

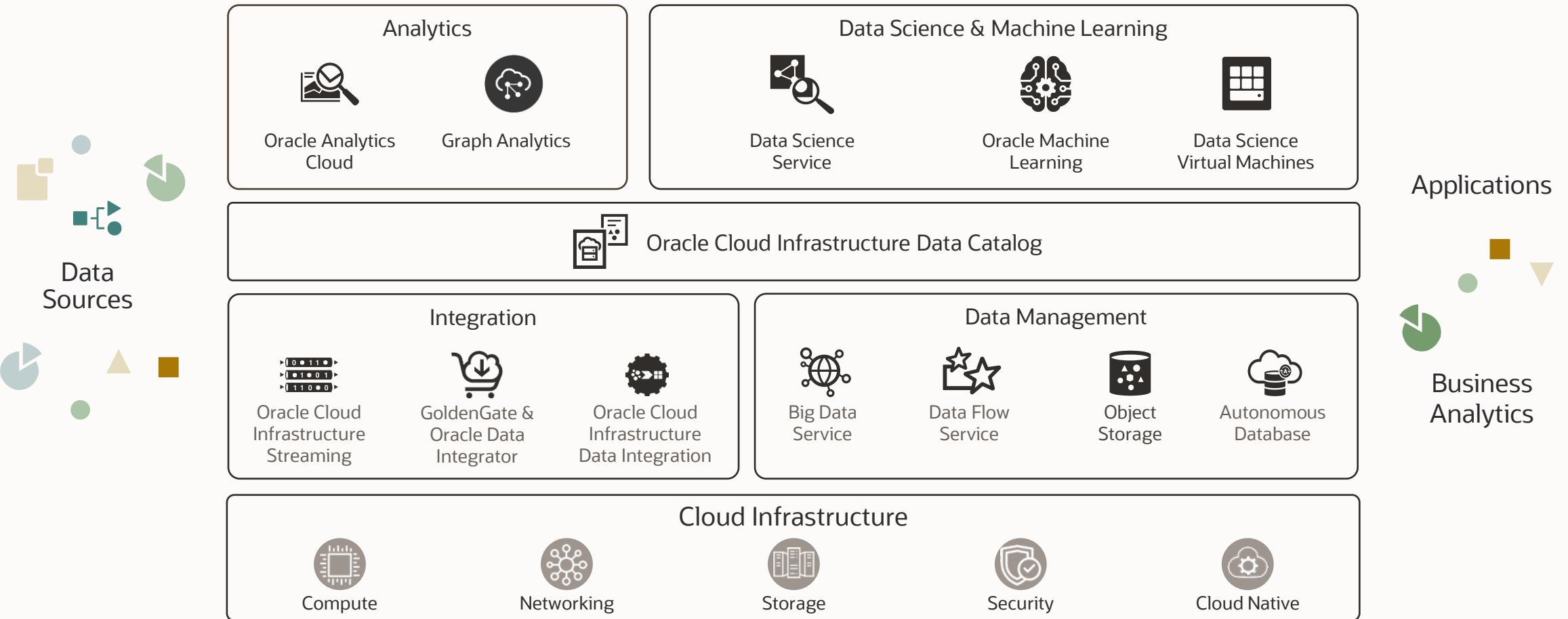


総まとめ！オラクルの機械学習 /Big Data系クラウドサービス

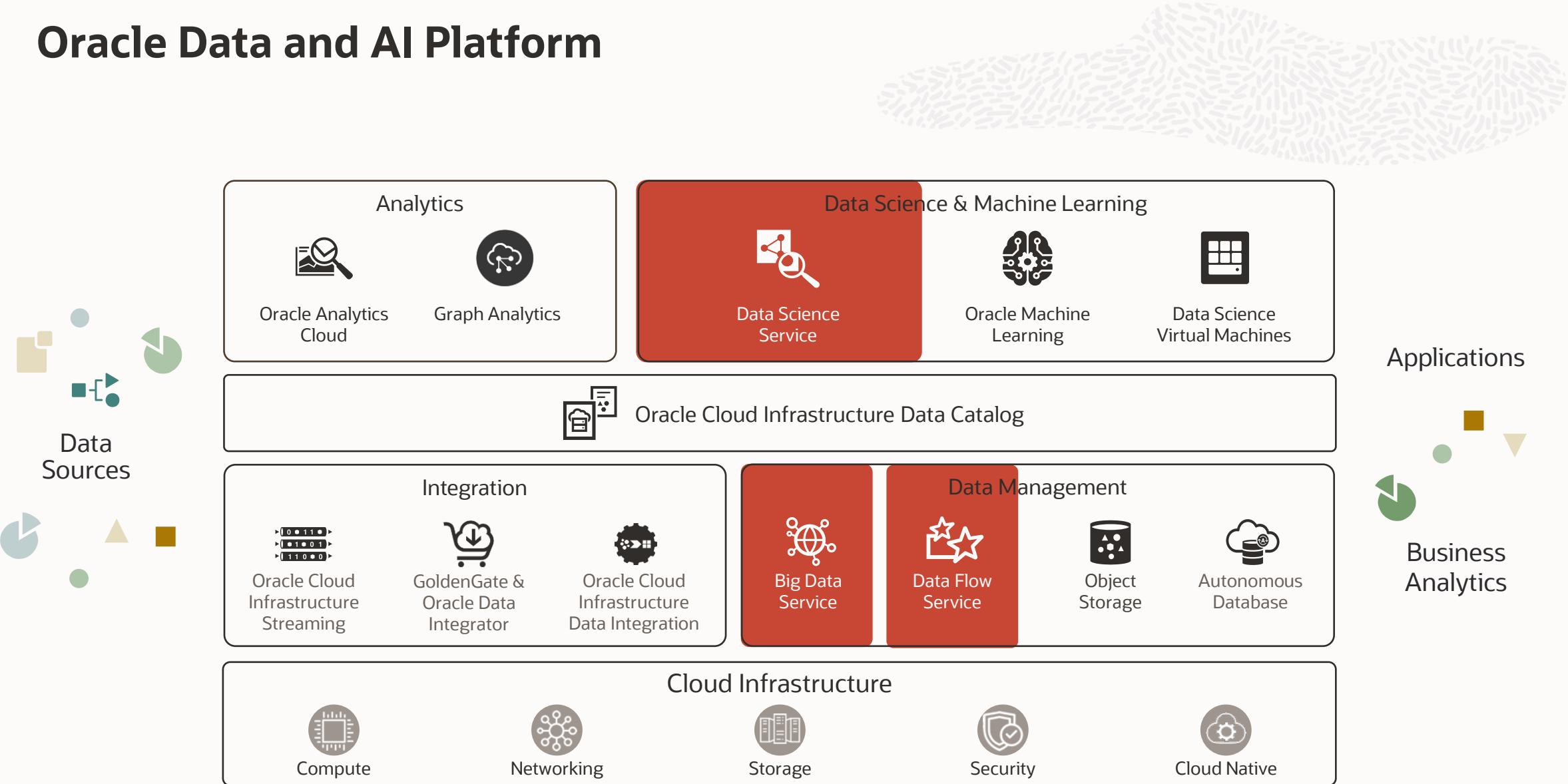
2020年12月18日

—
園田憲一
日本オラクル株式会社

Oracle Data and AI Platform



Oracle Data and AI Platform

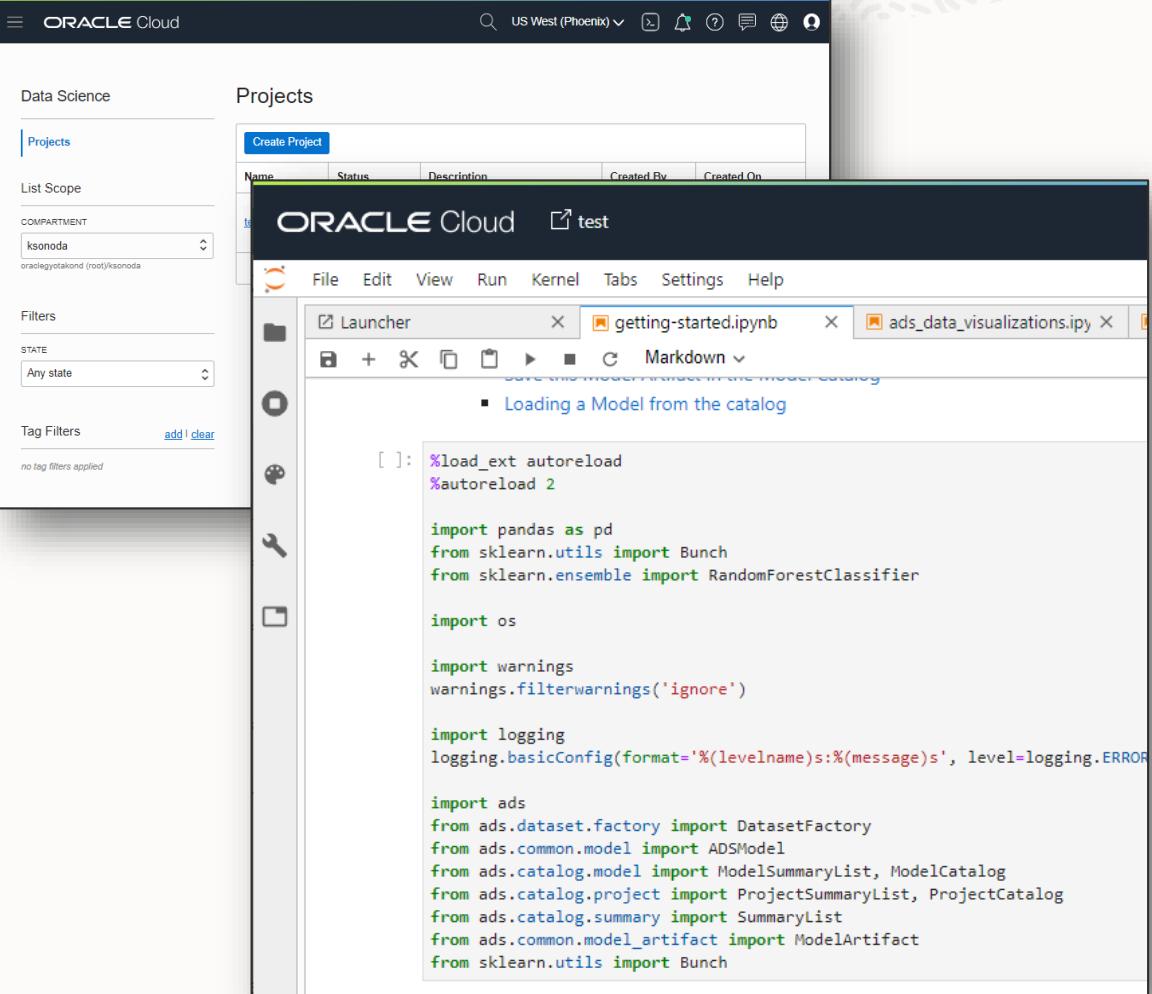




Data Science Service

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



The screenshot shows the Oracle Cloud Data Science service interface. On the left, there is a sidebar with 'Data Science' and 'Projects' tabs, and filters for 'List Scope', 'COMPARTMENT' (set to 'ksonoda'), 'Filters' (set to 'Any state'), and 'Tag Filters'. The main area is a Jupyter Notebook interface titled 'test'. It shows a code cell with the following Python code:

```
%load_ext autoreload
%autoreload 2

import pandas as pd
from sklearn.utils import Bunch
from sklearn.ensemble import RandomForestClassifier

import os

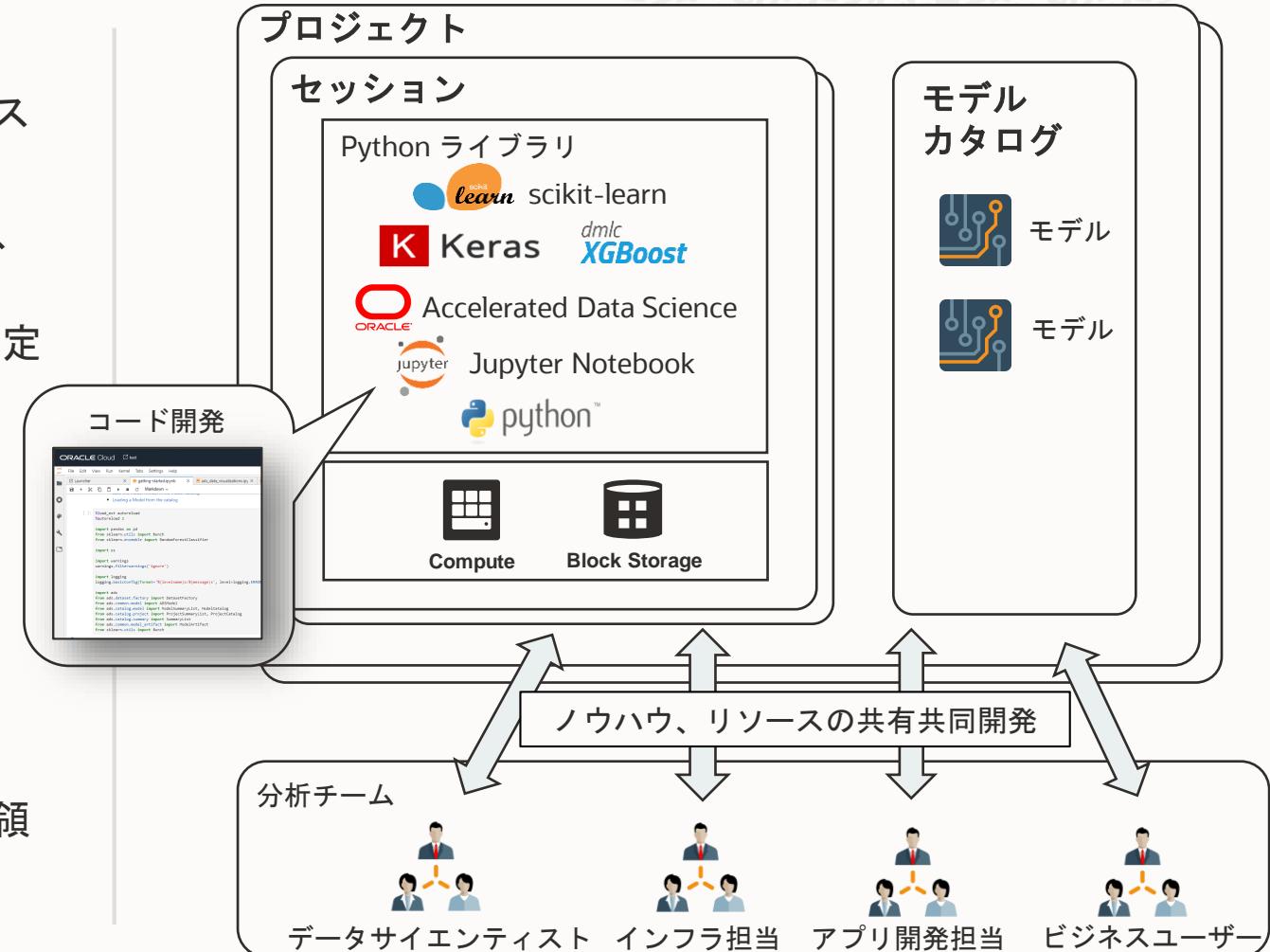
import warnings
warnings.filterwarnings('ignore')

import logging
logging.basicConfig(format='%(levelname)s:%(message)s', level=logging.ERROR)

import ads
from ads.dataset.factory import DatasetFactory
from ads.common.model import ADSModel
from ads.catalog.model import ModelSummaryList, ModelCatalog
from ads.catalog.project import ProjectSummaryList, ProjectCatalog
from ads.catalog.summary import SummaryList
from ads.common.model_artifact import ModelArtifact
from sklearn.utils import Bunch
```

