

# オラクルが構築を急ぐ、 データレイク/データサイエンスソリューションの 全貌とは

日本オラクル株式会社  
クラウド事業戦略統括  
クラウドソリューション推進本部  
早川 博

**Modern Cloud Day Tokyo**

次世代クラウドが変える日本のビジネス

以下の事項は、弊社の一般的な製品の方向性に関する概要を説明するものです。また、情報提供を唯一の目的とするものであり、いかなる契約にも組み込むことはできません。

以下の事項は、マテリアルやコード、機能を提供することをコミットメント（確約）するものではないため、購買決定を行う際の判断材料になさらないで下さい。オラクル製品に関して記載されている機能の開発、リリースおよび時期については、弊社の裁量により決定されます。

OracleとJavaは、Oracle Corporation 及びその子会社、関連会社の米国及びその他の国における登録商標です。文中の社名、商品名等は各社の商標または登録商標である場合があります。

# 自己紹介



 @hhiroshell



## 早川 博（はやかわ ひろし）

- Cloud Native / Data Lake & Data Science
- エンジニアコミュニティ 「Cloud Native Developers JP」 オーガナイザー
- Developers Summit 2018、Japan Container Days 12.18、CloudNative Days Tokyo 2019登壇
- 「Kubernetes実践ガイド - クラウドネイティブアプリケーションを支える技術」 著者（共著）

# 目次

- イントロダクション
- データレイク/データサイエンステクノロジーの理想像
- オラクルが提供するデータレイク/データサイエンスソリューション

# イントロダクション

# 企業におけるデータ活用のトレンド – データドリブン経営

**データに基づく客観的な意思決定が、企業経営の鍵として認識されるように**

- データドリブン経営

“ 企業経営における意思決定を、「データ」と「アルゴリズム」に基づいた客観的な分析結果に基づいて行う、経営の手法 ”

- 企業におけるデータの活用のための、さまざまな技術が発展
  - Hadoop, HDFS, Data Warehouse, Data Mart, Notebook, BI…



# Data Driven Transformations: Multi Billion Opportunity

**\$103B**

Market Revenue for Big Data  
by 2027

**4.5X**

The rate at which big data on cloud  
is outpacing on-premise

**79%**

of executives say that their companies will  
lose edge and face extinction without  
embracing big data



**FinTech**

Risk Analysis, High-Frequency Trading



**Retail**

Recommendation Engines, Customer Analysis



**Healthcare**

Drug Testing, Cancer Research

※ Oracle Modern Cloud Day Tokyo 2019 「Oracle Cloud Infrastructure データ管理 – Big Dataおよびデータベース」 / Usman Khan より

