

# On-premise DB와 Cloud Database 통합

신용환

Cloud Excellence, Cloud Platform SE Group

Oracle Korea

# Safe Harbor Statement

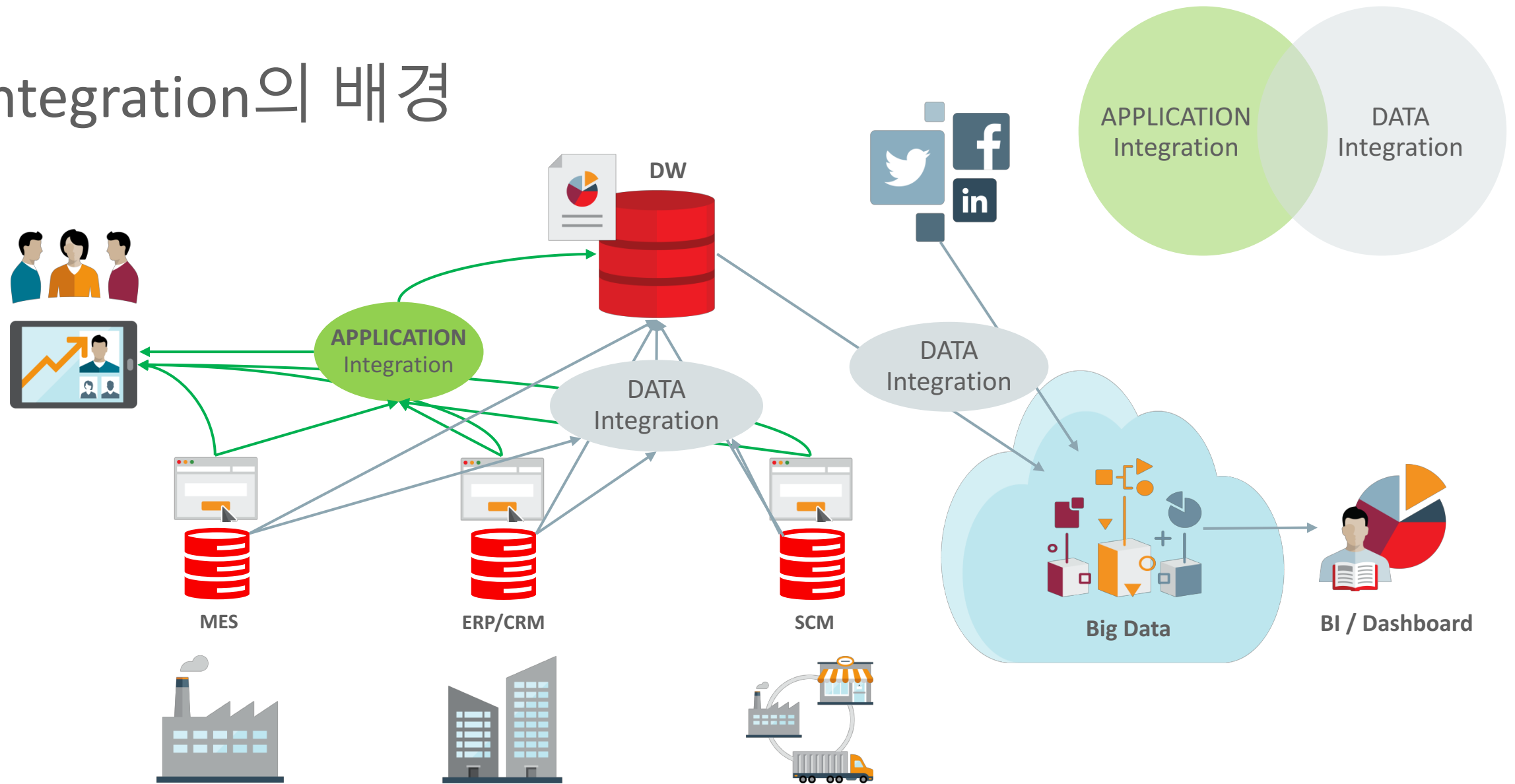
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

# Program Agenda

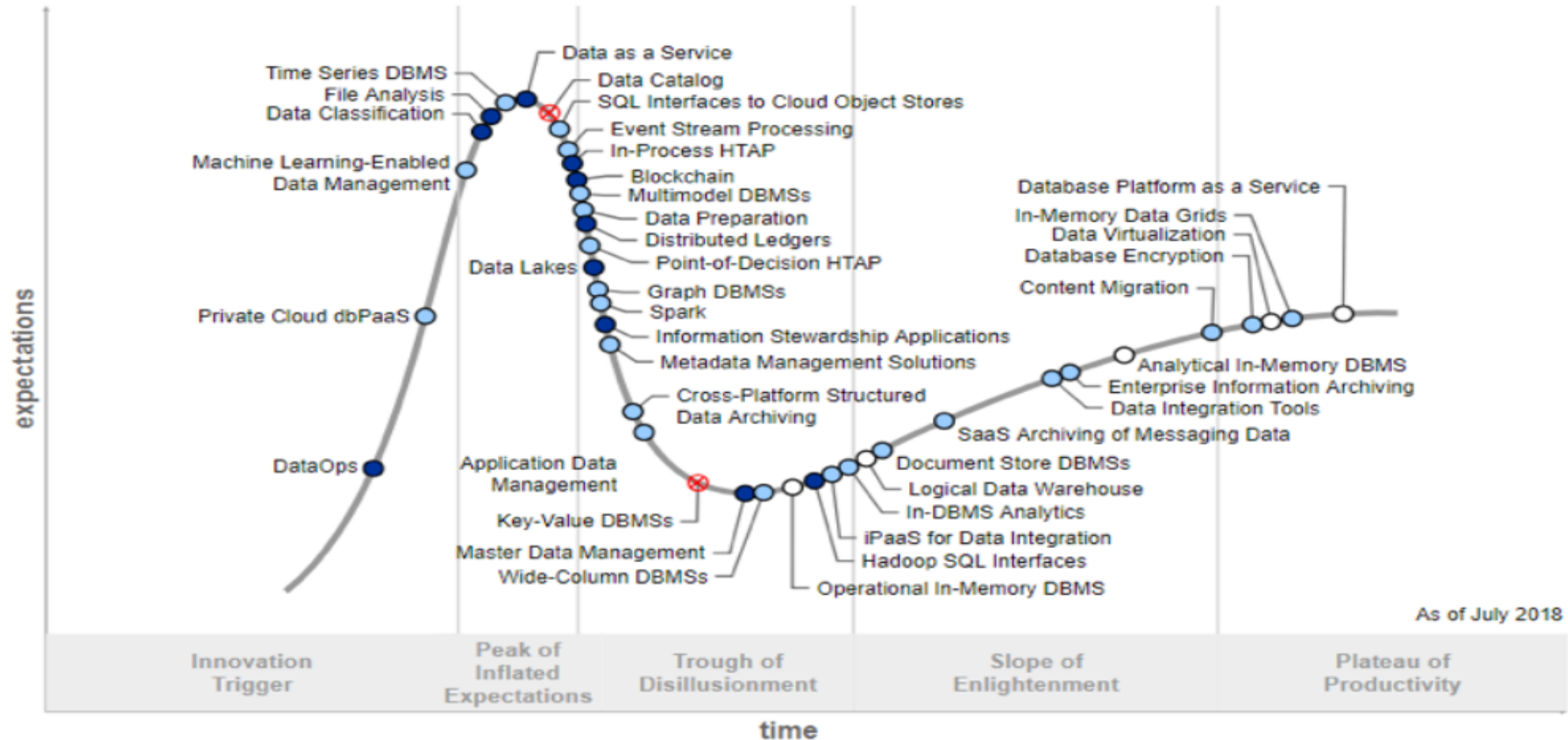
- 1 Concept & Trend
- 2 ETL & CDC
- 3 Data Management using Oracle Cloud

# Concept & Trend

# Integration의 배경



# Hype Cycle for Data Management, 2018, Gartner



# Data Management – Topics (Wikipedia)

## 1. Data Governance

- Data asset
- Data governance
- Data steward
- Data Ethics

## 2. Data Architecture

- Data architecture
- Data flows

## 3. Data modeling and Design

## 4. Database & Storage Management

- Data maintenance
- Database administration
- Database management system
- Business continuity planning

## 5. Data Security

- Data access
- Data erasure
- Data privacy
- Data security

## 6. Reference and Master Data

- Data integration
- Master data management
- Reference data

## 7. Data Integration and Inter-operability

- Data movement (Extract, transform, load )
- Data Interoperability

## 8. Documents and Content

- Document management system
- Records management

## 9. Data Warehousing and Business Intelligence

- Business intelligence
- Data analysis and Data mining
- Data warehouse and Data mart

## 10. Metadata

- Metadata management
- Metadata
- Metadata discovery
- Metadata publishing
- Metadata registry

## 11. Data Quality

- Data cleansing
- Data integrity
- Data enrichment
- Data quality
- Data quality assurance

Data Store

Movement

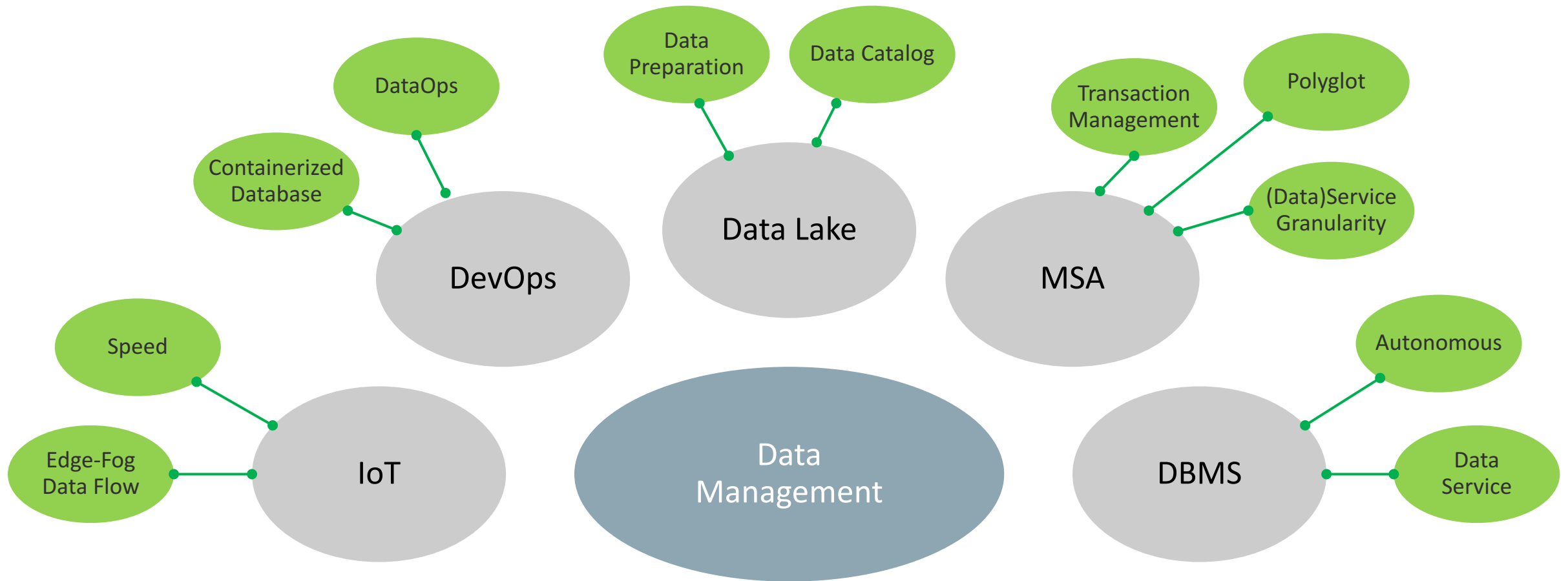
Data Governance

Data Quality

Metadata Mgmt.



