

ORACLE

総まとめ！オラクルの機械学習 /Big Data系クラウドサービス

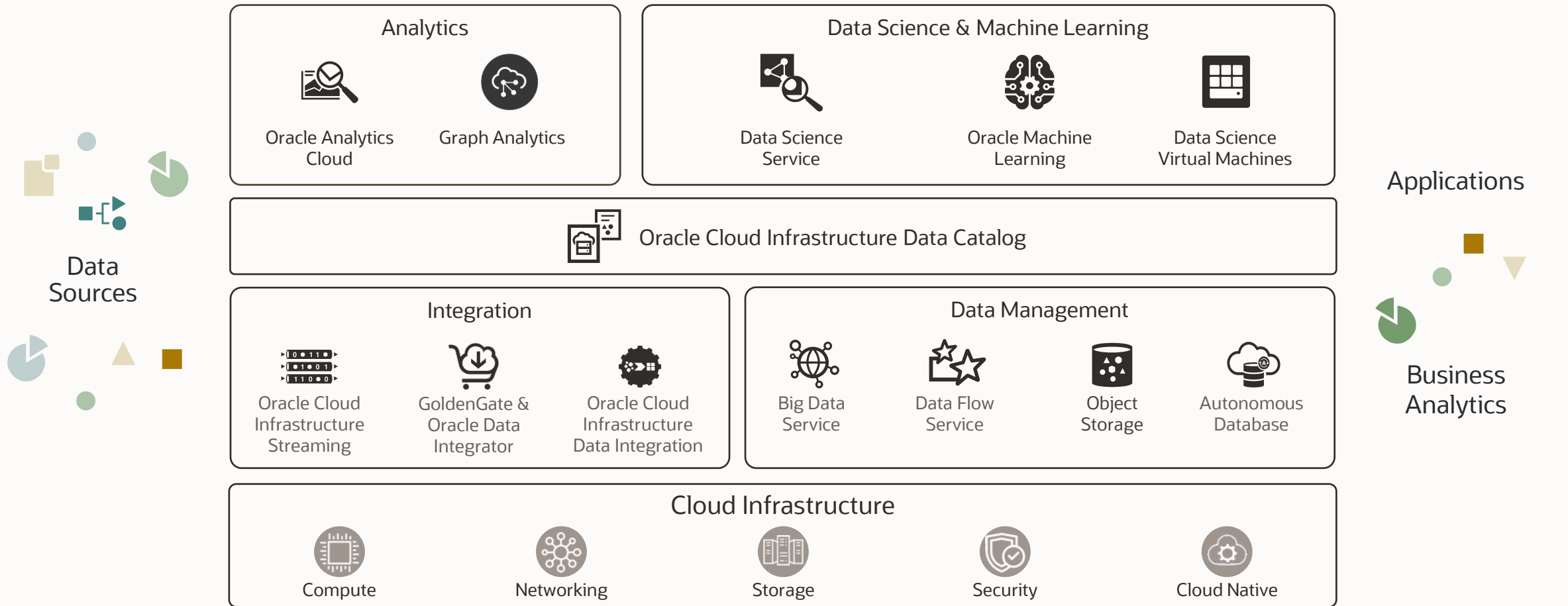
2020年12月18日

園田憲一

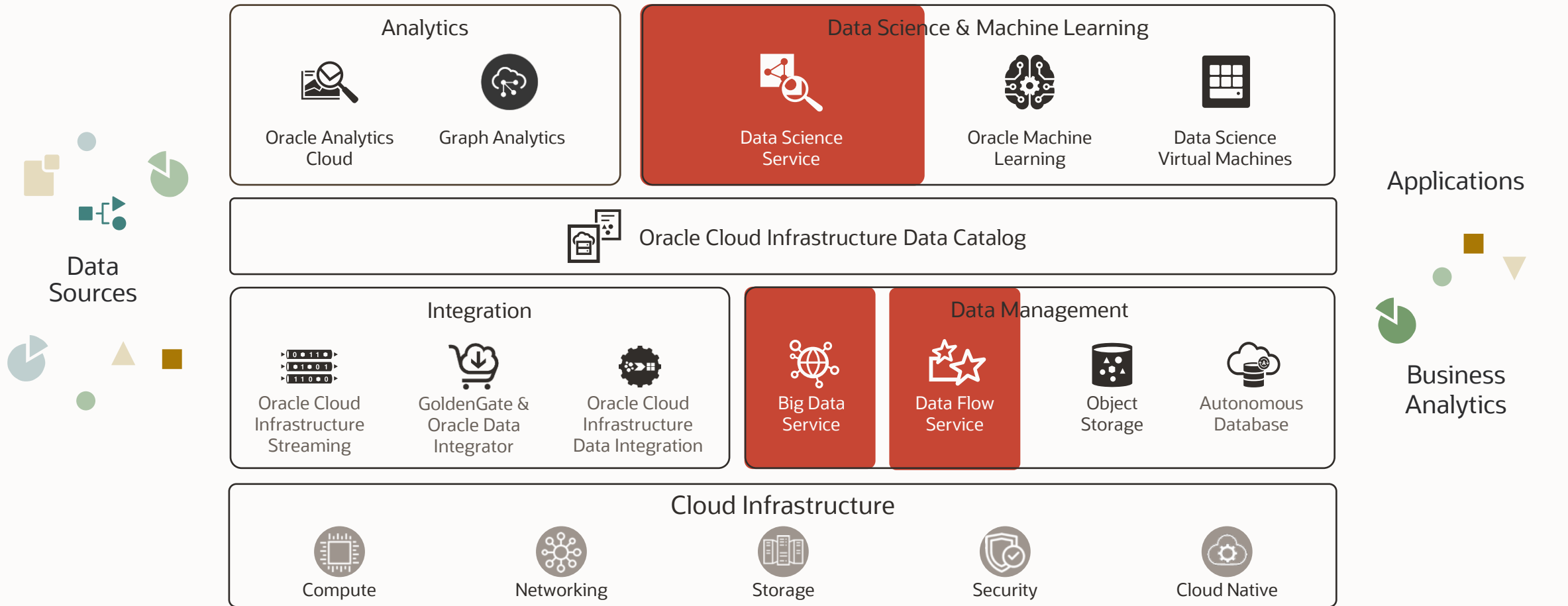
日本オラクル株式会社



Oracle Data and AI Platform



Oracle Data and AI Platform

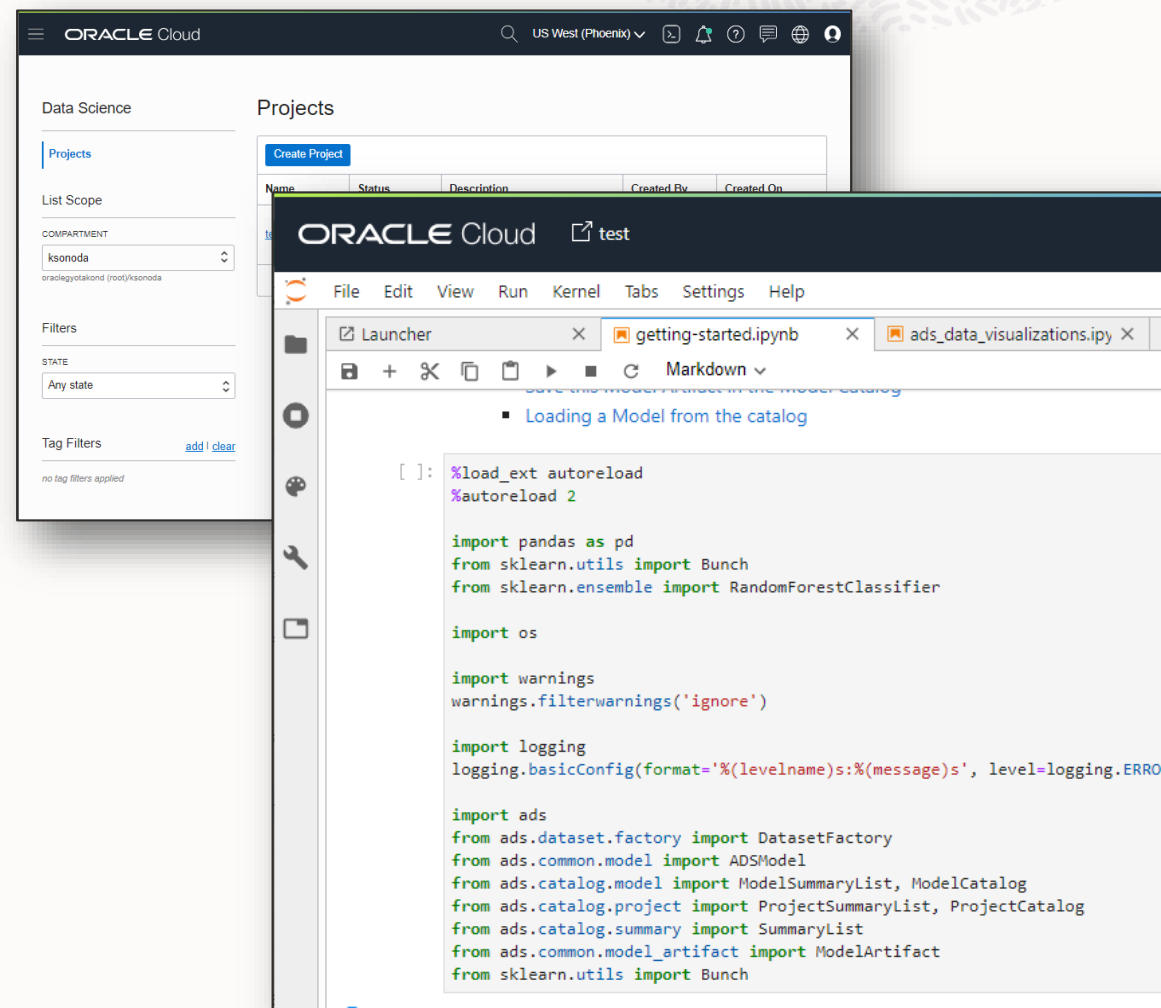




Data Science Service

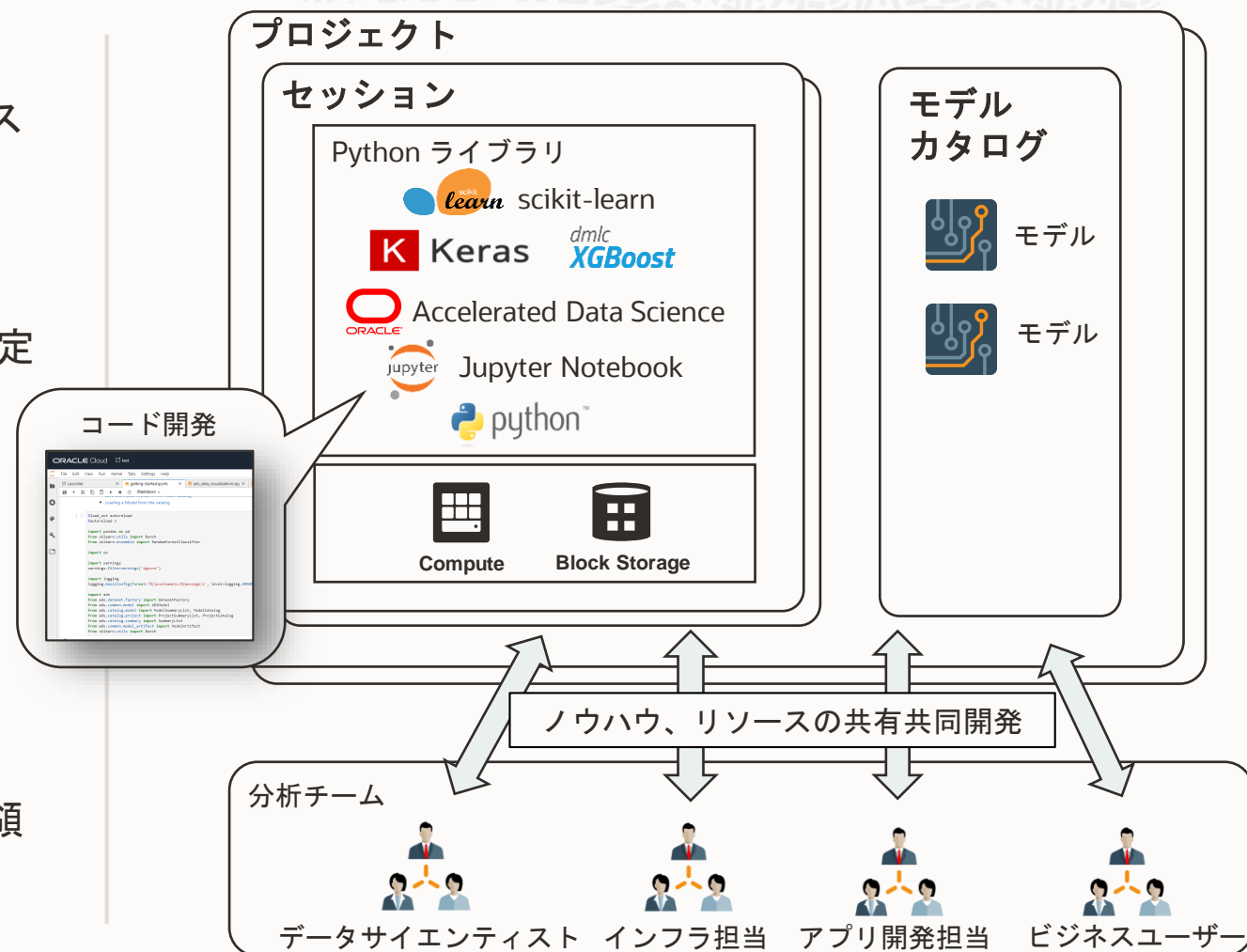
Data Science Service 概要

1. オールインワンの機械学習開発環境を提供するクラウドサービス
 - Pythonの実行環境
 - 機械学習ライブラリ(OSS)
 - Oracle Accelerated Data Science(ADS)
2. チームによる分析プロジェクトの推進
 - 分析プロジェクトを推進できる各部門の専門家がチームとして、共同ワークスペースを提供
 - 複数のデータサイエンティスト同志のノウハウの共有
 - ソースコード、学習用データ、予測モデル
3. PaaSとしては無償、IaaSのみの課金



Data Science Serviceの主要コンポーネント

1. プロジェクト
 - 全てのリソースを保持する共同ワークスペース
2. ノートブック・セッション
 - オープンソースのMLライブラリがプリインストールされたComputeインスタンス
 - Computeシェイプ、Block Volumeの容量を指定し作成、GPUインスタンス利用可
3. 主要ライブラリ
 - Python
 - Keras、scikit-learn、XGBoost
 - Oracle Accelerated Data Science(ADS)
 - Jupyter Notebook
4. モデルカタログ
 - 構築したモデルを登録、共有するストレージ領域