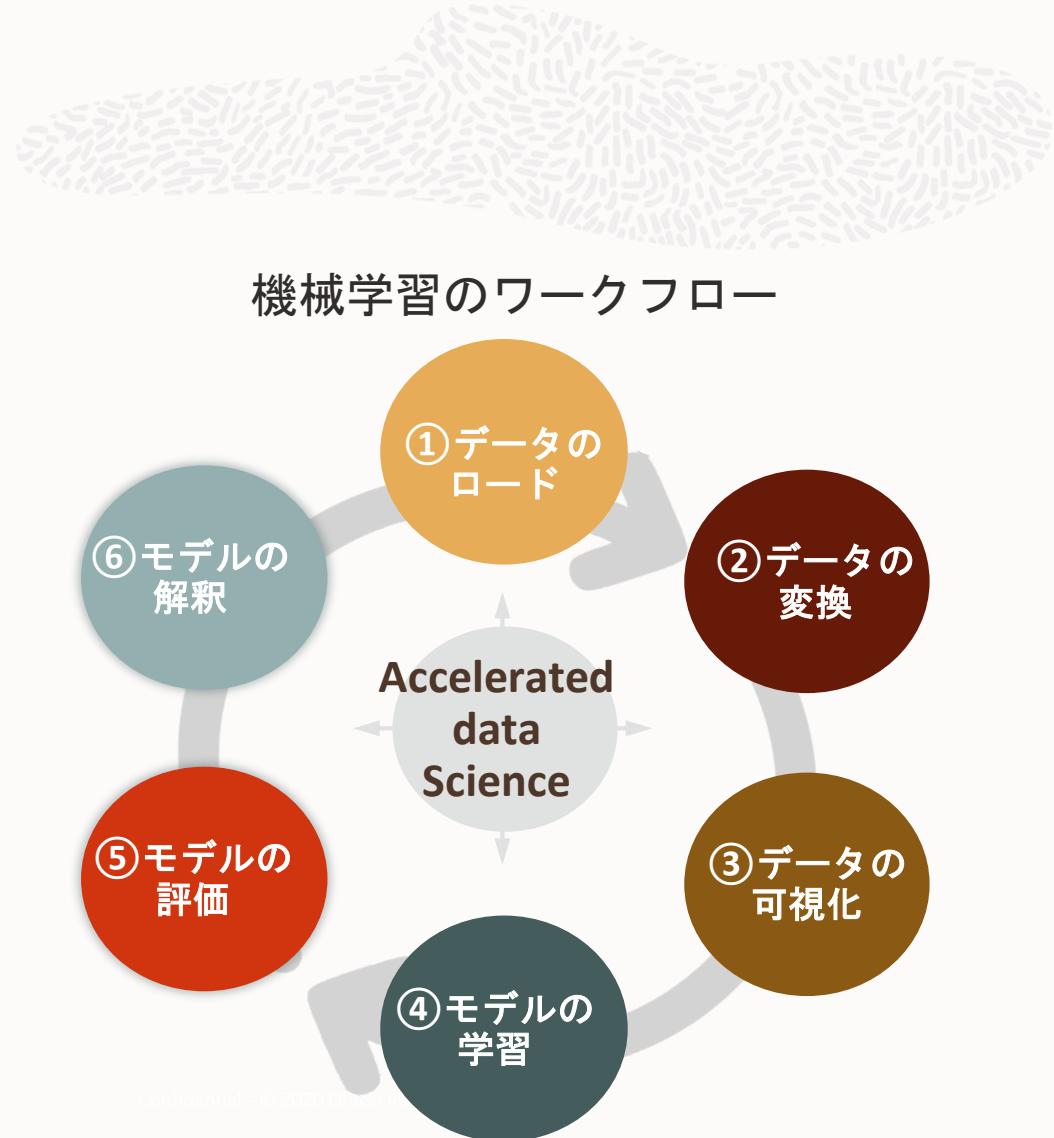
Data Science Serviceの主要コンポーネント

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

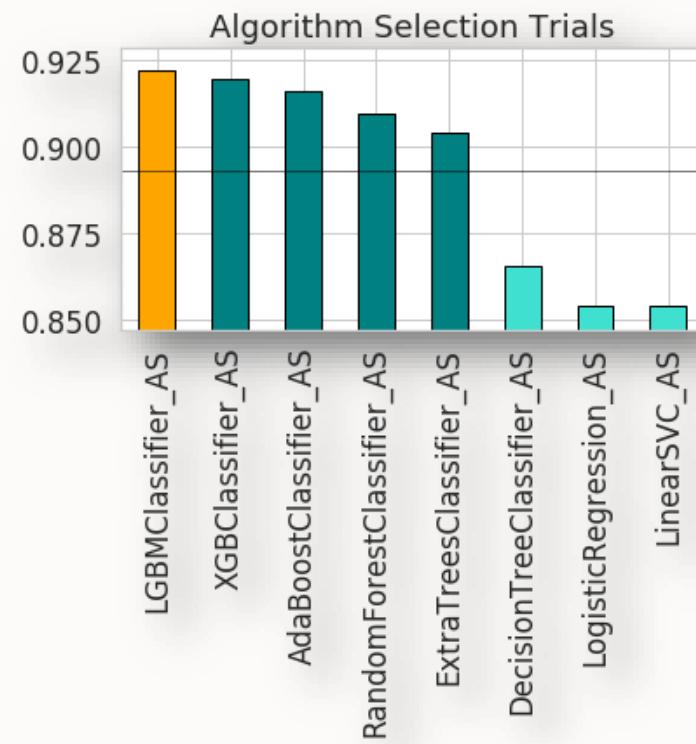


モデルの学習 Oracle AutoMLによる自動化



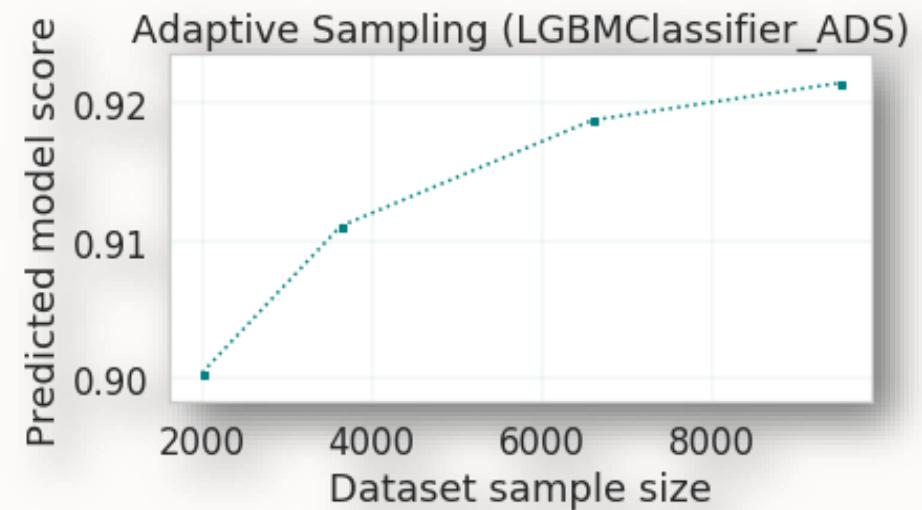
①最適なアルゴリズム選択の自動化

```
oracle_automl.visualize_algorithm_selection_trials()
```



②サンプリングの自動化

```
oracle_automl.visualize_adaptive_sampling_trials()
```

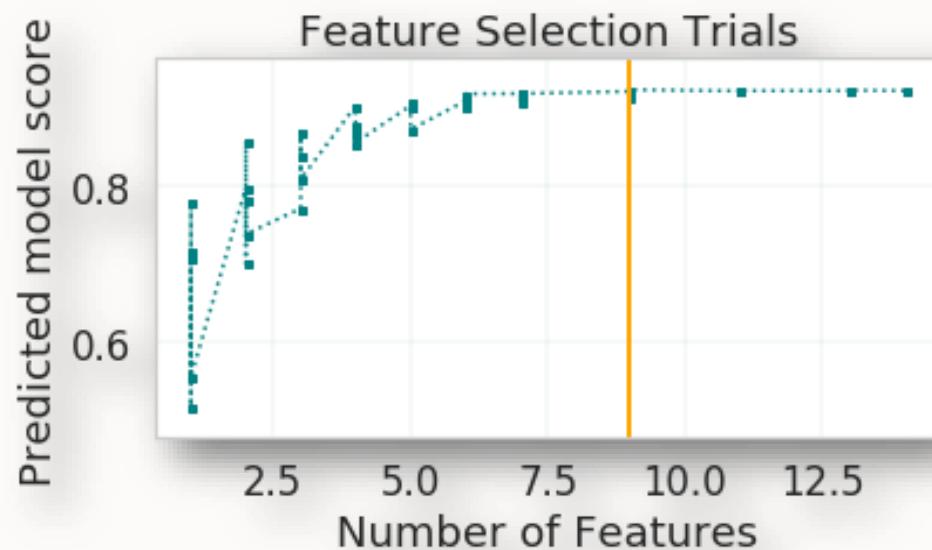


モデルの学習 Oracle AutoMLによる自動化



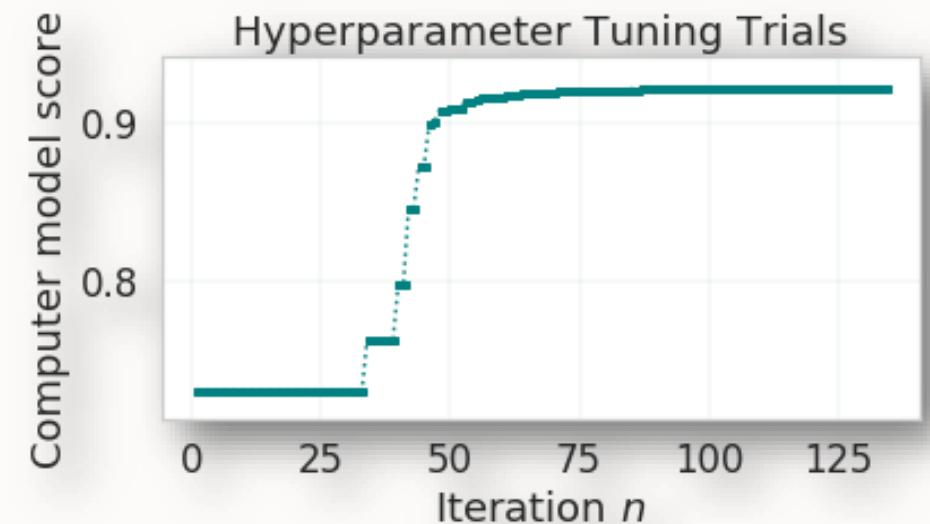
③最適な特徴量選択の自動化

```
oracle_automl.visualize_feature_selection_trials()
```