# データドリブン経営のその先へ

意思決定だけではない、新規ビジネス創出を見据えたデータ活用へ

## 企業の意思決定のためのデータ活用

企業経営の意思決定のためにデータとアルゴリズムを活用する

## 既存ビジネスの効率化、高付加価値化

既存ビジネスをデータによって改善し、より高いビジネス利益を得る

e.g.)  
スマートファクトリー、保険取引におけるリスク分析

## 新規ビジネスの創出

機械学習、AIなど、より高度なデータ活用によって、新たなビジネスを創出する

e.g.)  
AIチャットボット、企業人材分析コンサルティングサービス



# ユースケース

企業におけるデータ活用が、あらゆる分野・あらゆる部署で重要に



Government



Retail



Technology



Automotive



Consumer Goods



Media



Communications



Healthcare



Gaming



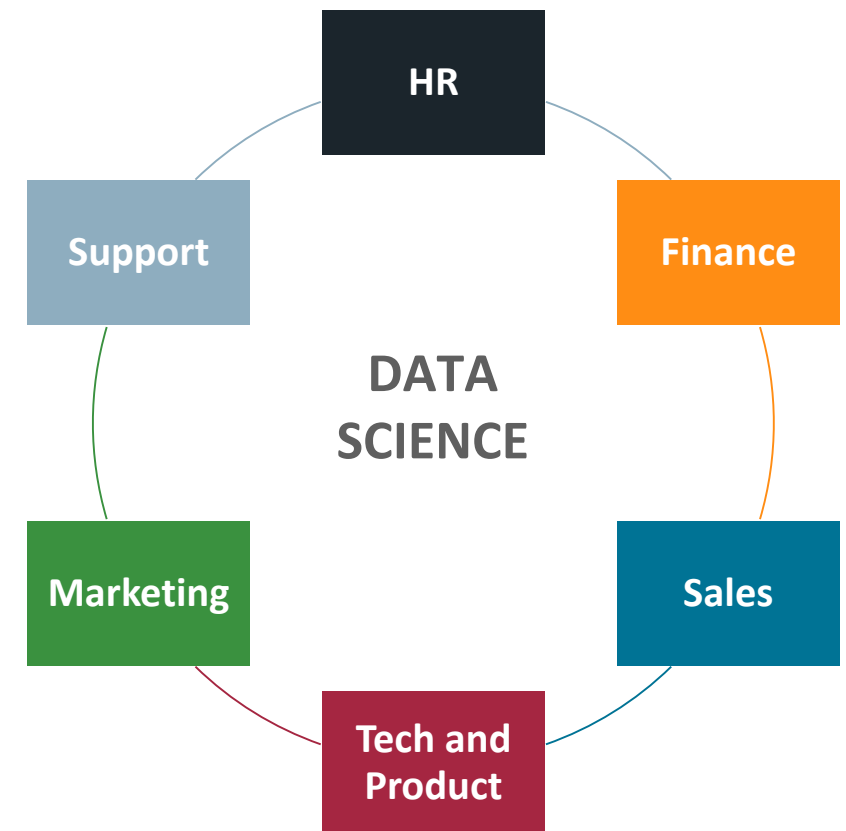
Travel



Energy



Finance & Insurance



# ユースケース

企業におけるデータ活用が、あらゆる分野・あらゆる部署で重要に



Marketing  
Response Models

Scheduled Jobs



Customer churn

APIs



Lifetime Value

Apps



Text sentiment  
analysis

Reports



Computer vision  
and image tagging

Apps



Forecasting

Reports



Risk management  
with machine  
learning

APIs



Recommendation  
engines

APIs



Transactional data  
ETL

Scheduled Jobs



Data discovery and  
auditing

Reports

# データによる新規ビジネス創出に求められるもの

## 機械学習、データサイエンスを導入するだけでは足りない

- 単発の分析だけでは、データ活用のための取り組みは成立しない
- データを継続的に収集・蓄積し、繰り返し学習し、アプリケーションに利用する仕組みが必要

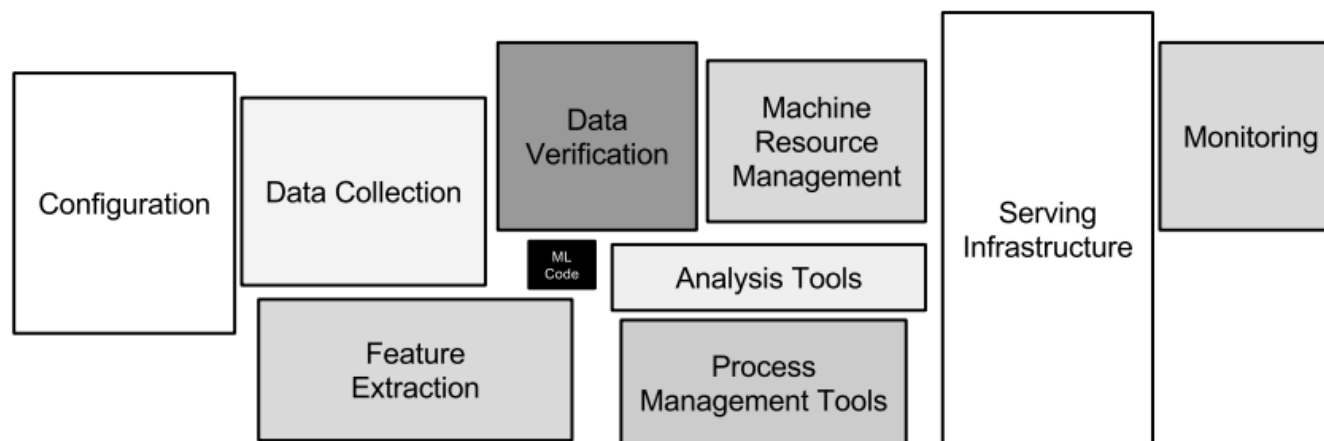


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

※ <https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>

# 企業におけるデータ活用に求められるもの

**蓄積・活用・維持 / 3つのフェーズを見据えたデータ活用の戦略が必要**

## ① 蓄積

- 企業が保有するあらゆる資産（データ）を、ビジネスインサイトの源泉としてできる限り活用する

## ② 分析

- あらゆるステークホルダーのデータ活用を支援し、ビジネスインサイト獲得の可能性を最大化する

## ③ 活用

- 分析結果を様々なアプリケーションで活用することで、ビジネス成果に結びつける

# データレイク/データサイエンス テクノロジーの理想像



# 【再掲】 企業におけるデータ活用に求められるもの

## 蓄積・活用・維持 / 3つのフェーズを見据えたデータ活用の戦略が必要

### ① 蓄積

- 企業が保有するあらゆる資産（データ）を、ビジネスインサイトの源泉としてできる限り活用する

### ② 分析

- あらゆるステークホルダーのデータ活用を支援し、ビジネスインサイト獲得の可能性を最大化する

### ③ 活用

- 分析結果を様々なアプリケーションで活用することで、ビジネス成果に結びつける

# データ活用を実現するテクノロジーの理想像

蓄積・活用・維持 / 3つのフェーズをテクノロジーでサポート

## ① 蓄積

- 企業が保有するあらゆる資産（データ）を、ビジネスインサイトの源泉としてできる限り活用する

## ② 分析

- あらゆるステークホルダーのデータ活用を支援し、ビジネスインサイト獲得の可能性を最大化する

## ③ 活用

- 分析結果を様々なアプリケーションで活用することで、ビジネス成果に結びつける



# データの蓄積におけるポイント

## あらゆるデータソースをスケーラブルな単一のデータレイクに統合する

- 企業活動で利用される様々なITシステムをデータの源泉とする
  - 業務システム(ERP, HCM…), 基幹システム、サードパーティーのデータ…
- 多様かつ大量のデータソースからのデータ収集し蓄積する
  - データソースへの対応・データ量への対応を、コスト性能比で考える



# データの分析におけるポイント 1/2

## 単一のデータレイクをあらゆるステークホルダーから利用可能にする

- 様々なツールからデータにアクセスできることで、あらゆるステークホルダーからデータレイクを利用可能にする

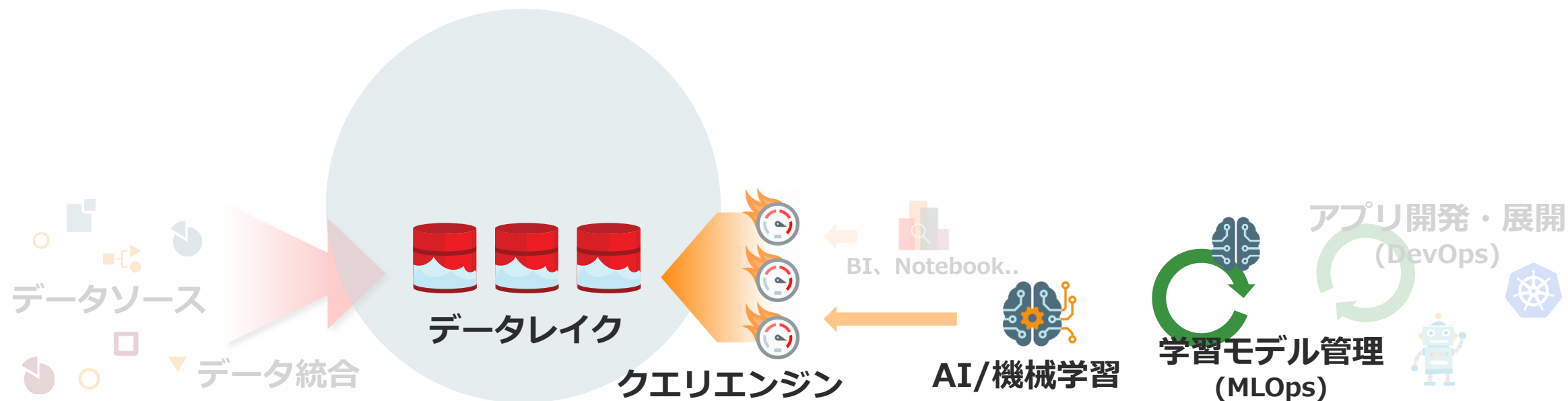
目的	ツール	ステークホルダー
対話的なデータの探索	Notebook	データエンジニア、データサイエンティスト
ビジネス的なデータの集計	BI, Analytics	LOB、アナリスト
機械学習、AI開発	機械学習ツール	データサイエンティスト、アプリケーション開発者



