

# 클라우드를 통한 새로운 통찰력의 확보

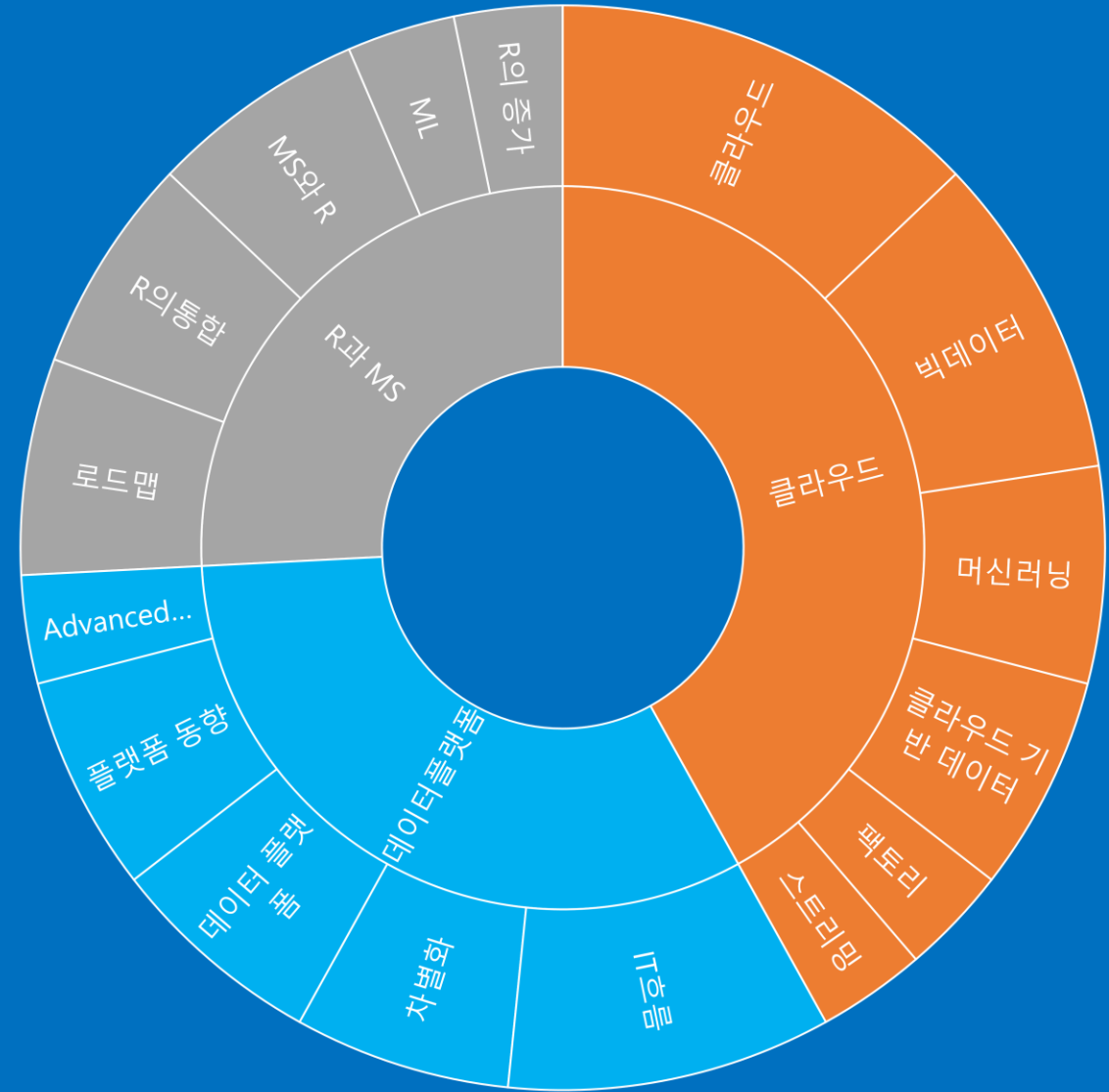
이건복 (kblee@microsoft.com)

Microsoft

# 목차

- 데이터 플랫폼
  - IT 흐름
  - 플랫폼 동향
  - 데이터 플랫폼
  - Advanced Analytics
  - 차별화
- 클라우드를 통한 분석의 진화
  - 클라우드 기반 데이터
  - 빅데이터
  - 스트리밍
  - 팩토리
  - 머신러닝
  - R의 증가
- R과 마이크로소프트
  - ML
  - R의 통합
  - 로드맵

plot()