# Modern Data Management Trend





# ETL & CDC

# Extract Transform Load (Wikipedia)

추출, 변환, 적재(Extract, transform, load, ETL)는 컴퓨팅에서 데이터베이스 이용의 한 과정으로 특히 데이터 웨어하우스에서 다음을 아우른다:

- 동일 기종 또는 타기종의 데이터 소스로부터 데이터를 추출한다.
- 조회 또는 분석을 목적으로 적절한 포맷이나 구조로 데이터를 저장하기 위해 데이터를 변환한다.
- 최종 대상(데이터베이스, 특히 운영 데이터 스토어, 데이터 마트, 데이터 웨어하우스)으로 변환 데이터를 적재한다.

## 상용 도구 [ 편집 ]

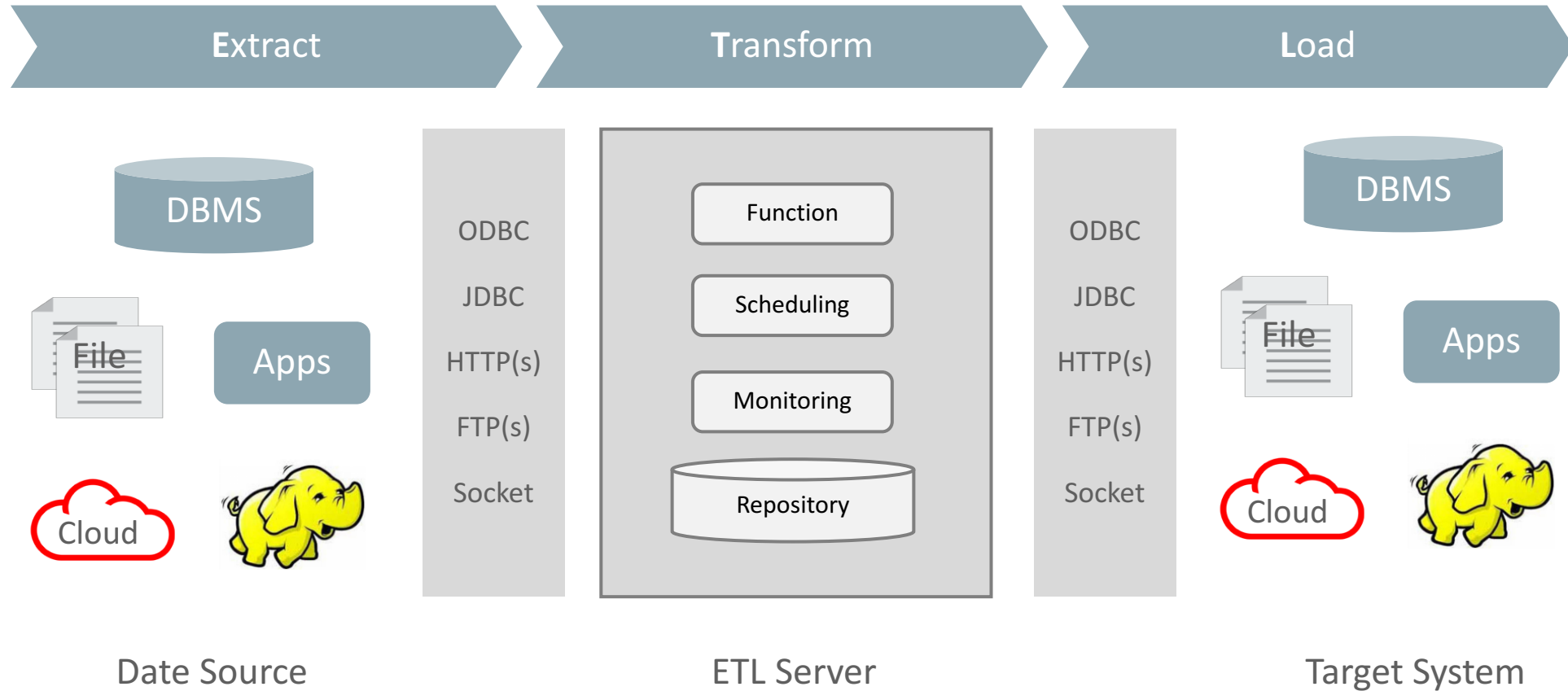
상업적으로 이용 가능한 ETL 도구는 다음을 포함한다

- Adeptia
- Alooma
- Analytics Canvas
- Anatella
- Alteryx
- Benetl (프리웨어)
- Boomi (델)
- CampaignRunner
- ESF Database Migration Toolkit
- Informatica PowerCenter
- Talend
- InnoQuartz
- IBM 인포스피어 데이터스태이지
- Ab Initio
- 오라클 데이터 인티그레이터 (ODI)
- 오라클 웨어하우스 빌더 (OWB)
- 마이크로소프트 SQL 서버 인티그레이션 서비스 (SSIS)
- Tomahawk Business Integrator (Novasoft Technologies)
- Pentaho Data Integration (or Kettle)
- Stambia
- Diyotta DI-SUITE for Modern Data Integration
- FlyData
- SAP 비즈니스 오브젝트 데이터 서비스
- SAS Data Integration Studio
- Skyvia
- SnapLogic
- Clover ETL
- SQ-ALL
- North Concepts Data Pipeline
- TeraStream (국산)

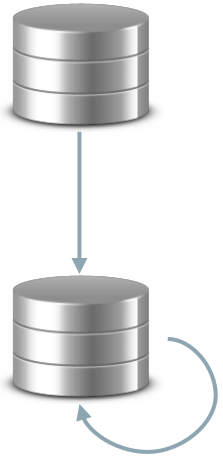
## 오픈 소스 도구 [ 편집 ]

- Talend Open Studio for Data Integration.
- Pentaho Data Integration
- Jaspersoft ETL
- KNIME
- Apache NIFI
- Rhino ETL
- StreamSets
- InnoQuartz ETL

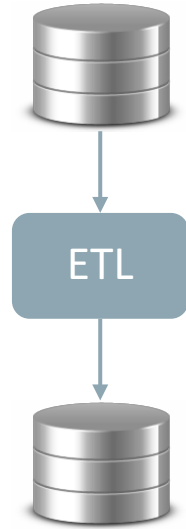
# ETL 아키텍처



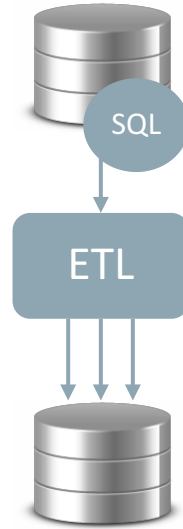
# Extract Transform Load 변천사



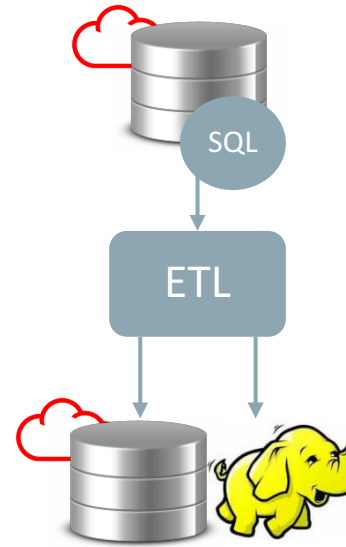
**SQL 기반 In-House**  
- File  
- Timestamp