# データの分析におけるポイント 2/2

## 機械学習の繰り返し実行と、そこから生まれる学習モデルの管理をおこなう

- 機械学習は繰り返し実行し、常に最新のデータに基づいた成果を利用できるようにする（一度学習モデルを作成して終わりではない）
  - データセットと学習モデルのライフサイクル管理が重要

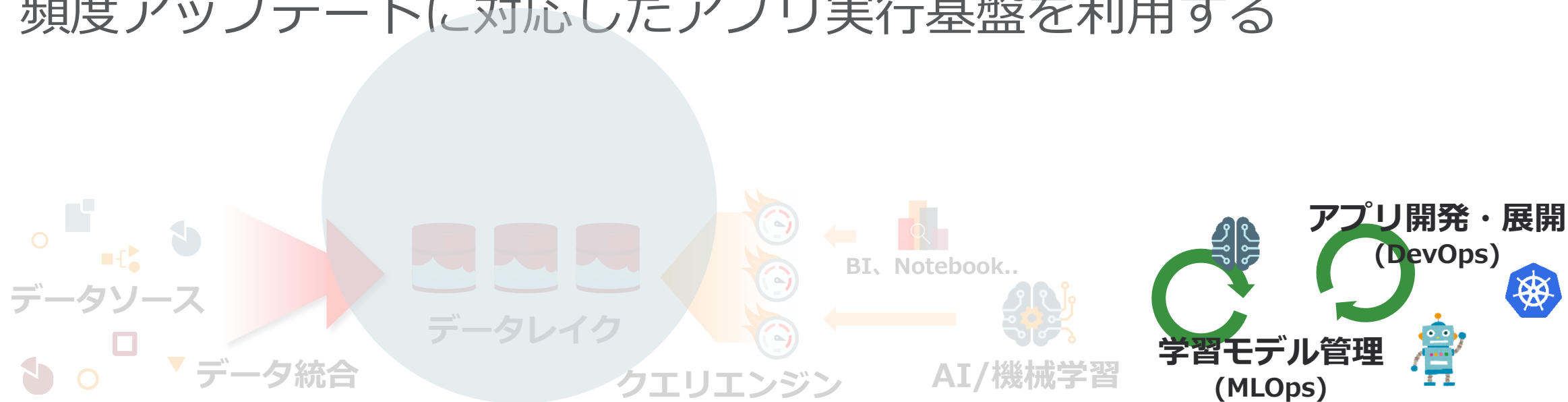




# データの活用におけるポイント

## 学習モデルをアプリケーションに活用し、ビジネス成果を最大化する

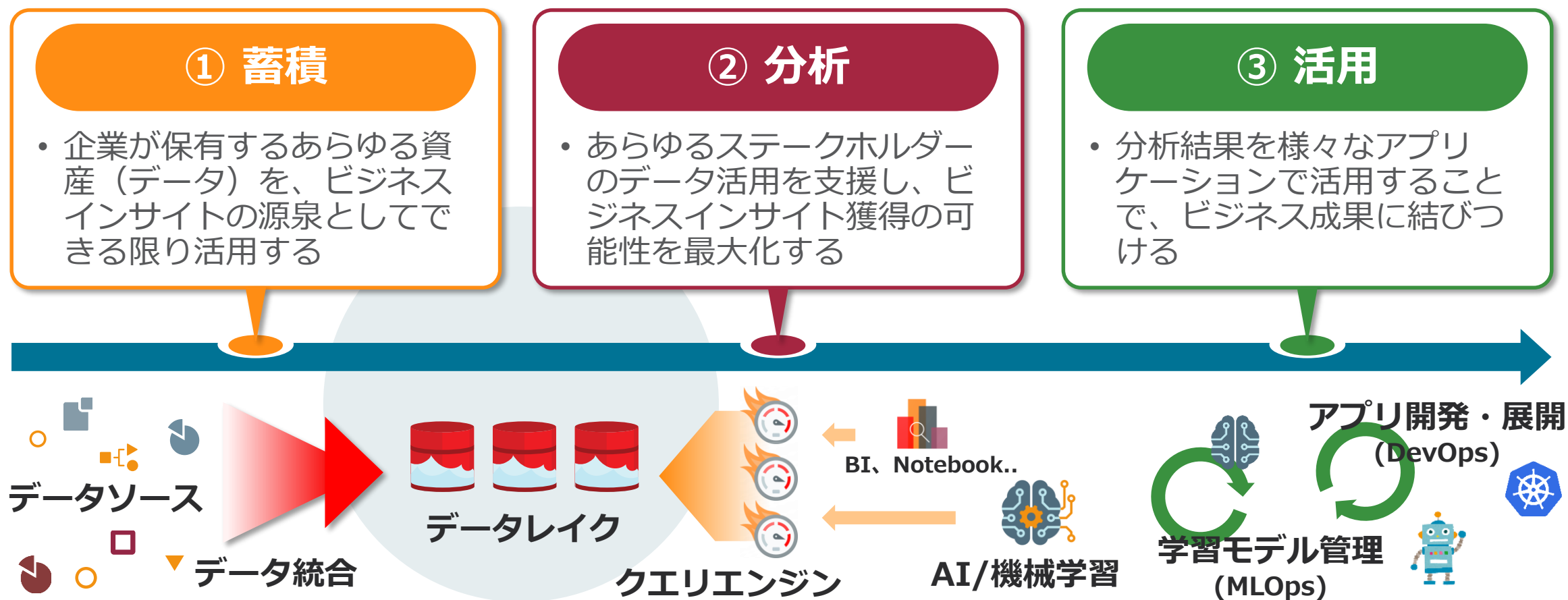
- 学習モデルをアプリケーションの機能に利用することで、より高い価値を得る
- 学習モデルの更新を繰り返しアプリケーションに取り入れるため、高頻度アップデートに対応したアプリ実行基盤を利用する