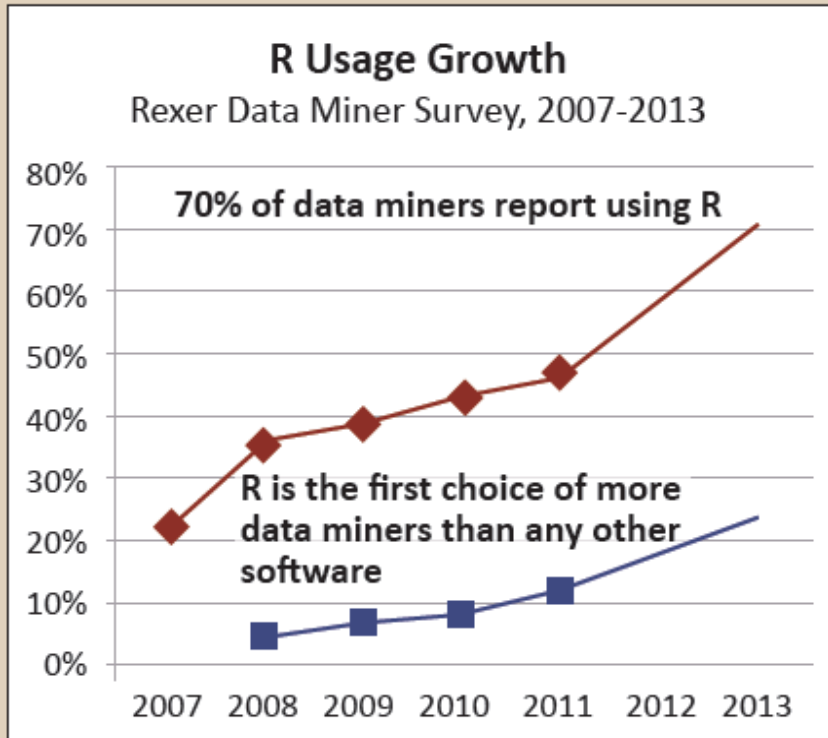


# IT비즈니스의 흐름



# R 사용의 증가

## R is Exploding in Popularity & Functionality



Source: [www.rexeranalytics.com](http://www.rexeranalytics.com)

*"I've been astonished by the rate at which R has been adopted. Four years ago, everyone in my economics department [at the University of Chicago] was using Stata; now, as far as I can tell, R is the standard tool, and students learn it first."*

– Deputy Editor for New Products at Forbes

*"A key benefit of R is that it provides near-instant availability of new and experimental methods created by its user base — without waiting for the development/release cycle of commercial software. SAS recognizes the value of R to our customer base..."*

– Product Marketing Manager SAS Institute, Inc.

- R 은 가장 높은 임금을 받을 수 있는 IT기술중에 하나이다.  
(Dice.com survey, January 2014)
- R 은 SQL이후 가장 많이 사용되는 데이터 과학용 기술  
(O'Reilly survey, January 2014)
- 이미 70%의 데이터 마이닝 전문가가 R 을 사용하고 있음  
(Rexer survey, October 2013)
- R 은 전체 프로그램 언어중에 15위에 위치하고 있다.  
(RedMonk language rankings, January 2014)
- R 은 다른 데이터 과학 언어보다 빠르게 성장하고 있다.  
(KDNuggets survey, August 2013)

