빅 데이터 플랫폼 아키텍처의 중요한 요소 중 하나에 불과합니다. 전체 빅 데이터 분석 파이프라인을 하나로 통합하는 것 자체가 하나의 과제처럼 보일 수 있습니다. 다행히, 시작하기 전에 전체 환경을 완전히 구축할 필요는 없으며, 시작부터 전체 전략에 지원하는 모든 개별 구성 요소를 통합할 필요도 없습니다.

Tableau는 플랫폼 간에 데이터를 이동하고, 필요에 따라 인프라를 조정하고, 새로운 데이터 유형의 이점을 활용하며, 새로운 사용자 및 사용 사례를 사용할 수 있는 기능인 유연성을 가장 우선순위로 두기 때문에 빅 데이터 패러다임에 잘 들어맞습니다. 빅 데이터 분석 솔루션을 배포하는 것은 인프라 또는 전략을 강요하는 것이 아니라, 빅 데이터 환경 내의 파트너 기술을 사용한 투자를 포함하여 고객의 기존 투자를 활용하도록 지원해야 한다고 믿습니다.

## 클라우드 인프라

조직은 점점 더 많이 비즈니스 프로세스와 인프라를 클라우드로 이동하고 있습니다. 클라우드 기반 인프라 및 데이터 서비스를 통해 온프레미스 Hadoop 데이터 레이크를 사용할 때 직면하는 중요한 몇 가지 어려움을 제거했으므로 클라우드 기반 빅 데이터 분석 솔루션은 그 어느 때보다도 더 구현 및 관리가 쉬워졌습니다.

Hadoop은 저비용, 확장형 저장소(Hadoop 분산 파일 시스템, HDFS), 전용 처리 엔진(MapReduce를 시작으로, 시간이 지남에 따라 Hive, Impala 및 Spark), 그리고 공유 데이터 카탈로그(Hive 메타스토어)의 강력한 조합으로 최신 데이터 레이크를 위한 기반을 구축했습니다.

현재 한때 함께 배치되었던 저장소 및 컴퓨팅 서비스는 필요에 따라 클라우드에서 독립적으로 확장 가능합니다. 리소스 또한 훨씬 더 쉽게 확장 및 축소되며 사용한 만큼 가격이 책정됩니다. 종합적으로 클라우드는 서비스의 더 나은 효율성, 관리 및 조직화에 기여합니다.

AtScale 제품 담당 부사장인 Josh Klahr의 [이 흥미로운 글](#)에서 더 많은 내용을 알아보세요.

Tableau는 [Amazon Web Services](#), [Google Cloud Platform](#) 및 [Microsoft Azure](#)를 비롯해 조직에서 이미 사용 중인 클라우드 기반 기술에 대한 핵심적인 통합을 제공합니다.

## 수집 및 준비

최신 수집 및 로드 설계 패턴에서 크기나 모양에 상관없이 원시 데이터는 통상적으로 정형이든, 반정형이든, 아니면 비정형이든 데이터의 기본 형식으로 광대한 양의 데이터를 저장하는 저장소 리포지토리인 데이터 레이크에 전송되어 저장됩니다. 데이터 레이크는 누구나 원시 데이터를 다양한 방식으로 신속하게 분석할 수 있도록, 더욱 빠르고 유연한 데이터 수집 및 저장을 통해 최신 빅 데이터 분석 요건을 충족합니다.

스트림 데이터는 소셜 네트워크, 스마트 계량기, 홈 자동화, 비디오 게임, IoT 센서 등 연결된 모든 기기 및 앱에서 계속해서 생성됩니다. 때로는 이러한 데이터가 반정형 데이터 경로를 통해 수집되기도 합니다. 실시간 분석 및 예측 알고리즘을 스트림에 적용할 수 있지만, 일반적으로 스트림 데이터는 람다 아키텍처(Lambda Architecture)를 사용해 원시 형식으로 전송/저장되거나 분석 용도로 Hadoop과 같은 데이터 레이크에 저장됩니다. 람다 아키텍처는 배치 및 스트림 처리 방법에서 각각의 장점을 살려 방대한 양의 데이터를 처리하도록 설계된 데이터 처리 아키텍처입니다. 이 아키텍처는 지연 시간, 처리량 및 오류 허용 문제의 균형을 맞추도록 설계되었습니다. 현재 스트리밍 데이터에는 Amazon Kinesis, Storm, Flume, Kafka, Informatica Vibe Data Stream과 같은 다양한 옵션이 있습니다.