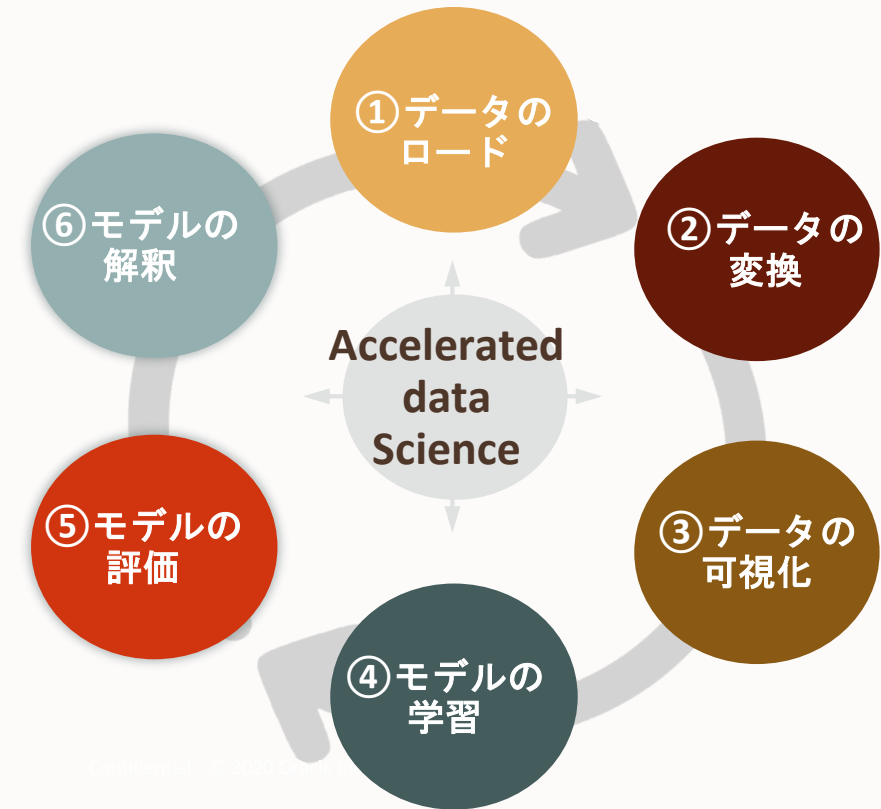


Accelerated Data Science

- Oracle Cloud Infrastructure Data Scienceの一部として機能するPythonライブラリ
- 機械学習のライフサイクル全てのフェーズで使いやすくシンプルなAPI
- Oracle Cloudのその他のサービスやAWS S3 Azure Blobなどのサービス連携が容易
- Oracle AutoML
 - ワークフロー内の処理の自動化
 - 処理時間の削減
 - モデル精度の向上



機械学習のワークフロー

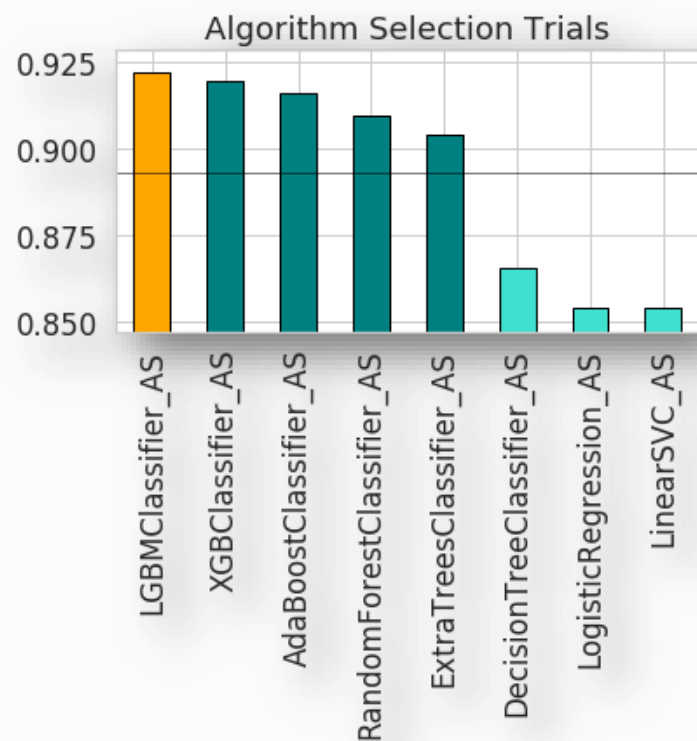


モデルの学習

Oracle AutoMLによる自動化

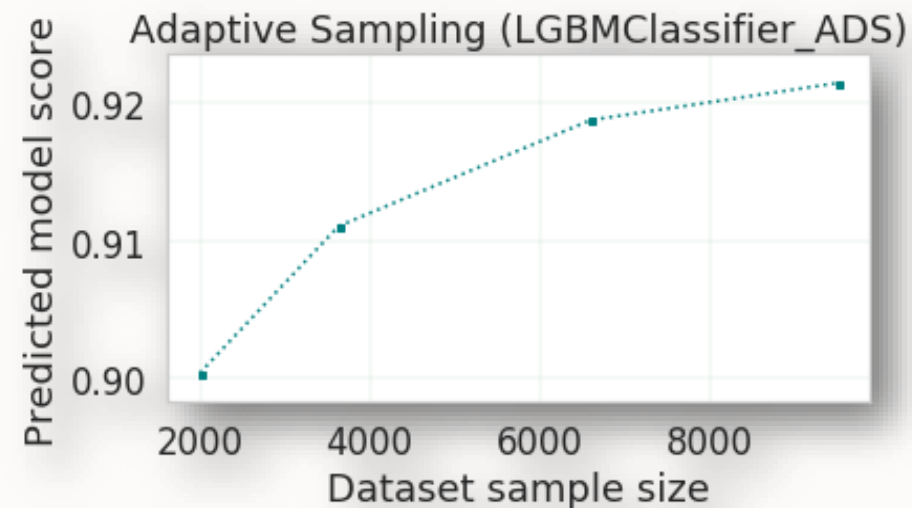
①最適なアルゴリズム選択の自動化

```
oracle_automl.visualize_algorithm_selection_trials()
```



②サンプリングの自動化

```
oracle_automl.visualize_adaptive_sampling_trials()
```

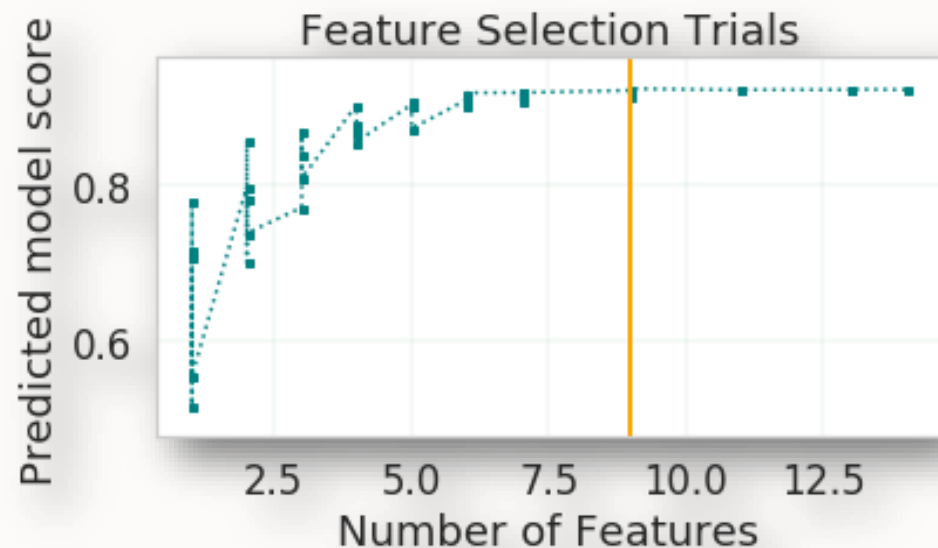


モデルの学習

Oracle AutoMLによる自動化

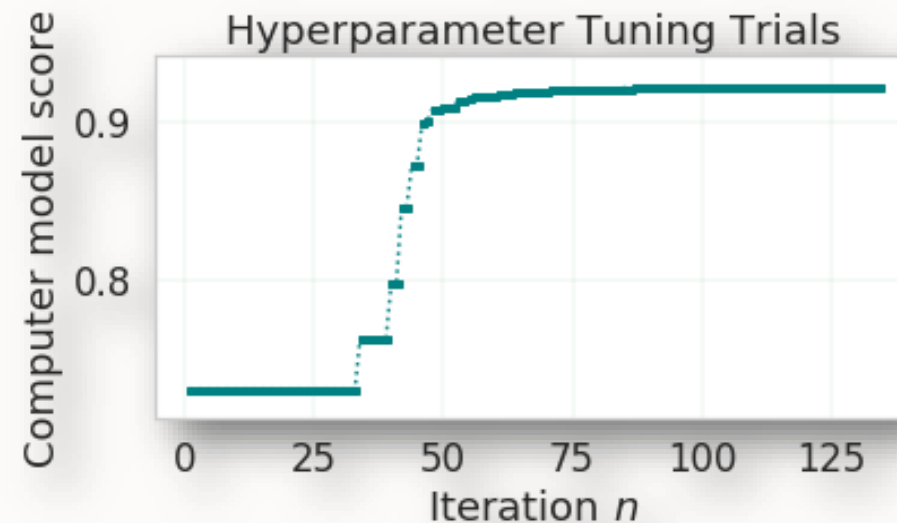
③最適な特徴量選択の自動化

```
oracle_automl.visualize_feature_selection_trials()
```



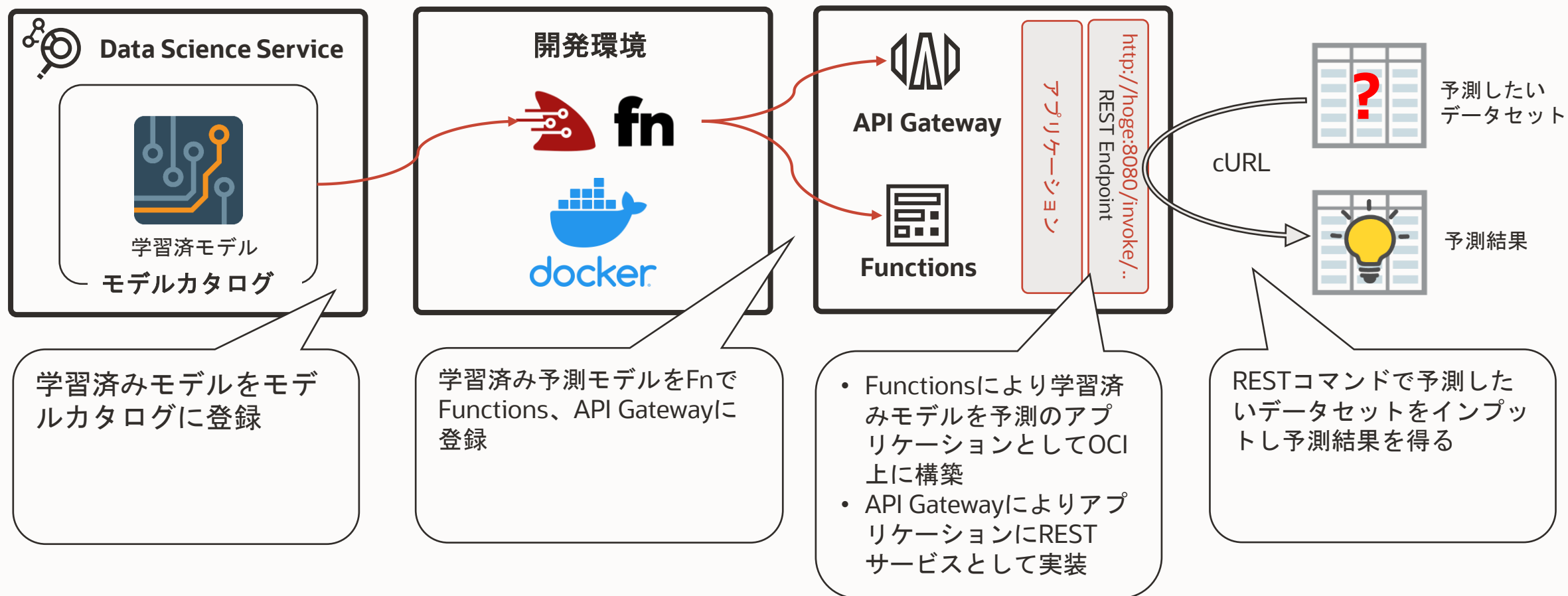
④ハイパーパラメータ・チューニングの自動化

```
oracle_automl.visualize_tuning_trials()
```



予測モデルをアプリケーションとしてクラウド上に実装

Data Science Service / Functions Service / API Gateway Serviceの連携



ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka) ▾

データベース

- MySQL
- NoSQLデータベース
- データベース関連サービス
- APEXアプリケーション開発
- データ・セーフ
- データおよびAI
- ビッグ・データ
- デジタル・アシスタント
- データ・カタログ
- データ・フロー
- データ統合
- データサイエンス
- ソリューションおよびプラットフォーム
- アナリティクス
- リソース・マネージャ
- 電子メール配信
- アプリケーション統合

主な概念および用語

ドキュメント

Oracle Cloud Infrastructureを開始するには、いくつかの重要な概念と用語をよく理解してください。

Oracle UniversityによるFREEトレーニング

プロジェクト

注目

すべてのシステムが動作可能

ヘルス・ダッシュボードの表示

アカウント・センター

ユーザー管理

テナンシへのユーザーの追加

請求

コストの分析

支払方法の管理

最新情報

Exadata DBシステムでのData Guardのトラブルシューティング

2020年12月5日

ファンクションの同時実行のためのデフォルト合計メモリの増加

2020年12月3日

Exadata Cloud Service: Oracle Database 19cのアップグレード機能が使用可能に

2020年12月3日

ADBS- マルチDBバージョンの機能強化 - Free Tierバージョンの選択

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka) ▾



データ・サイエンス

プロジェクト

プロジェクト

リスト範囲

コンパートメント

madhacks.ochacafe ▾
orasejapan (ルート)/project/madhacks.ochacafe

フィルタ

状態

任意の状態 ▾

タグ・フィルタ [追加](#) [クリア](#)

タグ・フィルタは適用されていません

① データ・サイエンスの前提条件
[詳細情報の表示](#)






プロジェクトの作成

名前	ステータス	説明	作成者	作成日
アイテムが見つかりませんでした。				
表示するプロジェクトなし < 1の1 >				



ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka)      

データ・サイエンス

プロジェクト

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

フィルタ

状態

任意の状態

タグ・フィルタ

追加

タグ・フィルタは適用されていません

プロジェクトの作成

[ヘルプ](#) [閉じる](#)

プロジェクトにより、チームのデータ・サイエンス作業を整理できます。

コンパートメント ⓘ

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

名前 オプション ⓘ

banner_ab_test

説明 オプション

ab test for banner

タグ

タグ付けとは、テナンシ内のリソースを整理およびトラッキングできるメタデータ・システムです。タグは、リソースにアタッチできるキーと値から構成されます。

[タグ付けの詳細](#)

タグ・ネームスペース	タグ・キー	値
なし(フリーフォー...		