# 플랫폼으로서의 데이터

데이터 시각화  
중요성 증가

클라우드  
기반의  
데이터 처리

전통적인  
데이터베이스  
시장의 포화

다양한 형태의  
데이터 증가

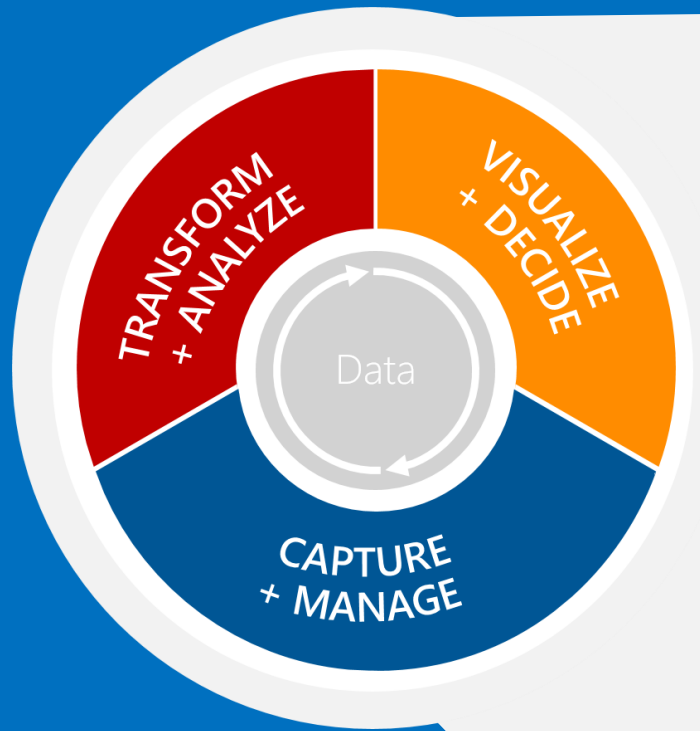
모바일  
소셜  
최적화

데이터의 빠른  
증가와 늘어난  
보관 주기

Business and Analytics Platforms

LEADERS





# 데이터 플랫폼

## 시각화 + 의사결정



응용 프로그램



보고서



대시보드



자연어 쿼리



모바일

## 변환 + 분석



데이터  
Orchestration



정보관리



이벤트  
데이터처리



모델링



머신러닝

## 저장 + 관리



관계형 데이터



비 관계형  
데이터



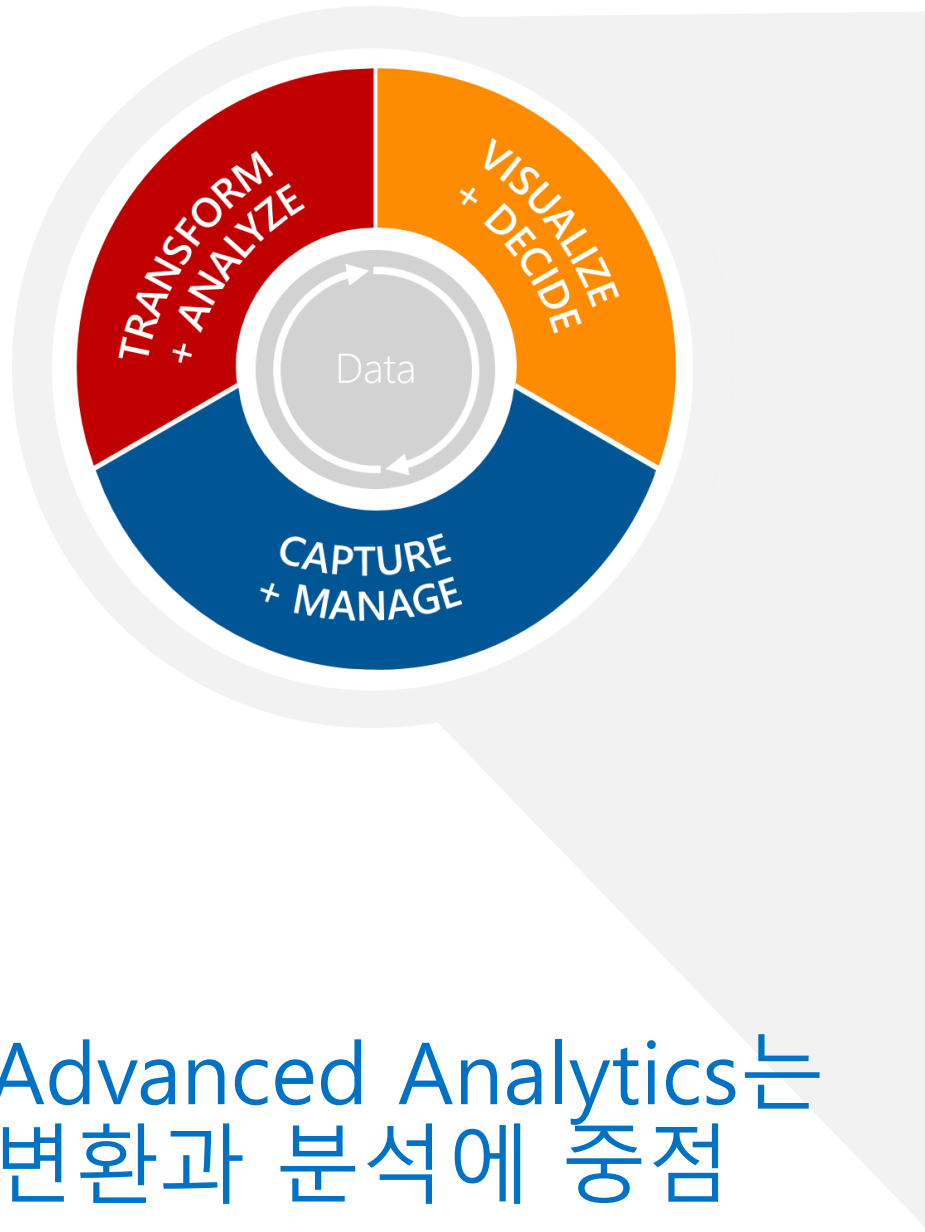
NoSQL



스트리밍



내부외부  
데이터 저장



Advanced Analytics는  
변환과 분석에 중점

## 시각화 + 의사결정



응용 프로그램



보고서



대시보드



자연어 쿼리



모바일