# オラクルが提供する データレイク/データサイエンス ソリューション

# 【再掲】 データ活用を実現するテクノロジーの理想像

## 蓄積・活用・維持 / 3つのフェーズをテクノロジーでサポート



# オラクルのデータ活用テクノロジーのコンセプト

- **1. Completed**

- データソースの整備から成果の刈り取り（アプリでの利用）までを包含
- データ活用を実現するテクノロジーの理想像を満たす

- **2. Enterprise**

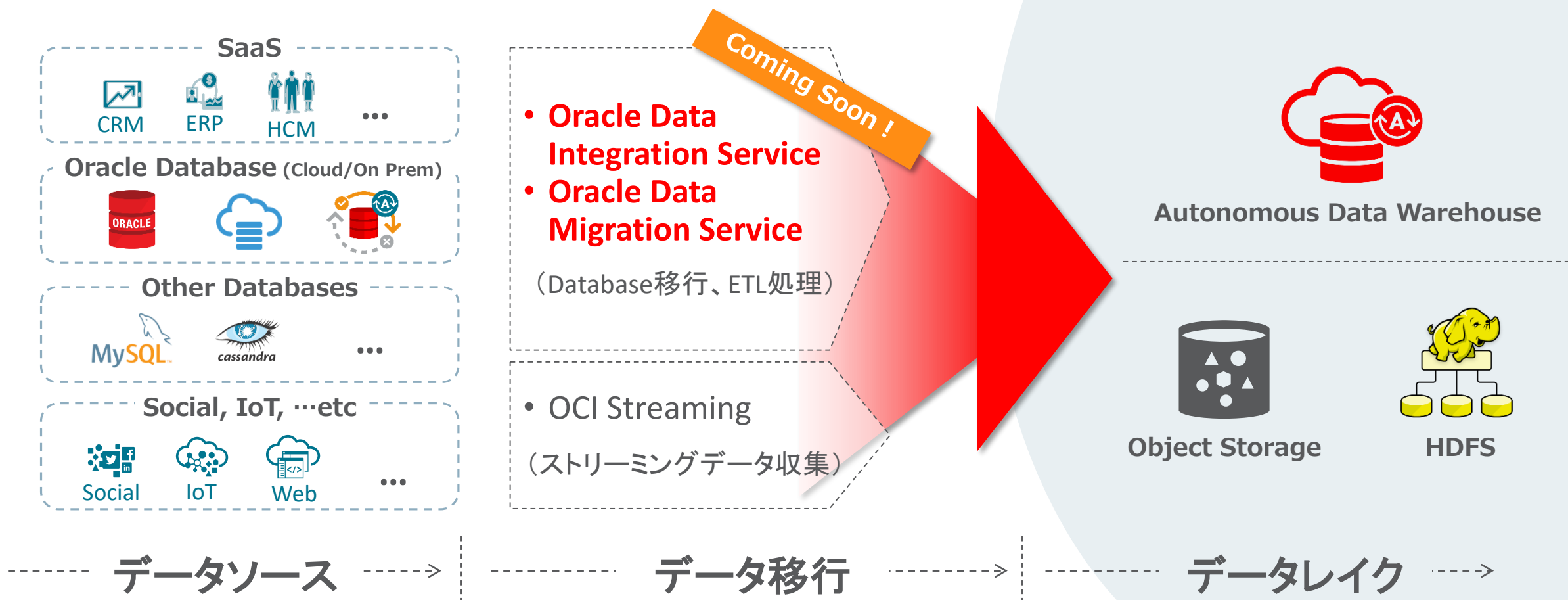
- 既存エンタープライズシステムをデータソースとして活用

- **3. High Performance**

- 高いパフォーマンスとスケーラビリティを実現する第2世代IaaS上にソリューションを構築

# 蓄積フェーズのテクノロジー

企業の既存ITシステムを含むあらゆるデータソースをデータレイクに統合





# Oracle Data Integration / Migration Service

## あらゆるデータソースをデータレイクに集約する、高信頼性ソリューション

- エンタープライズの既存ITシステムを含む幅広いデータソースに対応する移行ソリューション
  - Oracle Database (Cloud / OnPrem)
  - Oracle SaaS Services
  - Other Databases
  - …etc
- Oracle GoldenGateベースのテクノロジーで構築
  - 企業におけるリアルタイムのDBレプリケーション、ETL処理を高い信頼性で支えてきた基盤技術
  - エンタープライズでの豊富な導入実績

- Oracle Data Integration Service
- Oracle Data Migration Service  
(Database移行、ETL処理)

- OCI Streaming  
(ストリーミングデータ収集)

# OCI Object Storage

## エンタープライズグレードの高信頼性オブジェクト・ストレージ

- 高可用性設計（DCレベル、ファイルレベル）
  - データは、1つのリージョン内の複数アベイラビリティ・ドメインの複数ストレージ・サーバー上に冗長的に格納される
  - データ整合性はチェックサムを使用して継続的に監視され、破損したデータは自己修復される
- 高セキュリティ
  - Object Storage上のデータは自動的に暗号化される
  - 暗号化鍵は高頻度でローテート。AES-256でデータと鍵を暗号化
- 主要なObject Storageサービスとの互換性
  - 既存アプリケーションの高い移行性
  - AWS S3、OpenStack Swift互換

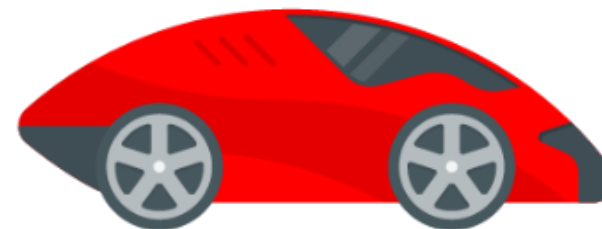


OCI Object Storage

# Autonomous Data Warehouse Cloud

## Oracle Database / Exadataのテクノロジーをクラウドで

- Oracle Autonomous Databaseサービス群における、データ分析処理向けの列指向型データベース
  - Oracle Database 18cとExadataの技術がベース
  - クラウドによる管理の省力化と柔軟な課金体型、スケーラビリティ
  - 本格的なエンタープライズ向けデータベースをより使いやすく提供



OLTP処理向けのAutonomous Transaction Processingもあります。

# Autonomous Data Warehouse Cloudの価値

## 迅速な立ち上がりと運用業務からの開放 → データの分析活用に注力

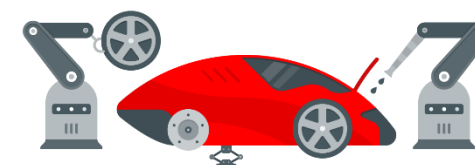
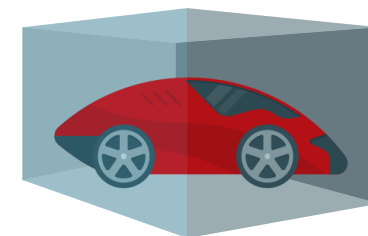
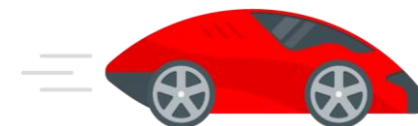
### • すぐに使える