데이터 레이크는 또한 '읽기 스키마' 기능으로 원시 데이터를 변환하는 API 또는 SQL 같은 언어를 통해 최적화된 처리 메커니즘을 제공합니다. 데이터가 데이터 레이크에 도달한 후에는 분석이 가능하도록 수집되고 준비되어야 합니다. Tableau는 이 프로세스를 지원하고 Tableau와 유연하게 협력하는 **Informatica**, **Alteryx**, **Trifacta**, **Datameer** 등의 파트너를 보유하고 있습니다. 또는 셀프 서비스 데이터 준비 용도로 **Tableau Prep**을 사용할 수 있습니다.

## 저장 및 처리

Hadoop은 복원력과 저렴한 비용, 확장형 저장소, 병렬 처리 및 클러스터링되는 워크로드 관리 덕분에 데이터 레이크에 사용되어 왔습니다. Hadoop은 빅 데이터 플랫폼으로 자주 사용되지만, 데이터베이스가 아닙니다. Hadoop은 범용 하드웨어 클러스터에서 데이터를 저장하고 응용 프로그램을 실행하기 위한 오픈 소스 소프트웨어 프레임워크입니다. Hadoop은 모든 종류의 데이터를 위한 대용량 저장소, 방대한 처리 능력과 엄청난 양의 동시 업무 또는 작업을 처리할 수 있는 기능을 제공합니다.

최신 분석 아키텍처에서 Hadoop은 오래된 기록 데이터를 데이터 웨어하우스에서 온라인 콜드 스토어로 이전하는 데 저렴한 비용으로 저장소 및 데이터 아카이브 기능을 제공합니다. 또한 IoT, 데이터 과학 및 비정형 분석 사용 사례에 사용됩니다. Tableau는 **Cloudera**(Impala), **Hortonworks**(Hive), 및 **MapR**(Apache Drill)을 사용하는 모든 중요한 Hadoop 배포와 직접 연결됩니다.

최신 분석 아키텍처에는 언제나 데이터베이스와 데이터 웨어하우스의 자리가 있을 것이며, 엔터프라이즈 전반의 셀프 서비스 보고를 위한 거버넌스, 정확성 및 적절한 차원형 데이터의 제공에 있어 지속적으로 중요한 역할을 수행할 것입니다. Hadoop, 데이터 레이크 등의 다른 기술을 채용한 회사도 보통 관계형 데이터베이스를 혼합된 데이터 원본의 일부로 유지합니다. **Snowflake**는 Tableau 지정 커넥터를 사용하는 클라우드 기본 SQL 기반의 엔터프라이즈 데이터 웨어하우스의 예입니다.

Amazon Web Services S3(Simple Storage Service), 유연한 스키마의 NoSQL 데이터베이스와 같은 개체 저장소도 데이터 레이크로 사용될 수 있습니다. Tableau는 Amazon S3에 연결되는 **Amazon의 Athena 데이터 서비스**를 지원하며 NoSQL 데이터베이스에 직접 연결할 수 있는 다양한 도구도 제공합니다. Tableau에서 자주 사용되는 NoSQL 데이터베이스에는 **MongoDB**, **Datastax**, **MarkLogic** 등이 있습니다.

데이터 과학 및 엔지니어링 플랫폼인 **Databricks**는 일괄 처리 중심 및 대화형의 두 가지 모든 확장형 데이터 처리를 위해 많이 사용되는 엔진인 Spark에서 데이터 처리 기능을 제공합니다. Spark 지정 커넥터를 통해 복잡한 기계 학습 모델의 결과를 Tableau의 Databricks에서 시각화할 수 있습니다.