## 변환 + 분석



데이터  
Orchestration



정보관리



이벤트 처리



모델링



머신러닝

## 저장 + 관리



관계형



비관계형



NoSQL



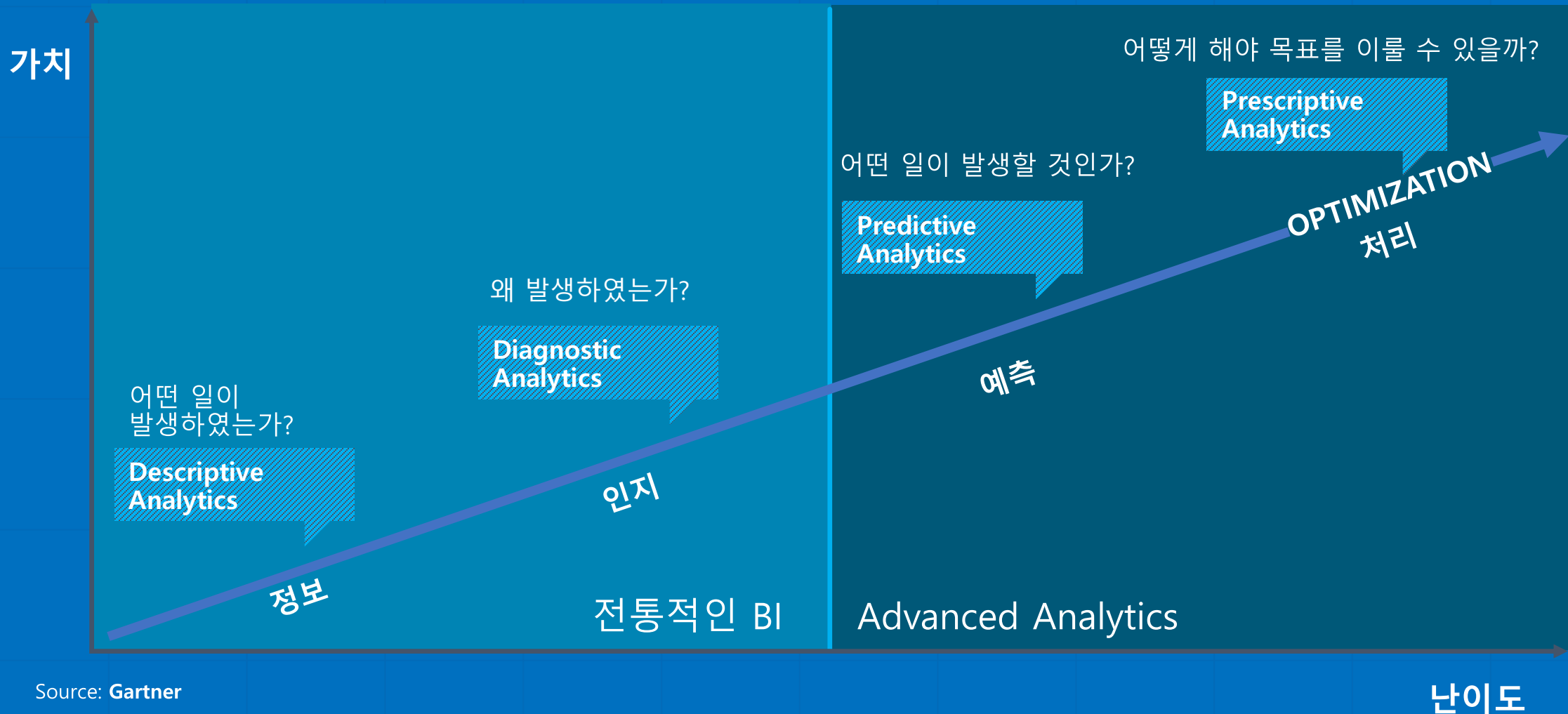
스트리밍



내부외부  
데이터 저장

# Advanced Analytics

## BI 그 이상의 의미





# 차별화

접근성

## 모형화

R, Python & SQLite  
데이터 모형구축

## 빠른 전개

단순한 배포  
모형의 서비스 제공

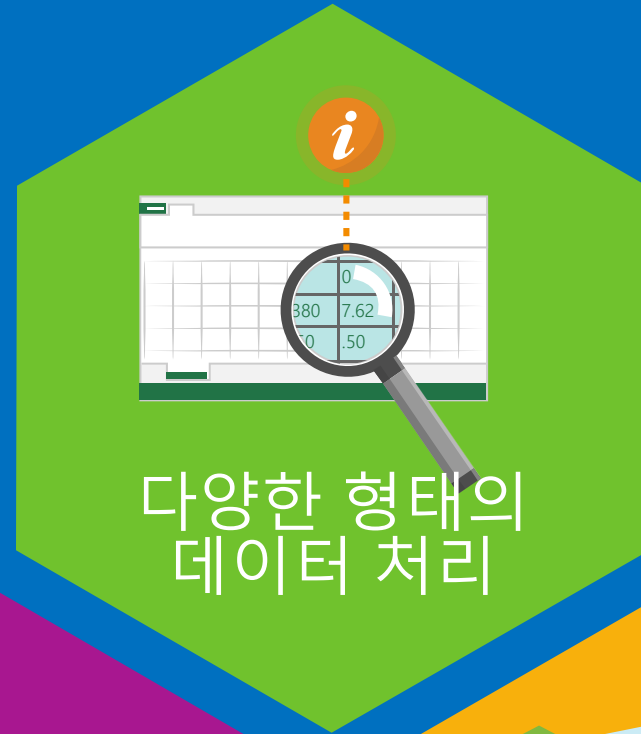
## 영역의 확장

시장기반의 비즈니스  
마켓플레이스

# 클라우드 플랫폼



글로벌  
데이터센터



다양한 형태의  
데이터 처리



기업의 모빌리티



IoT



응용 프로그램의  
개선

머신러닝

BI

검색

스트리밍 분석

HADOOP

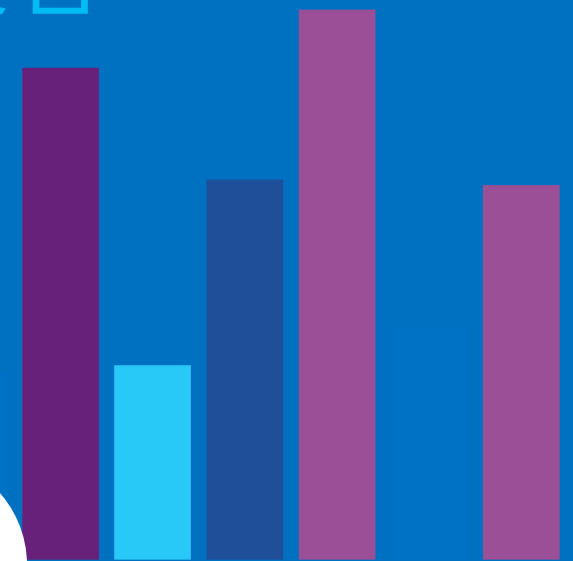
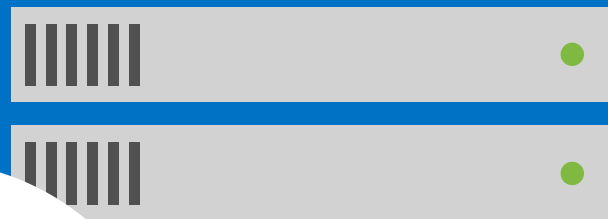
데이터 팩토리

