aws
SUMMIT
ONLINE
JAPAN | MAY 11-12, 2021

ビジュアルツールで簡単に始められるデータ加工

川村 誠

ソリューションアーキテクト, アナリティクス

アマゾン ウェブ サービス ジャパン株式会社

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

自己紹介

川村 誠

ソリューションアーキテクト

- データ分析サービスを担当
- 好きなサービス
  - Amazon EMR
  - AWS Glue
  - Amazon Athena

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

アジェンダ

- データ加工で困っていることはありませんか
- AWS Glue DataBrew
- デモ
- ユースケース
- 料金
- まとめ

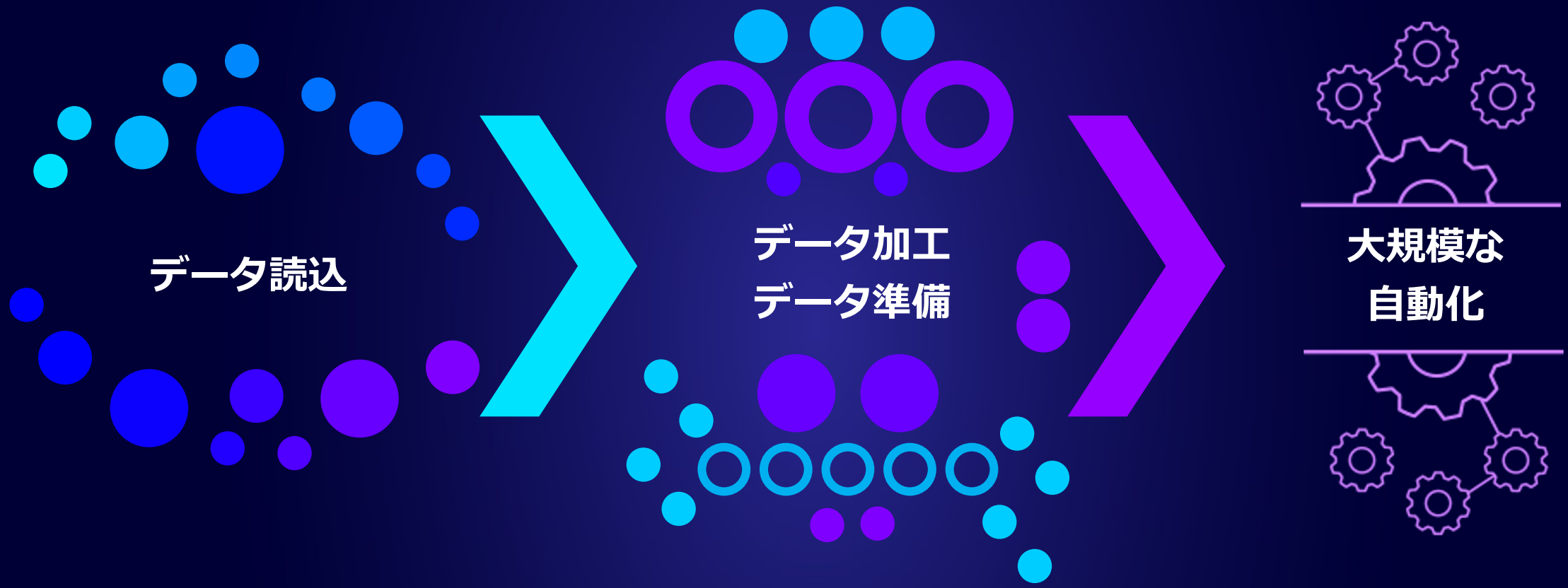
データ加工で困っていることはありませんか

データ加工で困っていることはありませんか

- 自部署の商品売り上げレポートを毎週まとめているけど、データ量が増えすぎて Excel で処理できなくなった
- 複数のデータソースを結合しているので処理の内容がよくわからなくなってしまった
- 手作業でレポートを作っており、毎週半日くらいがレポート作成に費やされている