- 早ければ30分でデータベース構築。構築コスト削減
- 初期投資不要



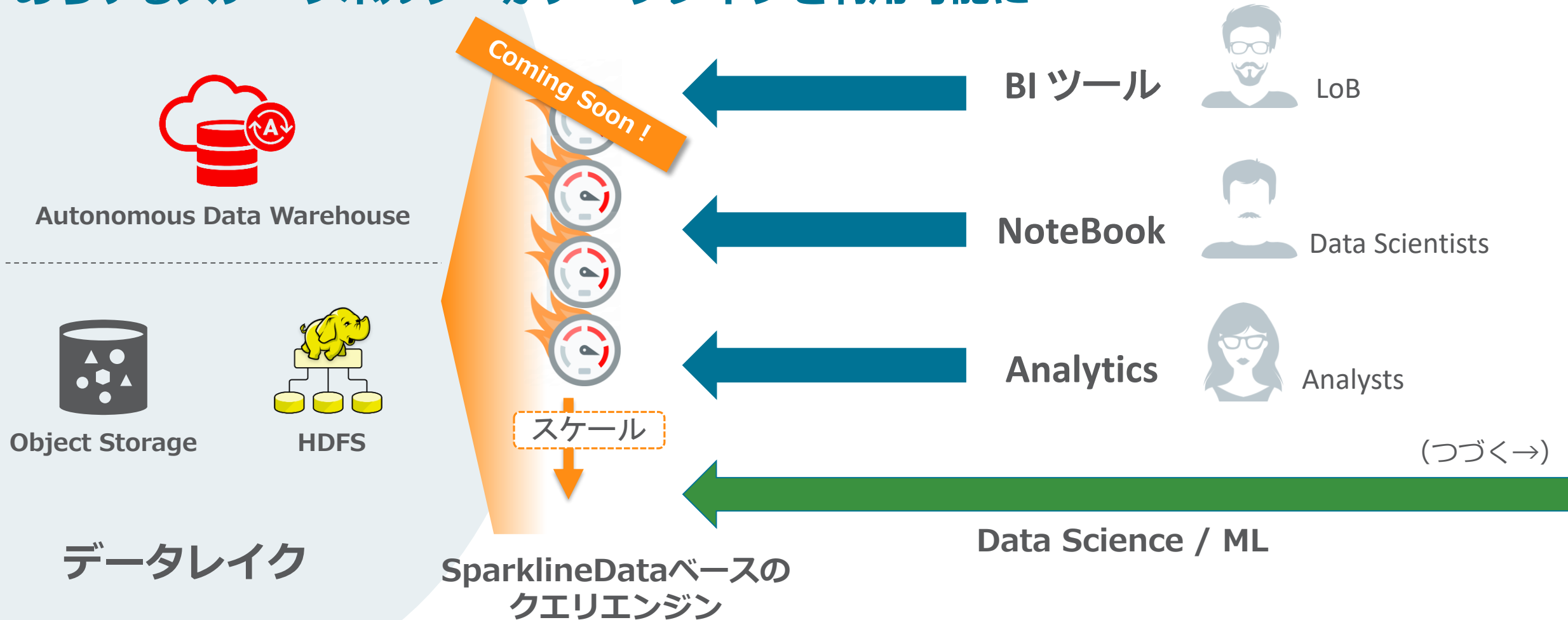
### • Autonomous

- Self-Driving（自己稼働）
  - 全ての管理・拡張・監視・チューニングの自動化
- Self-Securing（自己保護）
  - 外部からの攻撃や悪意のある内部ユーザからの防御
- Self-Repairing（自己修復）
  - 計画停止を含む全てのダウンタイムからの保護