데이터베이스

클라우드 데이터베이스

NOSQL

데이터 플랫폼



# 빅데이터

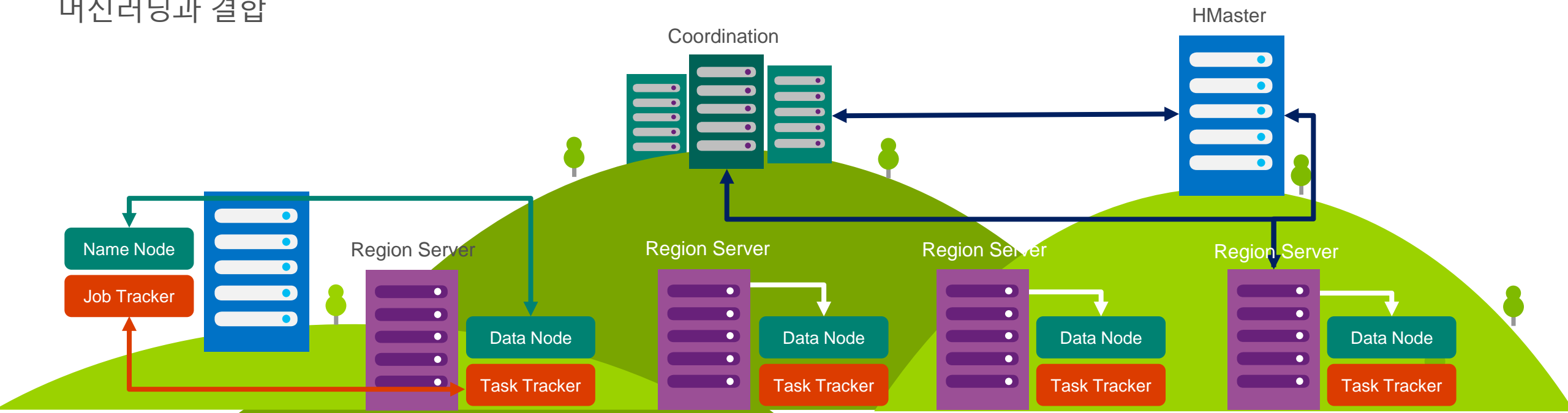
## 클라우드 기반 R-Hadoop

클라우드 기반의 엔진운영을 통한 인프라 비용의 절감

실시간에 가까운 분석 서비스 활용 (Storm)