…など

データ加工・準備には複雑な作業が必要



大規模に活用するためには複雑な処理の仕組み化も必要

AWS Glue DataBrew

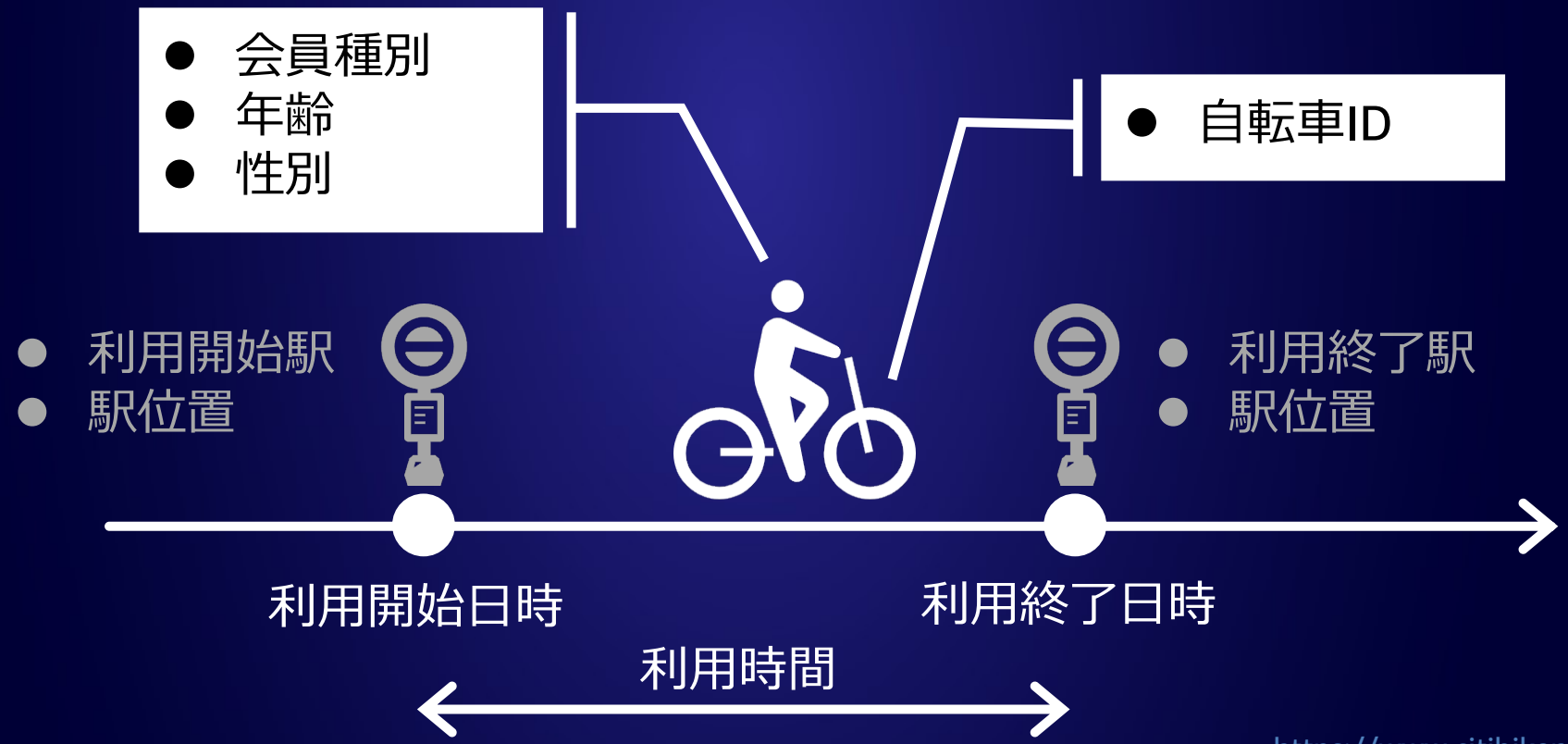
AWS Glue DataBrew

ビジネスユーザの方も簡単にすぐに始められて、
データの準備・加工にかかっていた手続きを高速化する
ビジュアルデータ準備ツール

デモ

デモシナリオ

- このデモでは、NY のレンタル自転車シェアプログラムの利用履歴データを使用し、データの加工整形や集計処理を実施します
- 新しい知見を得るため気象情報を追加し、集計処理を実施します



https://www.citibikenyc.com /system-data

AWS マネジメントコンソール

AWS のサービス

▼ 最近アクセスしたサービス

- S3
- CloudFormation
- Elastic Container Service
- IAM
- VPC
- Systems Manager
- Cloud9
- EMR
- AWS Compute Optimizer
- AWS Glue DataBrew
- Elastic Kubernetes Service
- Amazon MQ
- RDS
- EC2
- CloudWatch