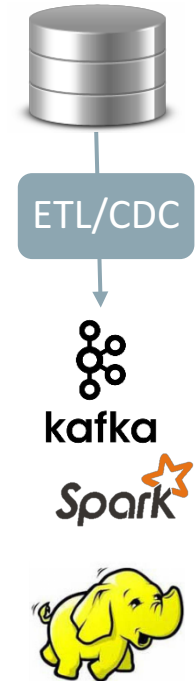
**ETL Tools**  
- Lots of Tools



**ETL Enhancement**  
- Parallel  
- Pushdown (ELT)  
- RealTime (CDC)



**Multi-Platform**  
- Bigdata  
- Cloud

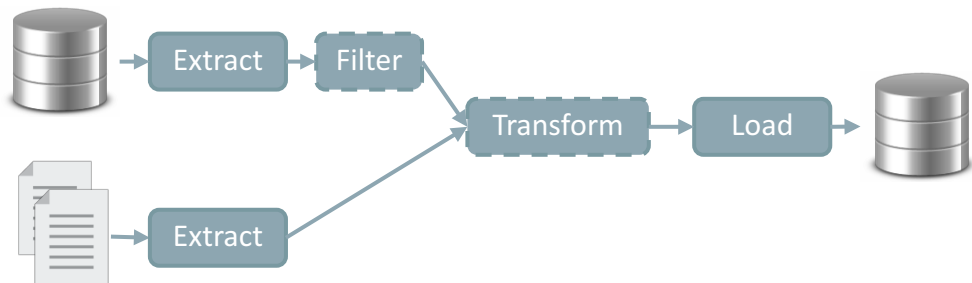


**Streaming**  
- Pipeline

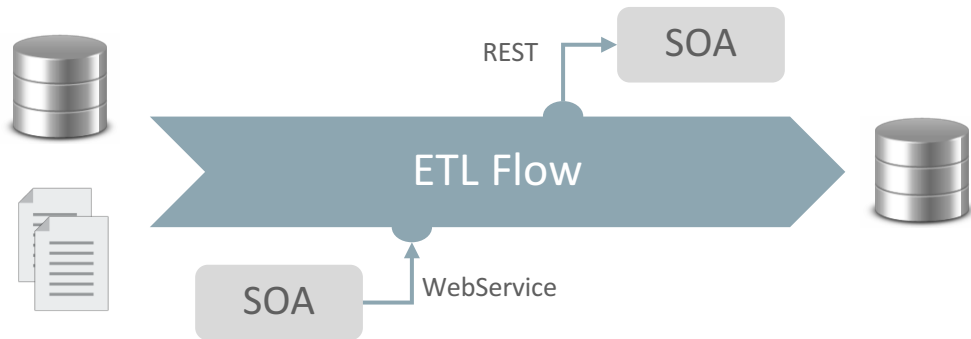
자동화, 관리 - 성능 - IT 환경 변화 - 실시간 분석

# ETL 관점의 아키텍처 패턴

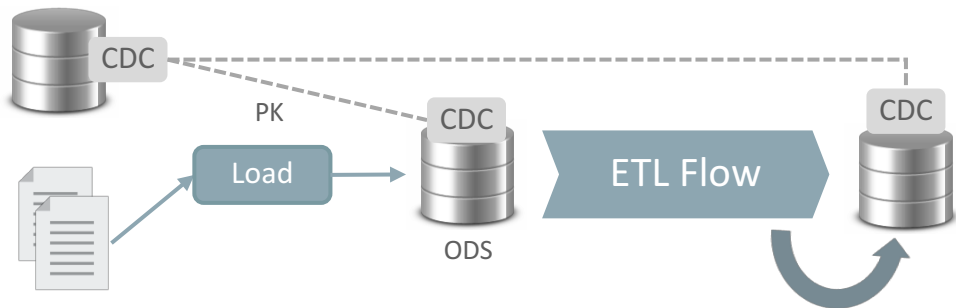
전통적 ETL (ETL Only)



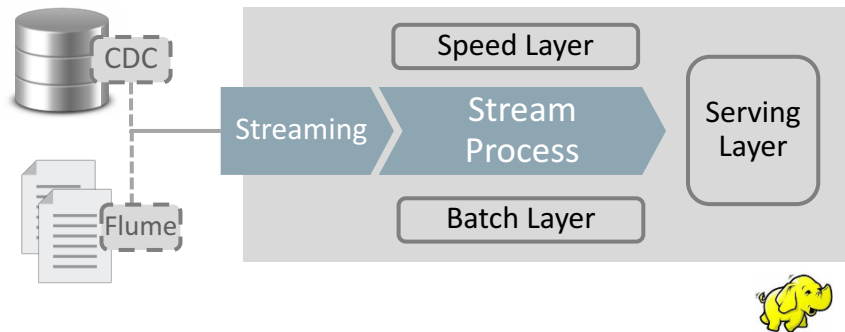
Application 연동



Real-Time Micro Batch (with CDC)



Streaming (with Bigdata)



# Change Data Capture

데이터베이스에서 변경 데이터 캡처 (CDC)는 변경된 데이터를 사용하여 작업을 수행할 수 있도록 변경된 데이터를 결정 (및 추적)하는 데 사용되는 일련의 소프트웨어 디자인 패턴입니다. CDC는 또한 엔터프라이즈 데이터 소스에 대한 **변경 사항의 식별, 캡처 및 전달을 기반으로 하는 데이터 통합 방식**입니다.

CDC 솔루션은 데이터웨어 하우스 환경에서 가장 자주 발생합니다. 시간 경과에 따른 데이터 상태를 캡처 및 보존하는 것이 데이터웨어 하우스의 핵심 기능 중 하나이기 때문에 모든 데이터베이스 또는 데이터 저장소 시스템에서 CDC를 활용할 수 있습니다. (by google 번역) - **Wikipedia**

CDC (Change data capture)

데이터복제 솔루션 시장에는

**퀘스트소프트의 웨어플렉스(SharePlex)**

**IBM의 데이터미러(Data Mirror)**

**골드게이트의 ER**

**티맥스소프트의 '프로싱크'**

가 경쟁하고 있지만 시장 점유율 면에서는 퀘스트의 웨어플렉스가 앞도적이다. 프로싱크는 운영 서버와 원격지에 있는 복제 서버(백업 서버) 또는 서로 다른 운영 서버간에 데이터의 변경 내용을 동기화 및 비동기 방식으로 목표 시스템으로 전송하거나 규칙에 맞게 데이터를 변환하는 등 다양한 기능을 수행하는 데이터 복제 솔루션이다.

**CDC(Change Data Capture) : 변경데이터 캡처**

2009.12 - <https://cafe.naver.com/iloveuljin/346>

2012.09 - <https://cafe.naver.com/sdk800402/160548>

데이터 소스(data source)로부터 변경 데이터(data)를 캡처(Capture) 하여 타겟 시스템(target system)에 전송 (transfer)하는 CDC(change data capture)를 구현하며 Redundant Data(기존의 타겟 데이터)는 전송되지 않으므로 데이터 관리의 효율성이 증대되며, 과부하 걸린 운영시스템의 로드(load) 감소하므로 전체적인 작업 생산성을 향상시킬 수 있다.

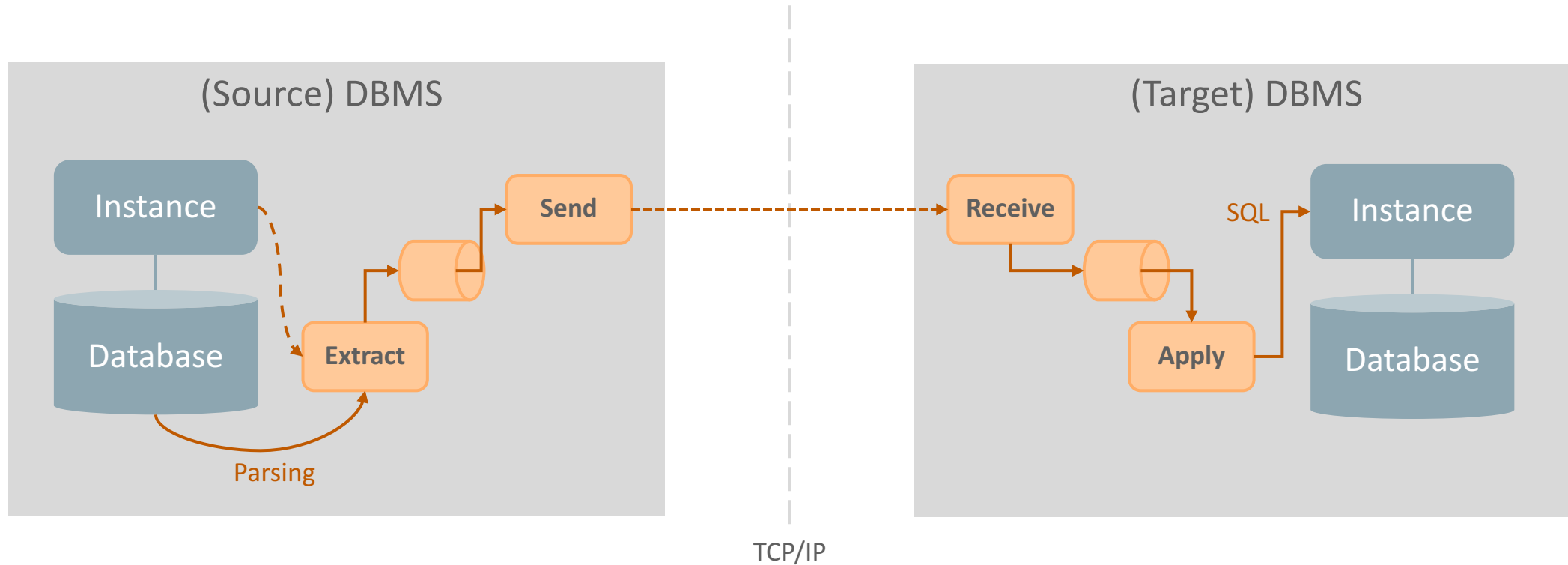
CDC(Change Data Capture)

2009.08 - 기술사 카페 <https://cafe.naver.com/81th/2875>

5. [DB]대용량의 데이터를 복제할 때 부하를 최소화 하기 위하여 CDC(Change Data Capture)의 방법이 활용된다.  
가. CDC의 개념과 특징을 설명하시오.  
나. CDC의 동작원리를 설명하시오.  
다. CDC의 활용방안을 제시하시오.

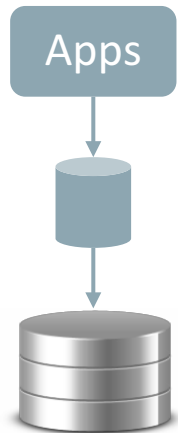
# CDC 아키텍처

1990's – Database HA/DR  
2000's – OLTP Replication  
2010 – Data Warehouse Appliances  
2015 – Data Lake

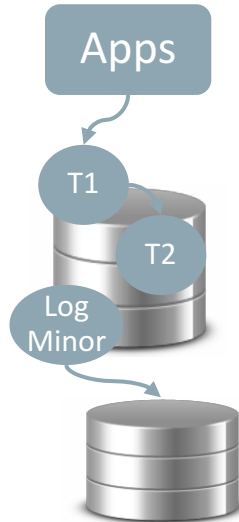




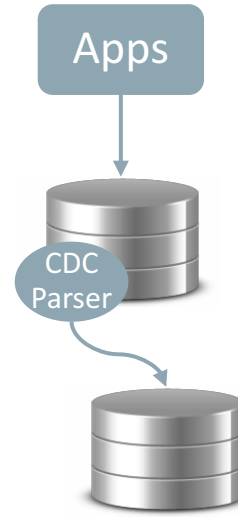
# Change Data Capture 변천사



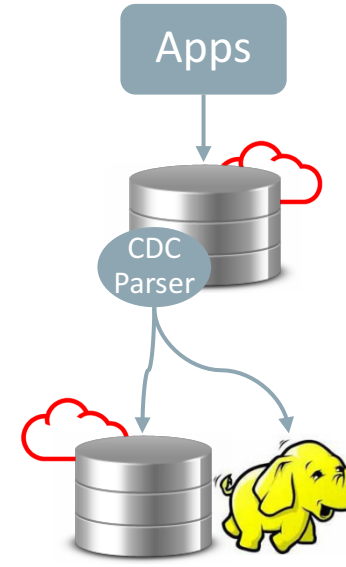
**Message Queue**  
- EAI



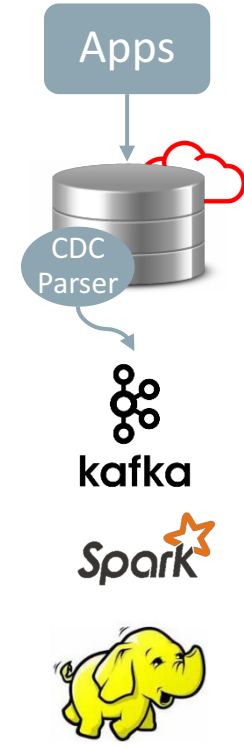
**Database**  
- Trigger  
- Log Minor



**Log Parsing**



**Multi Source/Target**  
- Bigdata  
- Cloud

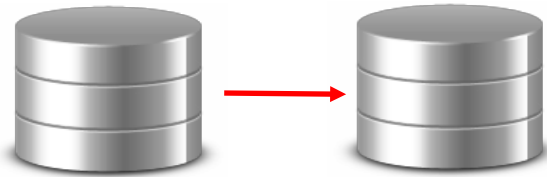


**Streaming**  
- Pipeline

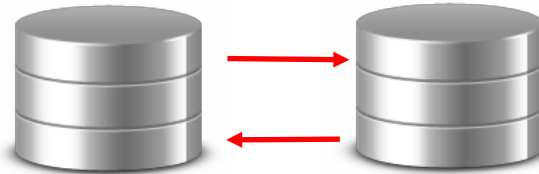
데이터연동 - DB 기능 활용 - 솔루션 활용 - IT환경변화 - 실시간 분석