## 쿼리 가속화

빅 데이터에서 기계 학습과 정서 분석을 수행할 수 있지만, 사람들은 일반적으로 대화형 SQL이 얼마나 빠르는지 가장 먼저 질문합니다. 결국 더 빠르고 반복적으로 KPI 대시보드뿐 아니라 탐색적 분석을 위해 빅 데이터를 사용하려는 비즈니스 사용자에게 SQL이 핵심 역할을 하기 때문입니다.

이러한 속도에 대한 요구는 **Exasol** 및 **MemSQL** 같은 인메모리 및 대규모 병렬 처리(MPP) 기술을 활용하는 더 빠른 데이터베이스와, Kudu 같은 Hadoop 기반 저장소, **Vertica** 같은 사전 처리를 통해 더 빨리 쿼리할 수 있는 기술의 채택을 촉진했습니다. Hadoop 엔진에서의 SQL(Apache Impala, Hive LLAP, Presto, Phoenix 및 Drill) 사용과 Hadoop 기술에서의 OLAP 사용(**AtScale**, **Jethro Data** 및 **Kyvos Insights**)으로 인해 이러한 쿼리 가속기는 기존 웨어하우스와 빅 데이터의 경계를 더욱 모호하게 만듭니다.

## 데이터 카탈로그

엔터프라이즈 데이터 카탈로그는 기본적으로 데이터 원본 및 공통 데이터 정의에 대한 비즈니스 용어집의 역할을 하므로, 사용자는 관리 및 승인된 데이터 원본으로부터 의사 결정을 위한 적절한 데이터를 더 쉽게 찾을 수 있습니다. 엔터프라이즈 데이터 카탈로그는 수집된 데이터 원본을 검사하여 테이블, 뷰 및 저장 프로시저의 메타데이터로 채워집니다. 데이터 큐레이션 작업은 기술 자료 정보와 웹 링크까지 포함하도록 확장되어 사용자가 데이터의 컨텍스트를 이해하고 더 지능화된 분류 및 자동 데이터 탐색이 가능하도록 지원합니다.

데이터 카탈로그는 시각적 분석 솔루션 내에 존재하는데, Tableau와의 원활한 통합을 위해 독립형 서비스로도 제공됩니다. 데이터 카탈로그 파트너에는 **Informatica**, **Alation**, **Unifi**, **Collibra**, **Waterline** 등이 있습니다.