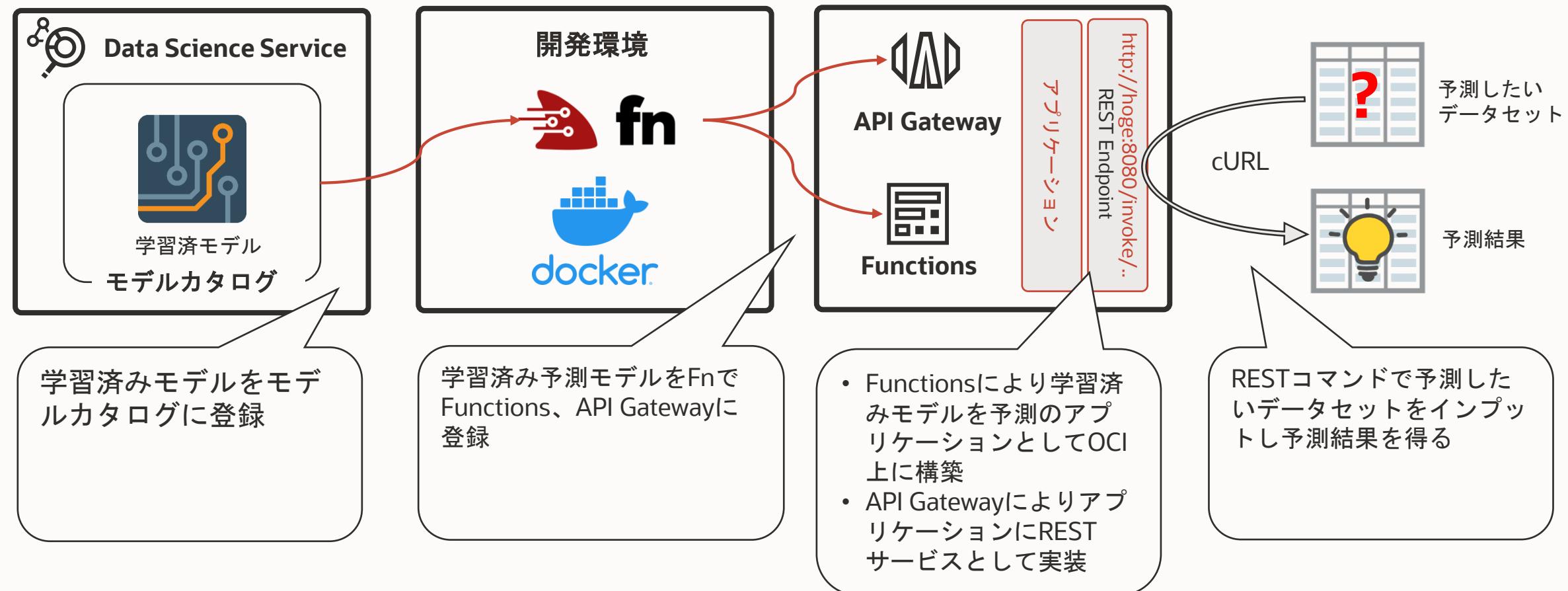
④ハイパーパラメータ・チューニングの自動化

```
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 注目

プロジェクト

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。

APExの概要

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
テナント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
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

データ・サイエンス

プロジェクト

プロジェクト



データ・サイエンスの前提条件

[詳細情報の表示](#)

リスト範囲

コンパートメント

madhacks.ochacafe

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 >				

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

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

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

任意のプロジェクト名

名前

banner_ab_test

説明

ab test for banner

タグ

タグ付けとは、テナント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

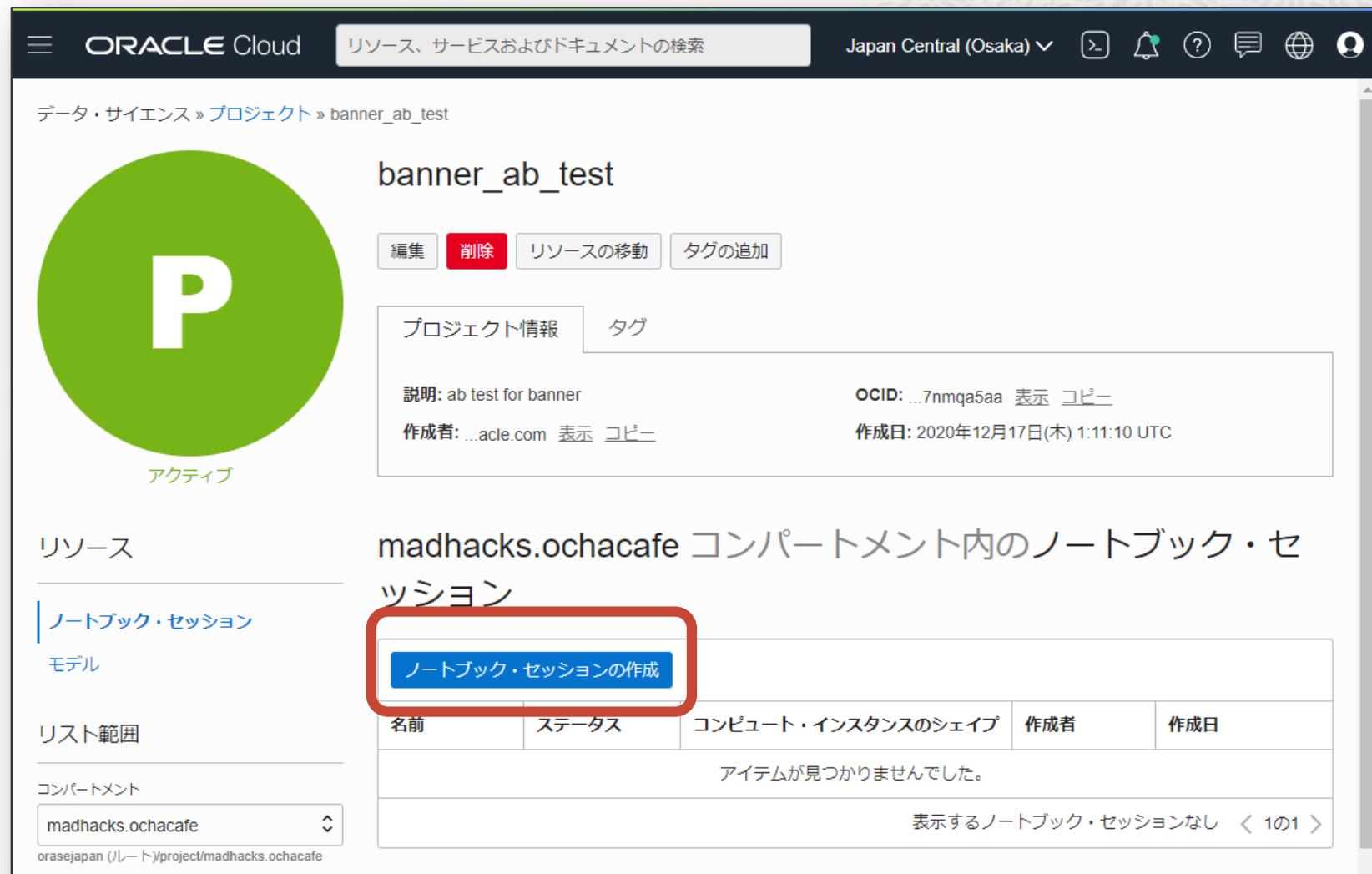
タグ・キー

値

なし(フリーフォーマット)

×

[+ 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](#)



☰ 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

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 OCID: ...7nmqa5aa 表示 コピー
作成者: ...acle.com 表示 コピー 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 (レート) / 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
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 >

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

開発環境のComputeインスタンスのスペックを選択

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

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

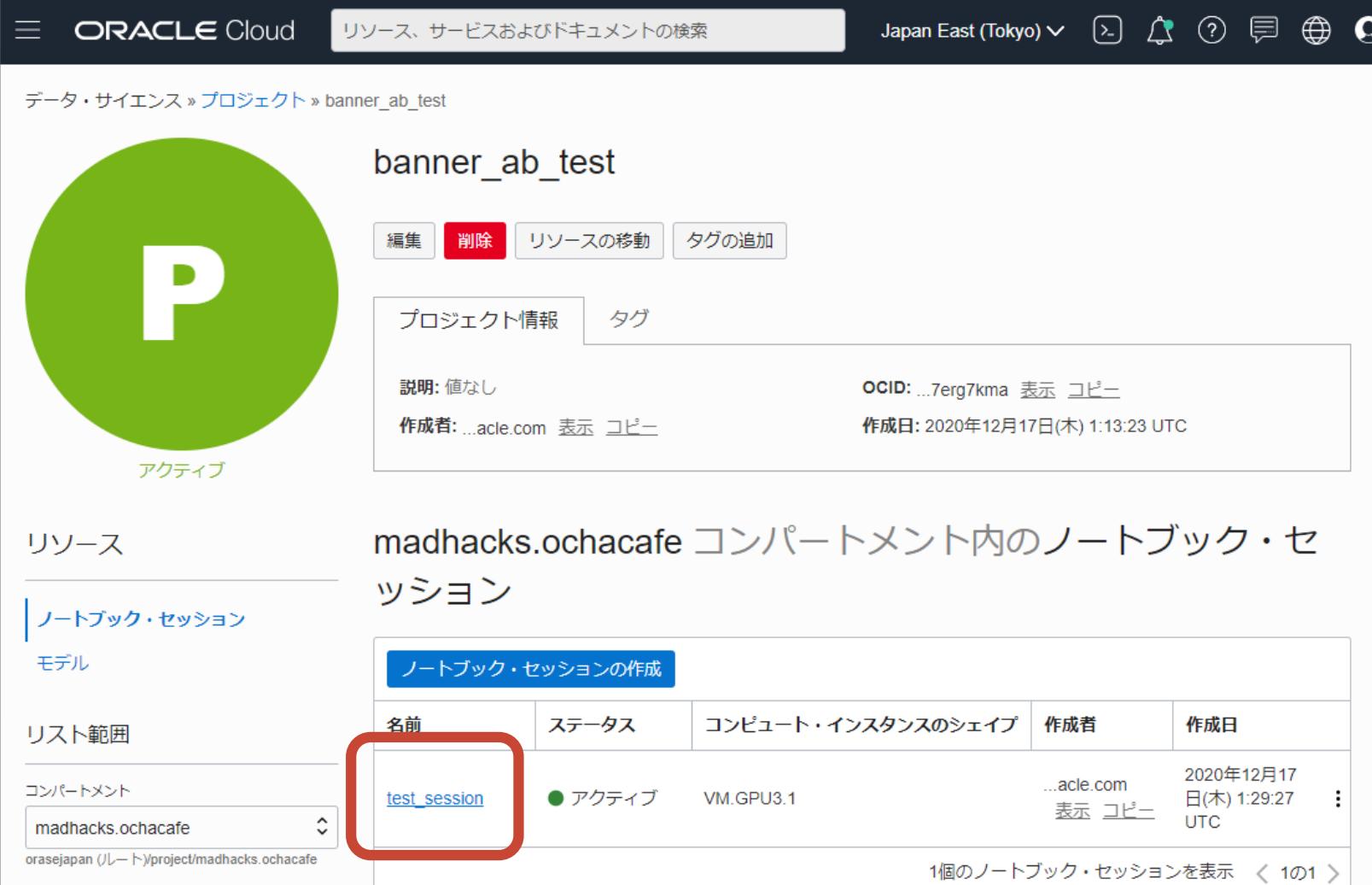
madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

フィルタ

作成

取消



データ・サイエンス » プロジェクト » banner_ab_test

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 OCID: ...7erg7kma 表示 コピー

作成者: ...acle.com 表示 コピー 作成日: 2020年12月17日(木) 1:13:23 UTC

リソース 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

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

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	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

test_session

開く (Red box)

非アクティブ化 終了 他のアクション ▾

ノートブック・セッション情報 タグ

OCID: ...jkzs3mfa 表示 コピー 作成日: 2020年12月17日(木) 1:29:27 UTC
作成者: ...acle.com 表示 コピー 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): 1,000 GB VCN: [bigdata-vcn](#)
サブネット: [Public Subnet-bigdata-vcn](#)

リソース メトリック

メトリック

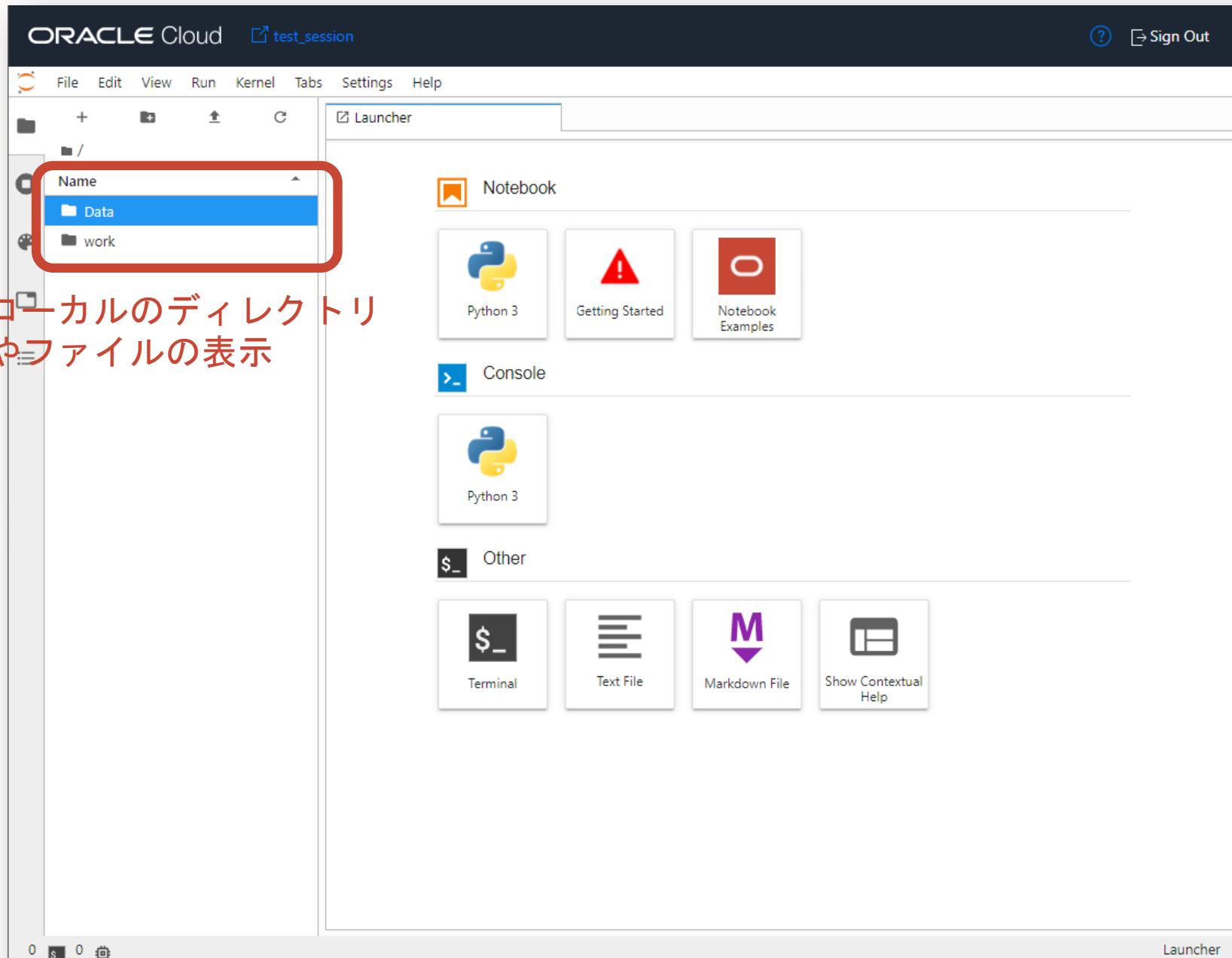
メトリックは、新規に作成またはアクティブ化されたノートブック・セッションに対してのみ使用可能です。新規ノートブック・セッションを作成するか、これを非アクティブ化してから再アクティブ化して、メトリックを有効にしてください。新規ノートブック・セッションのメトリックが表示されない場合は、メトリックを読み取る権限があることを確認します。 [詳細](#)

開始時間: 2020/12/17 0:39:30 終了時間: 2020/12/17 1:39:30 クイック選択: 過去1時間 ▾ チャートのリセット

CPU使用率 (i) オプション ▾

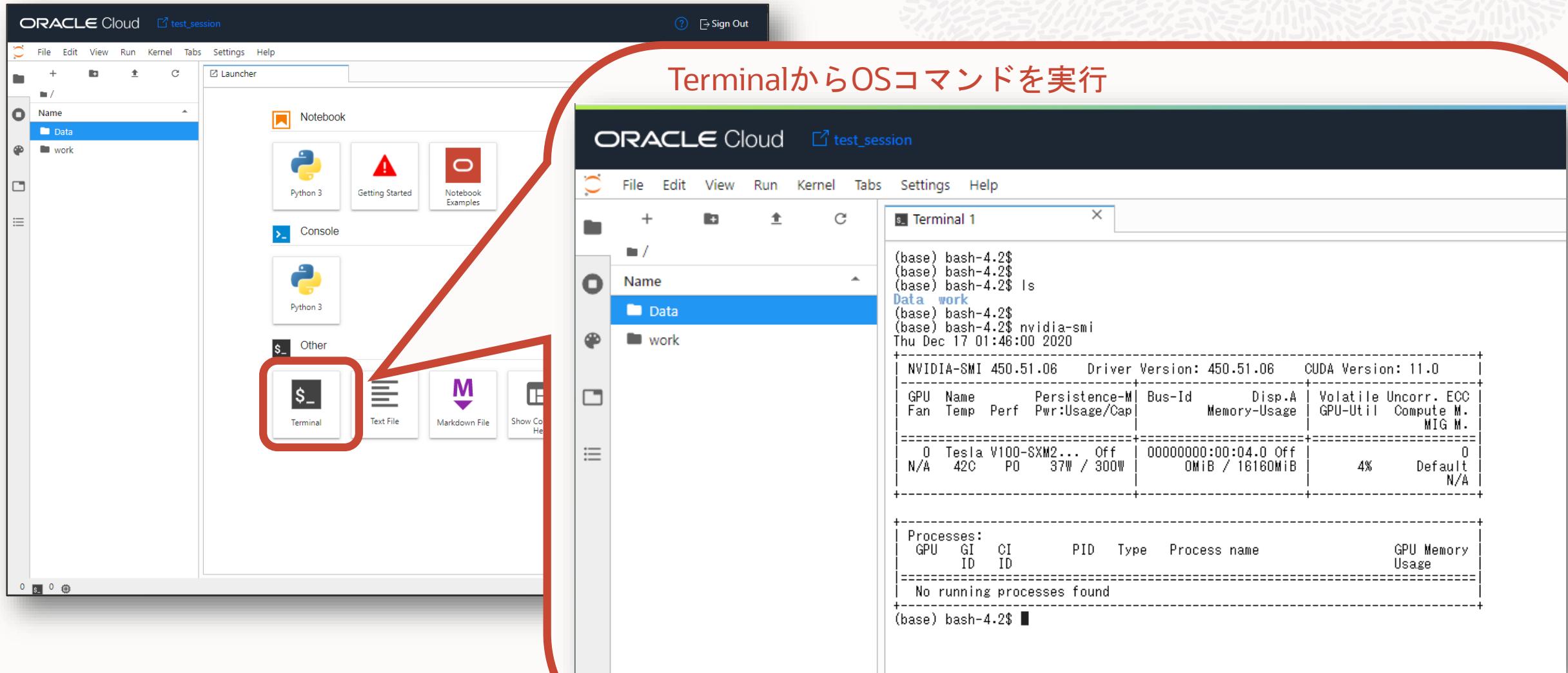
間隔: 1分 ▾ 統計: 平均 ▾

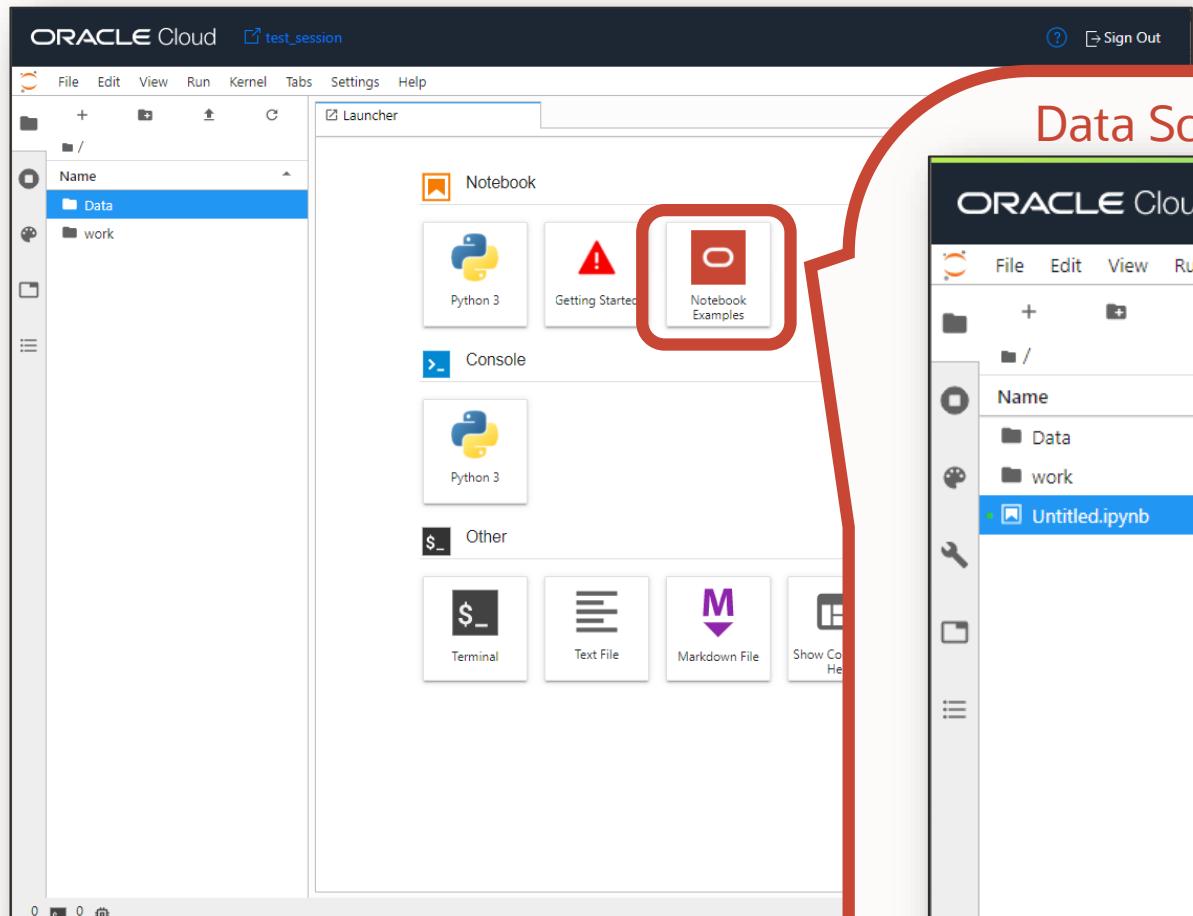




ローカルのディレクトリ
やファイルの表示







Data Science Serviceで提供されているサンプルコード

The screenshot shows a Jupyter Notebook titled 'Untitled.ipynb' in the Oracle Cloud Data Science Service. The notebook contains a single cell with the following Python code:

```
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')

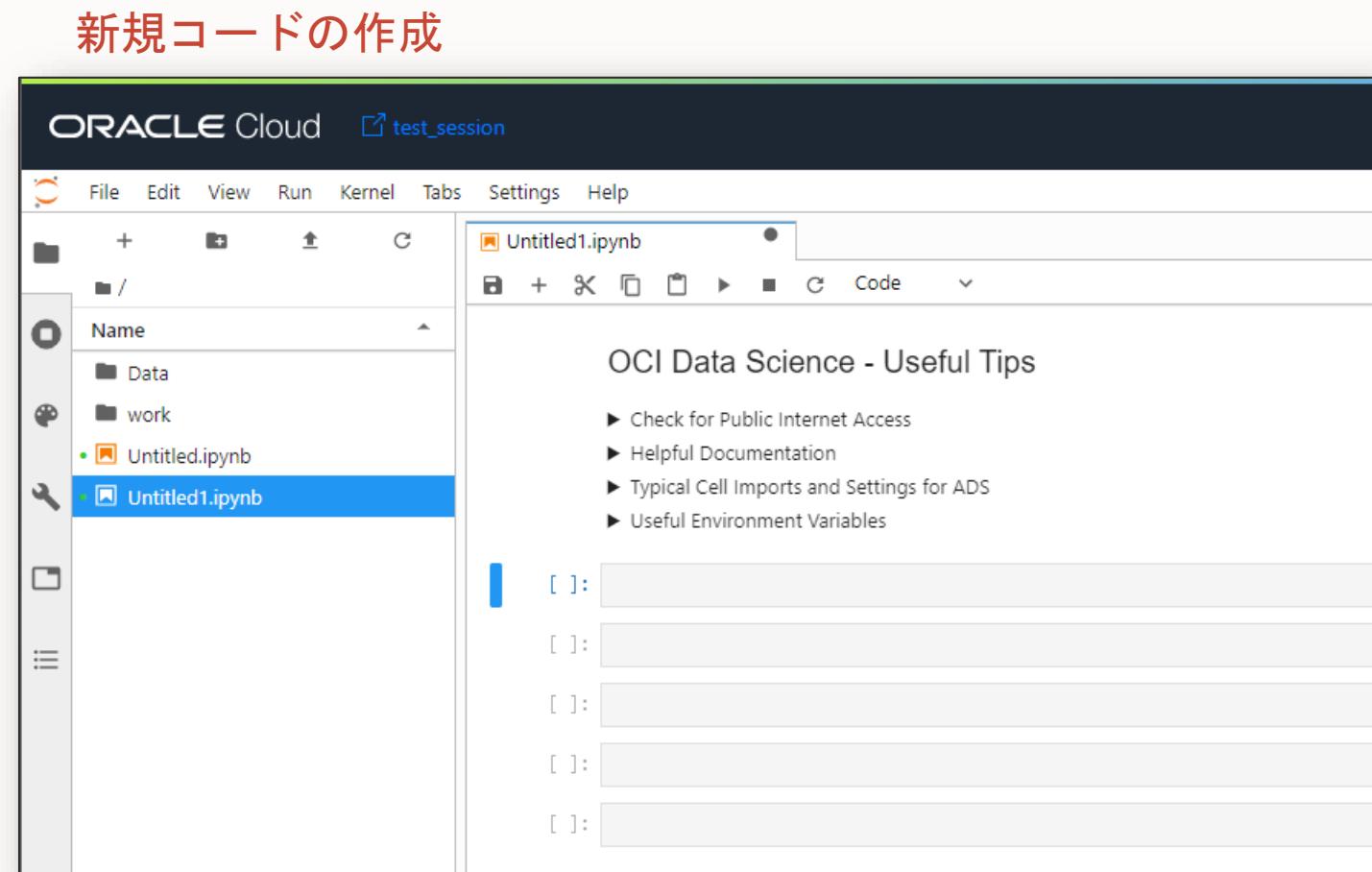
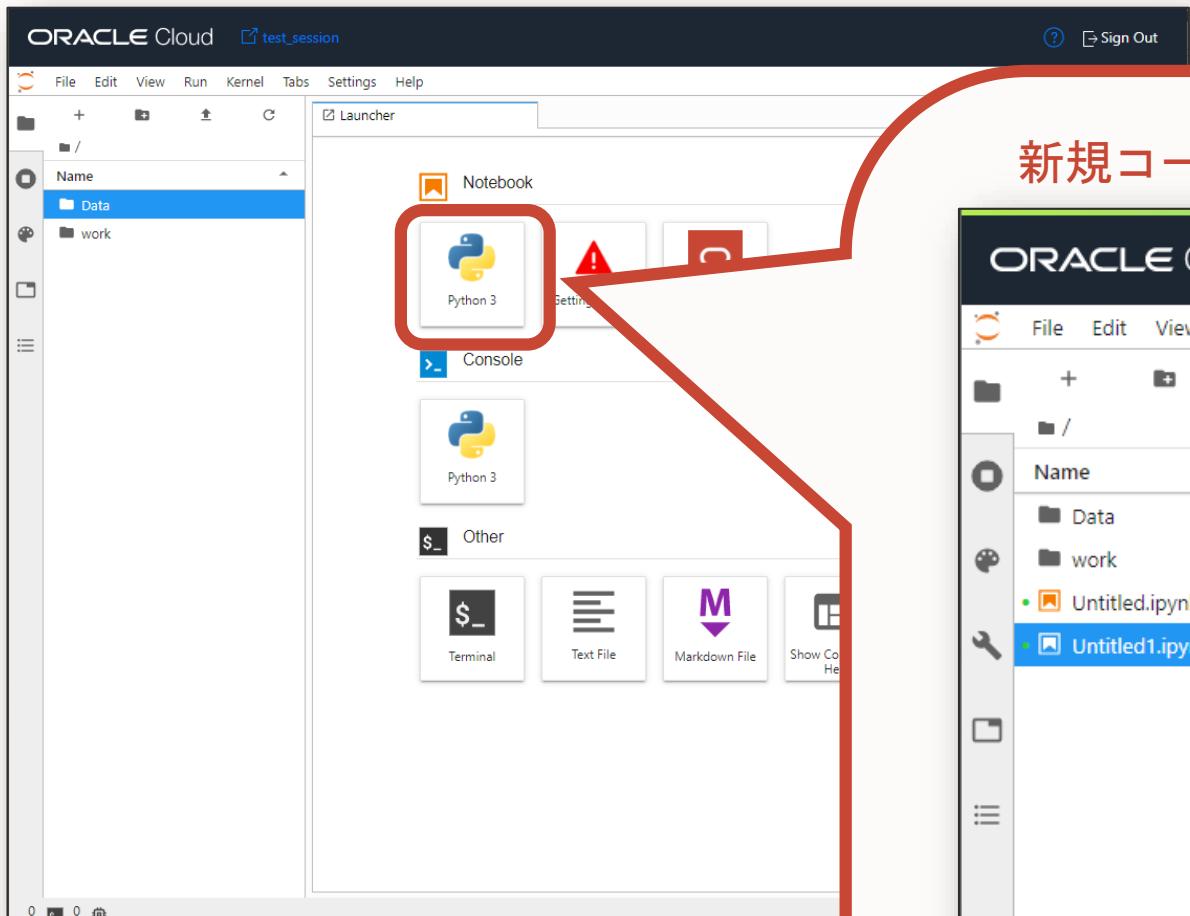
import logging
logging.basicConfig(format='%(levelname)s:%(message)s', level=logging.ERROR)

import pandas as pd

from ads.dataset.dataset_browser import DatasetBrowser
from ads.automl.provider import OracleAutoMLProvider
from ads.automl.driver import AutoML
from ads.evaluations.evaluator import ADSEvaluator
from ads.common.data import MLData

from sklearn.metrics import accuracy_score, make_scorer, precision_score, recall_score
from sklearn.preprocessing import LabelEncoder