# CDC 적용 패턴

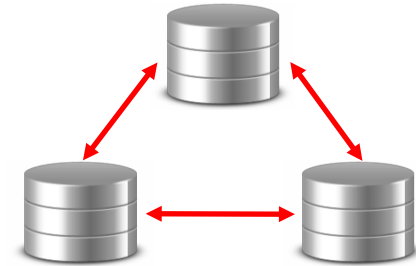
Uni-Directional



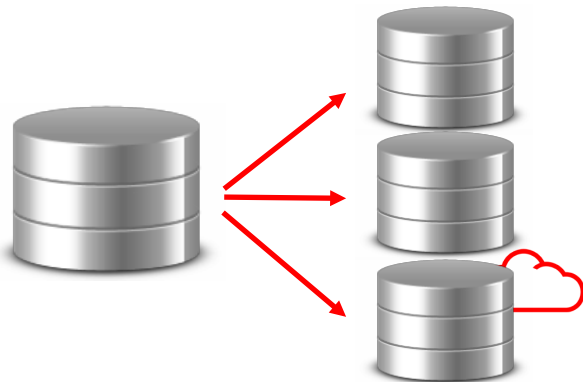
Bi-Directional



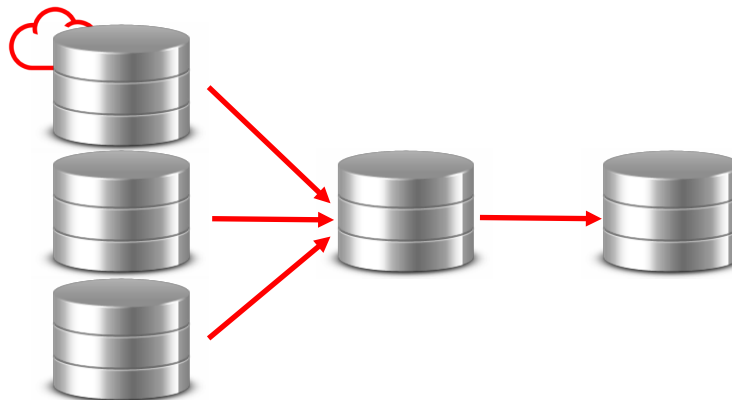
Peer-to-Peer



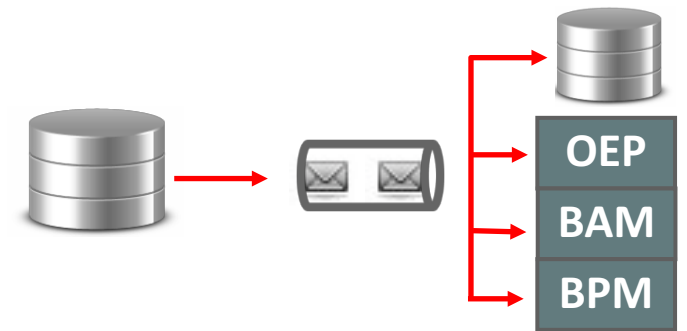
Broadcast



Consolidation

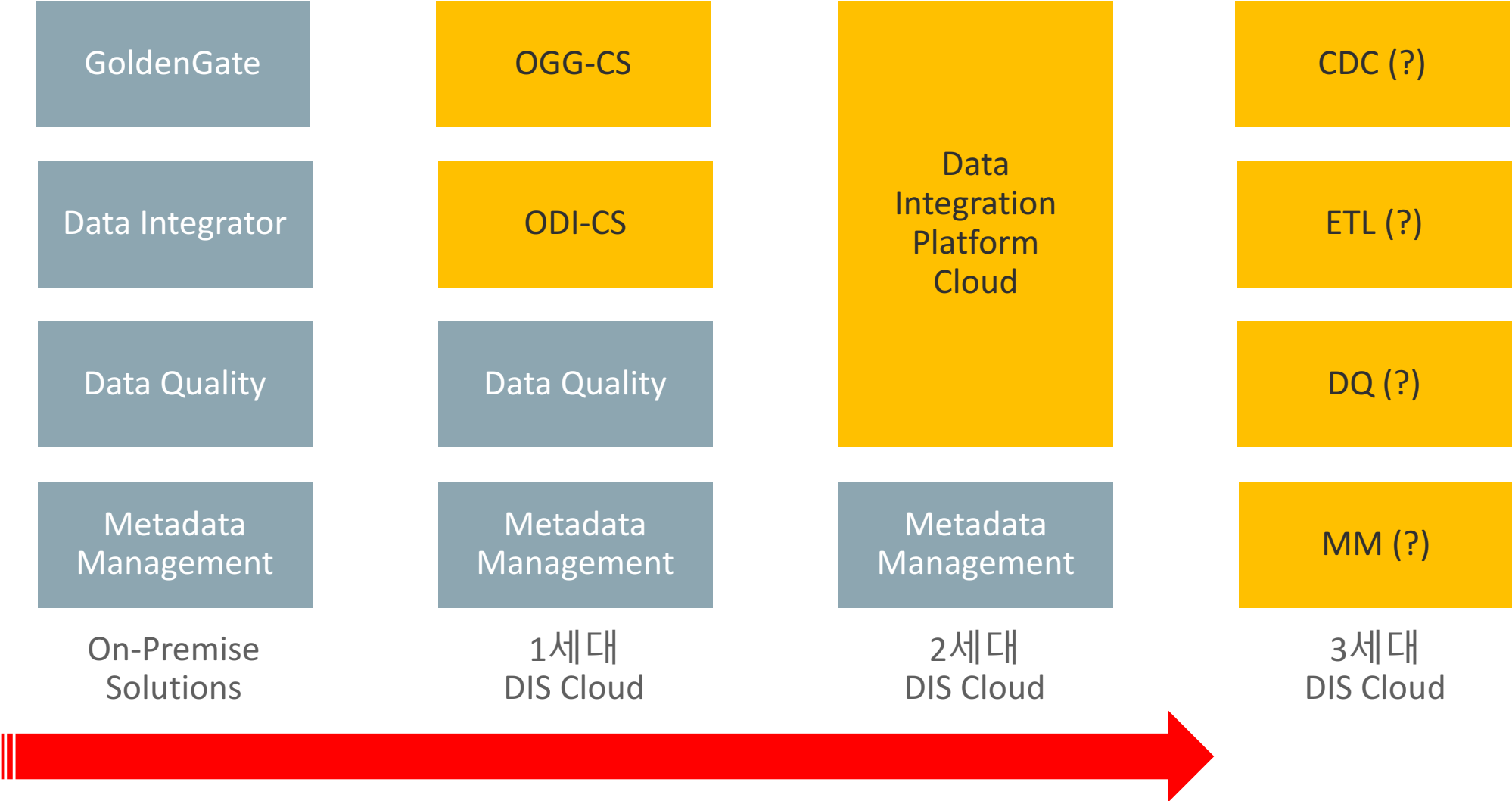


Data Distribution

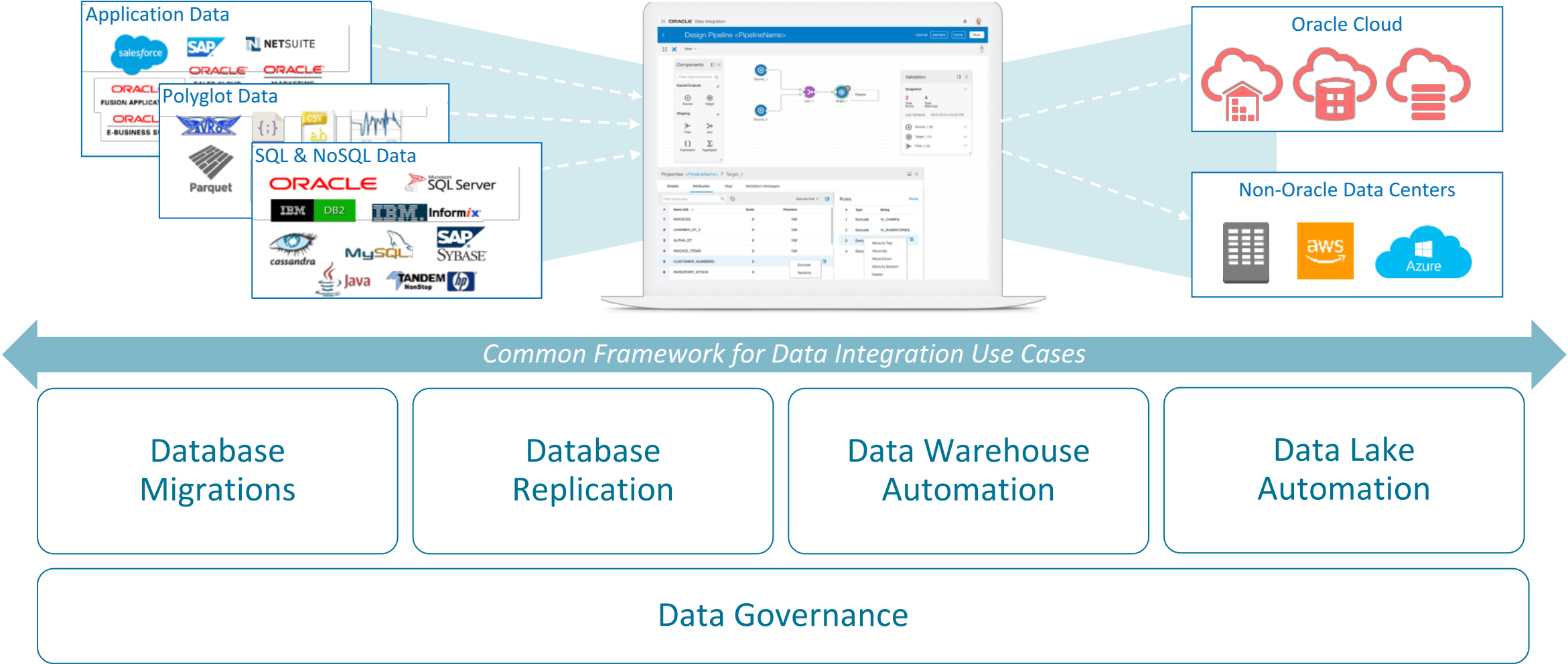


# Data Management using Oracle Cloud

# Oracle Cloud DIS Solutions



# Oracle Data Integration Platform



# Oracle Data Integration Platform Cloud

## Oracle Data Integration Platform Cloud **Standard** (ETL)

### Core Use Cases:

- 손쉬운 DW/DM 구축
- 자동화된 운영
- 손쉬운 데이터 마이그레이션

### Core Capabilities:

- 단순한 사용 환경
- 대용량 데이터 처리
- 변환
- Push-down 데이터 처리
- 기본 데이터 프로파일링

## Oracle Data Integration Platform Cloud **Enterprise** (ETL+CDC)

### Core Use Cases:

#### Standard Edition plus:

- Big Data 연계