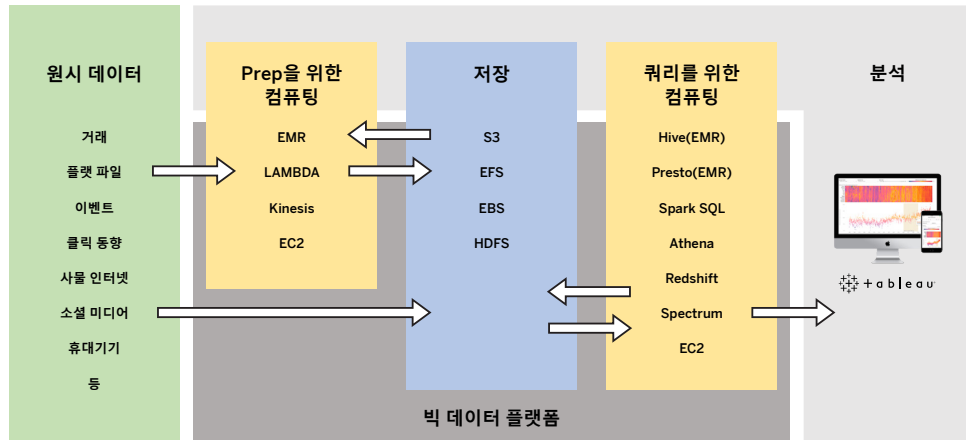
## 빅 데이터 분석 아키텍처

빅 데이터 아키텍처의 성공적인 구축에 관한 한 '모든 경우에 딱 들어맞는 방법'은 없다는 사실을 반드시 기억해야 합니다. Tableau 고객은 빅 데이터 분석을 위한 고객만의 맞춤 솔루션을 보유하고 있으며 데이터 파이프라인을 구성할 다양한 플랫폼과 도구를 갖추고 있습니다. 그렇지만, 이러한 빅 데이터 분석 플랫폼의 성공에 기여한 공통된 아키텍처 요소에 대한 몇 가지 관찰 결과를 이어서 보여드리려고 합니다.

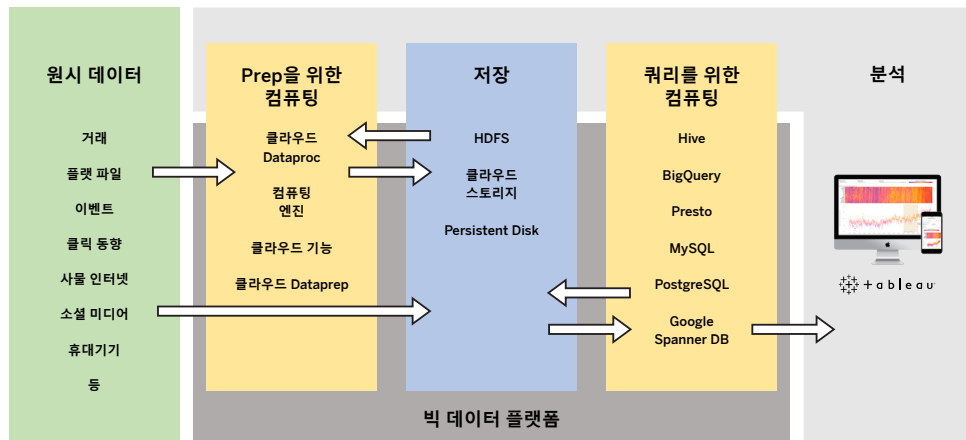
**고지 사항:** 아래의 예는 Tableau의 해석이며 클라우드 제공업체 또는 그들이 대표하는 고객이 계획했던 바는 아니었습니다. 가능한 경우, 원래 도표에 대한 링크를 포함하였습니다. 아래의 도표들은 다른 흐름 간의 핵심 요소의 유사성을 강조하기 위해 단순화한 일반화 도표입니다. 전체 빅 데이터 분석 플랫폼의 모든 요소를 반영하지 않을 수 있으며 특정 사용 사례만을 나타낼 수 있습니다. '준비를 위한 컴퓨팅'은 '프로세스/카탈로그'와 가장 유사하며 여기에서 '쿼리를 위한 컴퓨팅'은 '분석/모델'과 가장 유사합니다.