%load_ext autoreload
%autoreload 2
```



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

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

プロジェクト内に共有された学習済みモデル 1個のモデルを表示 < 1の1 >

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

成者:

作成日: 2020年10月1日(木) 7:30:34 UTC

OCID: ...hgtdkqma4q 表示 コピー

モデル: ocid1.datasciencemodel.oc1.ap-tokyo-1.amaaa

アーティファクト: aaassl65iqazjki5qt5im64zf5jvuk5cm5yfrpf5df3

ファクト: phgtdkqma4q.zip (28.56MB)

モデルの来歴

GitリポジトリURL: 値なし

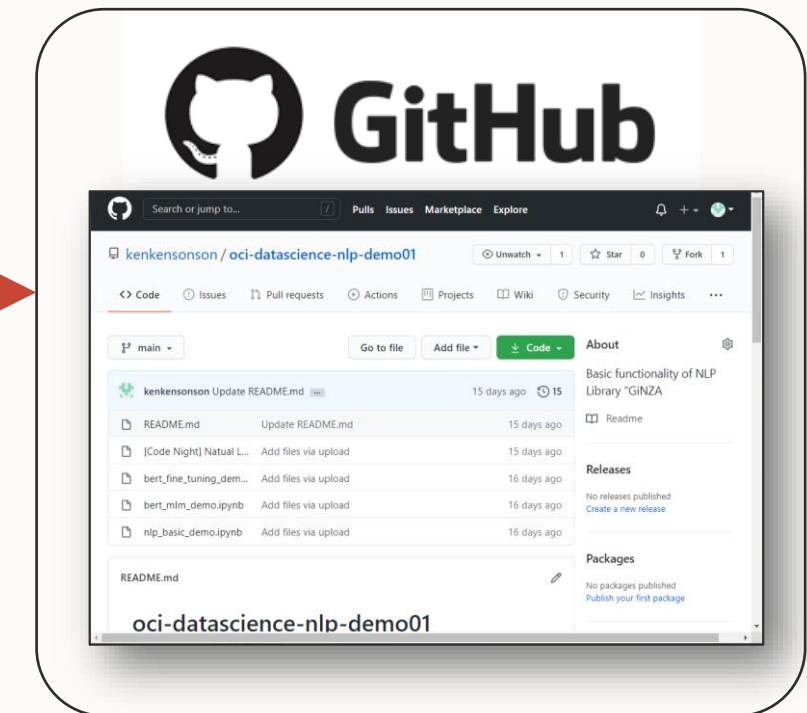
Gitコミット: 値なし

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

連携



使用条件およびプライバシーポリシー Cookieの設定

Copyright © 2020, Oracle and/or its affiliates. All rights reserved.

ociid1.datascience....zip

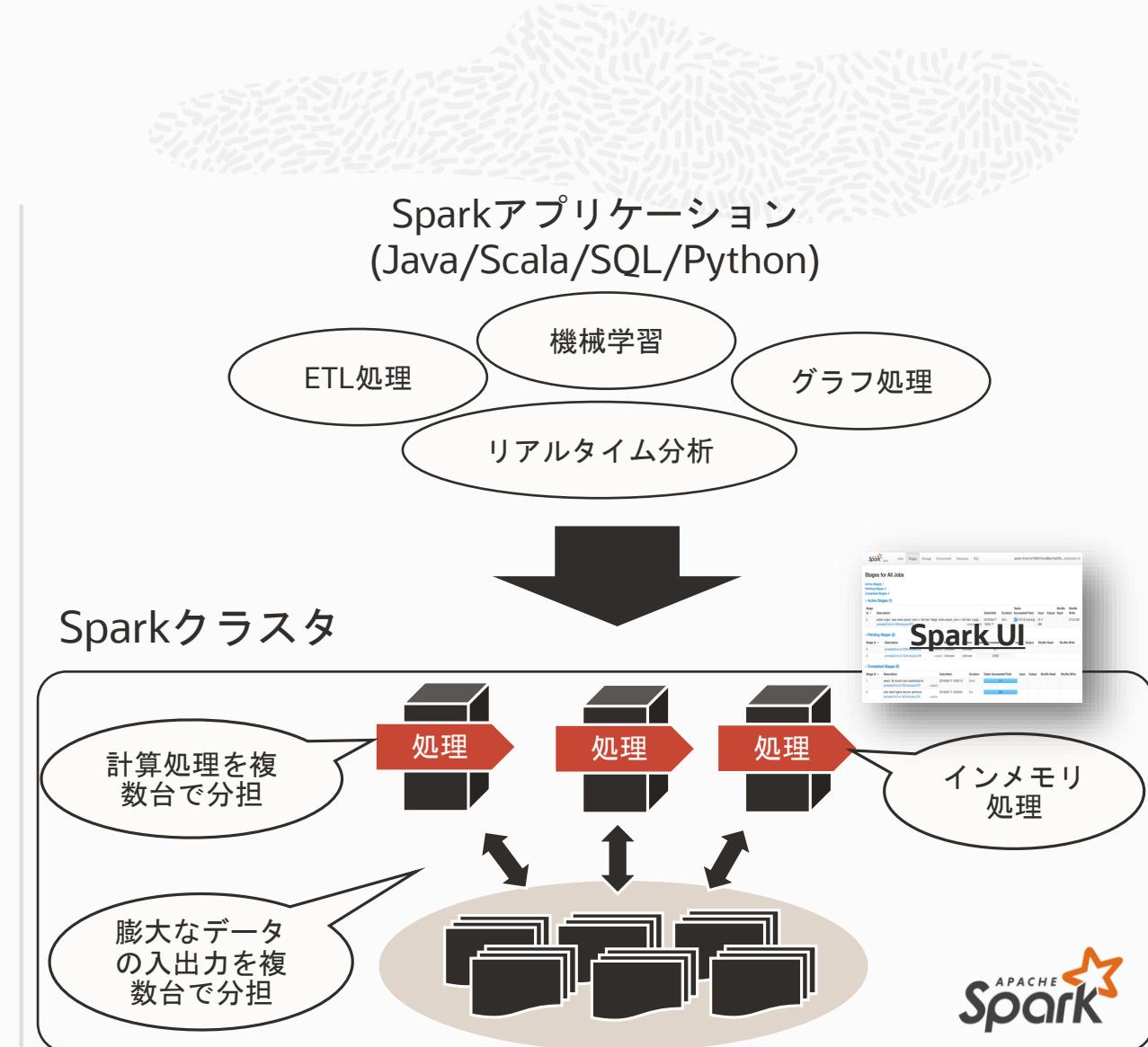




Data Flow Service

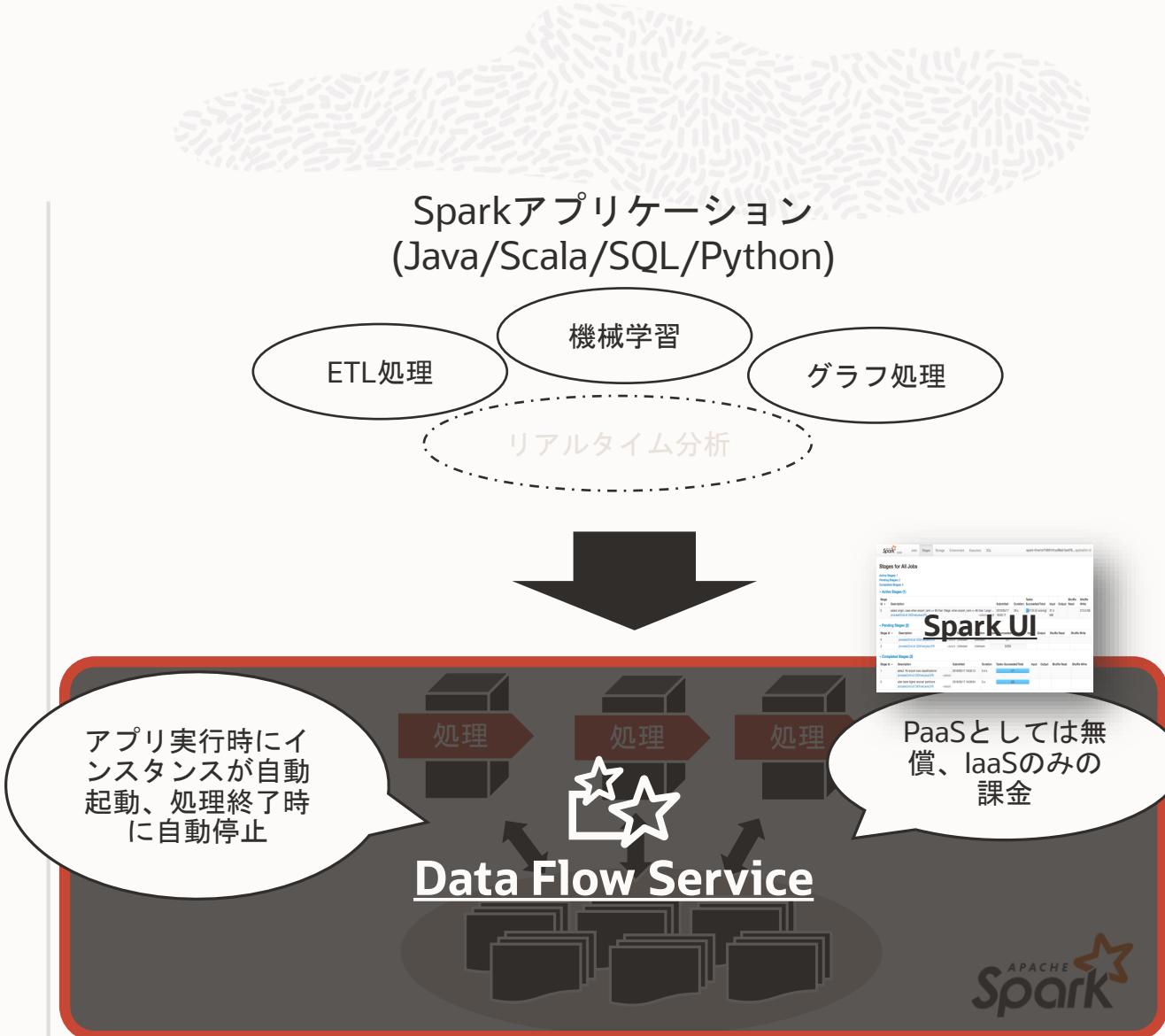
Apache Spark

- 並列分散処理のプラットホーム
 - 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

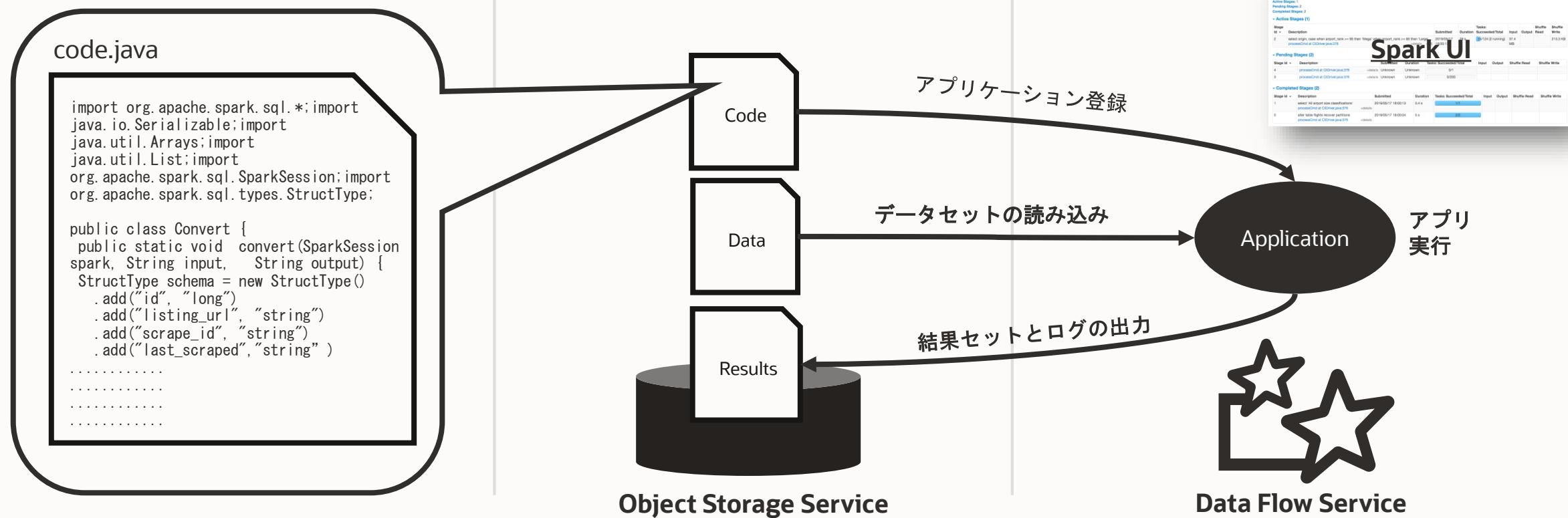


Data Flow Serviceのワークフロー

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

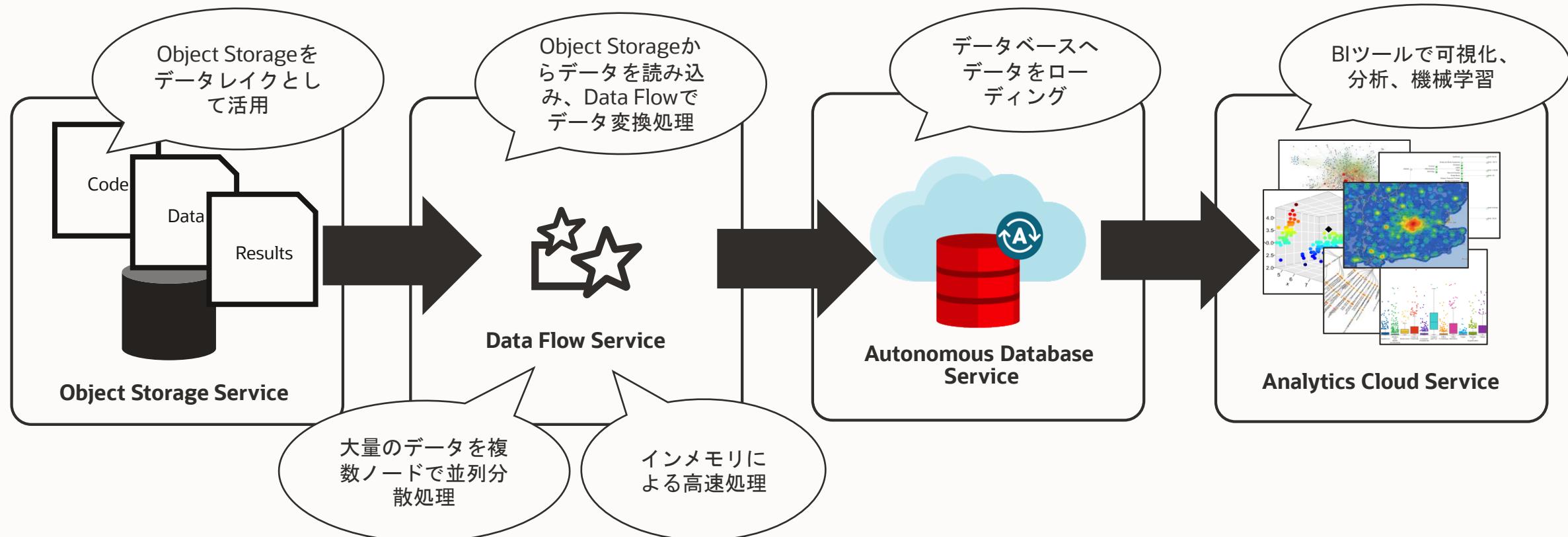
②コードのテキストファイルと対象のデータセットをオブジェクトストレージにアップロード

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



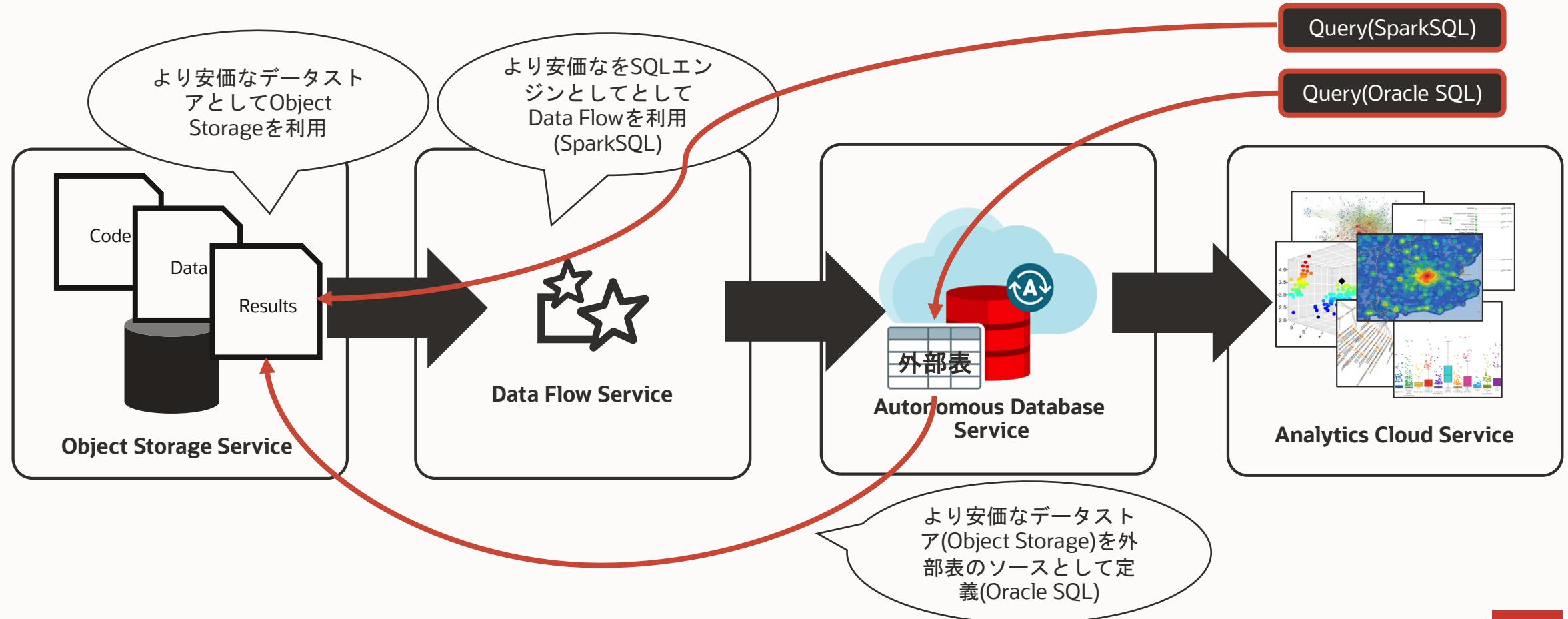
ユースケース① ETL

- 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



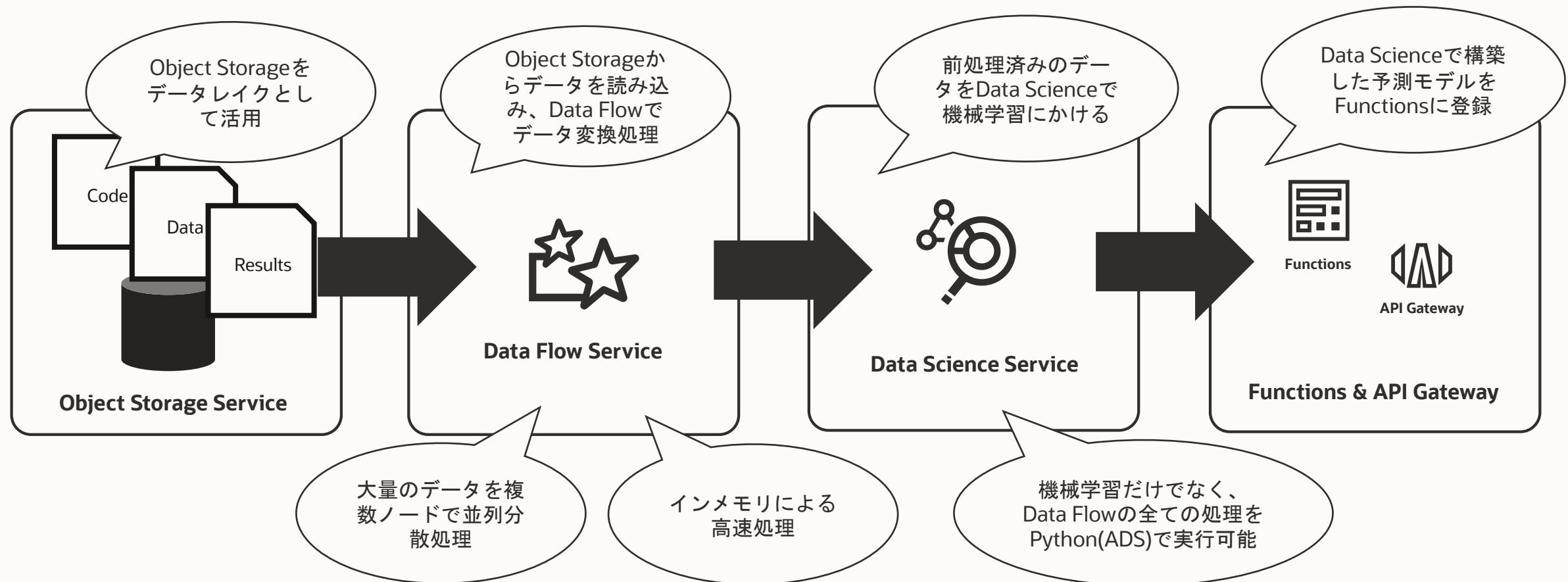
ユースケース② アクティブ・アーカイブ

- 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

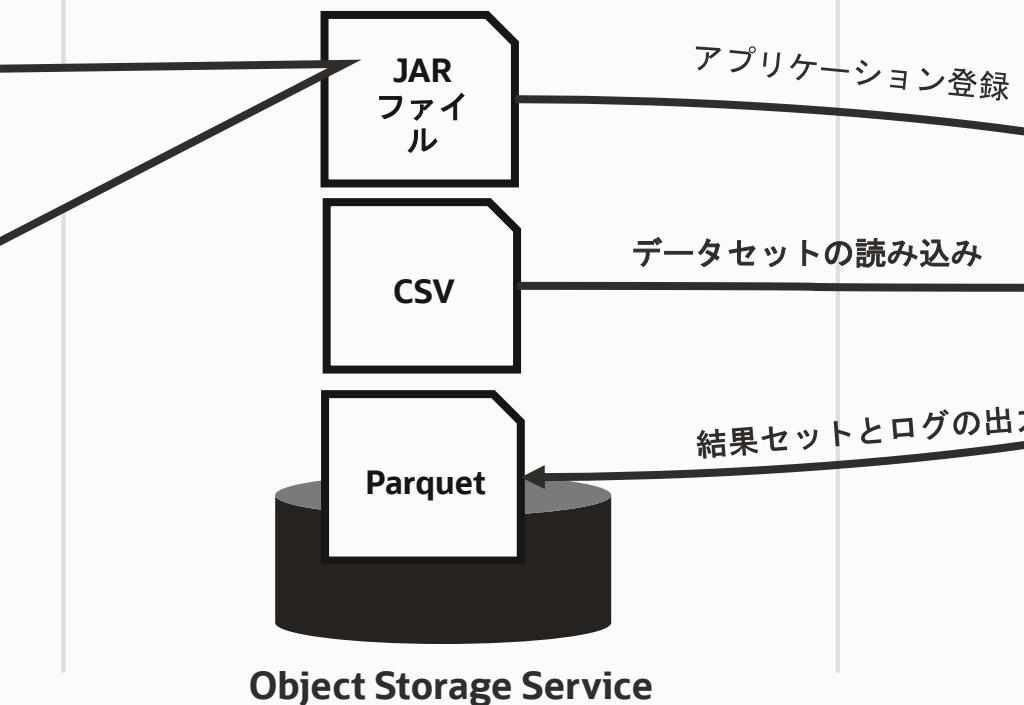


データ変換のデモンストレーション

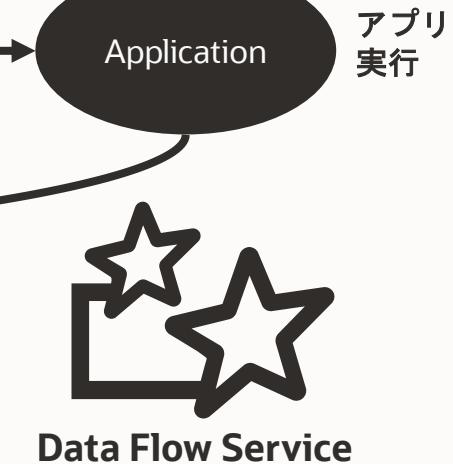
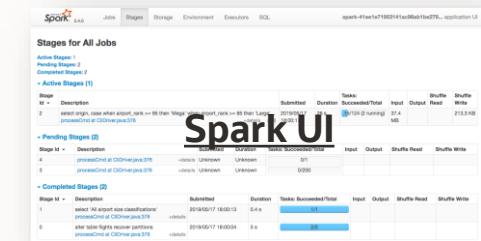
①Sparkアプリをコーディング
(Java, Scala, Python, SQL)

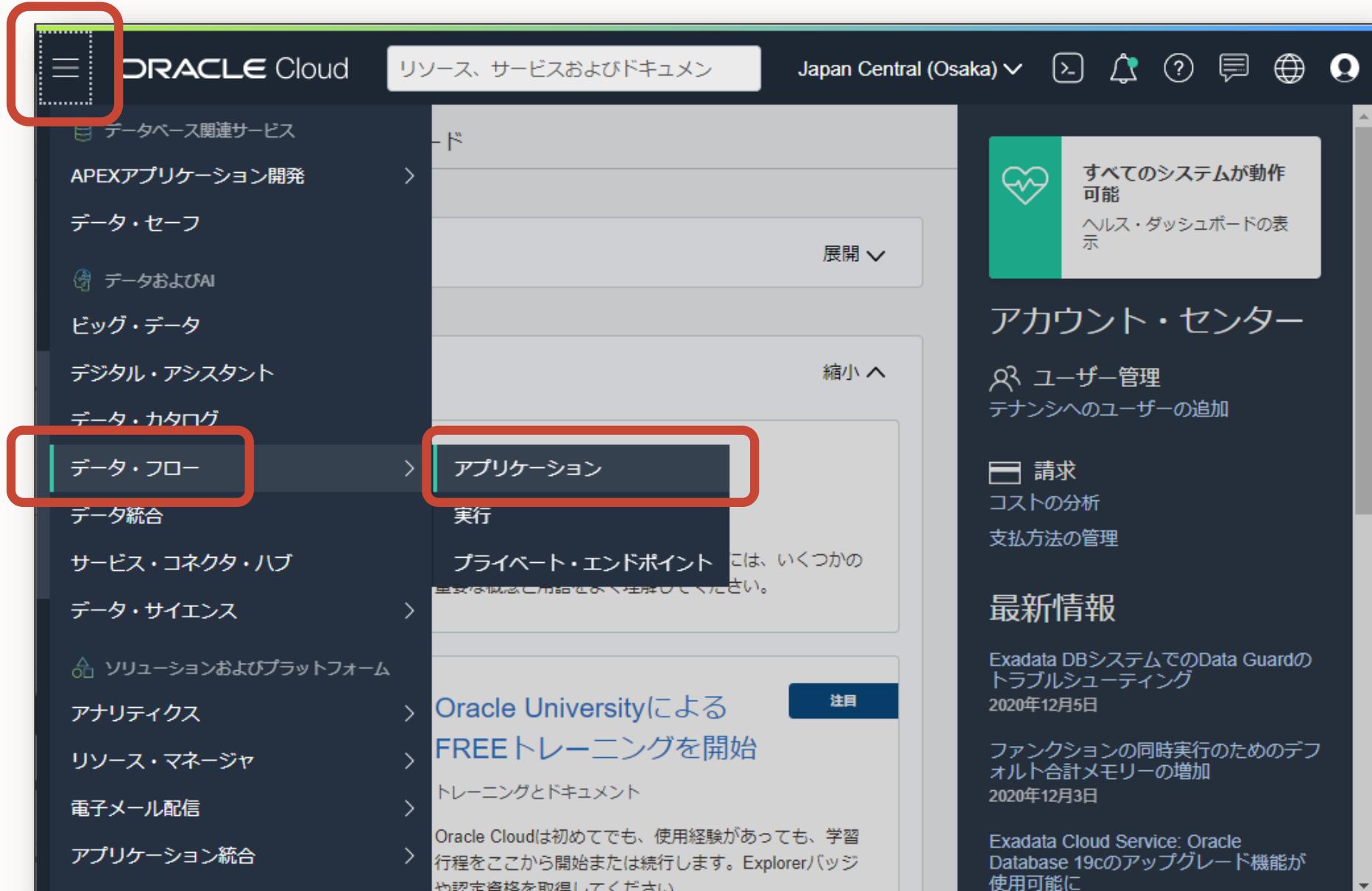


②コードのファイルと対象のデータセットをオブジェクトストレージにアップロード



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) ヘルプ ニュース ドキュメント ワークスペース

データベース関連サービス

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

データ・フロー

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 テナント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

三 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 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](#)

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

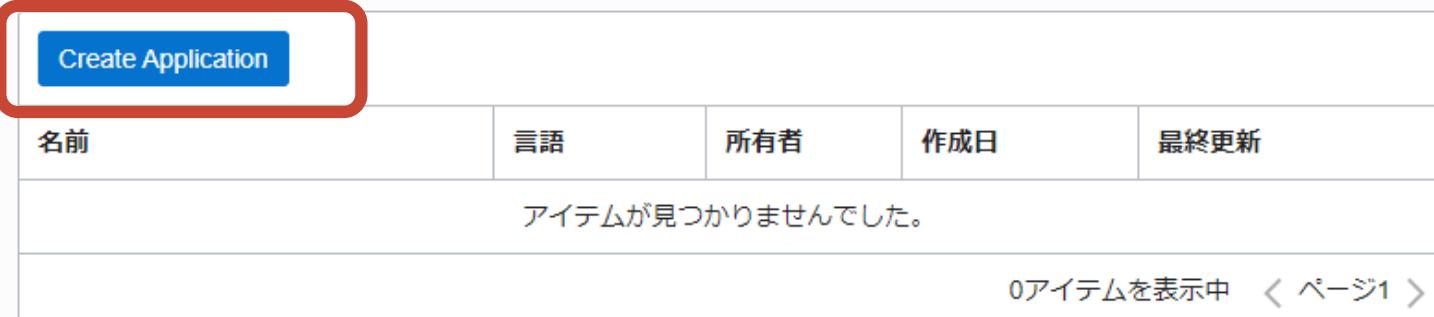
madhacks.ochacafe

タグ・フィルタ [追加](#) [クリア](#)

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 >



三 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) ヘルプ

Sparkアプリケーションの作成

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