+ 追加タグ

☒ 「作成」をクリックして詳細ページを表示

作成

取消

任意のプロジェクト名

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka)

データ・サイエンス » プロジェクト » banner_ab_test

P

アクティブ

banner_ab_test

編集

削除

リソースの移動

タグの追加

プロジェクト情報

タグ

説明: ab test for banner

作成者: ...acle.com

OCID: ...7nmqa5aa

作成日: 2020年12月17日(木) 1:11:10 UTC

リソース

ノートブック・セッション

モデル

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

madhacks.ochacafe コンパートメント内のノートブック・セッション

ノートブック・セッションの作成

名前	ステータス	コンピューター・インスタンスのシェイプ	作成者	作成日
アイテムが見つかりませんでした。				

表示するノートブック・セッションなし < 1の1 >

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan East (Tokyo)

データ・サイエンス » プロジェクト

ノートブック・セッションの作成

ヘルプ 閉じる

コンパートメント ⓘ

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

名前 オプション ⓘ

test_session

コンピュート・インスタンスのシェイプ

VM.GPU3.1

ブロック・ストレージ・サイズ(GB) ⓘ

1000

サイズは500GBから1,024GB (1TB)まで変更できます

MADHACKS.OCHACAFEのVCN ⓘ (コンパートメントの変更)

bigdata-vcn

MADHACKS.OCHACAFEのサブネット ⓘ (コンパートメントの変更)

Public Subnet-bigdata-vcn

タグ

タグ付けとは、テナンシ内のリソースを整理およびトラッキングできるメタデータ・システムです。タグは、リソースにアタッチできるキーと値から構成されます。

☒ 「作成」をクリックして詳細ページを表示

作成

取消

アクティブ

リソース

ノートブック・セッション

モデル

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

フィルタ

開発環境のComputeインスタンスのスペックを選択



ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan East (Tokyo)

データ・サイエンス » プロジェクト » banner_ab_test

P

アクティブ

banner_ab_test

編集

削除

リソースの移動

タグの追加

プロジェクト情報

タグ

説明: 値なし

作成者: ...acle.com

作成日: 2020年12月17日(木) 1:13:23 UTC

OCID: ...7erg7kma

リソース

ノートブック・セッション

モデル

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

madhacks.ochacafe コンパートメント内のノートブック・セッション

ノートブック・セッションの作成

名前	ステータス	コンピュート・インスタンスのシェイプ	作成者	作成日
test_session	● アクティブ	VM.GPU3.1	...acle.com 表示 コピー	2020年12月17日(木) 1:29:27 UTC

1個のノートブック・セッションを表示 < 1の1 >

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan East (Tokyo)

データ・サイエンス » プロジェクト » banner_ab_test: ノートブック・セッション » test_session

N

アクティブ

開く

編集

非アクティブ化

終了

他のアクション

test_session

ノートブック・セッション情報

タグ

OCID: ...jkzs3mfa 表示 コピー

作成者: ...acle.com 表示 コピー

ブロック・ストレージ・サイズ(GB): 1,000 GB

サブネット: [Public Subnet-bigdata-vcn](#)

作成日: 2020年12月17日(木) 1:29:27 UTC

コンピュータ・インスタンスのシェイプ: VM.GPU3.1

VCN: [bigdata-vcn](#)

リソース

メトリック

メトリック

メトリックは、新規に作成またはアクティブ化されたノートブック・セッションに対してのみ使用可能です。新規ノートブック・セッションを作成するか、これを非アクティブ化してから再アクティブ化して、メトリックを有効にしてください。新規ノートブック・セッションのメトリックが表示されない場合は、メトリックを読み取る権限があることを確認します。 [詳細](#)

開始時間

終了時間

クイック選択

2020/12/17 0:39:30

2020/12/17 1:39:30

過去1時間

チャートのリセット

CPU使用率

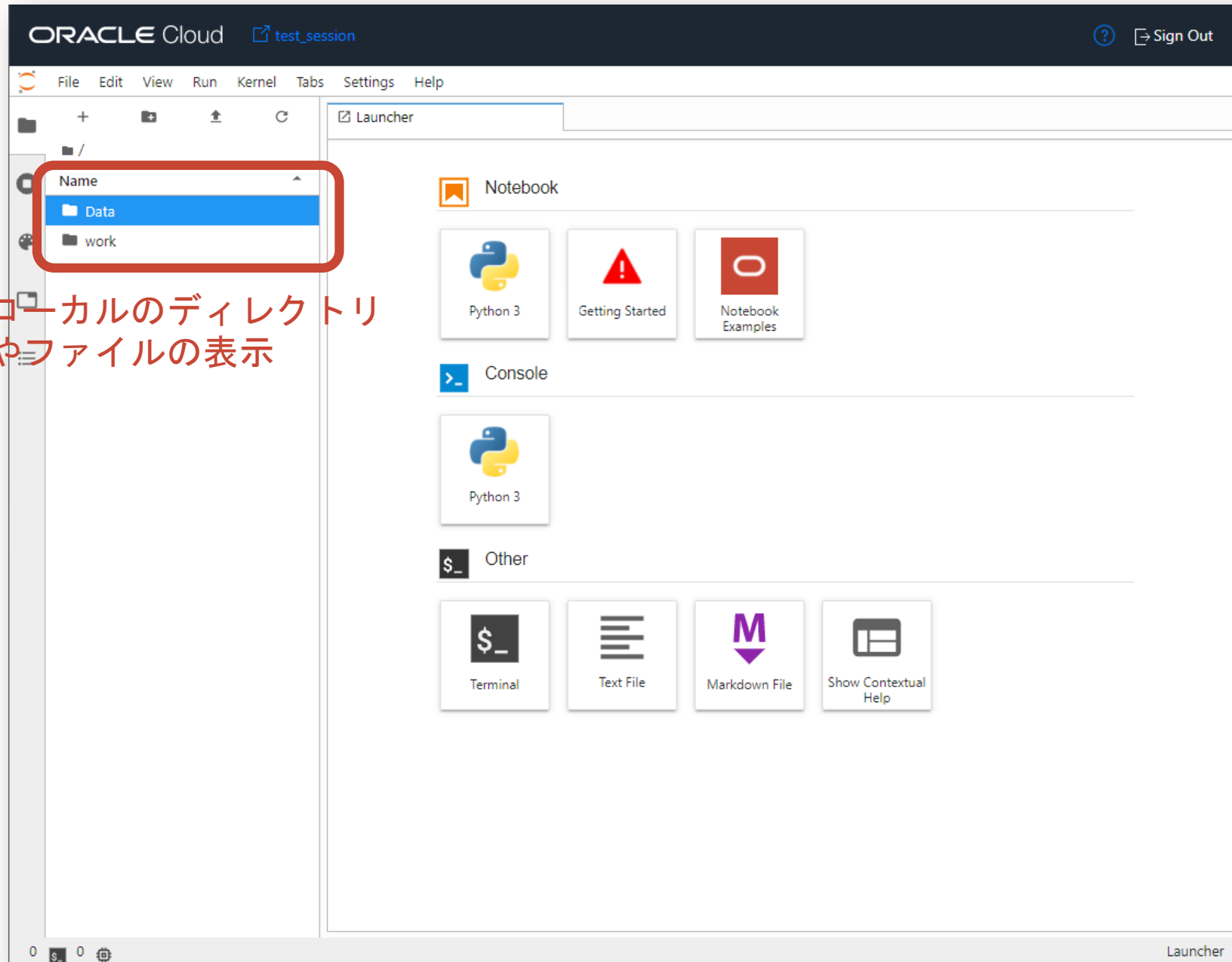
オプション

間隔

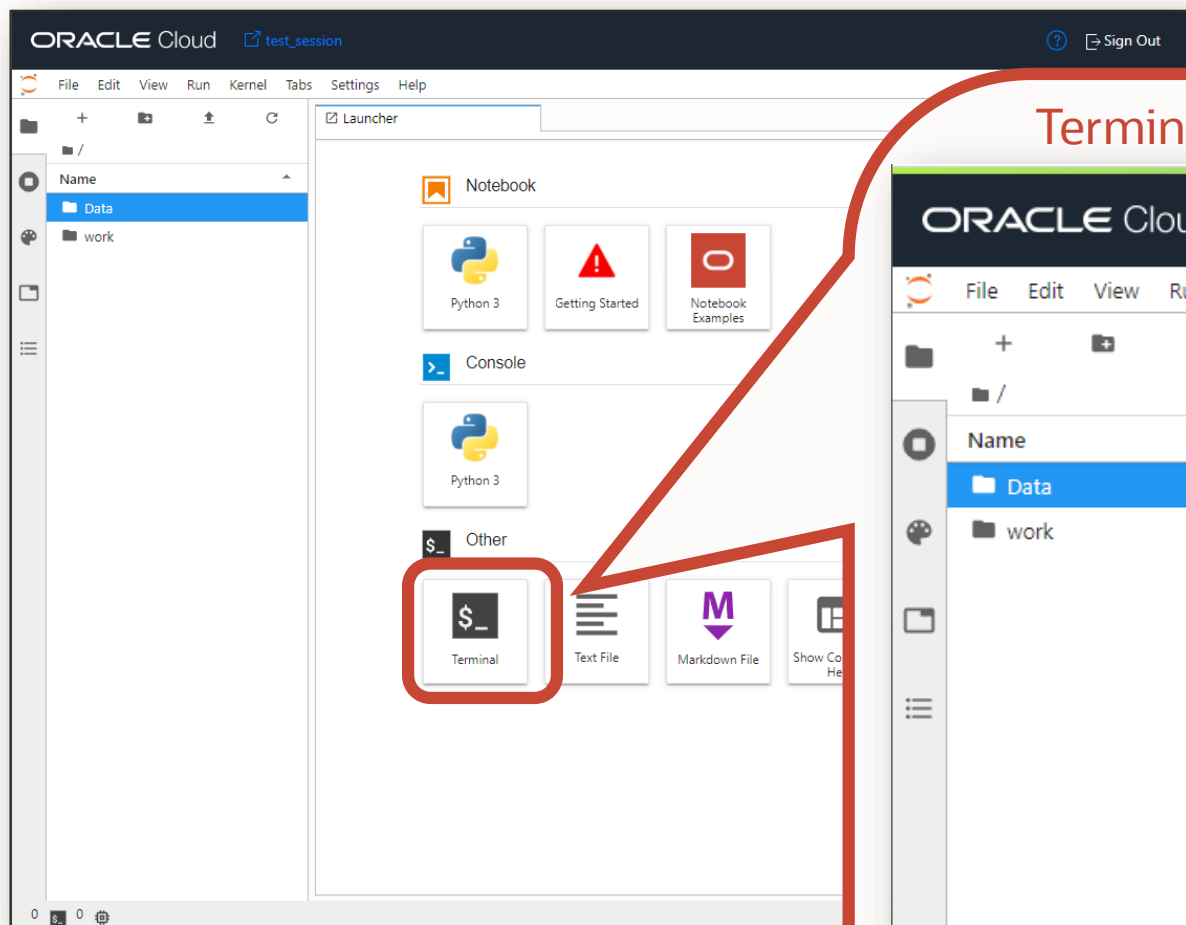
統計

1分

平均



ローカルのディレクトリ
やファイルの表示



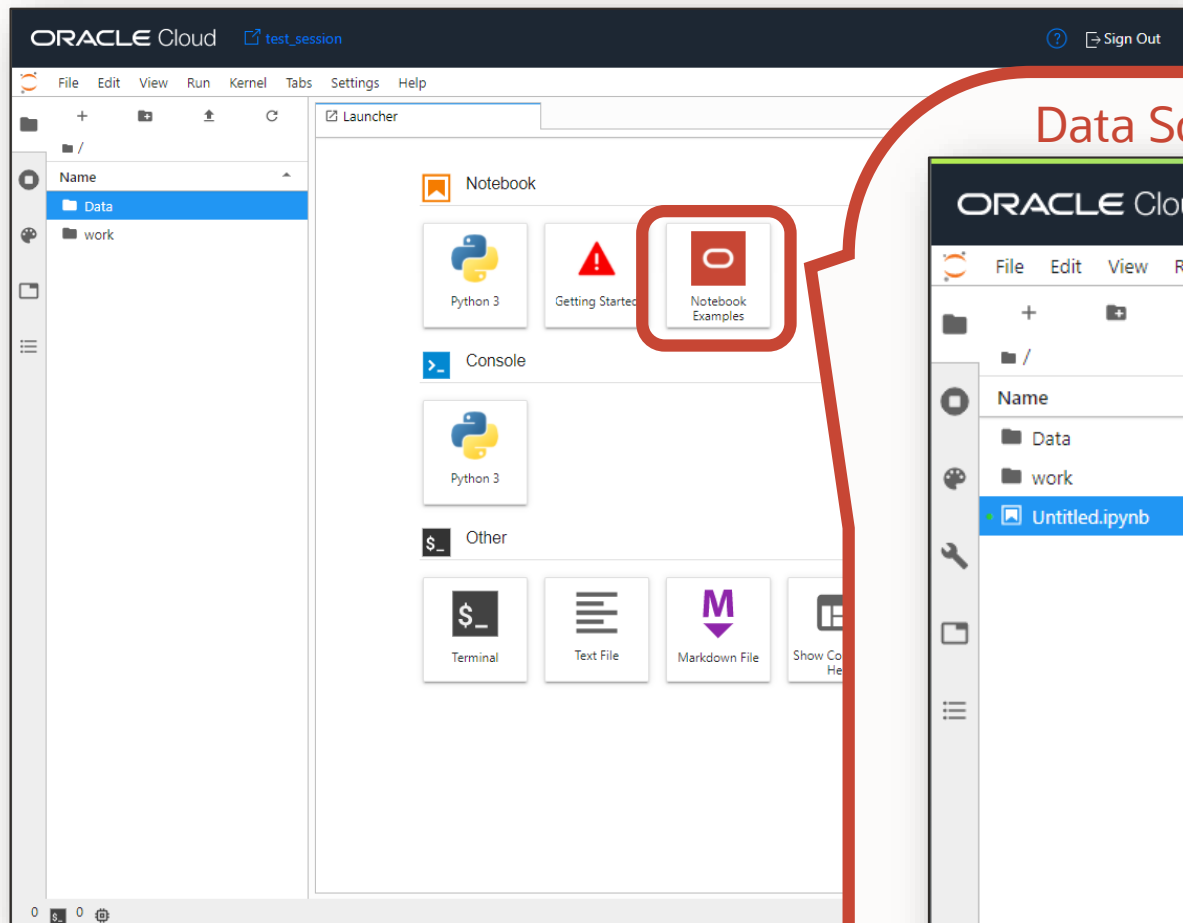
TerminalからOSコマンドを実行

The screenshot shows the Oracle Cloud interface with the 'Terminal' window open. The terminal displays the following commands and output:

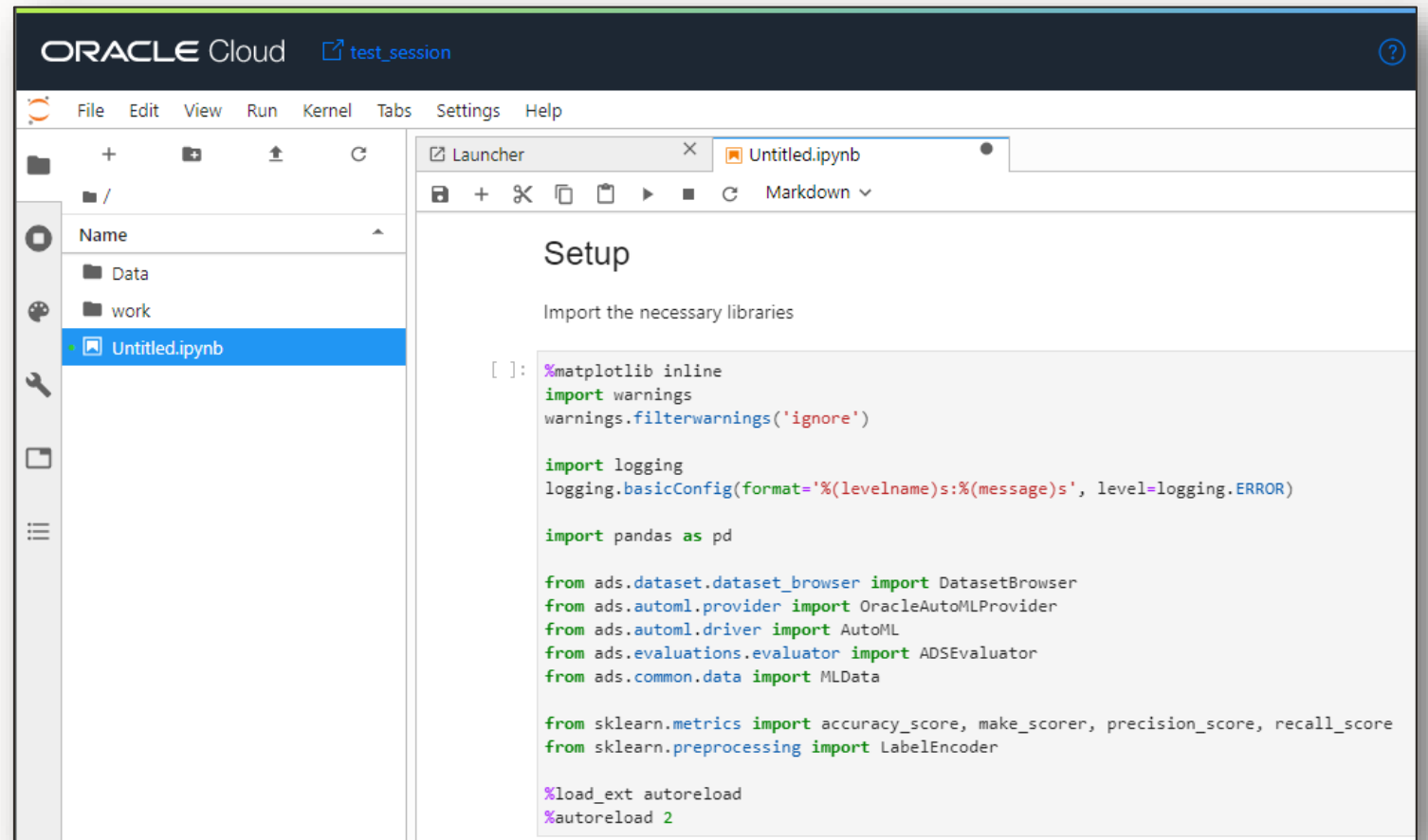
```
(base) bash-4.2$  
(base) bash-4.2$  
(base) bash-4.2$ ls  
Data work  
(base) bash-4.2$  
(base) bash-4.2$ nvidia-smi  
Thu Dec 17 01:46:00 2020
```

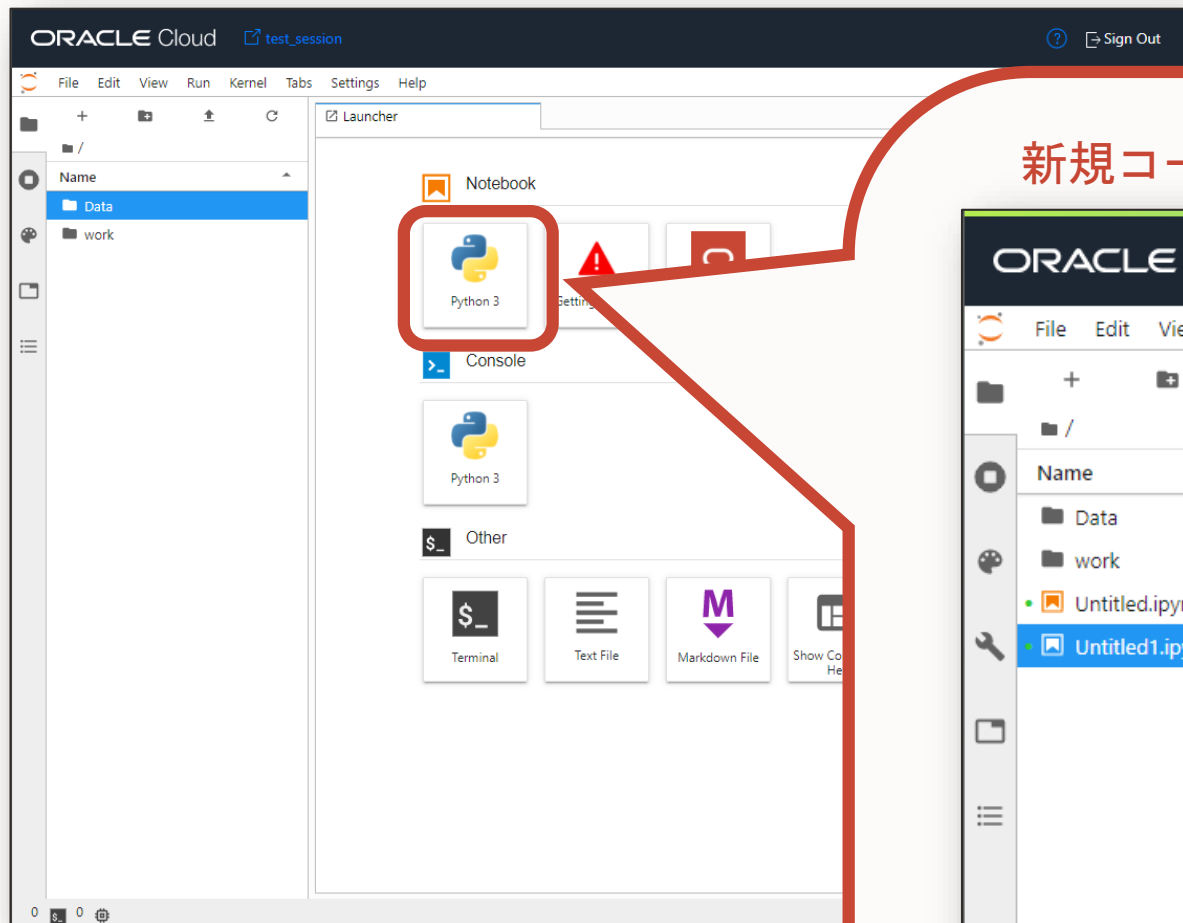
NVIDIA-SMI 450.51.06		Driver Version: 450.51.06		CUDA Version: 11.0	
GPU	Name	Persistence-M	Bus-Id	Disp.A	Memory-Usage
Fan	Temp	Perf	Pwr:Usage/Cap		
=====					
0	Tesla V100-SXM2...	Off	00000000:00:04.0	Off	
N/A	42C	P0	37W / 300W	0MiB / 16160MiB	4%
=====					

```
Processes:  
GPU  GI  CI  PID  Type  Process name  GPU Memory  
ID   ID   ID                   Usage  
-----  
No running processes found  
(base) bash-4.2$
```

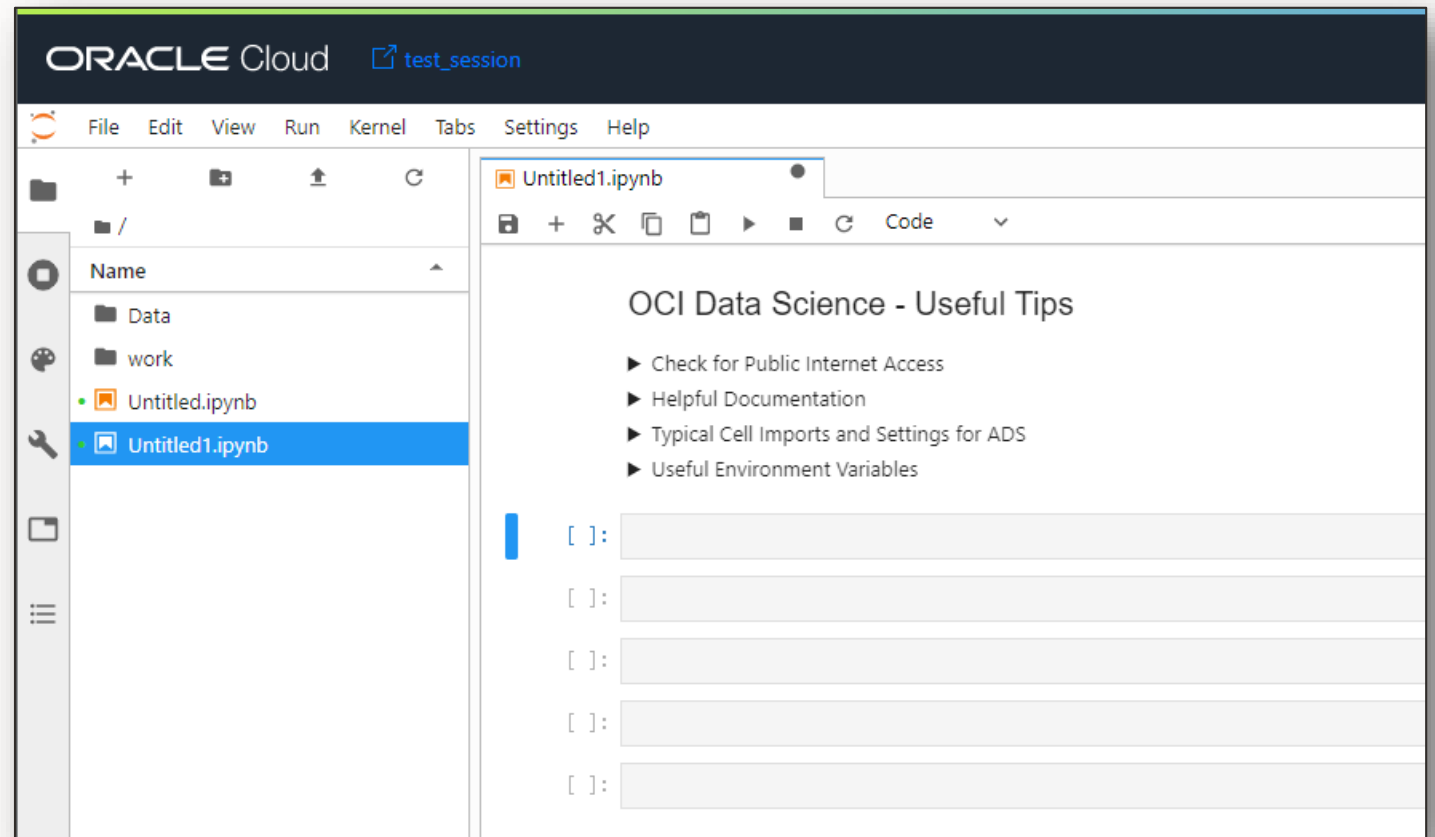


Data Science Serviceで提供されているサンプルコード





新規コードの作成



ORACLE Cloud リソース、サービスおよびドキュメントの検索 Japan East (Tokyo)

データ・サイエンス » プロジェクト » test

test

編集 削除 リソースの移動 タグの追加

プロジェクト情報 タグ

説明: test OCID: ...gtp2o3tq 表示 コピー
作成者: ...acle.com 表示 コピー 作成日: 2020年5月18日(月) 7:17:00 UTC

リソース madhacks.ochacafe コンパートメント内のモデル

モデルの作成

名前	ステータス	作成者	作成日
my_model	● アクティブ	oracleidentitycloudservice/kenichi.sonoda@oracle.com	2020年10月1日 (木) 7:30:34 UTC

1個のモデルを表示 < 1の1 >

プロジェクト内に共有された学習済みモデル

モデルカタログ

ノートブック・セッション

モデル

リスト範囲

コンパートメント

madhacks.ochacafe

ORACLE Cloud リソース、サービスおよびドキュメントの検索 Japan East (Tokyo)

データ・サイエンス » プロジェクト » test: モデル » my_model

my_model

編集 モデル・アーティファクトのダウンロード 非アクティブ化 削除 他のアクション

モデル情報 タグ

モデル情報

説明: test description

作成者: oracleidentitycloudservice/kenichi.sonoda@oracle.com

作成日: 2020年10月1日(木) 7:30:34 UTC

OCID: ...hgtdkqma4q 表示 コピー

モデル・アーティファクト: ocid1.datasciencemodel.oc1.ap-tokyo-1.amaaa...
aaassl65iqazjkii5qt5im64zf5jvuk5cm5yfrpf5df3
ファクト: phgtdkqma4q.zip (28.56MB)

モデルの来歴

GitリポジトリURL: 値なし
Gitコミット: 値なし
Gitブランチ: 値なし
モデル・ディレクトリ: 値なし
トレーニング・スクリプト: 値なし

使用条件およびプライバシー Cookieの設定

Copyright © 2020, Oracle and/or its affiliates. All rights reserved. Confidential. Internally Restricted/Highly Restricted

GitHub

Search or jump to... Pulls Issues Marketplace Explore

kenkensonson / oci-datascience-nlp-demo01

Code Issues Pull requests Actions Projects Wiki Security Insights

main Go to file Add file Code About

kenkensonson Update README.md 15 days ago 15

README.md	Update README.md	15 days ago
[Code Night] Natural L...	Add files via upload	15 days ago
bert_fine_tuning_demo...	Add files via upload	16 days ago
bert_mlm_demo.ipynb	Add files via upload	16 days ago
nlp_basic_demo.ipynb	Add files via upload	16 days ago

README.md

oci-datascience-nlp-demo01

Basic functionality of NLP Library "GINZA"