- 데이터 동기화
- 무중단 마이그레이션
- Active DR 구축
- 실시간 DW

### Core Capabilities:

#### Standard Edition plus:

- Streaming, Big Data ETL
- 실시간 데이터 연계

## Oracle Data Integration Platform Cloud **Governance** (ETL+CDC+DQ)

### Core Use Cases:

#### Enterprise Edition plus:

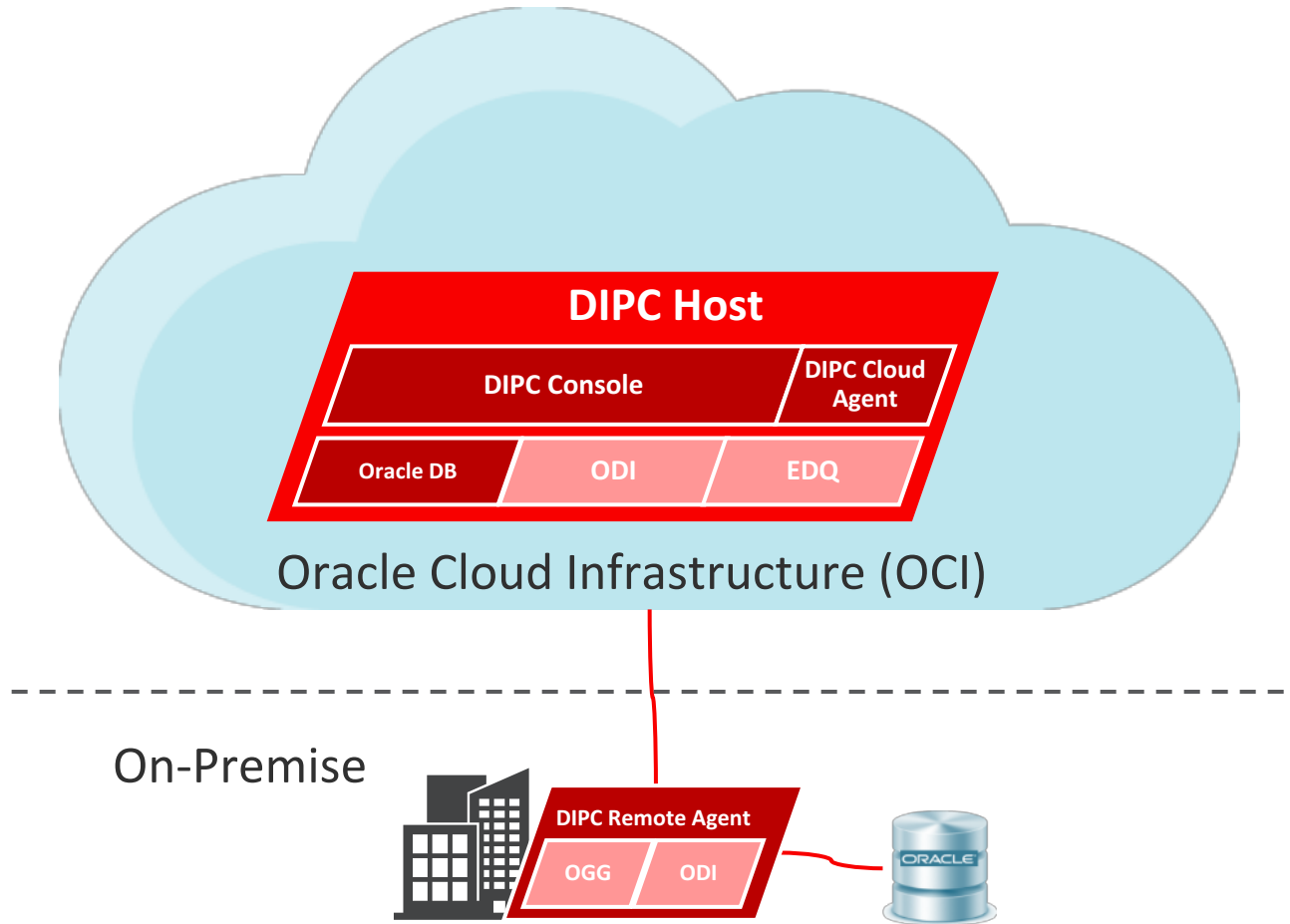
- 데이터 품질 관리
- 데이터 정제
- 데이터 계보

### Core Capabilities:

#### Enterprise Edition plus:

- 데이터 프로파일링 & 검증
- Match & Merge
- 용어사전 & 계보

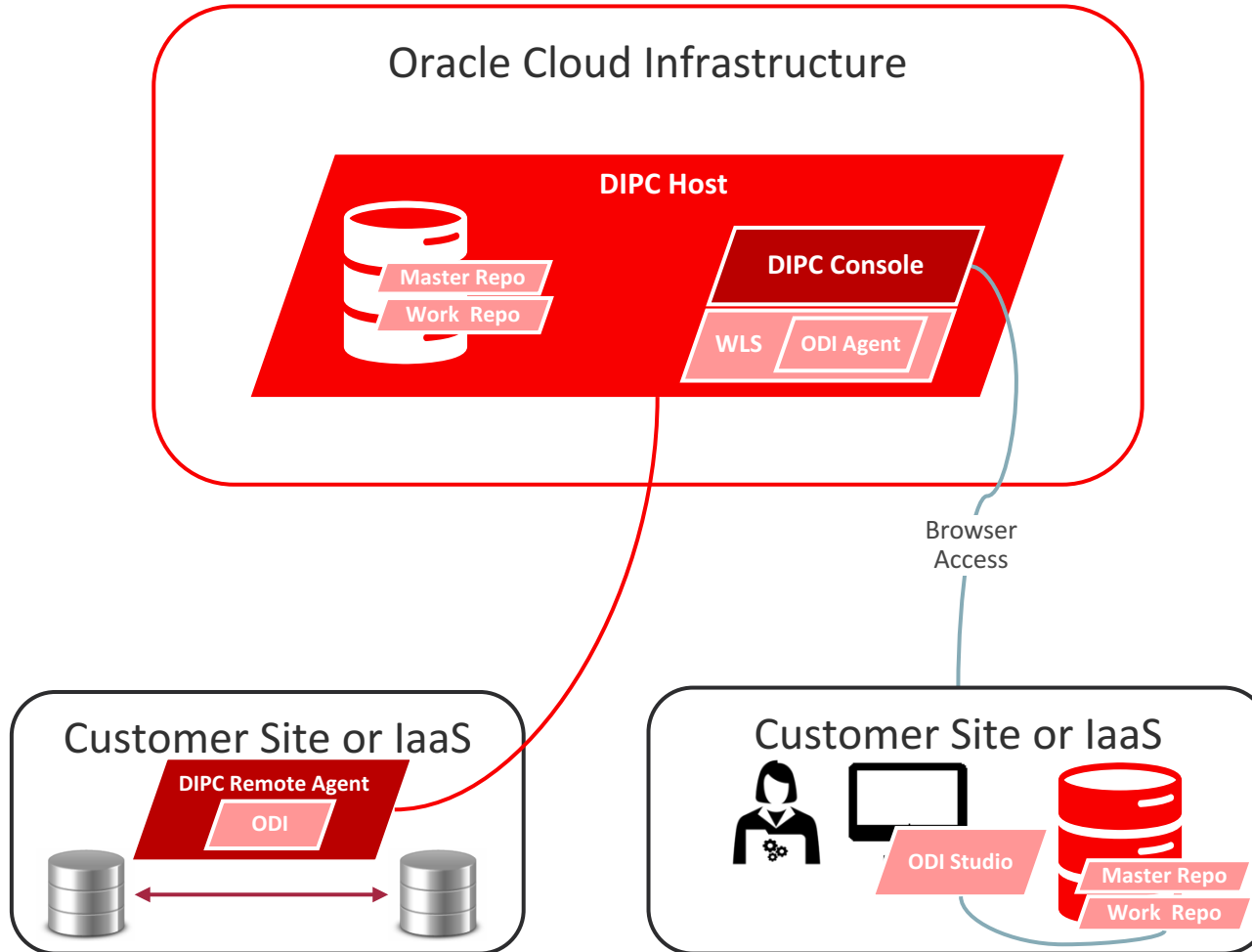
# DIPC Architecture



- **DIPC Components**
  - DIPC Console (web app)
  - DIPC Agent (local Agent running in DIPC instance)
  - DIPC Remote Agent for remote data access
- DIPC metadata stored in DB included on DIPC host along with Oracle Data Integrator (ODI) and Enterprise Data Quality (EDQ) repositories