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 Cookieの設定

madhacks.ochacafe

Data Flow

Create Application

名前

言語

Java

Python

SQL

Scala

アプリケーションに任意の名前を指定

名前 *i*

ETL

説明 オプション *i*

ETL

リソース構成

Sparkバージョン

Spark 2.4.4

ドライバ・タイプ *i*

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

エグゼキュータ・タイプ *i*

作成 取消

Copyright © 2020, Oracle and/or its affiliates. All rights reserved.

三 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

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

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)

エクセキュータの数 ⓘ

1

SparkのWorkerノードの数

アプリケーション構成

ファイルURL ⓘ

ow_2019_dataflow_lab@bigdatadatasciencelarge/usercontent/oow-lab-2019-java-etl-1.0-SNAPSHOT.jar

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)

メイン・クラス名 ⓘ

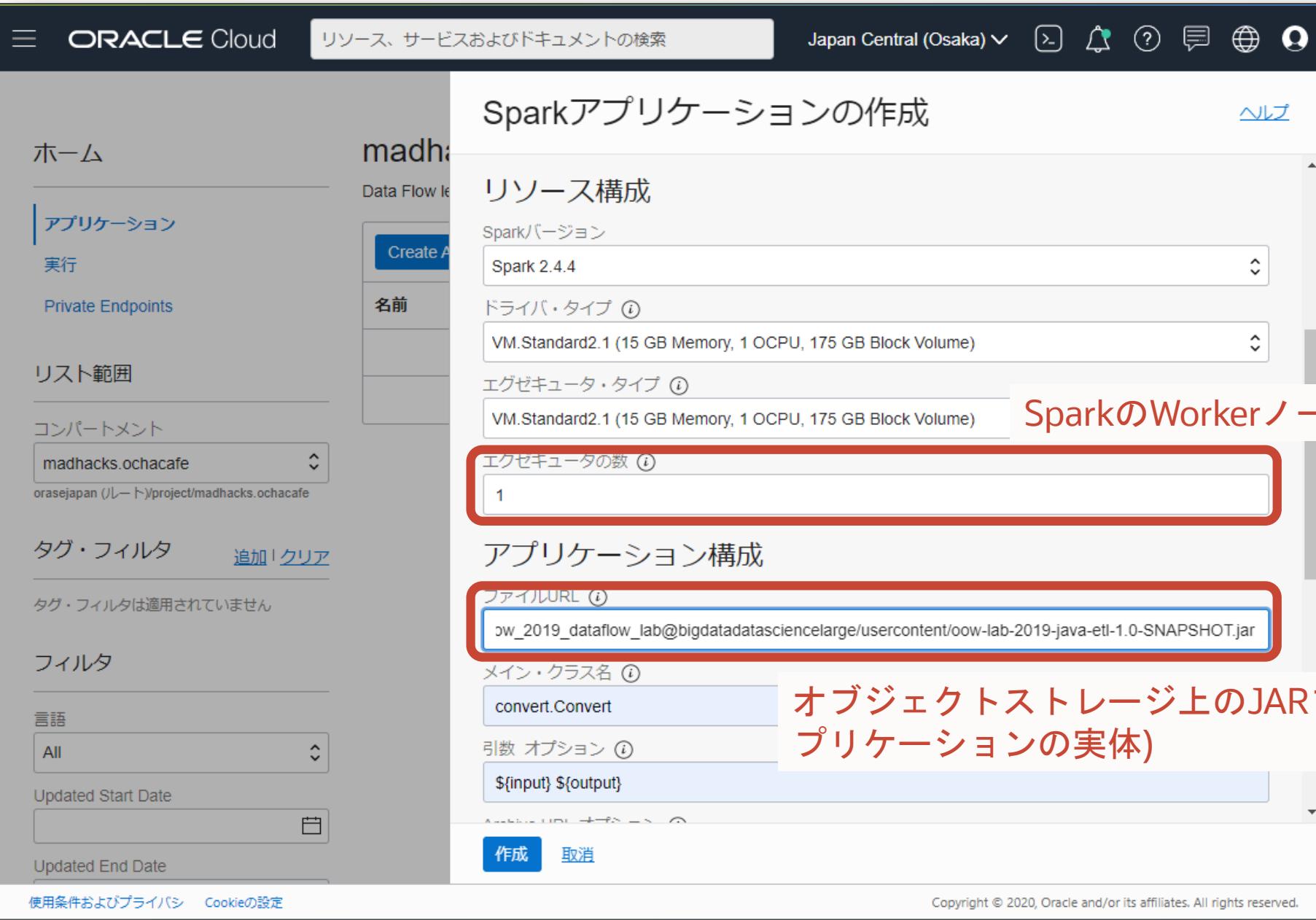
convert.Convert

引数 オプション ⓘ

\${input} \${output}

作成 取消

Copyright © 2020, Oracle and/or its affiliates. All rights reserved.



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

引数 オプション ⓘ

`${input} ${output}`

Archive URI オプション ⓘ

名前

名前 デフォルト値

input `oci://dataflow-data@orasejapan/optimized_listing`

output `oci://dataflow-data@orasejapan/optimized_listing`

処理対象のCSVファイル

タグ

タグ付けとは、テナント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です。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

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

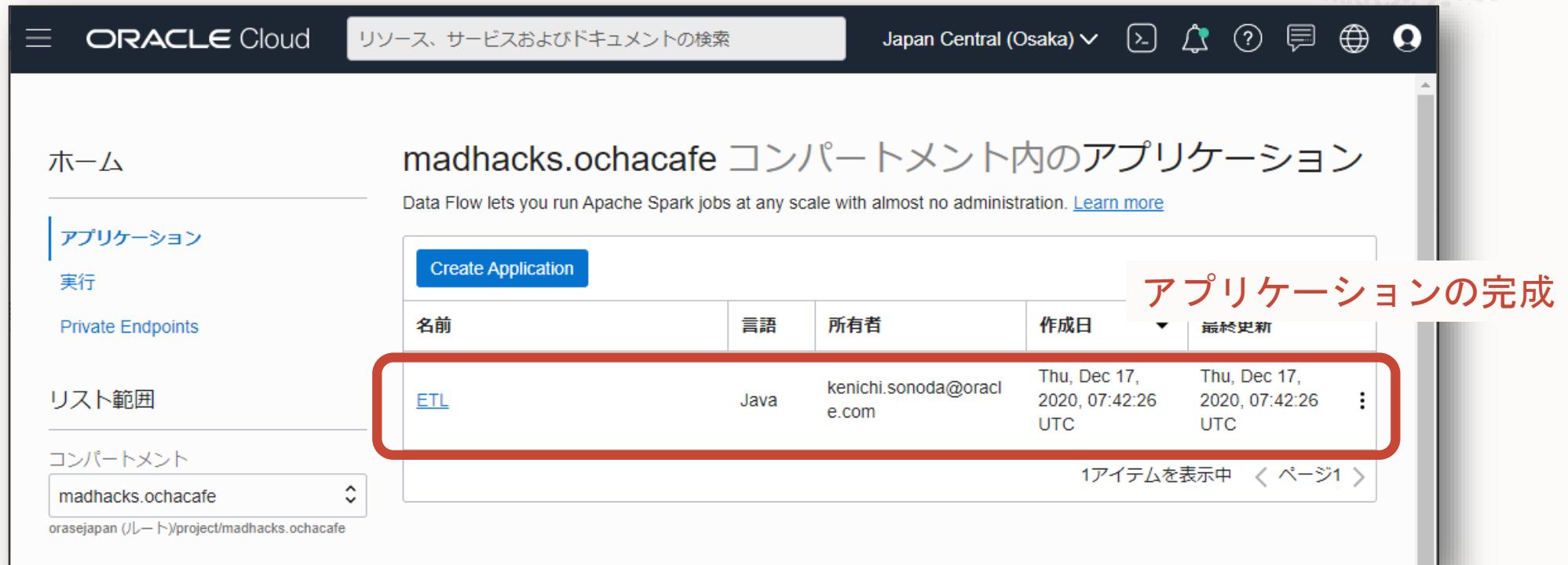
フィルタ

言語

All

Updated Start Date

Updated End Date



madhacks.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](#)

アプリケーションの完成

名前	言語	所有者	作成日	最終更新	操作
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

三 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@oracle.com	Thu, Dec 17, 2020, 07:42:26 UTC	View Details 編集
1アイテムを表示 実行				

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) ヘルプ ナビゲーション

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	データ読み取り	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	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 >

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 タグ・フィルタ 追加 | クリア

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

madhacks.ochacafe コンパートメント内の実行

実行

Private Endpoints

リスト範囲

コンパートメント

madhacks.ochacafe

タグ・フィルタ

追加 | クリア

名前	言語	ステータス	所有者	作成日	期間	OCPU 合計	データ読み取り	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	詳細	Spark UI