# 주요 클라우드 제공업체 예

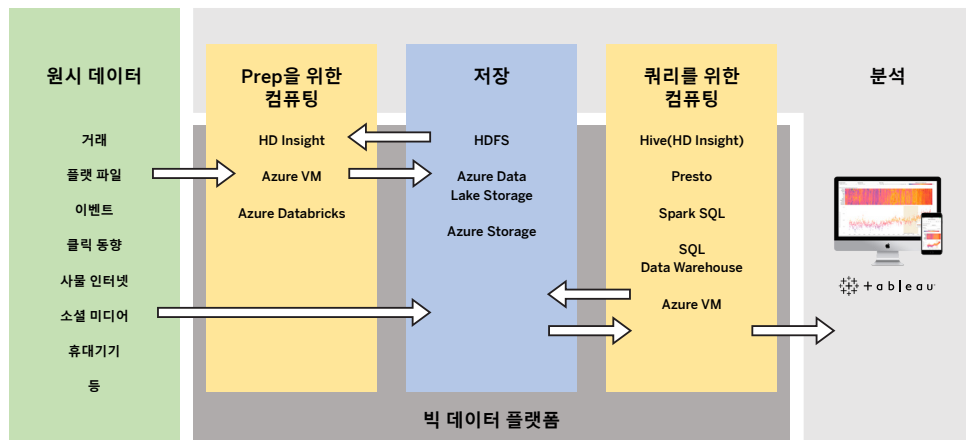
## Amazon Web Services



## Google Cloud Platform

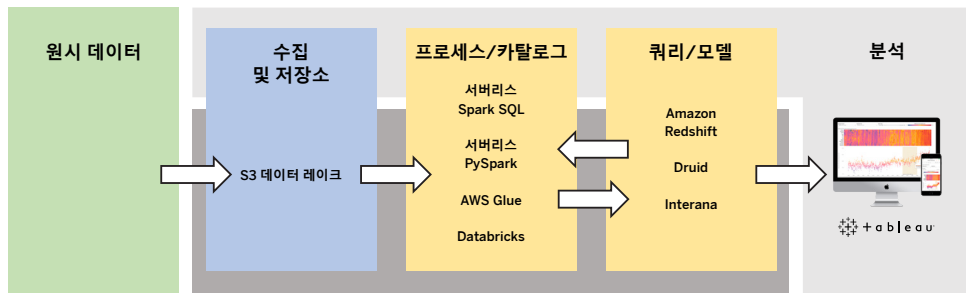


## Microsoft Azure

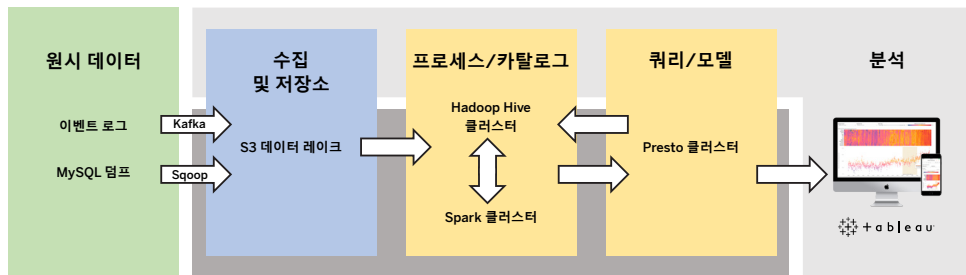


# Tableau 고객 사례

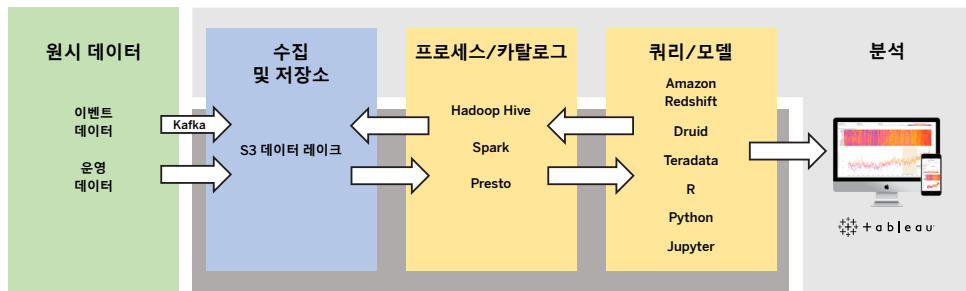
## Edmunds – 자세히 알아보기



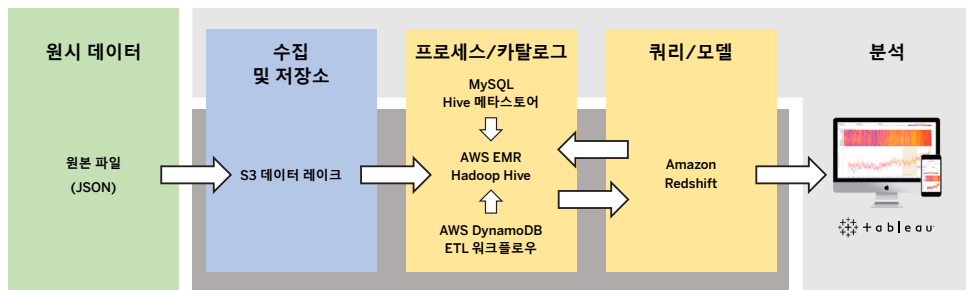
## Airbnb – 자세히 알아보기



## Netflix – 자세히 알아보기



## Expedia – 자세히 알아보기



## 공통 패턴

어떤 엔터프라이즈 아키텍처도 서로 같지 않지만 유사한 패턴과 공통으로 공유하는 요소에 주목하면 나에게 맞는 빅 데이터 분석 플랫폼 전략을 세우는 데 도움이 될 수 있습니다. 성공적인 빅 데이터 분석 아키텍처에서 일관성 있게 관찰되는 내용은 다음과 같습니다.

- **저장소 계층** — 흔히 데이터 레이크라고 합니다. 데이터 전략에는 복수의 저장소 환경이 필요할 수 있는데, 반드시 정형, 반정형 및 비정형 데이터로 구성되어야 합니다.
- **서버 및 서버리스 컴퓨팅 엔진** — 일부는 데이터 준비 및 분석용이고 그외 다른 엔진은 쿼리용입니다. 리소스를 사전 할당할 필요가 없는 서버리스 컴퓨팅의 동적 본질 덕분에 더 많은 유연성과 탄력성이 허용됩니다.
- **규모, 속도 및 다양성 지원** — 이는 데이터 뿐만 아니라 아직 발견되지 않은 예를 포함하여, 점점 더 복잡해지고 다양해지는 사용 사례에도 적용됩니다.