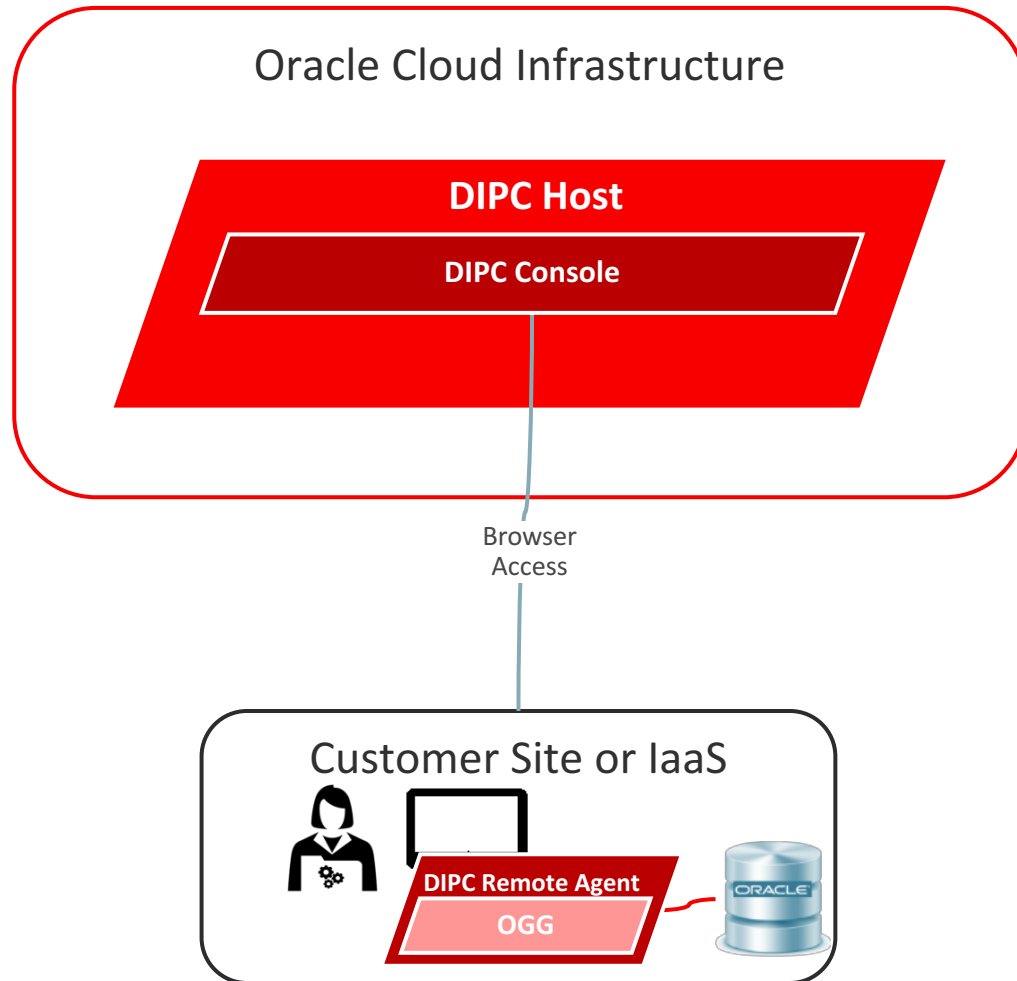


# DIPC Architecture – Standard (ODI)



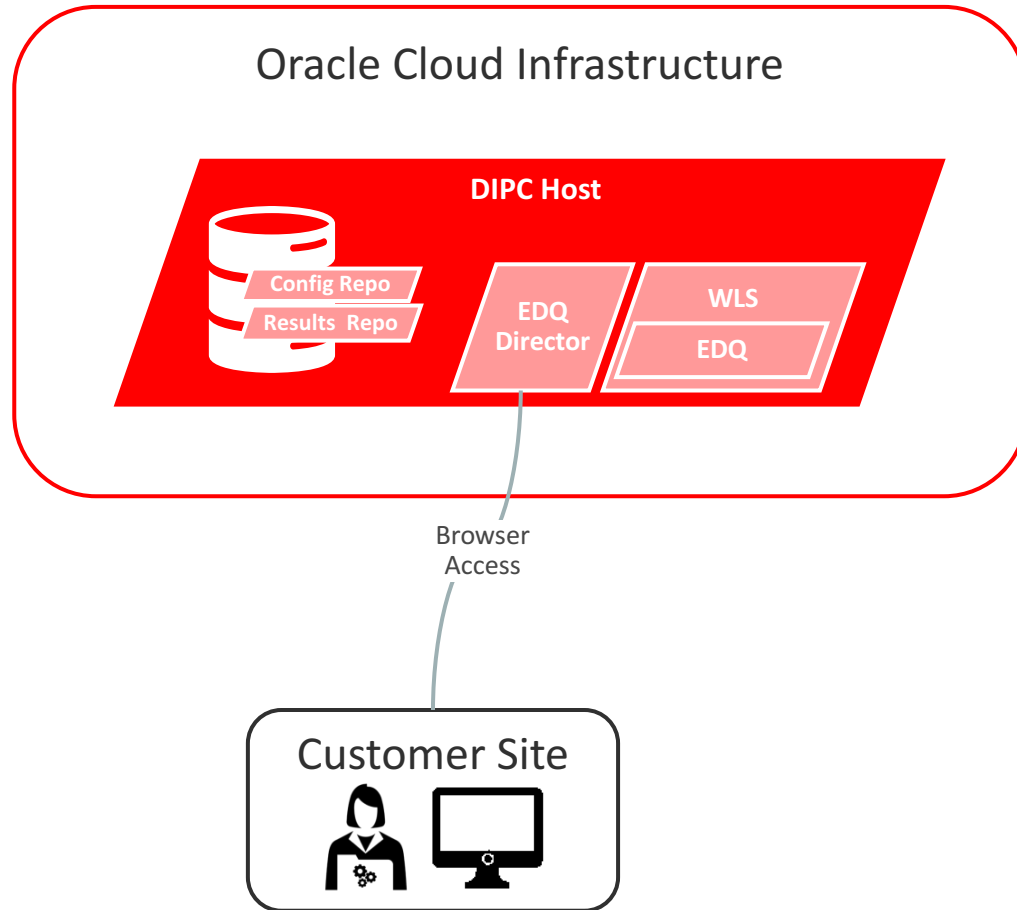
- 클라우드 상의 웹로직에 ODI EE Agent 제공
- 사용자를 위해 프로비저닝 단계에서 구성하여 제공
- 사용되는 레파지토리는 DIPC 호스트에 포함된 Oracle DB 사용
- DIPC 호스트에는 ODI Studio는 제공하지 않음
  - ODI 업무는 고객사 환경에서 개발
  - 개발된 업무를 DIPC 호스트에 배포후 운영
- Cloud Agent나 Remote Agent 에서 실행

# DIPC Architecture – Enterprise (+OGG)



- DIPC 호스트는 OGG가 설치되어 있지 않음
- DIPC Remote Agent에서 모든 실시간 복제 환경 구성
- On-Premise 연동시 데이터는 클라우드를 거치지 않음
- DIPC Remote Agent 는 다음을 포함 (선택)
  - OGG for Oracle 11g
  - OGG for Oracle 12c
  - OGG for Big Data

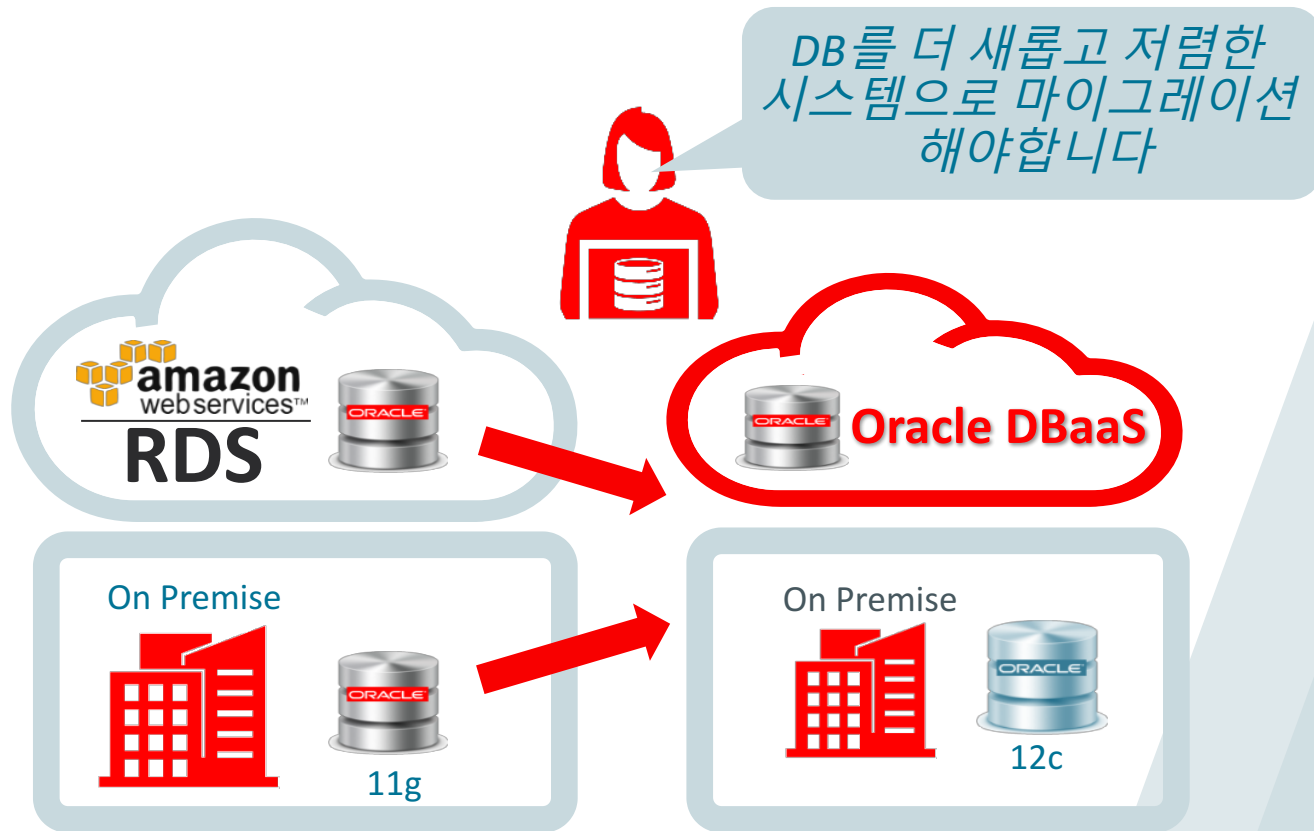
# DIPC Architecture – Governance (+OEDQ)



- EDQ 전체 설치
- 사용자를 위해 프로비저닝 단계에서 구성하여 제공
- 웹로직 서버에 EDQ 구성됨
- 사용되는 레파지토리는 DIPC 호스트에 포함된 Oracle DB 사용
- EDQ 디렉터(개발툴) 은 browser를 통해 원격 접근

# Solution Use case 1:

## Database Migrations (Cloud & On Premise)



※ OP-OP 데이터 복제시 데이터는 외부(Cloud)로 전송안됨 (소스 → 타겟)

### Database Migrations

- 이전 데이터베이스 또는 불완전하게 최적화 된 클라우드 서비스에서 원활하게 데이터를 마이그레이션 할 수 있습니다.
- Cloud에서 최신 Oracle DB 버전을 활용하십시오.