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

連携

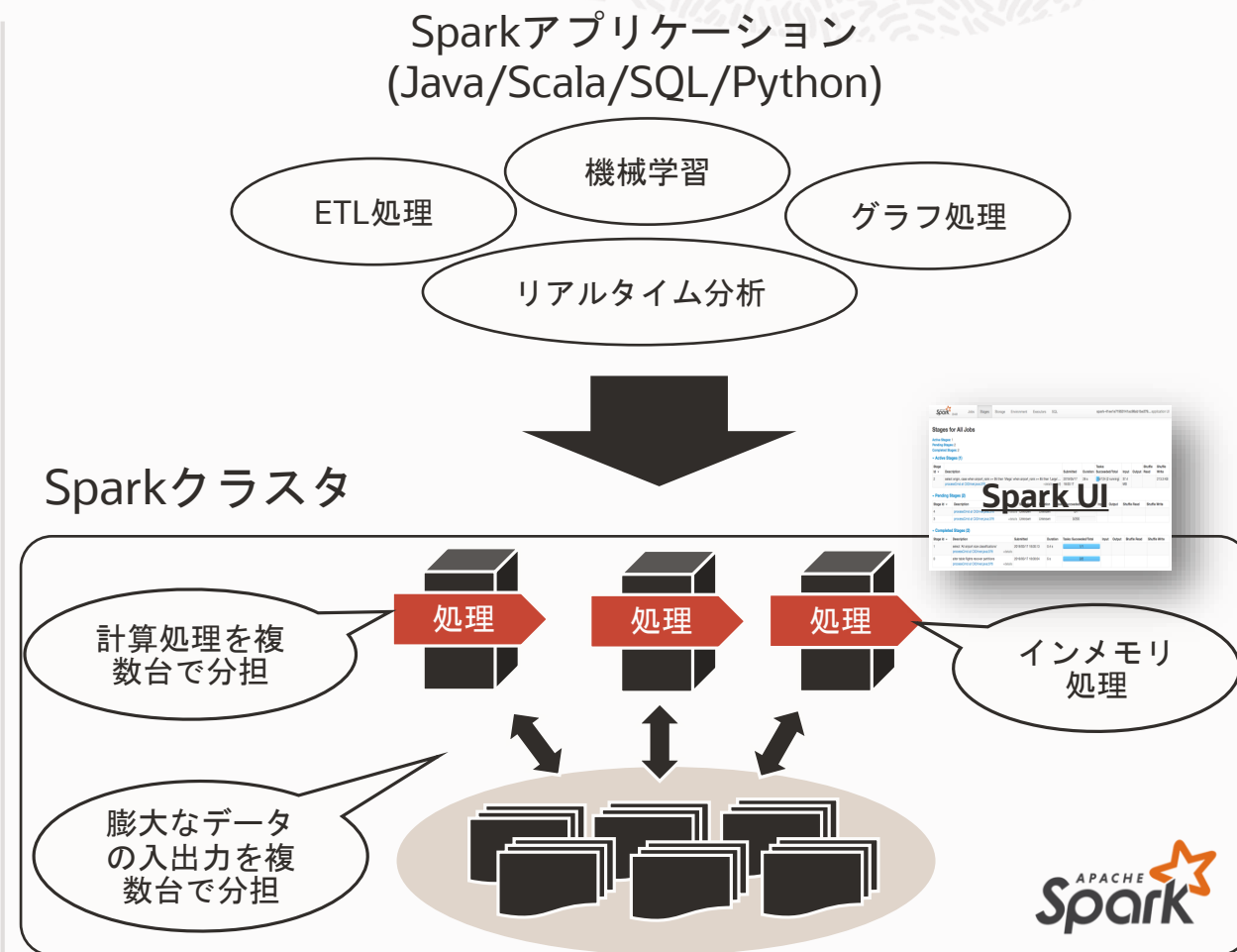




Data Flow Service

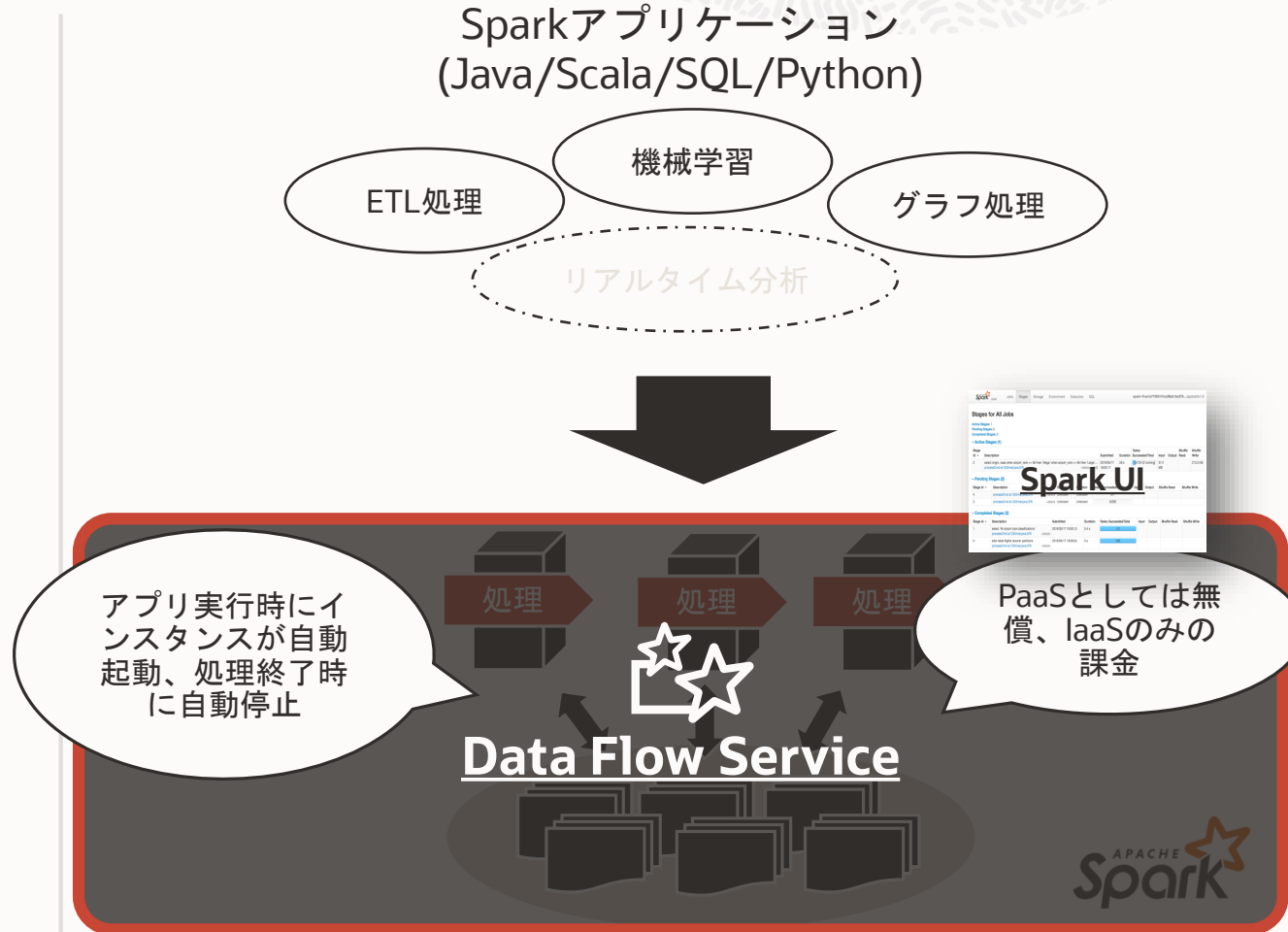
Apache Spark

- 並列分散処理のプラットフォーム
 - 大規模なデータや計算処理を複数のサーバーで分担
 - 同時並行で要件時間内に処理を完了させる
- インメモリの仕組みにより高速処理
- リソース管理ソフトウェアにより並列分散処理の複雑なコードが不要
 - タスクの分割、ノードのリソース管理、タスクスケジューリング、排他制御、ノード障害時のハンドリング、データの多重化
 - Spark Standalone、YARN、MESOS、Kubernetes
- ワークロード
 - ETL処理、リアルタイム分析、機械学習、グラフDB処理



Data Flow Service

- 複数ノードから構成されるSpark Clusterがマネージドサービスに
- Sparkクラスタ(複数ノード)のプロビジョニング、運用管理作業、ストレージ領域の容量管理、監視などが全て不要に
- アプリ実行時に、インスタンスが自動起動、アプリ終了時にインスタンスが自動停止
- PaaSとしての課金なし、IaaSのみの課金
 - アプリ実行時間のComputeインスタンスの課金
 - オブジェクトストレージの課金
- Java、Scala、SparkSQL、PySparkでのアプリ開発
- Spark Streaming、Structured Streamingは利用不可(リアルタイム分析は非推奨)
- 基本的にはバッチ処理用途に

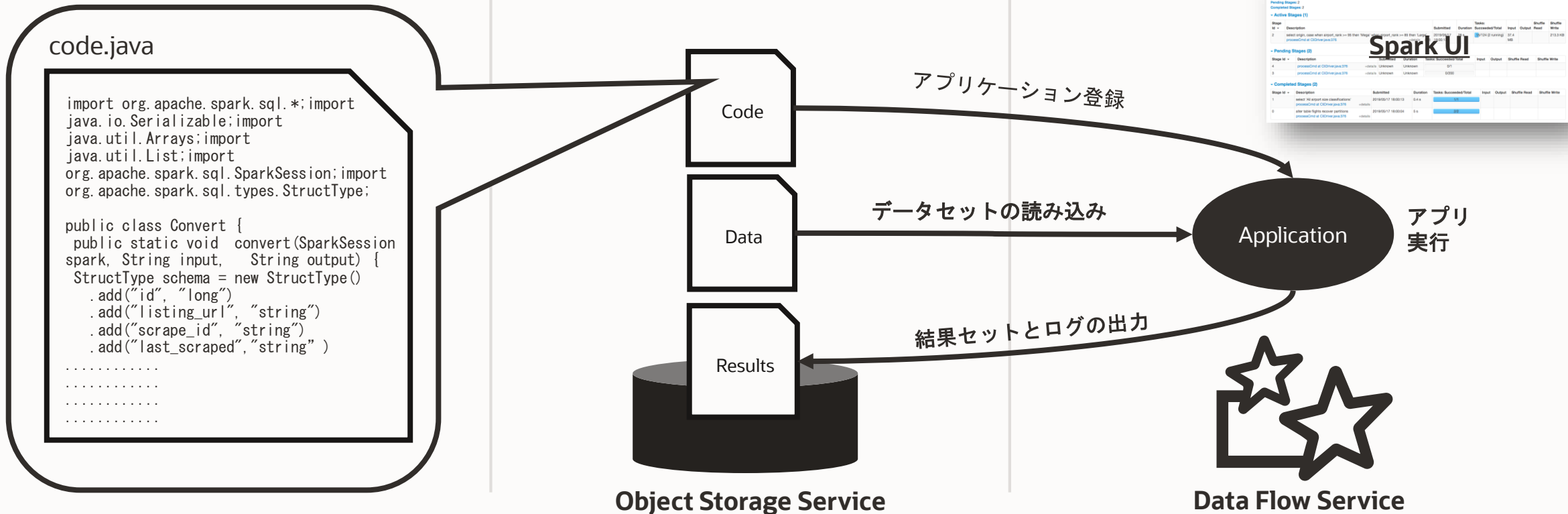


Data Flow Serviceのワークフロー

①Sparkアプリをコーディング
(Java, Scala, Python, SQL)

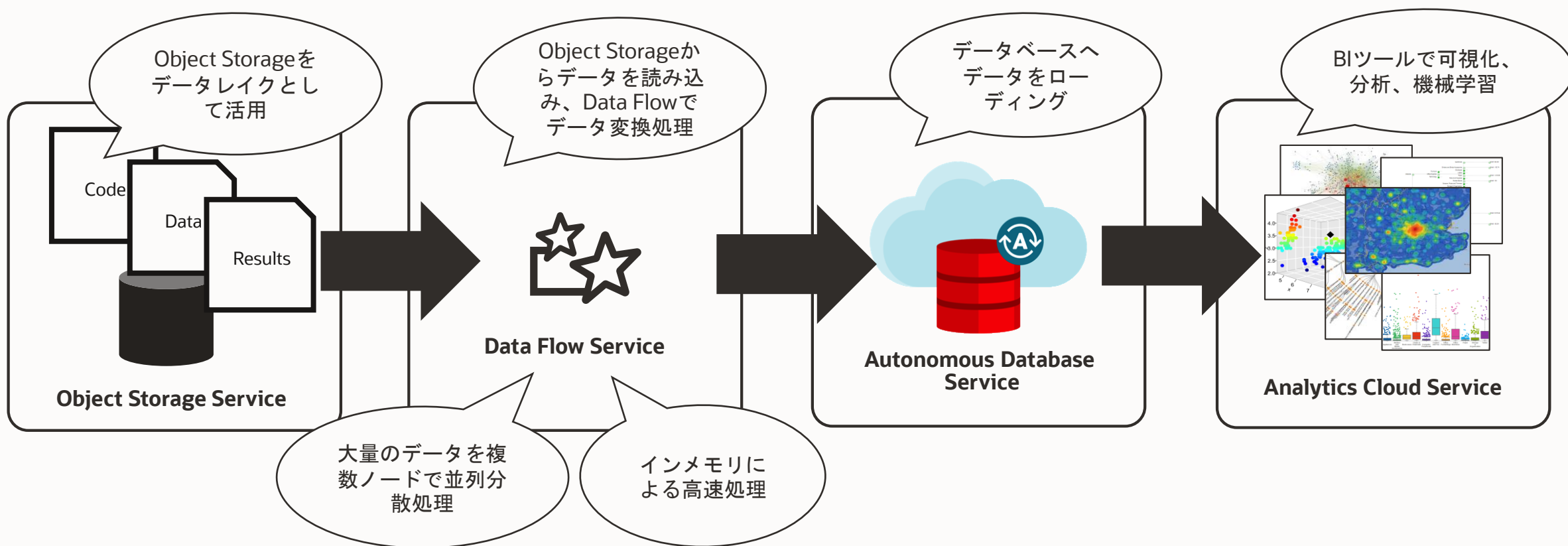
②コードのテキストファイルと
対象のデータセットをオブジェ
クトストレージにアップロード