- **작업에 맞는 도구** — 아키텍처의 요소를 조직의 고유한 데이터 전략에 맞추는 것도 중요하지만 변화하는 비즈니스 필요에 맞서는 대응력을 유지하는 것 또한 매우 중요합니다.
- **엔터프라이즈 수준 거버넌스 및 보안** — 이 영역에 관해서는 자세히 다루지 않았지만 보안 및 거버넌스는 확장성 및 데이터의 적절한 사용을 보장하기 위한 기반입니다.
- **비용 의식** — 빅 데이터 아키텍처에 필요한 능력과 유연성을 고려할 때 비용을 감안해야 합니다. 클라우드는 성장에 많은 탄력을 가하지만, 데이터 저장 및 처리, 동시성, 지연, 분석 사용 사례 등의 재정적 의미를 검토하는 것이 바람직합니다.

빅 데이터 환경이 계속 진화함에 따라, 모든 도전 과제에 공통인 한 가지 주제가 있습니다. 바로, 기업은 데이터의 규모가 크든 작든, 어느 위치에 있던 액세스할 수 있는 최신의 공통된 분석 플랫폼을 사용할 수 있어야 한다는 것입니다. 적절한 플랫폼, 프로세스, 그리고 직원에게 권한을 부여하는 프로그램을 통해 이뤄지는 데이터 기반 의사 결정이 엄청난 자산임이 입증될 것입니다.



## Tableau 정보

Tableau는 완전하고, 사용하기 쉬우며, 기업에서 바로 사용할 수 있는 시각적 비즈니스 인텔리전스 플랫폼으로서, 신속하고 규모에 맞는 셀프 서비스 분석을 통해 사람들이 데이터를 보고 이해할 수 있도록 지원합니다. Tableau는, 온프레미스나 클라우드, Windows나 Linux 상관없이 기존 기술 투자를 활용하고, 사용자의 데이터 환경이 변화하고 성장함에 따라 확장됩니다. 가장 가치 있는 자산인 데이터와 인적 자원을 효율적으로 활용하십시오.

## 추가 리소스

[최신 분석 플랫폼의 구성 요소](#)

[IT가 지원하는 Tableau 엔터프라이즈 분석](#)

[엔터프라이즈용 Tableau: IT 개요](#)

[Tableau 무료 평가판](#)