### Key Benefits – Online/Offline Migrations

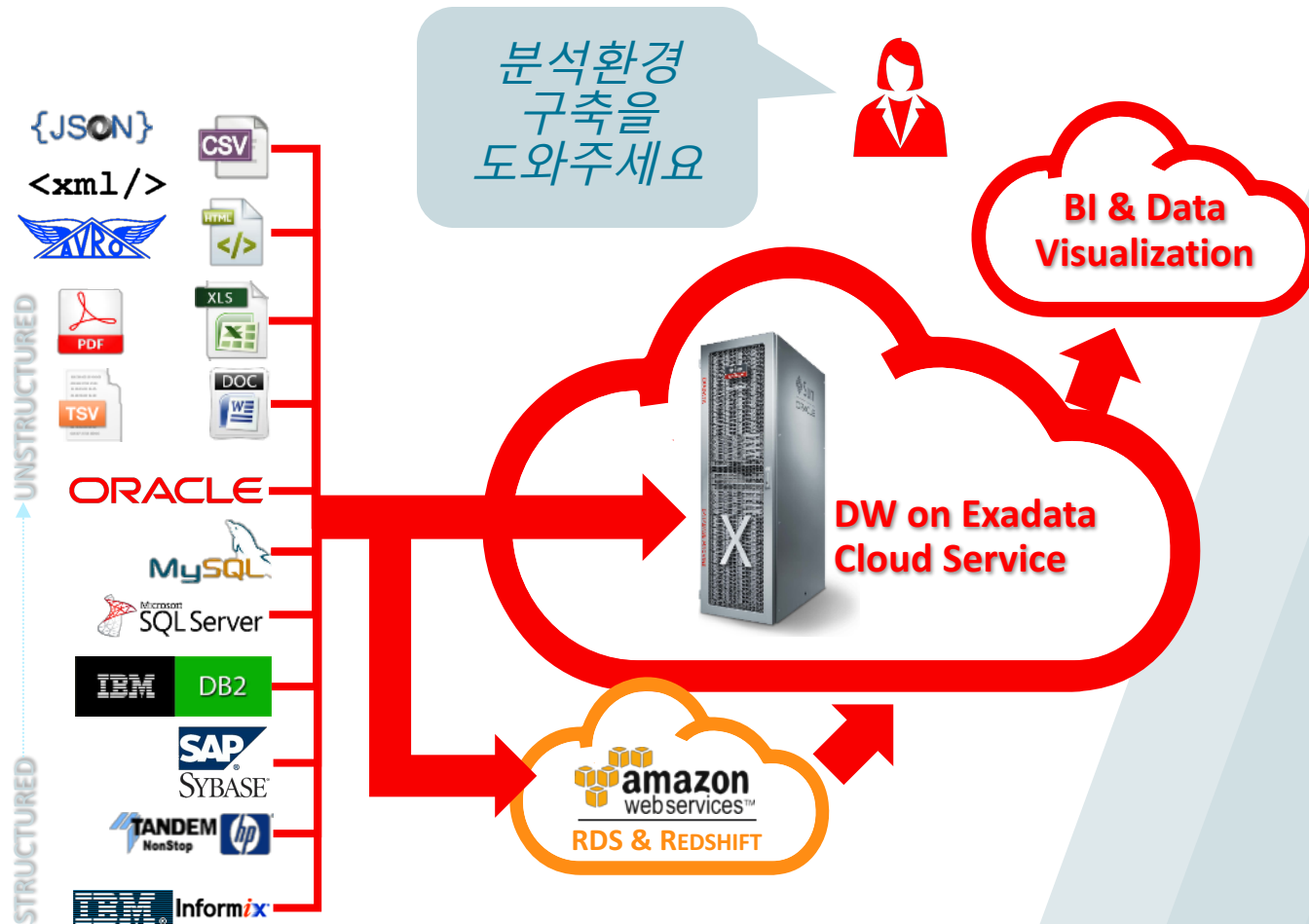
- 단계적 / 온라인 마이그레이션 데이터 손실없음
- 프로비저닝에 대한 강력한 자동화 적용
- 신뢰할 수 있고 검증 된 핵심 GoldenGate 기술.

### 3<sup>rd</sup> Party Cloud Support

- Amazon RDS 인스턴스를 Oracle로 마이그레이션
- 온 프레미스에서 타사 클라우드로의 데이터 전달

# Solution Use case 2:

## Database Warehouse Integration (OAC, ADWC)



### Data Warehouse Integration

- 클라우드의 운영 또는 고급 분석에는 데이터의 신뢰가 보장되어야 합니다.
- 온 프레미스 데이터 원본 또는 다른 클라우드 원본에서 오는 데이터로 작업하십시오.

### Key Capabilities: Comprehensive Solution

- 실시간으로 데이터 복제: DW / Marts에 대한 복제 및 스트리밍 처리 지원 기능이 내장 됩니다.
- 데이터 변환 및 정리: 혁신적인 접근 방식으로 워크로드를 다른 PaaS 리소스에서 실행할 수 있습니다.

### 3<sup>rd</sup> Party Cloud Support

- Amazon RDS 데이터베이스 및 Redshift 지원
- Kinesis 및 EMR을 통한 대용량 데이터 지원

# Solution Use case 3: Dev & Test Systems



내 DBA 작업을보다  
쉽고 생산적으로  
만들고 싶어요.

On Premise



10g & 11g



Oracle DBaaS  
12.1 & 12.2

## Database Dev-Test Environments

- 최소한의 인프라 또는 프로세스 오버 헤드로 지속적으로 개발 및 테스트 환경을 동기화 해야 합니다.
- 수동 동기화가 아닌 자동화 활용.

## Key Benefits: Automate Data Movement

- 낮은 오버 헤드로 인해 수동 및 복합 스크립팅 감소
- 반복 가능한 접근 방식으로 DBA 프로세스 표준화
- 간단한 프레임 워크는 다양한 DB에서 사용할 수 있습니다.

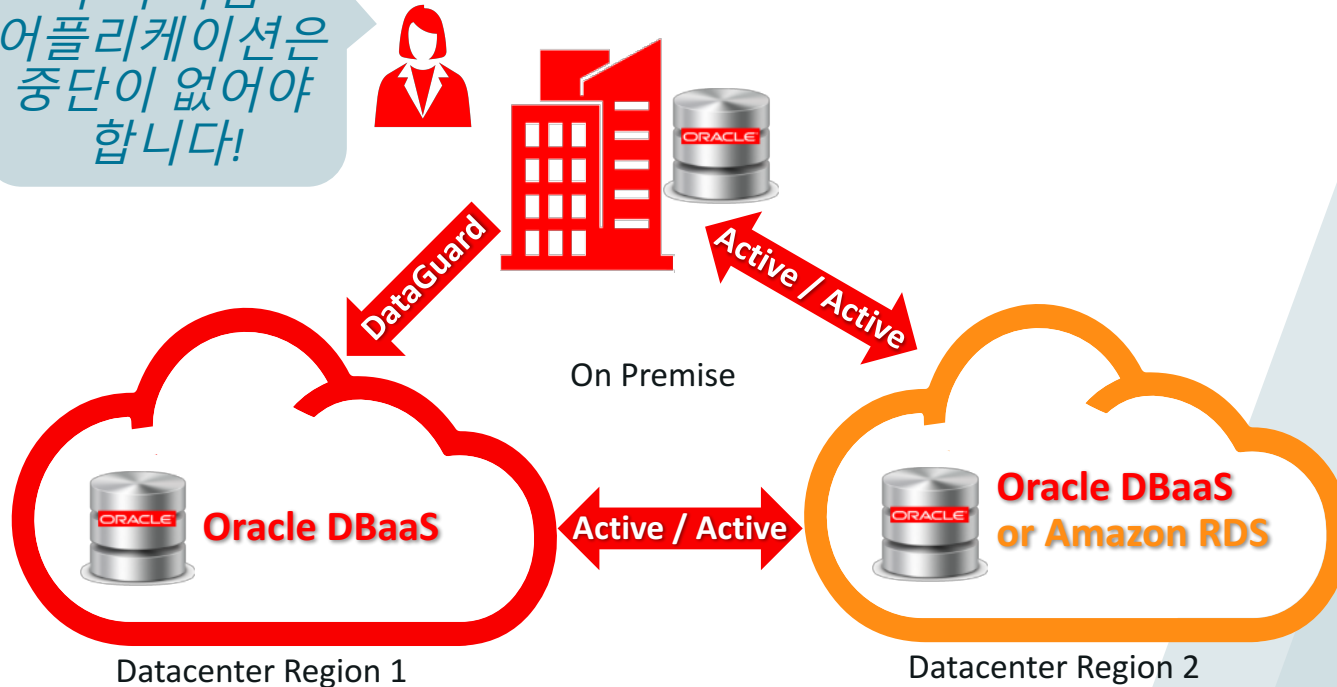
## Oracle Database as a Service

- 온 프레미스에서 실행중인 항목과 상관없이 지원되는 모든 Oracle 버전
- 믹스 앤 매치 버전 또는 패치 레벨

# Solution Use case 4:

## High Availability for DBaaS in the Cloud

우리 기업  
어플리케이션은  
중단이 없어야  
합니다!



### Regional High Availability

- 중요한 업무용 데이터베이스를 사용할 여유가 없을 때 다운 타임이 발생합니다.
- 지역 데이터 센터에서 재난 복구가 필요한 경우 특히 필요합니다.