③コードをアプリケーションと
してData Flowに登録、アプリ
実行



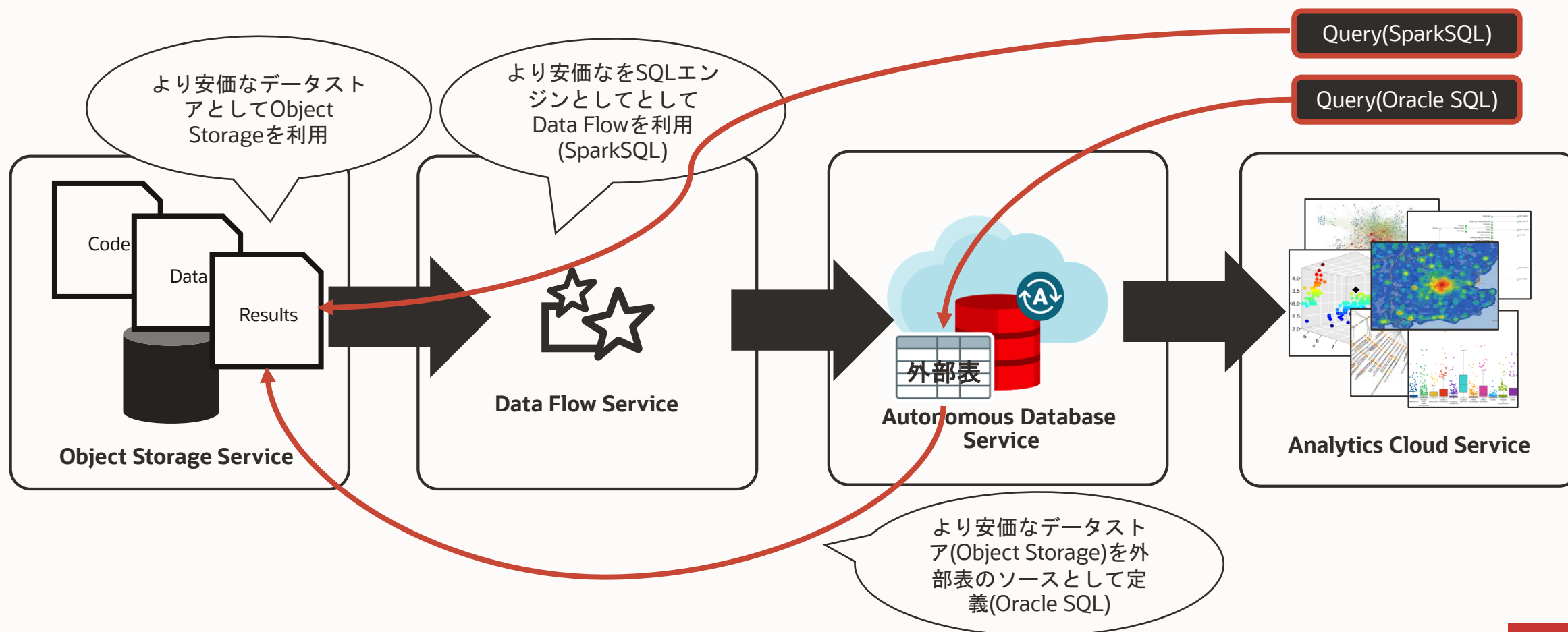
ユースケース① ETL

- データ変換とDWHへのデータロードを、インメモリ処理および複数ノードによる分散処理で高速に実行
- RDBのCPU、ストレージ課金のコストセーブに
- Object Storage / Data Flow / Autonomous Database / Analytics Cloud の連携



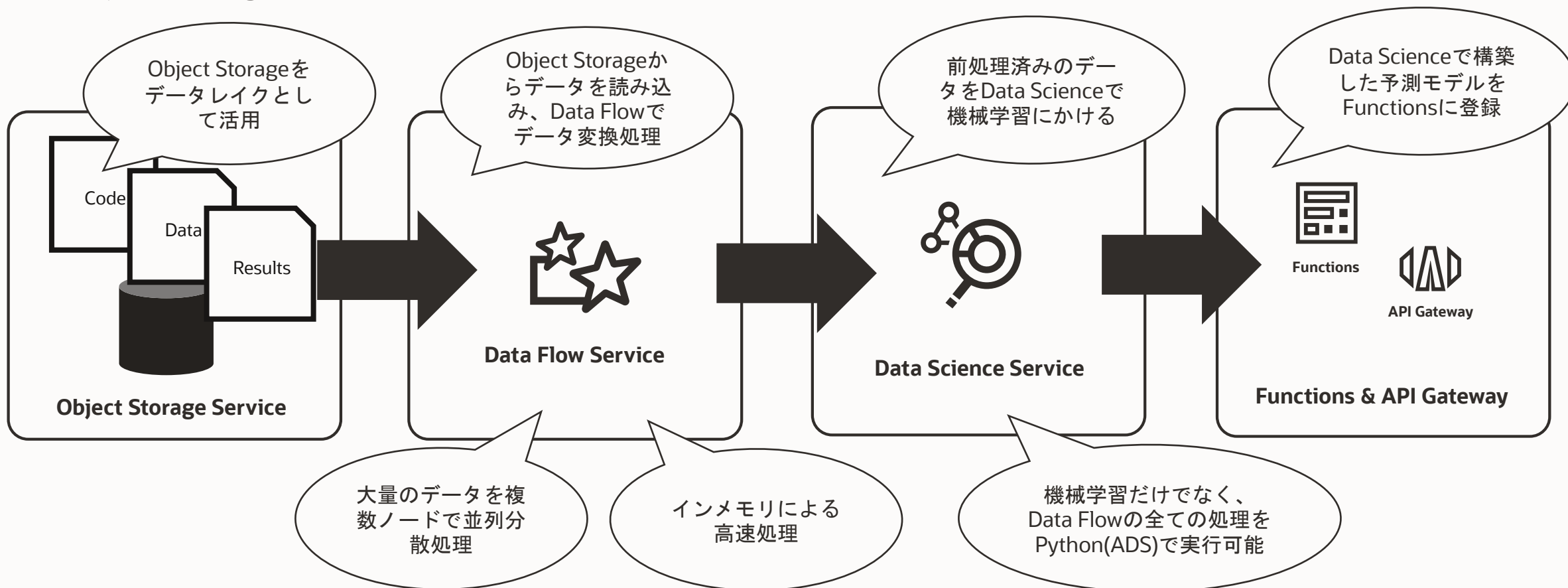
ユースケース② アクティブ・アーカイブ

- リアルタイムレスポンスが不要なクエリ処理をData Flowにオフロード
- RDBのCPU、ストレージ課金のコストセーブに



ユースケース③ 機械学習の前処理

- 機械学習のデータの前処理をData Flow Serviceにオフロード
- インメモリ処理および複数ノードによる分散処理で高速に実行
- Object Storage / Data Flow / Data Science

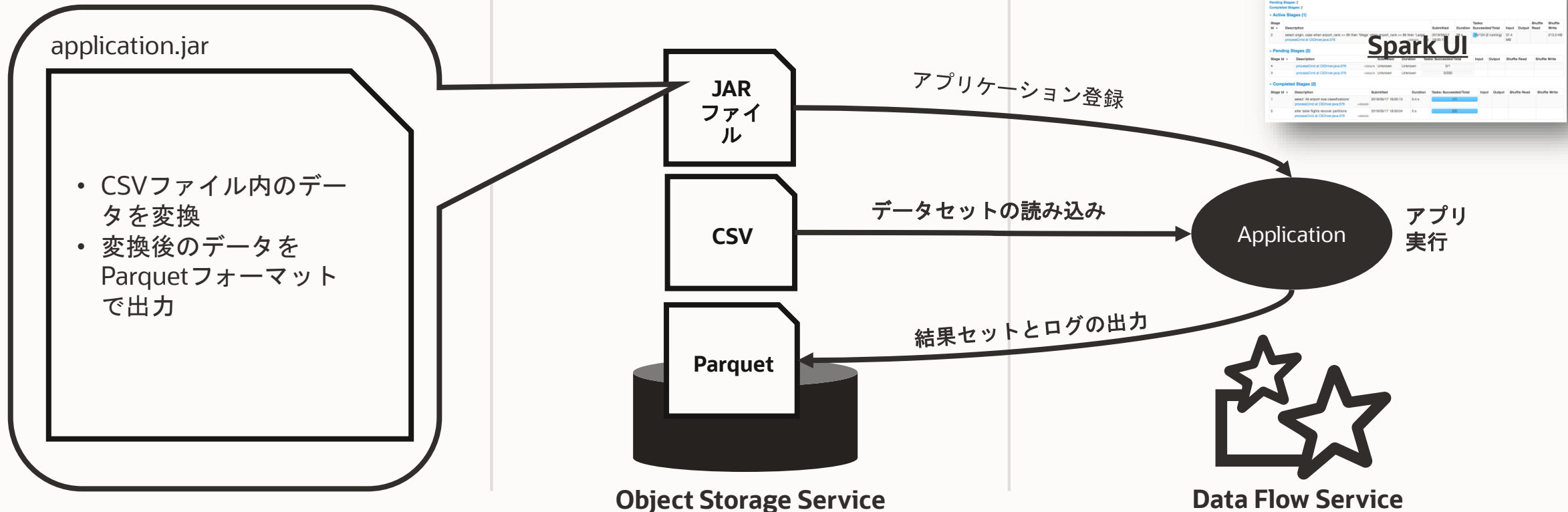


データ変換のデモンストレーション

①Sparkアプリをコーディング
(Java, Scala, Python, SQL)

②コードのファイルと対象の
データセットをオブジェクトス
トレージにアップロード

③コードをアプリケーションと
してData Flowに登録、アプリ
実行



ORACLE Cloud

リソース、サービスおよびドキュメン

Japan Central (Osaka)

データベース関連サービス

APEXアプリケーション開発

データ・セーフ

データおよびAI

ビッグ・データ

デジタル・アシスタント

データ・カタログ

データ・フロー

データ統合

サービス・コネクタ・ハブ

データ・サイエンス

ソリューションおよびプラットフォーム

アナリティクス

リソース・マネージャ

電子メール配信

アプリケーション統合

アプリケーション

実行

プライベート・エンドポイント

Oracle Universityによる
FREEトレーニングを開始

トレーニングとドキュメント

Oracle Cloudは初めてでも、使用経験があっても、学習
行程をここから開始または続行します。Explorerバッジ
や認定資格を取得してください。

すべてのシステムが動作
可能
ヘルス・ダッシュボードの表
示

アカウント・センター

ユーザー管理
テナンシへのユーザーの追加

請求
コストの分析
支払方法の管理

最新情報

Exadata DBシステムでのData Guardの
トラブルシューティング
2020年12月5日

ファンクションの同時実行のためのデフ
ォルト合計メモリーの増加
2020年12月3日

Exadata Cloud Service: Oracle
Database 19cのアップグレード機能が
使用可能に



ホーム

アプリケーション

実行

Private Endpoints

リスト範囲

コンパートメント

madhacks.ochacafe ▾

orasejapan (ルート)/project/madhacks.ochacafe

タグ・フィルタ

[追加](#) | [クリア](#)

タグ・フィルタは適用されていません

madhacks.ochacafe コンパートメント内のアプリケーション

Data Flow lets you run Apache Spark jobs at any scale with almost no administration. [Learn more](#)[Create Application](#)

名前

言語

所有者

作成日

最終更新

アイテムが見つかりませんでした。

0アイテムを表示中 < ページ1 >



ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka)

ホーム

アプリケーション

実行

Private Endpoints

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

タグ・フィルタ

追加 | クリア

タグ・フィルタは適用されていません

フィルタ

言語

All

Updated Start Date

Updated End Date

使用条件およびプライバシー

Cookieの設定

madhacks

Data Flow

Create Application

名前

Sparkアプリケーションの作成

言語

☒ Java

☐ Python

☐ SQL

☐ Scala

一般情報

名前

ETL

説明 オプション

ETL

リソース構成

Sparkバージョン

Spark 2.4.4

ドライバ・タイプ

VM.Standard2.1 (15 GB Memory, 1 OCPU, 175 GB Block Volume)

エグゼキュータ・タイプ

作成

取消

アプリケーションに任意の名前を指定

ORACLE Cloud リソース、サービスおよびドキュメントの検索 Japan Central (Osaka)

Sparkアプリケーションの作成

リソース構成

Sparkバージョン
Spark 2.4.4

ドライバ・タイプ ⓘ
VM.Standard2.1 (15 GB Memory, 1 OCPU, 175 GB Block Volume)

エグゼキュータ・タイプ ⓘ
VM.Standard2.1 (15 GB Memory, 1 OCPU, 175 GB Block Volume)

エグゼキュータの数 ⓘ
1

アプリケーション構成

ファイルURL ⓘ
ow_2019_dataflow_lab@bigdatadatasciencelarge/usercontent/ow-lab-2019-java-etl-1.0-SNAPSHOT.jar

メイン・クラス名 ⓘ
convert.Convert

引数 オプション ⓘ
\${input} \${output}

作成 取消

SparkのWorkerノードの数

オブジェクトストレージ上のJARファイル(アプリケーションの実体)

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka)

ホーム

アプリケーション

実行

Private Endpoints

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

タグ・フィルタ

追加 | クリア

タグ・フィルタは適用されていません

フィルタ

言語

All

Updated Start Date

Updated End Date

madhacks

Data Flow

Create Application

名前

Sparkアプリケーションの作成

ヘルプ

引数 オプション ⓘ

`${input} ${output}`

Archive URI オプション ⓘ

パラメータ オプション ⓘ

名前

デフォルト値

input

C

tent/kaggle_berlin_airbnb_listings_summary.csv

output

C

oci://dataflow-data@orasejapan/optimized_listing

タグ

タグ付けとは、テナンシ内のリソースを整理およびトラッキングできるメタデータは、リソースにアタッチできるキーと値から構成されます。

タグ付けの詳細

タグ・ネームスペース

タグ・キー

値

なし(フリーフォーム・タグ...)

