



ちょっぴりDD

AWS SDK for Pandas

Kensuke Fukumoto

IISV / SaaS SA
AWS Japan G.K

自己紹介

◆ 所属

- ◆ ISV / SaaS SA
- ◆ ソフトウェアベンダーや SaaS 事業者の支援

◆ 関心領域

- ◆ SaaS, ML, ストリーミングアプリケーション

◆ 好きなAWSサービス

Amazon Simple Queue Service (SQS)
Amazon Kinesis Data Streams



◆ 趣味

バイクで郊外の銭湯に行ってサウナに入ること。皇居ラン。

どんな環境で
データ分析していますか？



様々なデータ分析環境

ローカルPC



仮想化されたマシンを共有



クラウド



それぞれが慣れた環境で慣れたツールを利用

こんな課題ありませんか？



データ分析を普段の業務で行なっていて

ローカル PC



仮想化されたマシンを共有



リソース不足で業務効率が低下

クラウドでスケーラブルな環境を手に入れる



クラウド
使いませんか??



いつものツールが
使いたくて...

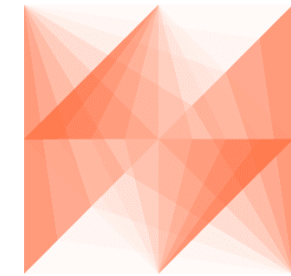
そんな時の解決策は...



あるよ

AWS SDK for pandas

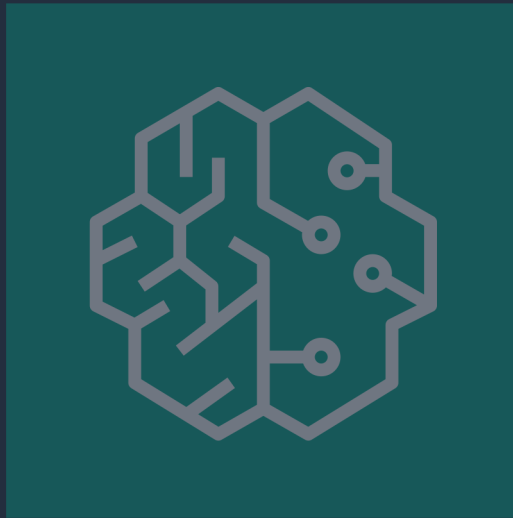
<https://aws-sdk-pandas.readthedocs.io/en/stable/>



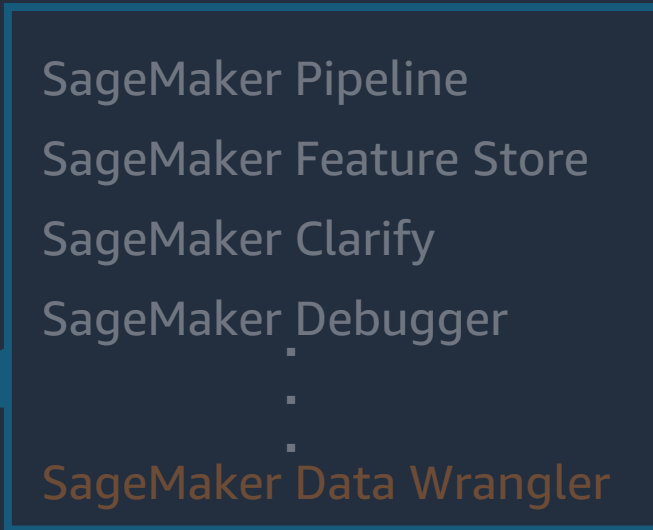
AWS SDK for pandas

AWS SDK for pandas (旧 AWS Data Wrangler)

AWS のデータと分析サービスと Pandas DataFrame 間のやり取りが簡単に行える Python ライブラリ
(インストール時に指定する名前は引き続き awswrangler)