### Key Benefits for DBaaS

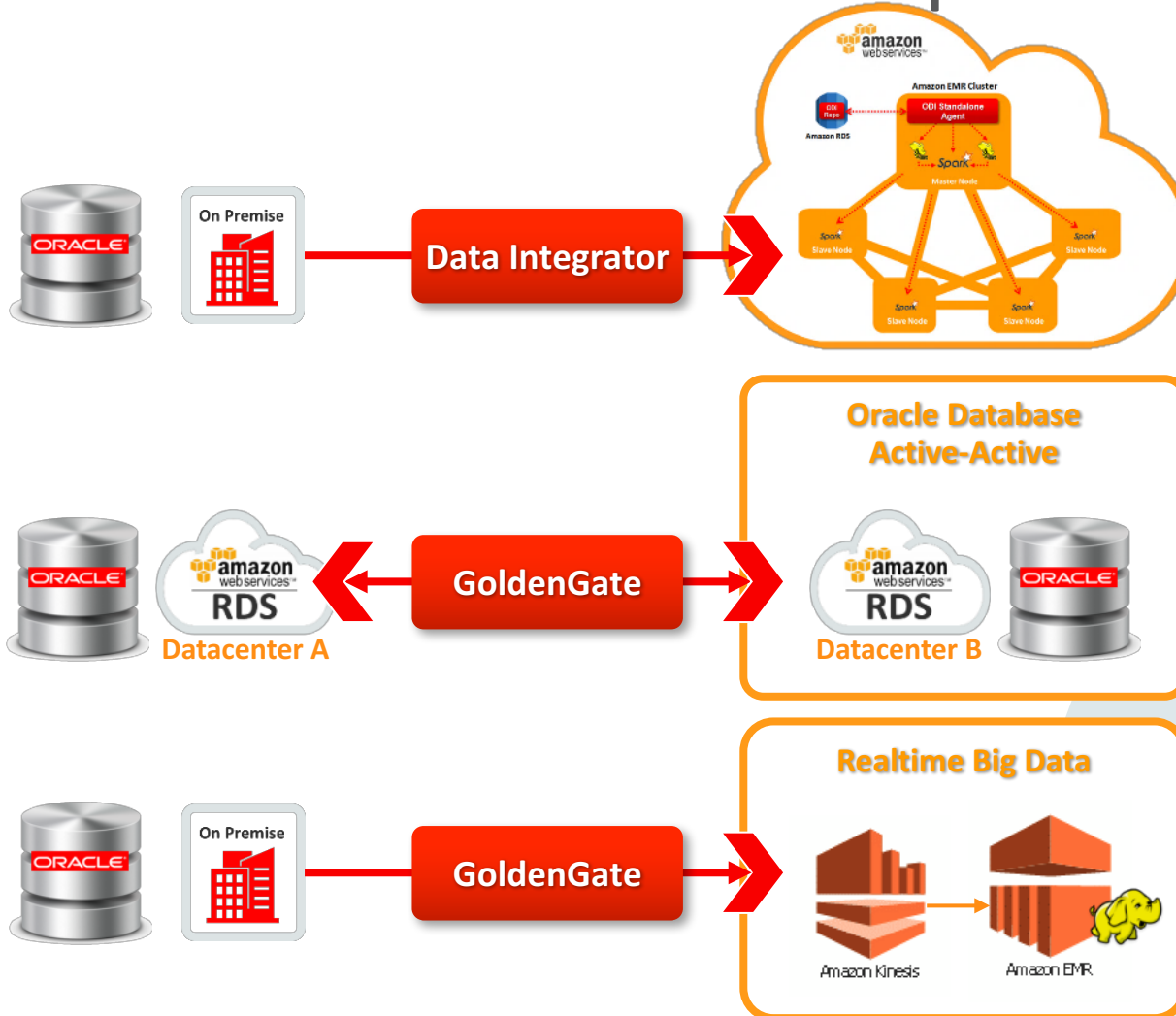
- 클라우드의 Active Data Guard를 사용하여 대기 데이터베이스를 항상 가동 할 수 있습니다.
- Active-Active를 사용하면 모든 데이터베이스에서 읽기 / 쓰기 트랜잭션을 수행 할 수 있습니다.

### Powerful Cloud Solution

- 강력한 서비스 프로비저닝 및 청구 자동화.
- Oracle 또는 Amazon Cloud는 지역 데이터 센터의 모든 클라우드에서 활성 - 활성 Oracle Database를 운영합니다.



# Solution Use case 5: Make **Amazon AWS** Enterprise Grade



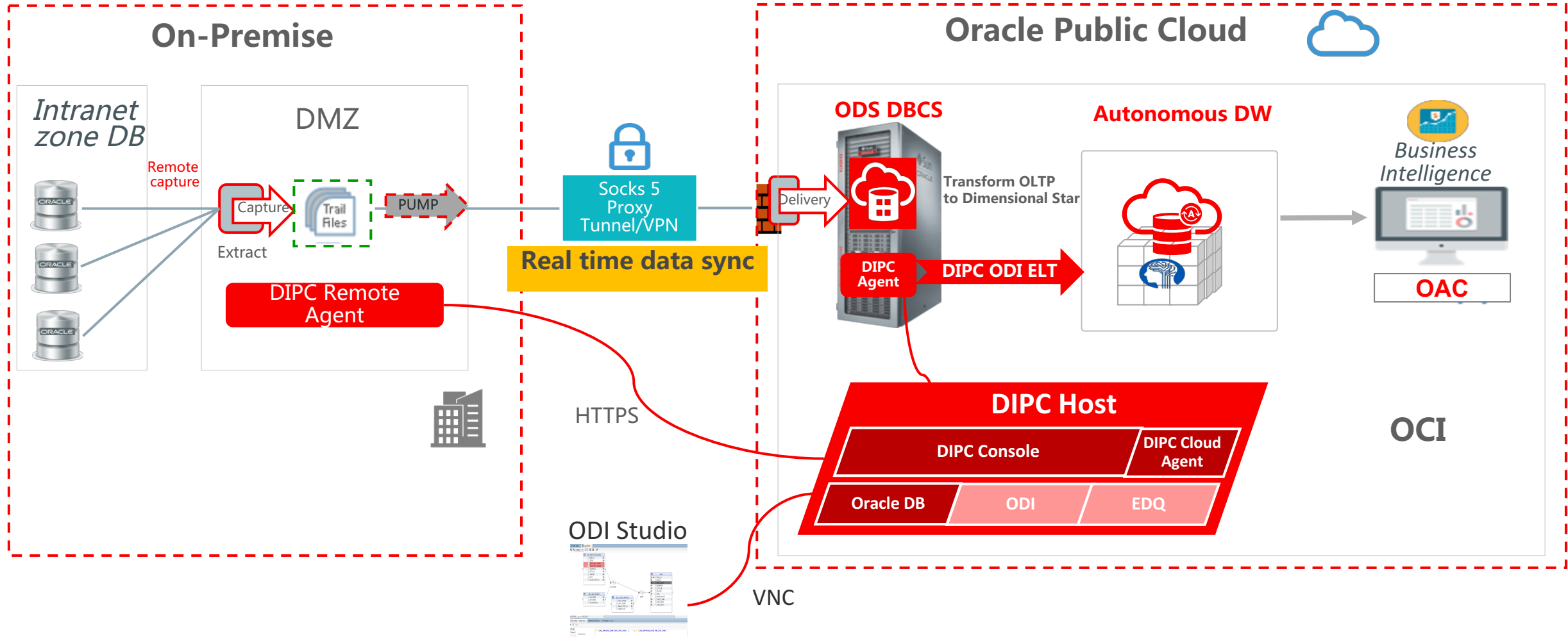
## Bring Enterprise-strength to AWS

- Kinesis, EMR, Redshift 및 RDS를위한 고급 데이터 통합 및 가용성 기능
- 초당 수천 건의 DB 트랜잭션의 입증 된 데이터 전송률
- 보안 컨트롤은 Amazon의 보안 데이터 센터에서 올바르게 작동
- 유연한 솔루션은 AWS 서비스 내에서 DB 또는 빅 데이터를 활용
- 가장 신뢰할 수 있는 데이터 복제 엔진
- 전세계 시장에서 가장 인기있는 ETL / ELT 툴
- 대용량 데이터에 대해 가장 빠르게 증가하는 데이터 통합

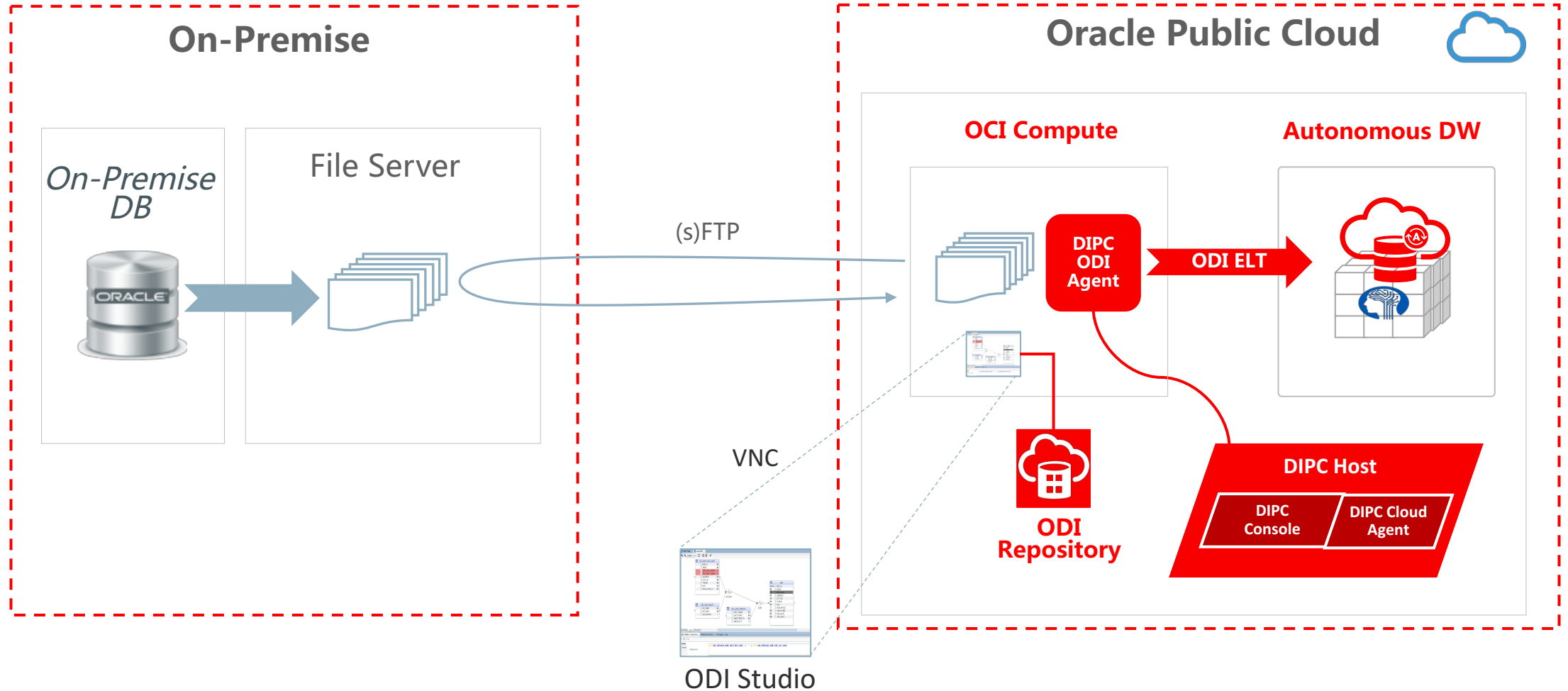
## Heterogeneous and Complete Solution

- Oracle Data Integration은 서로 다른 데이터 플랫폼 및 서로 다른 클라우드 플랫폼에서 작동하는 이기종 기능입니다.

# 아키텍처 예시 #1 - DIPC with ADW



# 아키텍처 예시 #2 - DIPC for On-Premise File



# **Hardware and Software** **Engineered to Work Together**

ORACLE®