×

+ 追加タグ

Show Advanced Options

作成

取消

処理対象のCSVファイル

変換処理後のParquetファイル

ORACLE Cloud リソース、サービスおよびドキュメントの検索 Japan Central (Osaka)

madhacks.ochacafe コンパートメント内のアプリケーション

Data Flow lets you run Apache Spark jobs at any scale with almost no administration. [Learn more](#)

Create Application

名前	言語	所有者	作成日	最終更新
ETL	Java	kenichi.sonoda@oracle.com	Thu, Dec 17, 2020, 07:42:26 UTC	Thu, Dec 17, 2020, 07:42:26 UTC

1アイテムを表示中 < ページ1 >

ホーム

アプリケーション

実行

Private Endpoints

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

アプリケーションの完成

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka)

ホーム

アプリケーション

実行

Private Endpoints

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

タグ・フィルタ

追加 | クリア

タグ・フィルタは適用されていません

フィルタ

madhacks.ochacafe コンパートメント内のアプリケーション

Data Flow lets you run Apache Spark jobs at any scale with almost no administration. [Learn more](#)

Create Application

名前	言語	所有者	作成日	最終更新
ETL	Java	kenichi.sonoda@orac le.com	Thu, Dec 17, 2020, 07:42:26 UTC	<div>View Details</div> <div>編集</div> <div>実行</div>

1アイテムを

アプリケーションの実行

リソースの移動

削除

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka)

ホーム

アプリケーション

実行

Private Endpoints

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

タグ・フィルタ

追加 | クリア

madhacks.ochacafe コンパートメント内の実行

名前	言語	ステータス	所有者	作成日	期間	OCPU 合計	データ読 取り	データ書 込み
ETL	Java	● 成功	kenichi.sonoda @oracle.com	Thu, Dec 17, 2020, 07:44:48 UTC	1分 50 秒	2	68 MB	31 MB
TestApp	Java	● 成功	kenichi.sonoda @oracle.com	Thu, Dec 17, 2020, 07:28:55 UTC	1分 54 秒	2	68 MB	31 MB

2アイテムを表示中 < ページ1 >

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka)

ホーム

アプリケーション

実行

Private Endpoints

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

タグ・フィルタ

追加 | クリア

madhacks.ochacafe コンパートメント内の実行

名前	言語	ステータス	所有者	作成日	期間	OCPU 合計	データ読取り	データ書き込み
ETL	Java	● 成功	kenichi.sonoda@oracle.com	Thu, Dec 17, 2020, 07:44:48 UTC	1分 50 秒	2		
TestApp	Java	● 成功	kenichi.sonoda@oracle.com	Thu, Dec 17, 2020, 07:28:55 UTC	1分 54 秒	2		
2アイテムを表示								

詳細

Spark UI

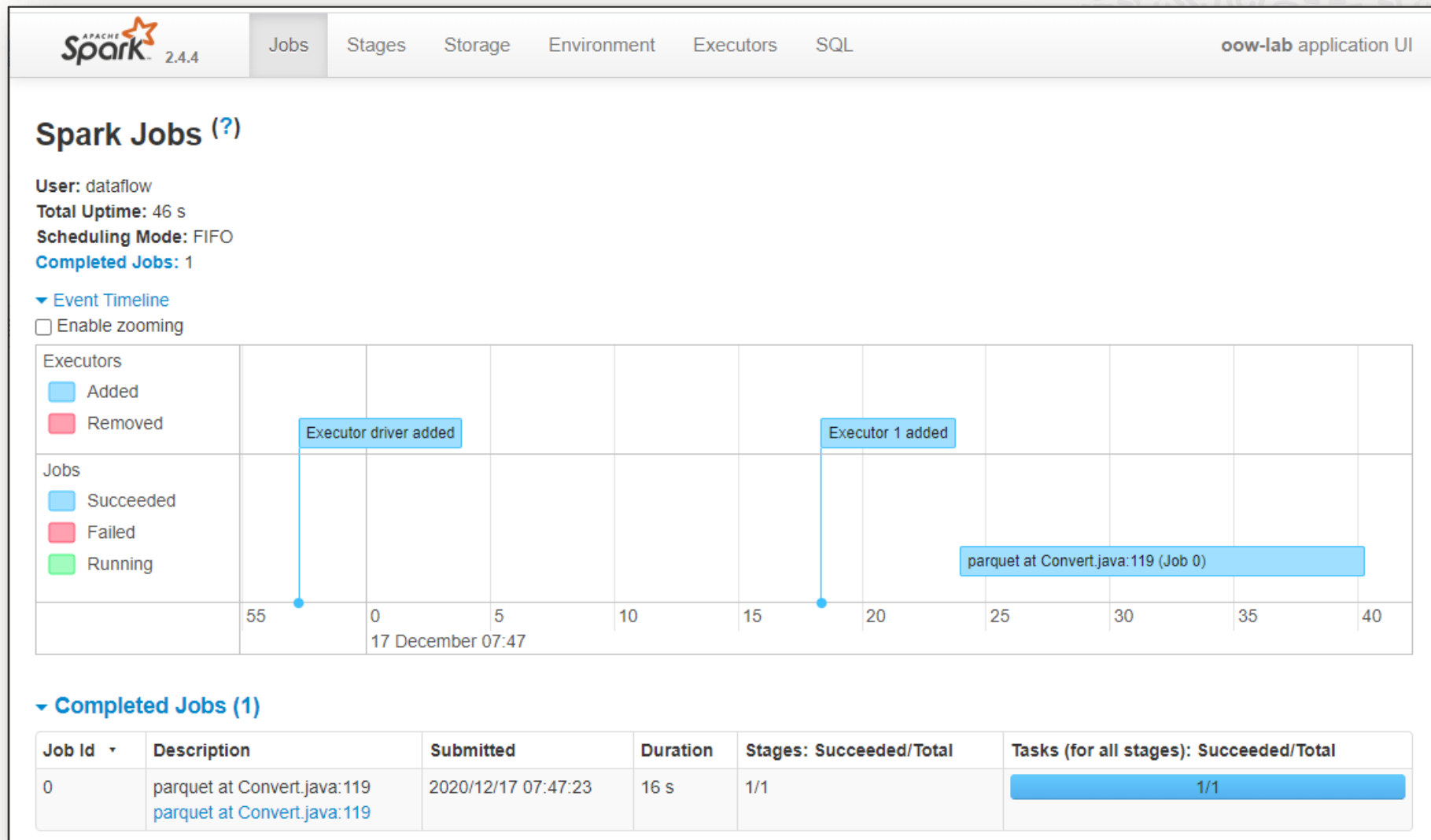
再実行

タグの追加

タグの表示

リソースの移動





ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka)

B

バケット情報

タグ

可視性: プライベート

暗号化キー: Oracle管理キー 割当て

ネームスペース: orasejapan

作成日: 2020年12月17日(木) 7:38:19 UTC

デフォルト・ストレージ層: 標準

コンパートメント: [madhacks.ochacafe](#)

近似カウント: 5オブジェクト

近似サイズ: 30.52MiB

ETag: 7a15340e-04eb-4a03-a96f-7e5ea9cf9288

オブジェクト・イベントの出力: ● 無効 編集

OCID: ...qhftvofa 表示 コピー

オブジェクト・バージョンング: ● 無効 編集

リソース

オブジェクト

メトリック

事前認証済リクエスト

作業リクエスト

ライフサイクル・ポリシー・ルール

レプリケーション・ポリシー

保持ルール

ログ

アップロード

他のアクション

接頭辞で検索

	名前	最終変更	サイズ	ステータス
	optimized_listings	-		
	_SUCCESS	20:		
	part-00000-262ab54d-4429-4807-9b27-b88f76c4afcc-c000.snappy.parquet	2020年12月17日(木) 7:47:40 UTC	30.52MiB	使用可能

オブジェクトストレージ上に出力された変換後のParquetファイル

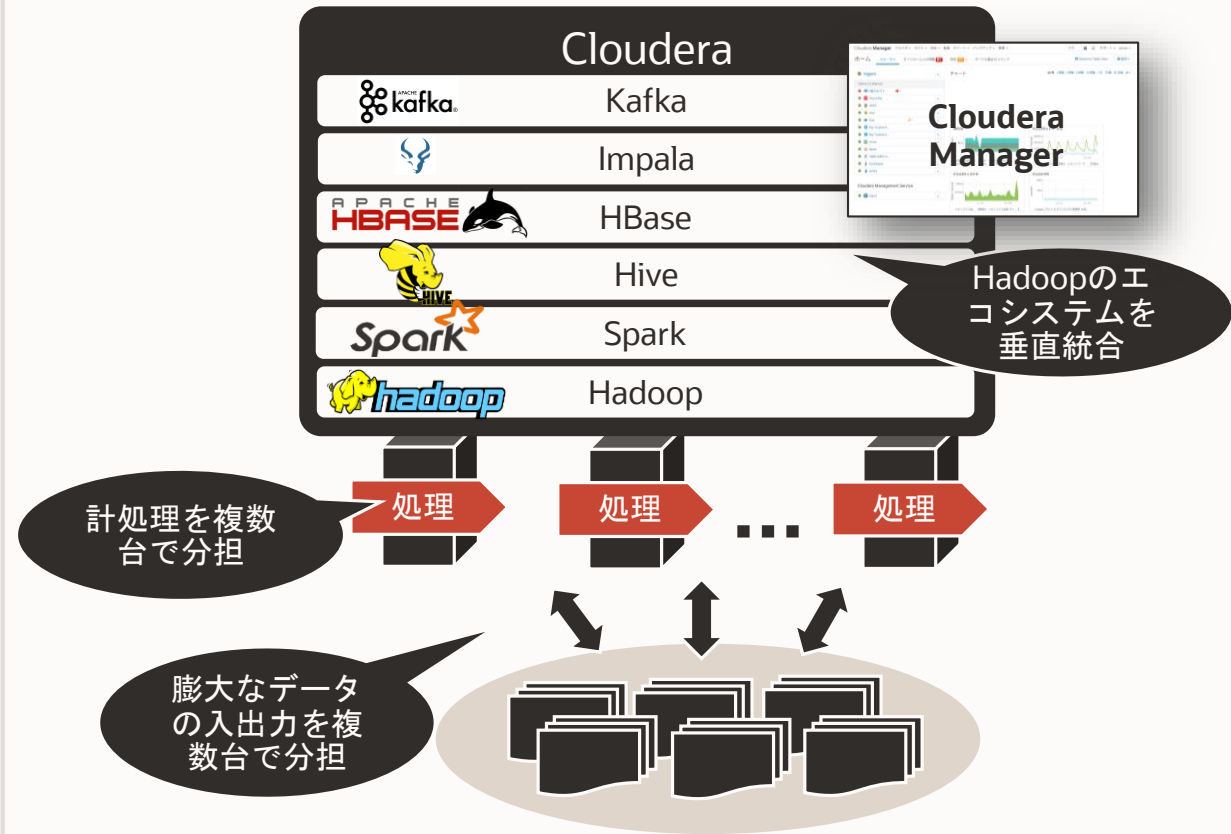


Big Data Service

Cloudera Enterprise Data Hub

- Cloudera社から提供されているHadoopのディストリビューション
- Apache Hadoopとそのエコシステムによるデータレイク、分析システム基盤
 - データレイク
 - データウェアハウス
 - リアルタイム分析
 - Business Intelligence
 - ETL
 - 機械学習
- Hadoop周辺のエコシステムを複数のノードに垂直統合
- 大規模な分散処理基盤における、ソフトウェアスタック、ハードウェアスタックを統合管理

Clouderaクラスタの構成

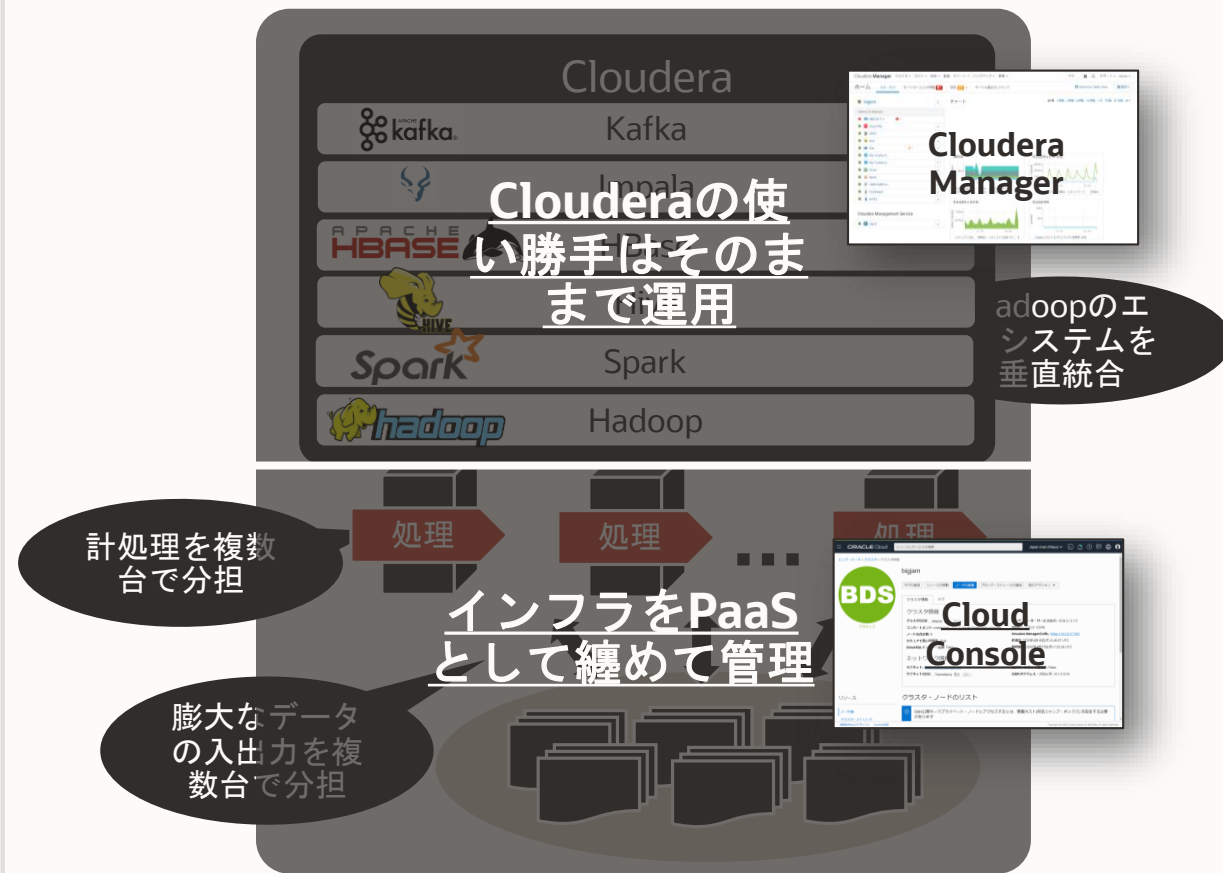


Big Data Service

- Clouderaのエコシステムをフル・マネージドで提供するPaaS
- 任意のサイズのクラスタを容易にプロビジョニング、運用、管理、構成変更
- Clouderaのベストプラクティスに沿った高可用性クラスタを容易に構築
- Cloudera ManagerなどClouderaの使い勝手はそのままに
- Cloud SQL
 - Hadoop領域にOracleのSQLでアクセス可能