# 分析フェーズのテクノロジー 1/2

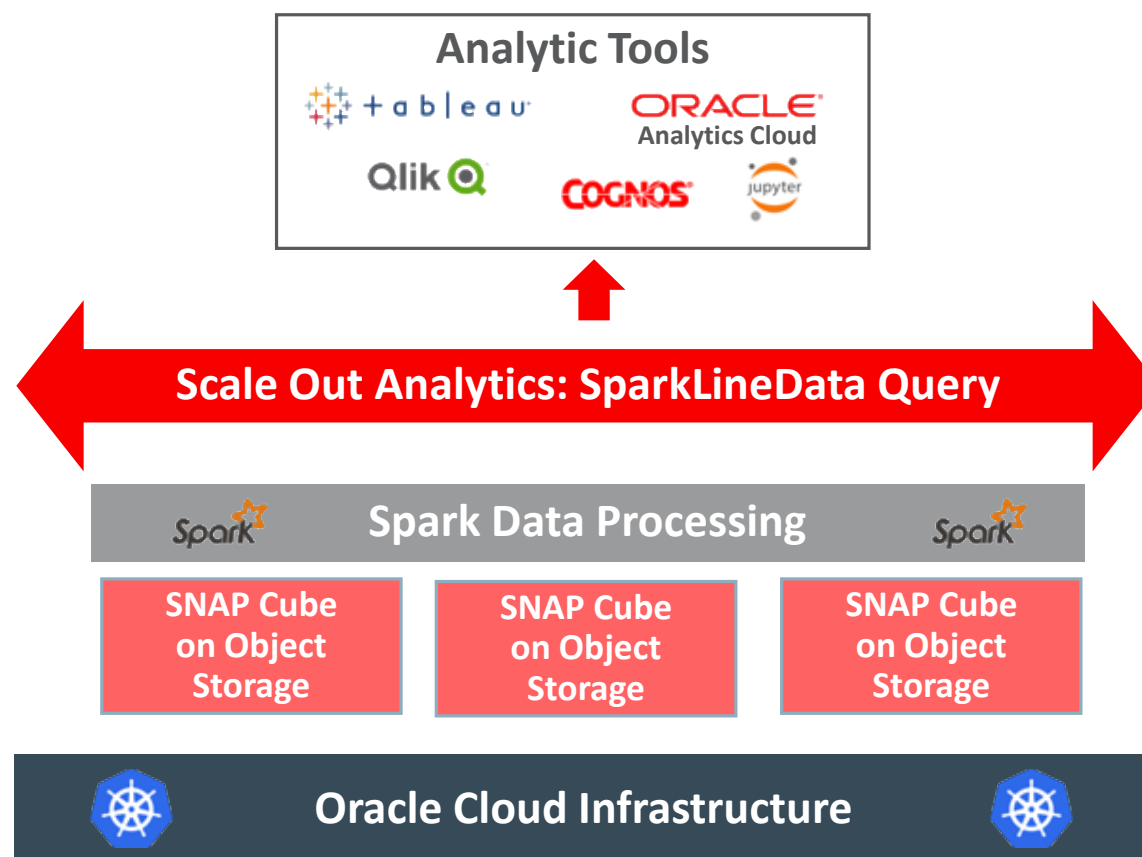
あらゆるステークホルダーがデータレイクを利用可能に



# SparklineData SNAPベースのクエリエンジン

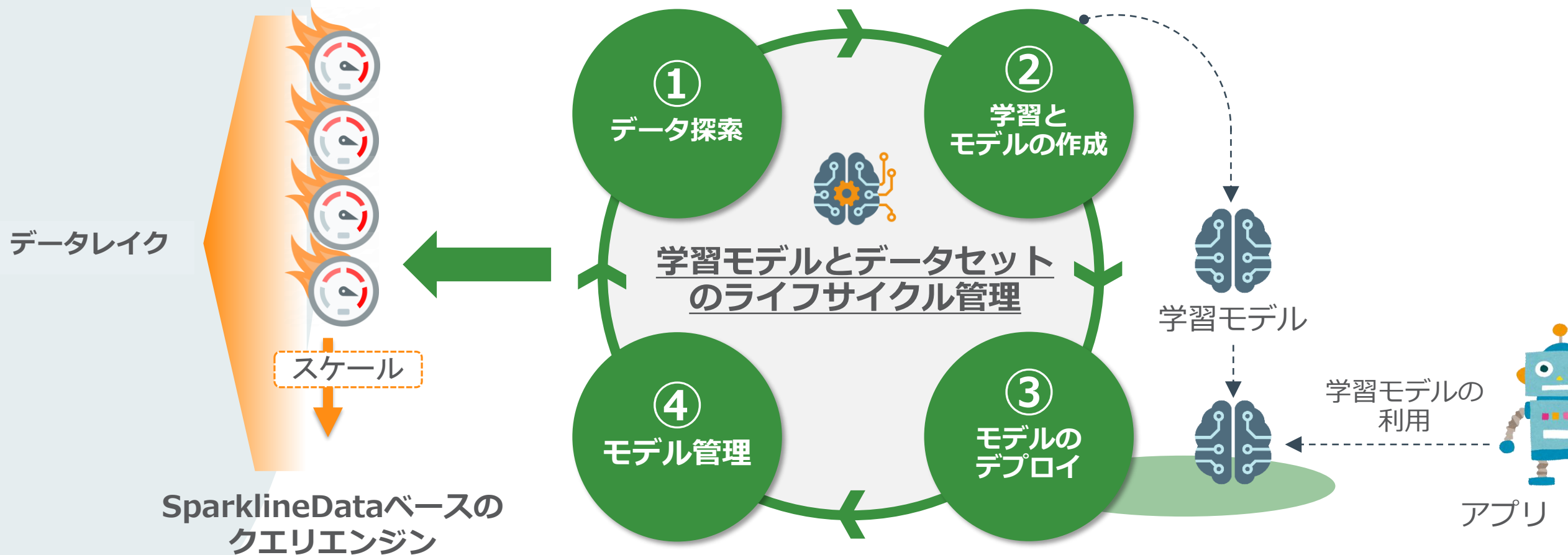
## 大規模データレイクのためのスケーラブルな高速クエリエンジン

- 様々なBIツールに対応
  - Tableau, Qlik, Cognos, BI EE, OAC
- 従来のBIツールから、データレイクをダイレクトに利用可能
- データレイク上で従来型データマートにおけるSubject Areaを実現
- オープンソースのSparkで構築
  - APIや運用モデルは共通