최신 버전의 Hadoop 적용

머신러닝과 결합



# 스트림 분석

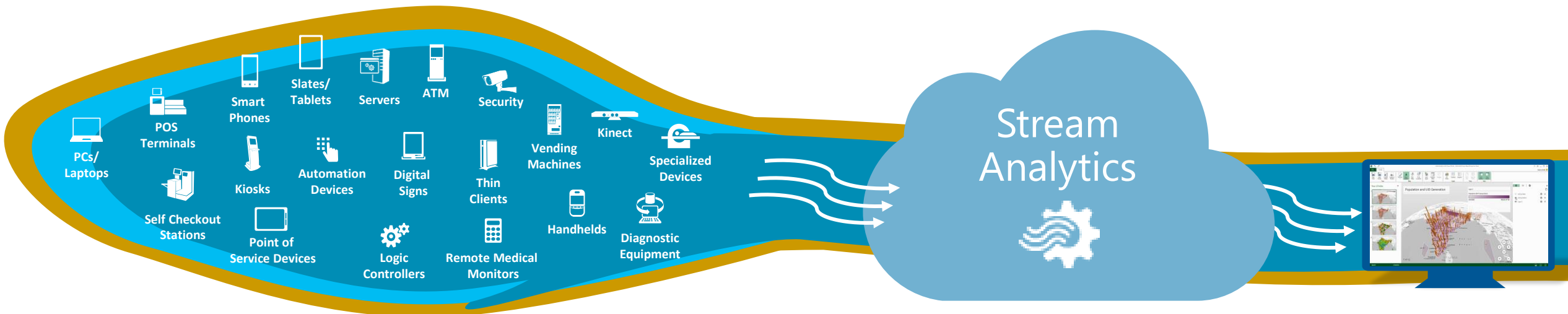
## 실시간 데이터의 처리

IoT장치 또는 센서와 응용 프로그램에서 발생하는 수 많은 이벤트 데이터의 저장과 처리

시간에 따른 분석을 SQL언어를 통한 처리

다양한 소스의 데이터의 동시 스트림 처리

결과를 클라우드내 저장 및 데시보드를 통한 실시간 정보 제공



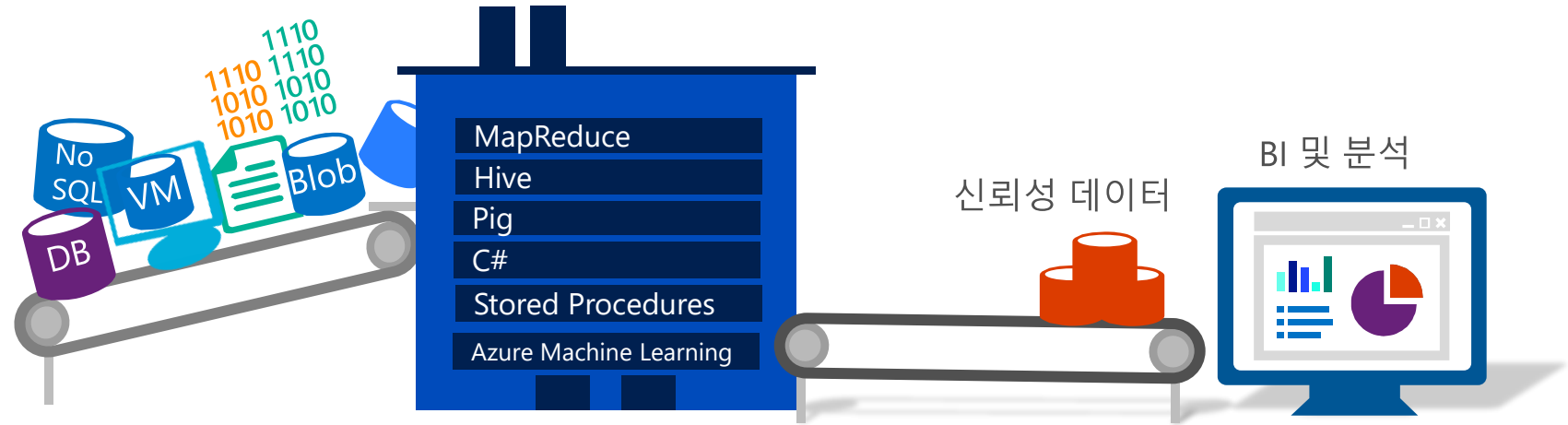
# 데이터 팩토리 (Data Factory)

## 다양한 데이터의 통합처리

관계형 및 비관계형 데이터에  
대한 통합 처리 공간

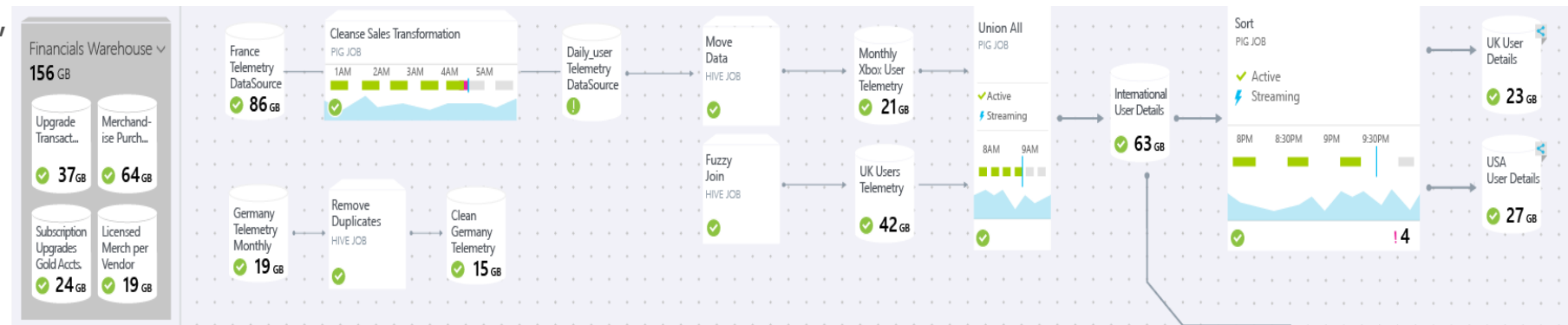
데이터의 전송과 변환에 대한  
오케스트레이션

BI사용자에게 검색가능한  
데이터로 배포



데이터 워크플로우 관리  
(워크플로우 일정, 데이터 관리,  
디버그)