TestApp	Java	● 成功	kenichi.sonoda @oracle.com	Thu, Dec 17, 2020, 07:28:55 UTC	1分 54 秒	2	再実行	タグの追加

2アイテムを表示

リソースの移動

Spark Jobs (?)

User: dataflow

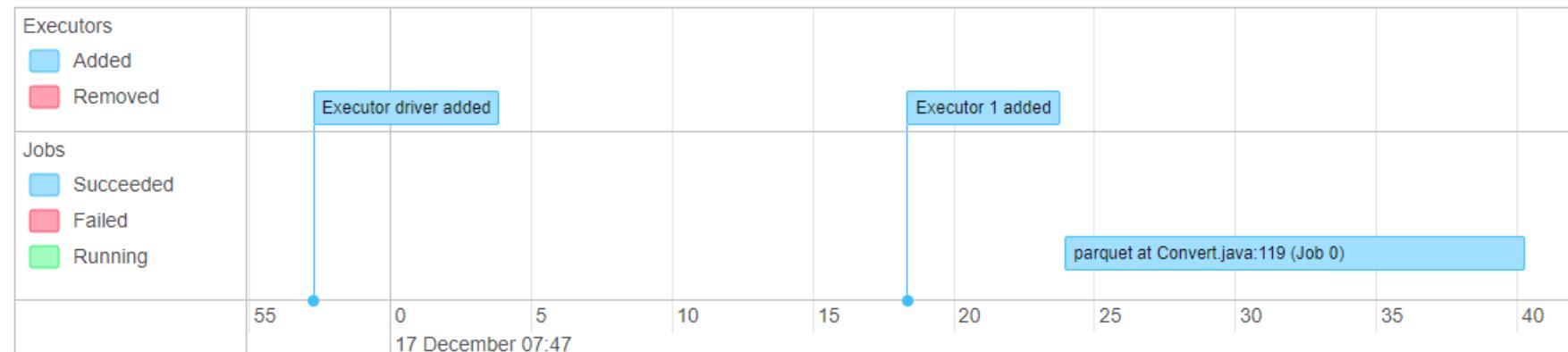
Total Uptime: 46 s

Scheduling Mode: FIFO

Completed Jobs: 1

▼ Event Timeline

Enable zooming



▼ Completed Jobs (1)

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
0	parquet at Convert.java:119 parquet at Convert.java:119	2020/12/17 07:47:23	16 s	1/1	1/1



リソース、サービスおよびドキュメントの検索

Japan Central (Osaka) ▾

Oracle Cloud

パケット情報 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

コンバートメント: [madhacks.ochacafe](#)

近似カウント: 5オブジェクト [i](#)

近似サイズ: 30.52MiB [i](#)

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

オブジェクト・イベントの出力: ● 無効 [編集](#) [i](#)

OCID: ...qhftvofa [表示](#) [コピー](#)

オブジェクト・バージョニング: ● 無効 [編集](#) [i](#)

リソース オブジェクト

オブジェクト

アップロード 他のアクション ▾

接頭辞で検索

<input type="checkbox"/>	名前	最終変更	サイズ	ステータス
<input type="checkbox"/>	optimized_listings	2020年12月17日(木) 7:47:40 UTC	30.52MiB	使用可能
<input type="checkbox"/>	_SUCCESS			
<input type="checkbox"/>	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

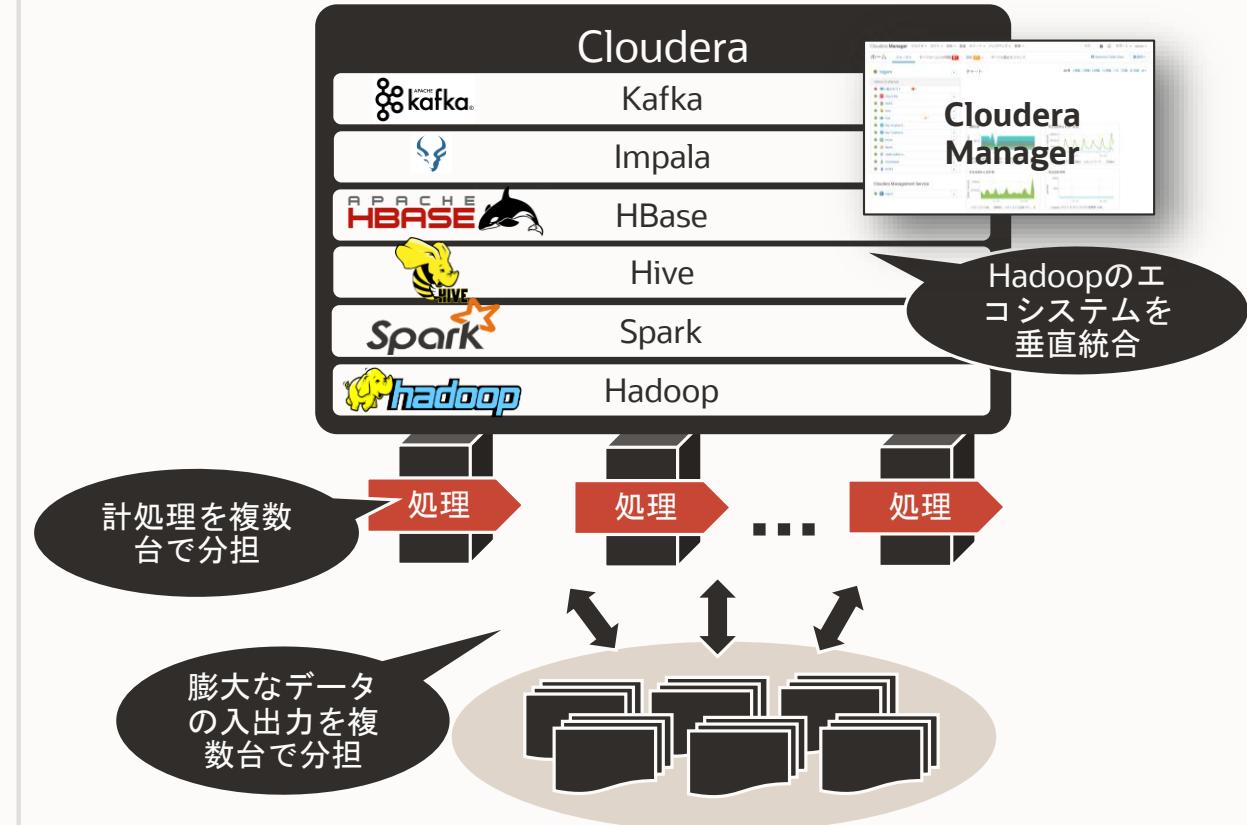


Big Data Service

Cloudera Enterprise Data Hub

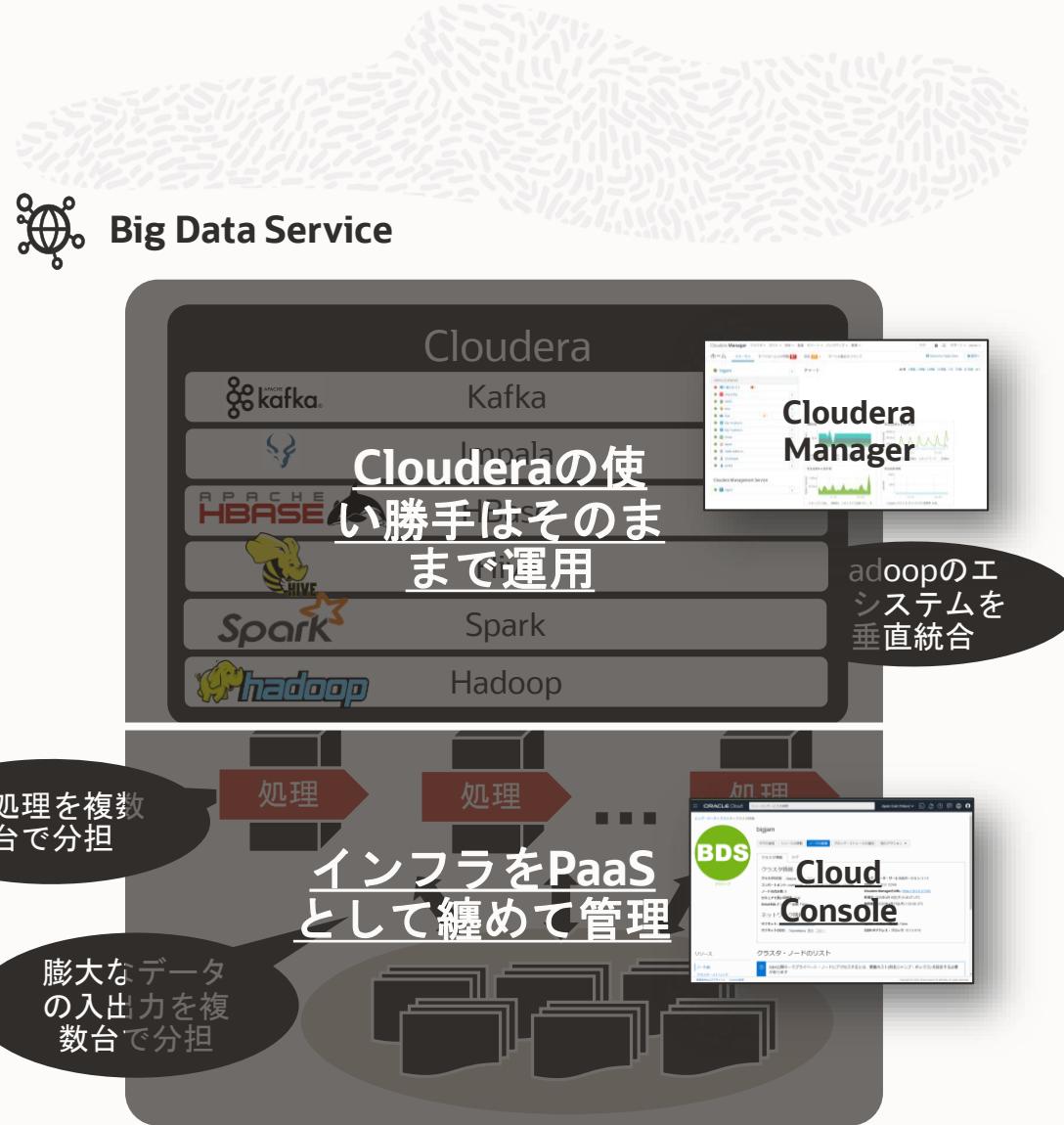
- Cloudera社から提供されているHadoopのディストリビューション
- Apache Hadoopとそのエコシステムによるデータレイク、分析システム基盤
 - データレイク
 - データウェアハウス
 - リアルタイム分析
 - Business Intelligence
 - ETL
 - 機械学習
- Hadoop周辺のエコシステムを複数のノードに垂直統合
- 大規模な分散処理基盤における、ソフトウェアスタック、ハードウェアスタックを統合管理

Clouderaクラスタの構成



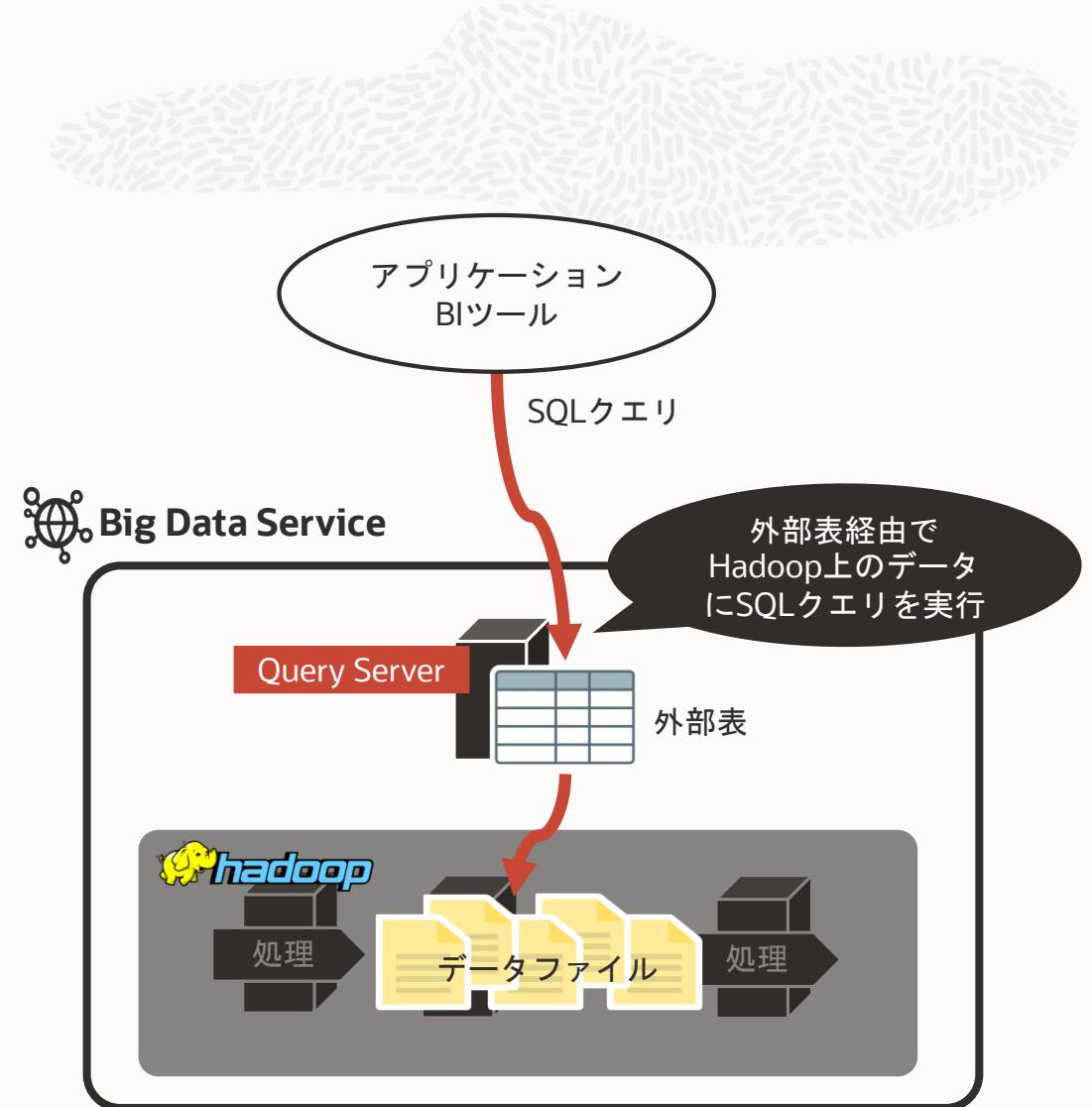
Big Data Service

- 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



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



The screenshot shows the Oracle Cloud interface. The top navigation bar includes the Oracle Cloud logo, a search bar, and a location dropdown set to "Japan East (Tokyo)". The left sidebar is highlighted with a red box and contains the following navigation items:

- ☰ メニュー
- Oracle Database
- 概要
 - Autonomous Data Warehouse
 - Autonomous JSON Database
 - Autonomous Transaction Processing
 - ペア・メタル、VMおよびExadata
 - Exadata Cloud@Customer
- 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
- サービス・コネクタ・ハブ
- データ・サイン

The main content area shows a notebook session named "test_session" with the following details:

- OCID: ...jkzs3mfa 表示 コピー
- 作成者: ...acle.com 表示 コピー
- プロック・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