# 分析フェーズのテクノロジー 2/2

## 学習モデルとデータセットのライフサイクル管理が肝



# Oracle Data Science Cloud

## データサイエンス・プロジェクトのライフサイクル全域をカバー

- コラボレーションワークスペース
- Notebookベースの探索ツール



- ✓ LoB
- ✓ IT管理者
- ✓ アナリスト
- ✓ データサイエンティスト

- モデルのパフォーマンスの監視と最適化



- オープンソース・フレームワークによる学習とモデル生成

- モデルをスケーラブルなAPIとして運用。学習の成果をアプリケーションで利用

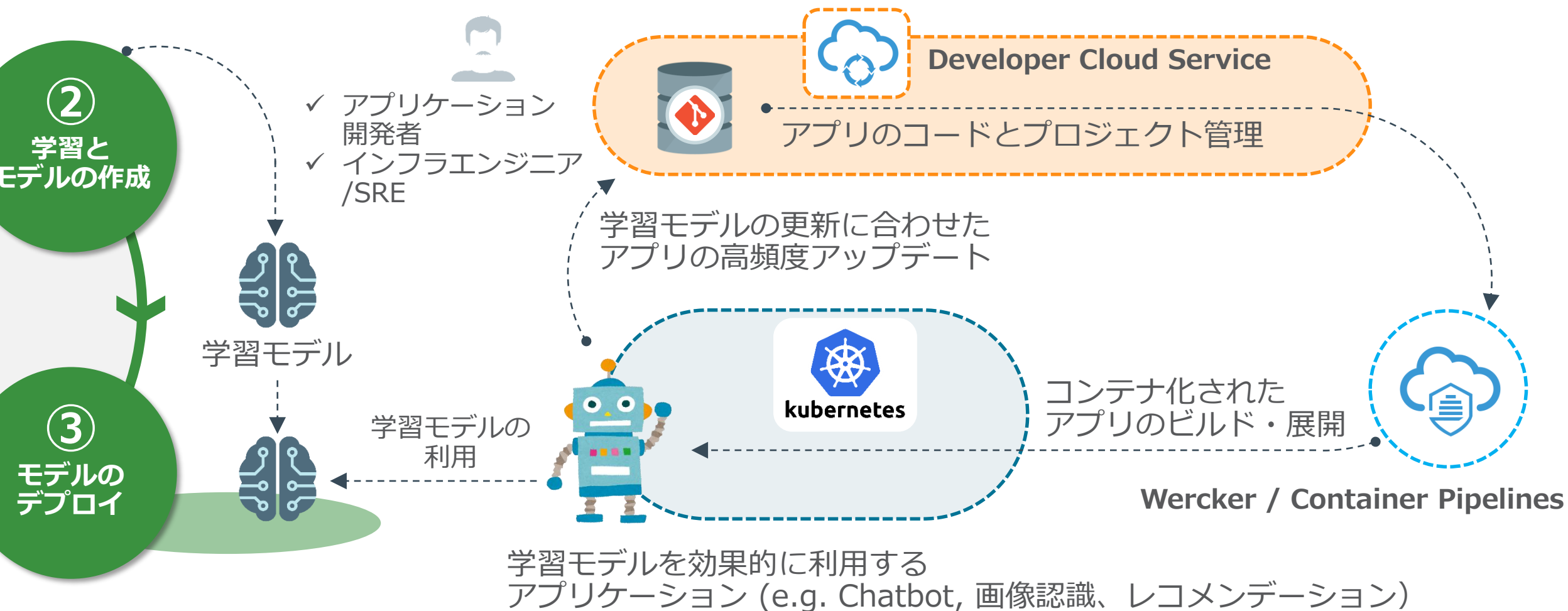


- ✓ アプリケーション開発者



# 活用フェーズのテクノロジー

## 学習モデルの更新に対応して価値の高いアプリケーション開発につなげる



# Oracle Container Engine for Kubernetes

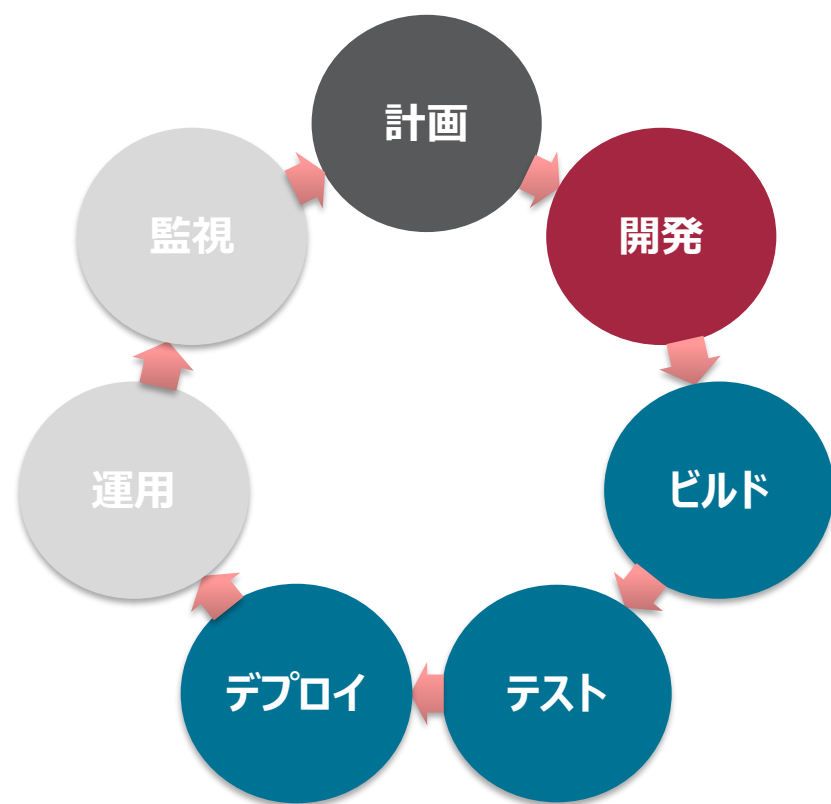
## エンタープライズグレードのマネージドKubernetesサービス

- **1. カスタマイズなしのピュアなKubernetes実装**
  - CNCF Certified Kubernetes Program適合
- **2. Oracle Cloud統合**
  - GUI / コマンドラインツール / 構成管理ツールによるクラスターのプロビジョニング、アップグレード
  - OCIのストレージ、ロードバランサーとの統合
  - Service BrokerによるOracle Cloud各種サービスとの自動接続
- **3. エンタープライズグレード**
  - 第2世代IaaSがもたらすパフォーマンス
  - Cross AD(Availability Domain)構成によるMaster/Workerの高可用性



# Oracle Developer Cloud Service

## DevOpsを支援する事前統合済みのクラウド型チーム開発プラットフォーム



- 迅速なアプリケーション開発のために統合された開発プラットフォーム
  - 開発支援ツール各種をオラクルが一括でクラウドで提供する開発プラットフォーム
  - コード管理 / ビルド管理 / タスク管理 などのツールを数ステップで一括導入
  - 開発からデプロイまでの工程でシームレスなCI/CD（継続的インテグレーション/デリバリー）を実現

# Wercker | Oracle Container Pipelines

## コンテナベースの開発に対応するCI/CDサービス

- オランダ発のスタートアップが発祥。2017年にOracleが買収してサービスを継続
- 成果物はコンテナとしてビルド
- パイプライン実行環境として任意のコンテナを導入でき、簡単に幅広いカスタマイズが可能
- 直感的で効率的なWorkflow作成
  - 初学者でも直感的に操作しやすいGUI



# Oracle Digital Assistant

## エンタープライズ向けチャットボットのための包括的なプラットフォーム



### マルチ・チャネル対応

メッセージング・サービスやアプリには非依存。Facebook Messengerとは事前統合済み



### 会話フロー・エンジン

コンテキストに沿った会話の流れを宣言的に定義するためのGUIツールを提供



### バックエンドとの連携

バックエンド・システムにセキュアにアクセスできるコンポーネントを提供



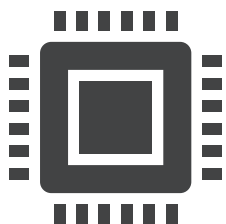
### AI エンジン

ユーザーからのリクエスト内容を理解するための自然言語解析や機械学習のモデルが構成済み

# Compute / Storage / Network

## ビッグデータ処理のワークロードのために最適化されたIaaS

### High Performance Compute



- フレキシブルなインスタンスの選択肢
  - VMs / Bare Metal
  - Intel / AMD EPYC
- 競合を大きく凌ぐ価格性能比

### Scalable, Independent Storage



- NVMe SSD, Object, Block and Archive Storage
- 高性能なHDFSの実現

### Guaranteed Network Performance



- オーバーサブスクライブなし。ノード間通信 25 Gbps
- ネットワーク性能のSLAをデフォルトで持つ唯一のクラウド
- アウトバウンド通信の課金なし

# Complete Cloud Infrastructure ポートフォリオ

## COMPUTE

### Bare metal/VM, CPUs/GPUs



Up to 64 CPU cores, 8 GPUs, 768 GB RAM, 51 TB local NVMe SSD, 5M IOPS, AMD and Intel processors

## STORAGE

### NVMe, Block, File, Object, Archive



Predictable IOPS Block Storage for up to 98% less, storage for whole lifecycle

## AUTONOMOUS DATABASE

### Transactions, Data Warehouse



Fast provisioning. Automatic tuning, patching, securing. 99.995% availability.