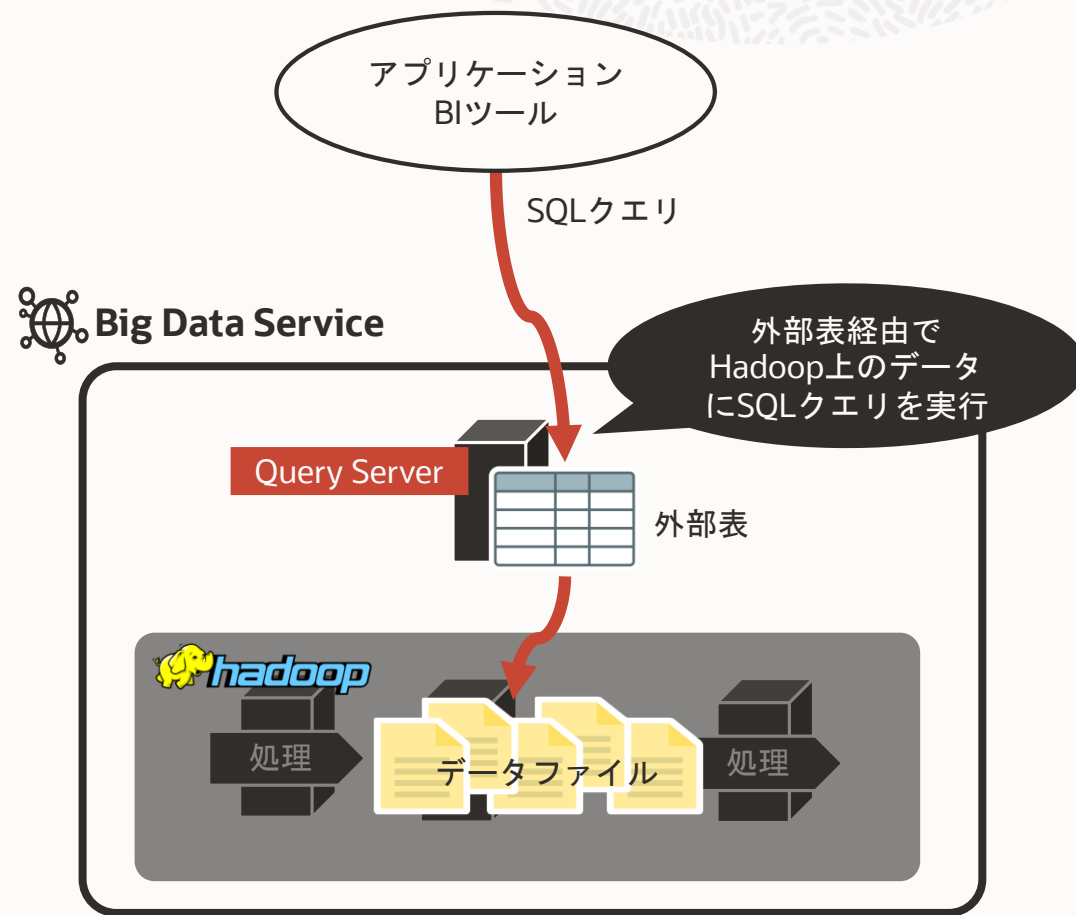


Big Data Service



Cloud SQL

- Hadoop上のデータにOracleのSQLクエリを実行
- クライアントアプリからQuery Server上の外部表を経由し、様々なデータソースにOracleのSQLでアクセス可能
- 対応データソース
 - HDFS、Hive、Kafka、HBase、Oracle NoSQL DB
- アプリケーションからQuery ServerへJDBC/SQL*Net接続
- Query Server
 - Oracle Database 18c互換クエリエンジン
 - 表統計、v\$ビューなど利用可



Oracle Cloud

リソース、サービスおよびドキュメントの検索

Japan East (Tokyo)

test_session

ノートブック・セッション情報

OCID: ...jkzs3mfa 表示 コピー

作成者: ...acle.com 表示 コピー

ブロック・ストレージ・サイズ(GB): 1,000 GB

サブネット: [Public Subnet-bigdata-vcn](#)

作成日: 2020年12月17日(木) 1:29:27 UTC

コンピュート・インスタンスのシェイプ: VM.GPU3.1

VCN: [bigdata-vcn](#)

メトリック

メトリックは、新規に作成またはアクティブ化されたノートブック・セッションに対してのみ使用可能です。新規ノートブック・セッションを作成するか、これを非アクティブ化してから再アクティブ化して、メトリックを有効にしてください。新規ノートブック・セッションのメトリックが表示されない場合は、メトリックを読み取る権限があることを確認します。 [詳細](#)

終了時間

クイック選択

2020/12/17 0:13:02

2020/12/17 1:13:02

過去1時間

チャートのリセット

CPU使用率

オプション

間隔 1分

統計 平均

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka)

ビッグ・データ

クラスタ

クラスタ

リスト範囲

コンパートメント

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

フィルタ

状態

任意の状態

タグ・フィルタ

追加

クリア

タグ・フィルタは適用されていません

クラスタ要件: [Big Data Serviceの準備](#)

[詳細情報の表示](#)

クラスタの作成

フィルタするクラスタ

名前	状態	ノード数	セキュアで高い可用性	作成日
クラスタが見つかりません				

0アイテムを表示中 < 1の1 >

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan Central (Osaka)

ビッグ・データ

クラスタ

リスト範囲

状態

任意の状態

タグ・フィルタ

追加

クリア

タグ・フィルタは適用されていません

クラスタの作成

Oracle Big Data Serviceを使用して、Cloudera Distribution Including Apache Hadoop (CDH)をベースとしたHadoopクラスタを作成します。クラスタを作成するには、インスタンス・シェイプとストレージを選択し、顧客およびクラスタ・プライベート・ネットワークを選択します。ワンクリックで、クラスタの可用性および安全性を高められます。

Hadoopノード

Master/Utilityノード

インスタンス・タイプの選択

仮想マシン

仮想マシンは、物理的なベア・メタル・ハードウェアの上位で稼働する、独立したコンピューティング環境です

ベア・メタル

ベア・メタル・コンピュータ・インスタンスによって、最高のパフォーマンスと強力な分離を備えた専用の物理サーバーにアクセスできます

Master/Utilityノード・シェイプの選択

VM.Standard2.4

Master/Utilityノード当たりのブロック・ストレージのサイズ(GB単位)

1000

50GBの増分で150GBから48TBの間。

Master & Utility ノードの数 読み取り専用

クラスタの作成

取消

- Master/Utilityノードのスペック
- Workerノードのスペックとノード数
- ストレージ容量
- その他

ORACLE Cloud

リソース、サービスおよびドキュメントの検索

Japan East (Tokyo)



ビッグ・データ

クラスタ

クラスタ

リスト範囲

コンパートメント

フィルタ

状態

タグ・フィルタ

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

任意の状態

追加 | クリア

タグ・フィルタは適用されていません

クラスタ要件: [Big Data Serviceの準備](#)

[詳細情報の表示](#)

クラスタの作成

フィルタするクラスタ

名前	状態	ノード数	セキュアで高い可用性	作成日
ks-cluster	アクティブ	6	False	2020年12月8日(火) 4:08:04 UTC

1アイテムを表示中 < 1の1 >



ORACLE Cloud リソース、サービスおよびドキュメントの検索 Japan East (Tokyo)

ビッグ・データ » クラスタ » クラスタ詳細

ks-cluster

ノードの追加 ブロック・ストレージの追加 Cloud SQLの追加 シェイプの変更 他のアクション

クラスタ情報 Cloud SQL情報 タグ

クラスタOCID: ...hzo3pa 表示 コピー

顧客ネットワーク情報 サブネット: Public Subnet-bigdata-vc

ノードの追加

ks-clusterクラスタにワーカー・ノードを追加します。新しいノード・インスタンス・シェイプとアタッチされたブロック・ストレージの量は既存のワーカー・ノードと同じです。

ノード・シェイプ 読取り専用

VM.Standard2.1

ノード・シェイプは既存のノードと同じである必要があります。

ノード当たりのブロック・ストレージ(GB単位) 読取り専用

150

ワーカー・ノードの数

1

クラスタ管理パスワード

追加 取消

ブロック・ストレージの追加

ブロック・ストレージをワーカー・ノードから削除することはできません。

既存の各ワーカー・ノードにストレージを追加します。すでにワーカー・ノードにアタッチされているストレージに影響はありません。

ノード当たりの現在のブロック・ストレージ(GB単位): 150

ノード当たりの追加のブロック・ストレージ(GB単位)

1000

50GBの増分で150GBから48TBの間。

クラスタ管理パスワード

追加 取消

Cloud SQLの追加

Cloud SQLは、別途課金されるアイテムです。

クラスタにCloud SQLサポートを追加します。この操作により、クラスタに問合せサーバー・ノードが追加され、すべてのワーカー・ノードにビッグ・データ・セル・サーバーが作成されます。

問合せサーバーのノード・シェイプ

VM.Standard2.4

問合せサーバー・ノードのブロック・ストレージ(GB単位)

2000

50GBの増分で1TBから48TBの間。

クラスタ管理パスワード

追加 取消

シェイプの変更

シェイプを変更するノードのタイプを選択してから、新規シェイプを選択します。選択したタイプのすべてのノードが新規シェイプで更新されます。

標準シェイプでのみノードをリシェイプできます。

ノード・タイプの選択

MASTER / UTILITY

既存のシェイプ: VM.Standard2.4

新規シェイプ

VM.Standard2.8

VM.Standard2.8

VM.Standard2.16

VM.Standard2.24

シェイプの変更 取消

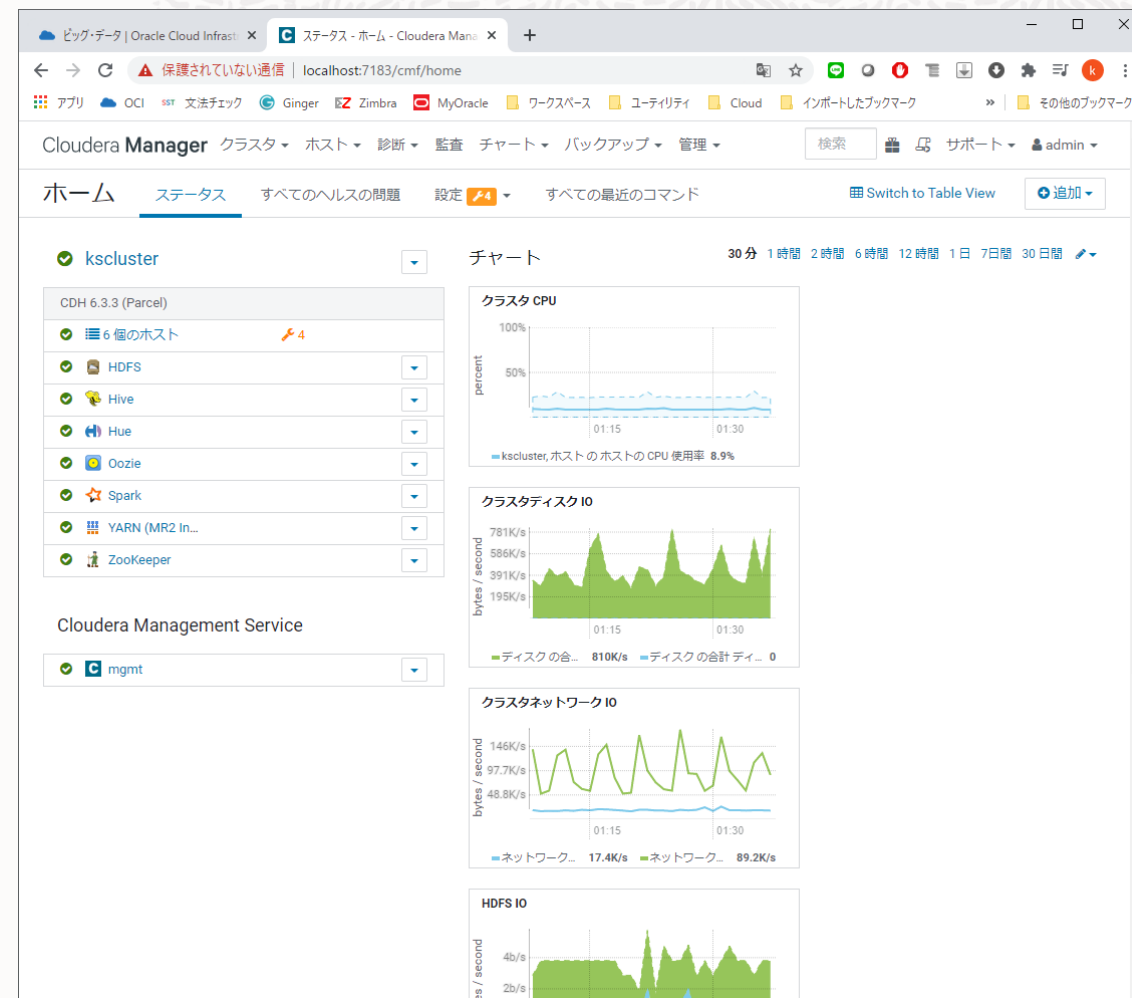


Utilityノードに接続してCloudera Managerにログイン

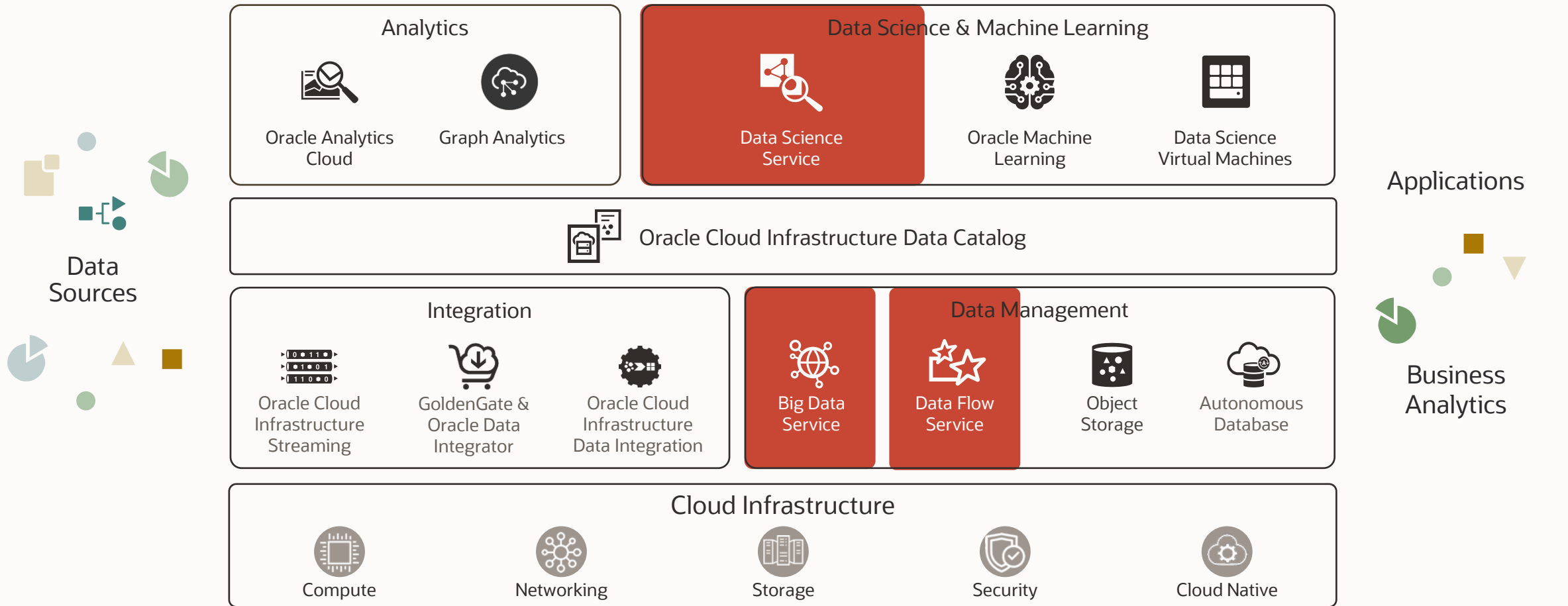
Cloudera Manager

サポートポータル ヘルプ

☐ 資格情報の保存



Oracle Data and AI Platform



ORACLE