Below this, a "メトリック" (Metrics) section is shown, also highlighted with a red box. It includes a tooltip explaining that metrics are only available for active or newly created sessions. The tooltip text is:

① 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

The metrics section includes filters for "終了時間" (End Time) and "クイック選択" (Quick Selection) set to "過去1時間" (Last 1 hour), and a "チャートのリセット" (Reset Chart) button.

三 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

① 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 >



ビッグ・データ

クラスタ

リスト範囲

- 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

クラ



クラスタの作成

ヘルプ

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・コンピュート・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 読取り専用

クラスタの作成

取消



三 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

madhacks.ochacafe

orasejapan (ルート)/project/madhacks.ochacafe

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 >

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

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



最終更新: 2020年12月16日(水) 13:01:53 UTC

リソース

クラスタ・ノードのリスト

ノード(6)

クラスタ・メトリック

作業リクエスト

SSH公開キーでプライベート・ノードにアクセスするには、要塞ホスト(別名ジャンプ・ボックス)を設定する必要があります
[要塞ホストの設定の詳細](#)

名前	ステータス	ノード・タイプ	シェイプ	IPアドレス	作成日	⋮
ksclustmn0	● アクティブ	Master	VM.Standard2.24	10.0.0.33	2020年12月8日(火) 4:08:04 UTC	⋮
ksclustuno	● アクティブ	Utility	VM.Standard2.24	10.0.0.37	2020年12月8日(火) 4:08:04 UTC	⋮
ksclustwn0	● アクティブ	Worker	VM.Standard2.1	10.0.0.36	2020年12月8日(火) 4:08:04 UTC	⋮
ksclustwn1	● アクティブ	Worker	VM.Standard2.1	10.0.0.34	2020年12月8日(火) 4:08:04 UTC	⋮
ksclustwn2	● アクティブ	Worker	VM.Standard2.1	10.0.0.35	2020年12月8日(火) 4:08:04 UTC	⋮
ksclustwn3	● アクティブ	Worker	VM.Standard2.1	10.0.0.39	2020年12月11日(金) 7:19:07 UTC	⋮

6アイテムを表示中

プロビジョニングされた
Masterノード
Utilityノード
Workerノード



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 ▾

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

クラスタ情報

Cloud SQL情報

タグ

クラスタOCID: ...hzo3pa 表示 コピー

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

顧客ネットワーク情報

サブネット: [Public Subnet-bigdata-vc](#)

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

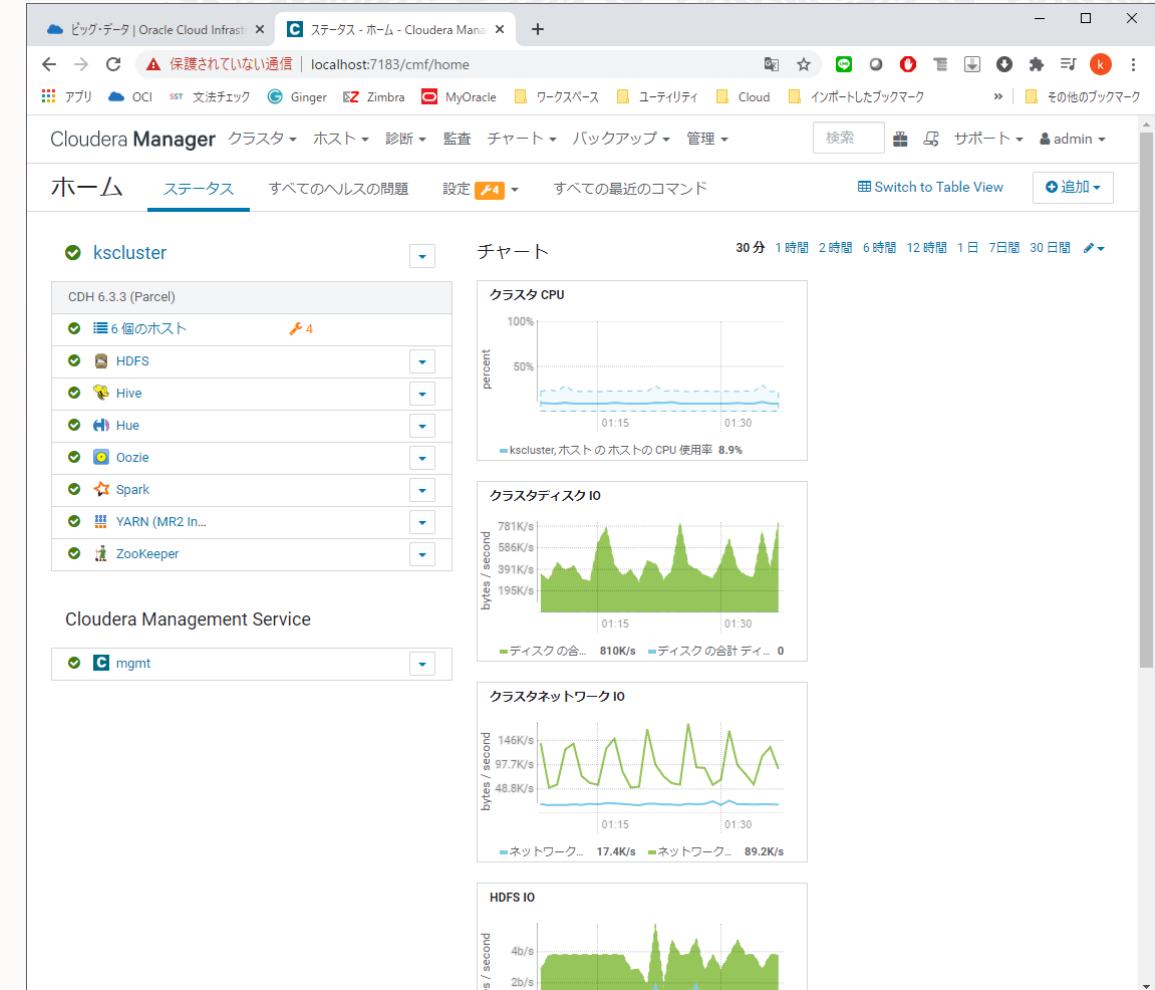
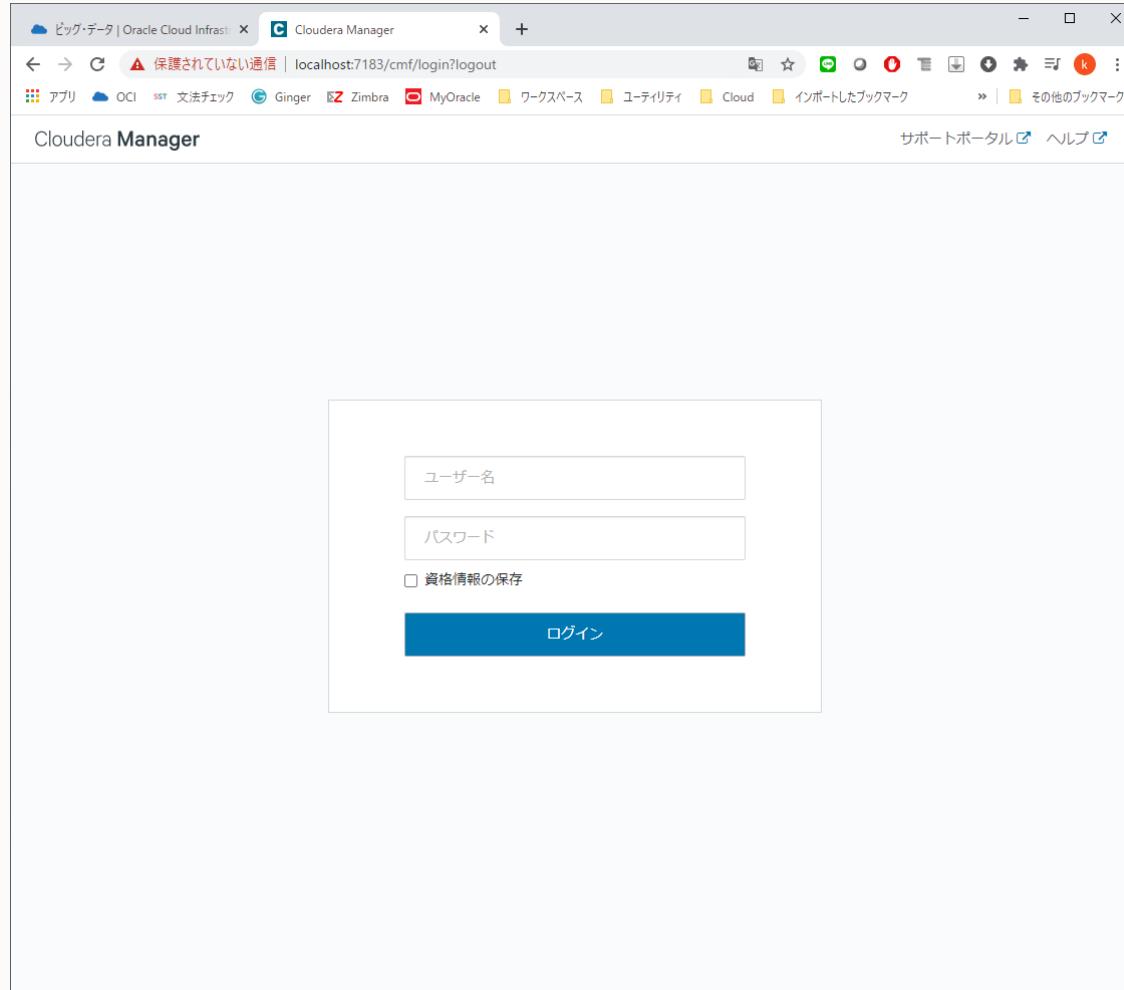
VM.Standard2.8

VM.Standard2.16

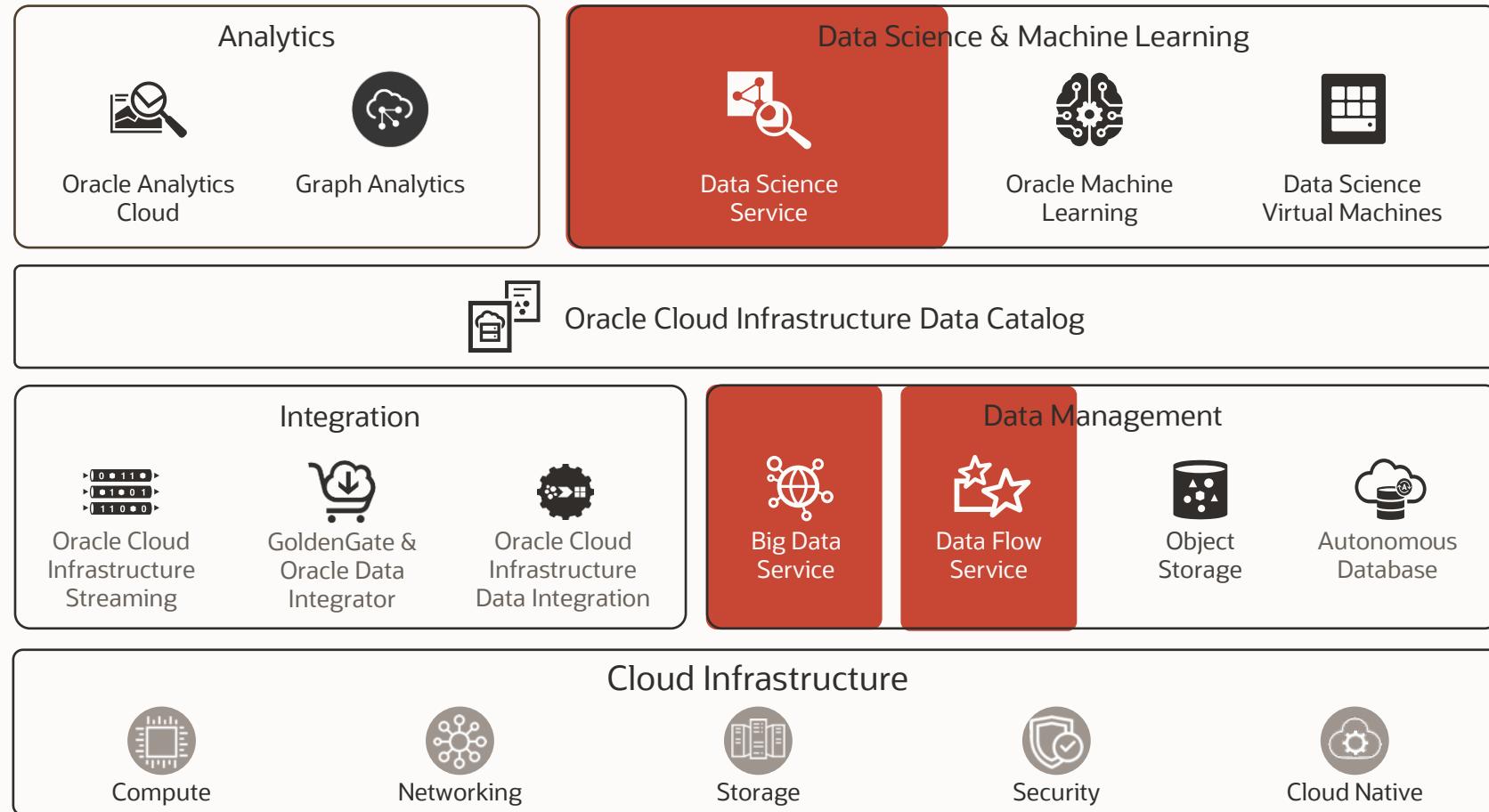
VM.Standard2.24

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



Oracle Data and AI Platform



Applications

Business Analytics

ORACLE

O