## SECURITY

### IAM, Audit, KMS, CASB



Integrated security services to protect data and to control and monitor access

## CONTAINERS

### Containers and Kubernetes



Fully managed, certified Kubernetes service with Docker containers

## NETWORKING

### VCN, LBaaS, FastConnect, VPN



Isolated networks with reserved IPs, security lists, firewalls, lowest cost private connectivity

## DATABASE

### Bare metal, VMs, Exadata



Millions of TPS; Full RAC and Active Data Guard support

## EDGE

### DNS, WAF, DDoS, Email



Global DNS, application protection, bot management, DDoS protection, email delivery

## MIGRATION

### Migrate Apps, KVM



Move on-prem environments to the cloud with full hardware performance

## OCI AT CUSTOMER

### IaaS, PaaS, Exadata on-premises



Subscription-priced cloud infrastructure, PaaS, and database managed by Oracle

## DATA MOVEMENT

### Storage appliance, Data Transfer



Software NAS gateway, data ingest service with full chain of custody (HDD or appliance)

## GOVERNANCE

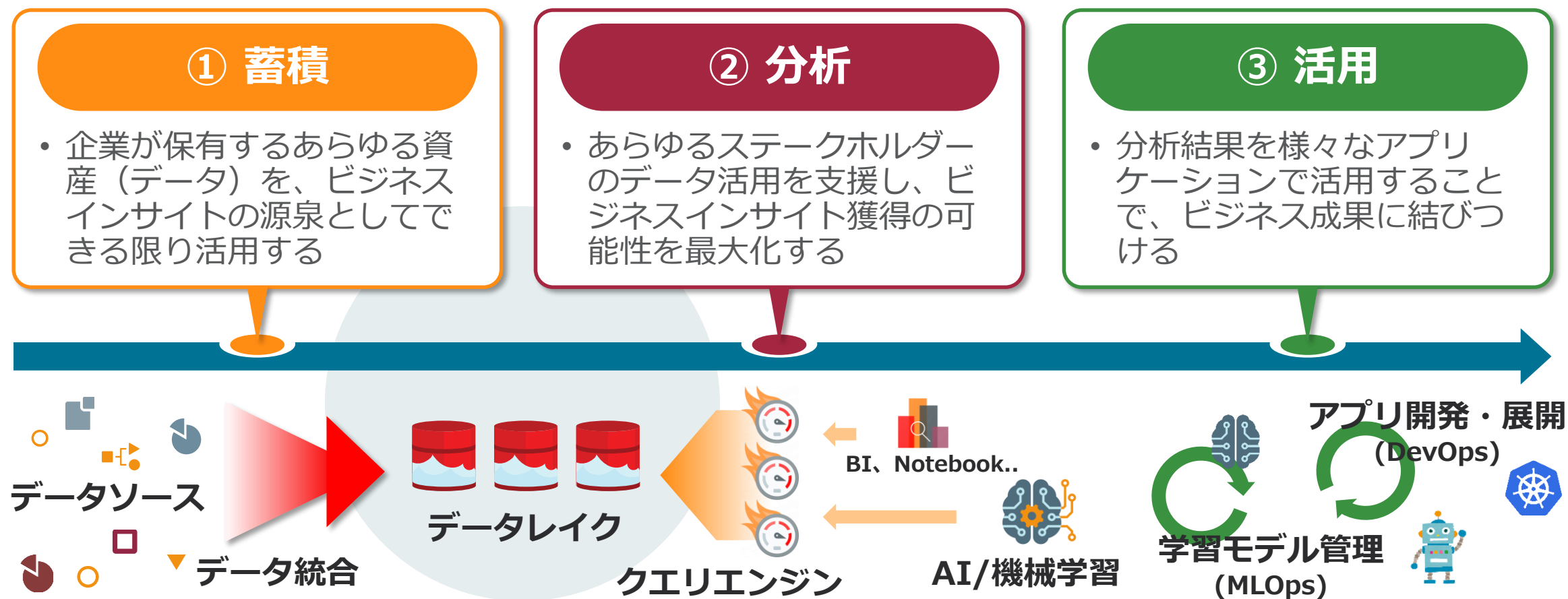
### IAM, Tagging, Cost Management



Logical separation and tagging of resources for simplified management

# まとめ - データ活用を実現するテクノロジーの理想像

## 蓄積・活用・維持 / 3つのフェーズをテクノロジーでサポート





# まとめ - オラクルのデータ活用テクノロジーのコンセプト

- **1. Completed**

- データソースの整備から成果の刈り取り（アプリでの利用）までを包含
- データ活用を実現するテクノロジーの理想像を満たす

- **2. Enterprise**

- 既存エンタープライズシステムをデータソースとして活用

- **3. High Performance**

- 高いパフォーマンスとスケーラビリティを実現する第2世代IaaS上にソリューションを構築

こんな時、かけこむ会社が増えています。



ビジネスプロセスを  
改善したい!



今のシステムは  
使いにくい!



システムコストを  
下げたい!



パフォーマンスを  
良くしたい!



経営分析を  
したいのだが...



どんなソリューションが  
あるの?



見積りはどれくらい  
なんだろう?



楽に管理を  
したい!

**Oracle Digitalは、オラクル製品の導入をご検討いただく際の総合窓口。  
電話とインターネットによるダイレクトなコミュニケーションで、どんなお問い合わせにもすばやく対応します。  
もちろん、無償。どんなことでも、ご相談ください。**



**お問い合わせは電話またはWebフォーム**

**☎ 0120-155-096**

受付時間 月～金 9:00-12:00 / 13:00-17:00  
(祝日および年末年始休業日を除きます)

**<http://www.oracle.com/jp/contact-us>**

ORACLE®