데이터 라이프사이클 관리 및  
모니터링



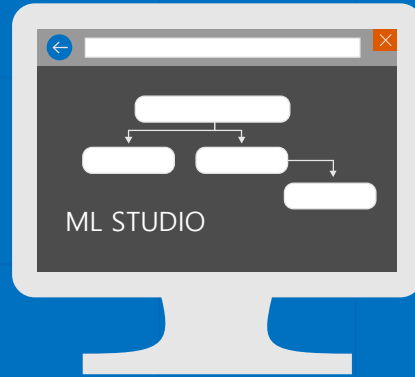
# 머신러닝

데이터 -> 예측모형 -> 웹 API를 이용한 처리

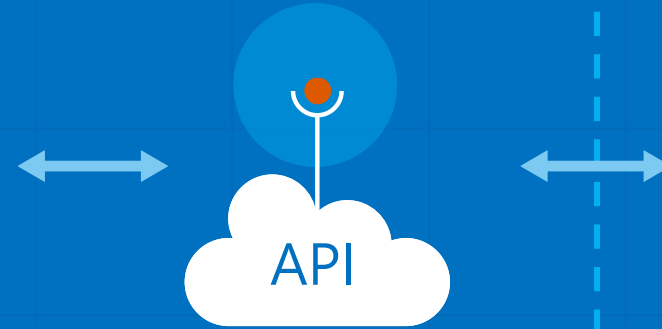
데이터



클라우드 공간에 수집된 데이터  
Hadoop  
관계형 DB



예측 모형 디자이너  
예: Azure ML



API호출이 가능한  
형태의 모형



마켓플레이스에서  
서비스형태로 제공

클라이언트



# 마이크로소프트와 Revolution Analytics

Revolution Analytics ∈ Microsoft == TRUE

분석 서비스의 메인  
영역으로의 확장 및  
클라우드와의 결합을 통한  
새로운 분석 서비스의  
제공



오픈소스 프로젝트는 지속적으로 제공

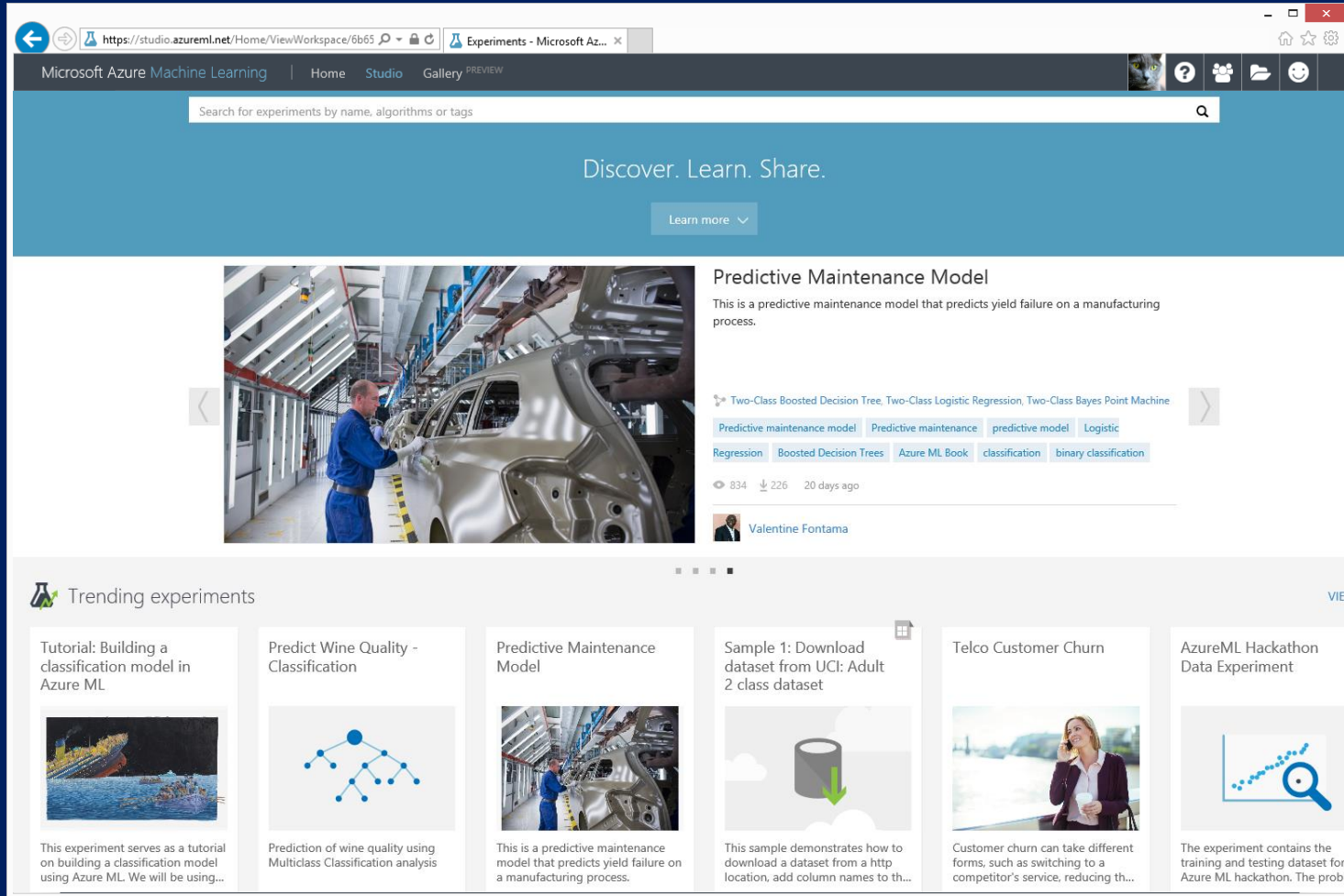
기존 마이크로소프트 제품과 통합

“R” 전문 영역의 확대

마이크로소프트 클라우드 환경과 호환



# R과 ML을 통한 분석 및 예측 Advanced Analytics



다양한 데이터의 처리

- 관계형 데이터
- 비구조화된 데이터
- IoT 데이터

데이터 분석 언어 및 마켓플레이스

- R
- Python
- Marketplace

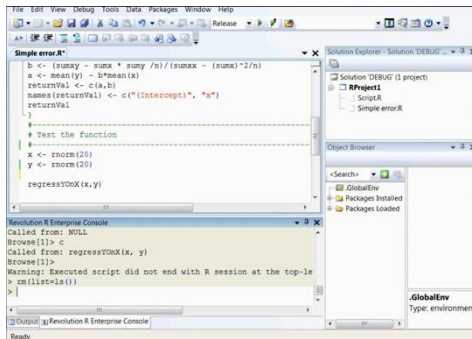
클라우드기반의 DB와의 연동

- 실시간 분석 기능