Amazon SageMaker



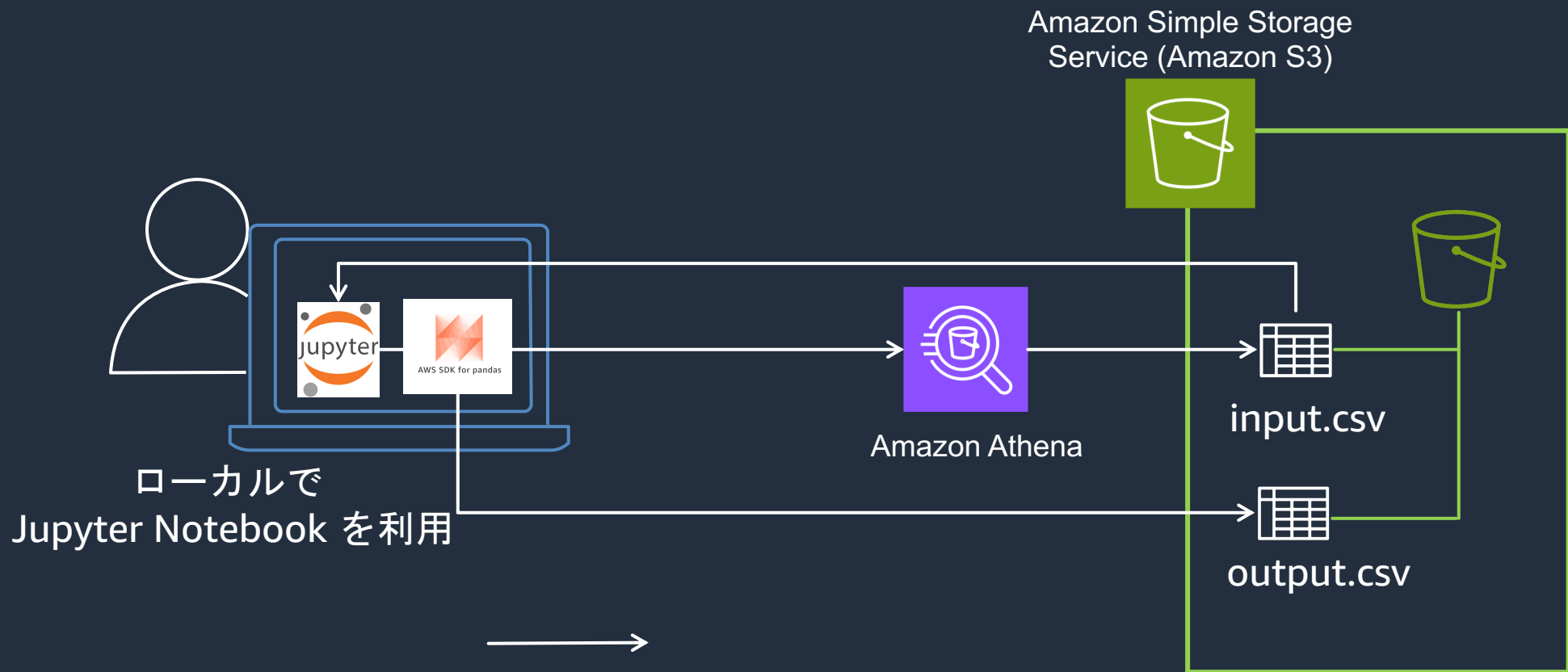
SaeMaker の機能群の 1 つ



ライブラリ

今日はこちらの話

シンプルで典型的なユースケース



デモのちよっぴり説明



デモ環境など

❖ 実行環境

- ❖ SageMaker Notebook (ローカル環境の代わり)

- ❖ Conda_python3 (Python 3.10.10)

❖ AWS SDK for Pandas

- ❖ awswrangler-3.2.1

❖ デモ内で使うデータ

- ❖ New York City Taxi and Limousine Commission (TLC) Trip Record Data (2018)

<https://registry.opendata.aws/nyc-tlc-trip-records-pds/>

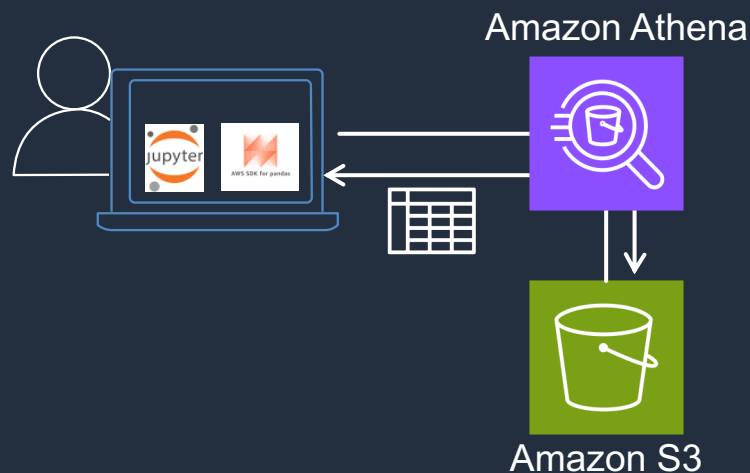
デモの内容

- デモ ① -



ローカルとクラウドで
データフレームをやり取り

- デモ ② -



Athena でのクエリを
シームレスに実施

- デモ ③ -



その他時間があれば...

まとめ

- ✓ AWS SDK for Pandas の紹介をしました。
 - ✓ 使い慣れたツールとシームレスにクラウドを統合
 - ✓ バックエンドはさまざまなサービスや機能、データタイプと連携
- ✓ ちょっぴりとデモを行いました。
 - ✓ Pandas データフレームとS3とのデータ連携
 - ✓ ローカルの処理を Athena にオフロードして実行環境をスケール



Thank you!