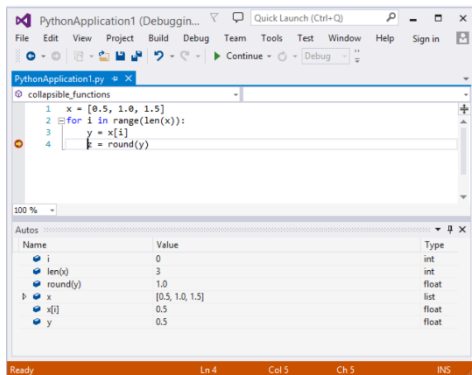
# R 사용자 도구

## 개발

 Visual Studio와 R 도구의 통합



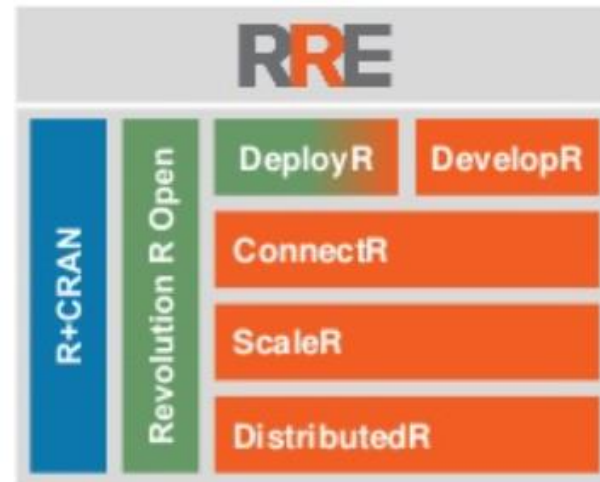
 Visual Studio Python 도구와 유사한 모델



## 배포

R기반의 기능의 제공.

- 클라우드 환경 : AzureML
- 온프레미스 환경:
  - T-SQL 에서 R 스크립트 호출
  - Revolution Analytics **DeployR**

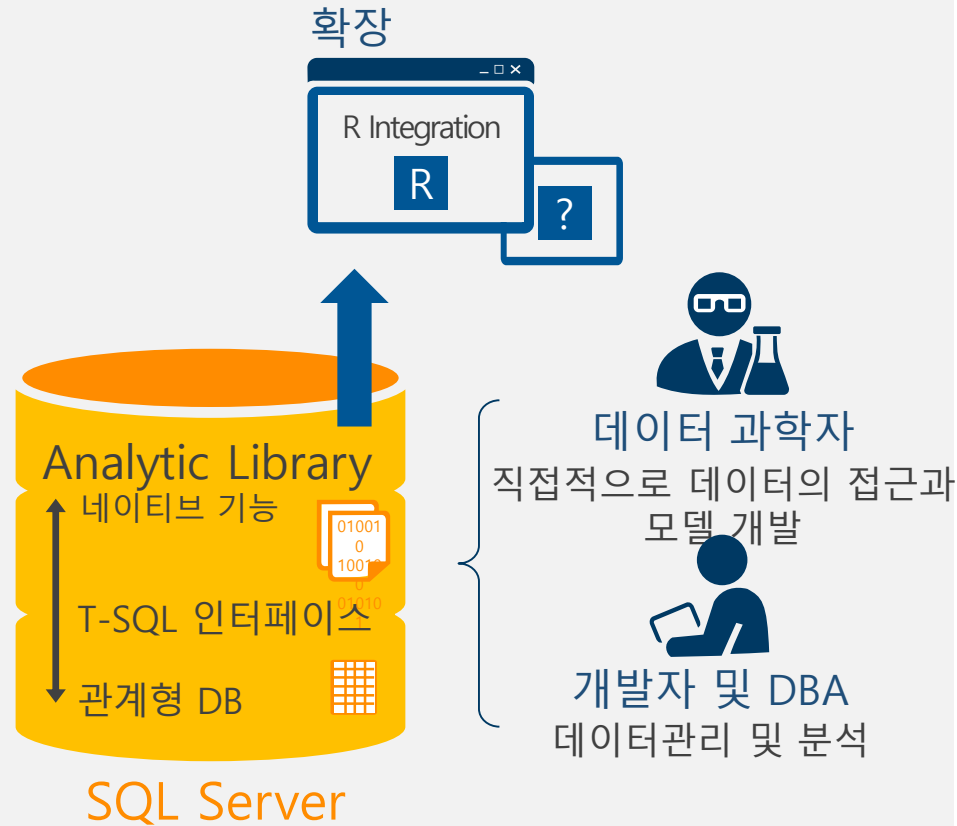


**DeployR** 서버는 어떠한 R기반의 분석 솔루션이 웹 서비스로서 동작할 수 있도록 함. 이 웹 서비스는 기존의 IT환경에 있는 자산들과 자연스럽게 연동이 될 수 있음

# Advanced Analytics을 위한 빌트인 기능

## 예제 솔루션

- 위변조 탐지
- 예상 매출
- 웨어하우스 효율성
- 관리 항목 예측



## 기능

SQL서버와 통합을 기반으로 TSQL을 이용한 ML모형의 개발 및 실행

## 장점

- ➔ ML모형의 빠른 배포와 사용
- ➔ 데이터의 이동을 최소화
- ➔ R 메모리 및 확장성 문제의 최소화

감사합니다!



© 2014 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries. The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.