▶ すべてのサービス

ソリューションの構築

シンプルなウィザードと自動化されたワークフローで作業を開始しましょう。

仮想マシンを起動する
EC2 を使用
2〜3 分

ウェブアプリを構築する
Elastic Beanstalk を使用
6 分

仮想サーバーを使用して構築する
Lightsail を使用
1〜2 分

ドメインを登録する
Route 53 を使用
3 分

IoT デバイスを接続する
AWS IoT を使用
5 分

AWS への移行を開始する
CloudEndure Migration を使用する
1〜2 分

外出先でも AWS リソースに常時接続

AWS コンソールモバイルアプリを iOS または Android モバイルデバイスにダウンロードします。 詳細はこちらから

AWS を試す

Amazon Redshift
データレイクにクエリを拡張できる、高速かつシンプルで、費用対効果の高いデータウェアハウス。
詳細はこちらから

AWS Fargate を使ってサーバーレスコンテナを実行
AWS Fargate が実行され、サーバーやクラスターを管理することなく、コンテナをスケーリングします。
詳細はこちらから

Amazon S3 によるスケーラブルで耐久性に優れたセキュアなバックアップと復元
他のお客様が AWS でどのようにコストを削減するバックアップおよび復元ソリューションを構築しているかご覧ください。 詳細はこちらから

AWS Marketplace
AWS で実行される人気のソフトウェア製品を検索、購入してデプロイしましょう。 詳細はこちらから

フィードバックがおありですか?

© 2008 - 2021, Amazon Web Services, Inc. またはその関連会社。無断転用禁止。

ユースケース

ユースケース１

- 外部から買ってきたマーケティングデータを追加したい
- 商品名の変更や販売店舗店舗移転があったので、手元にあるデータパッチを当てたい

新しい情報の追加

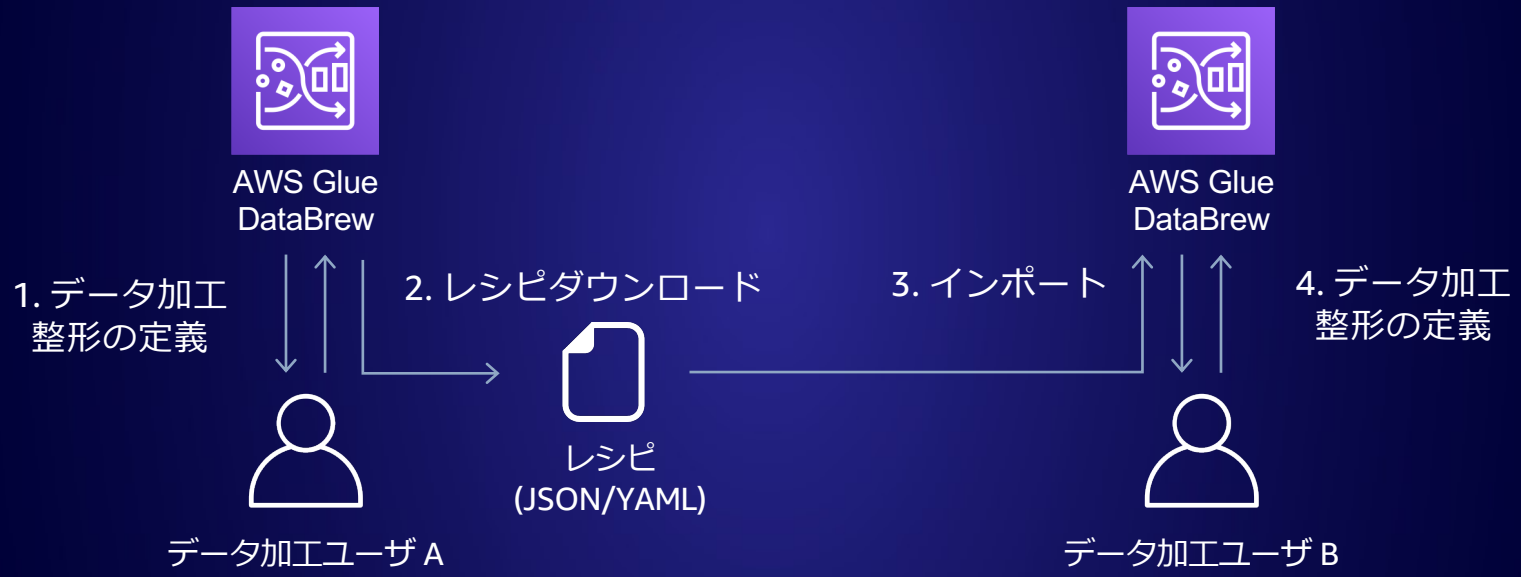
ユースケース 1: 新しい情報の追加

ユースケース 2

- 別プロジェクトの加工整形手順をベースに新しい加工整形手順を作りたい
- あの時の加工整形手順を使いまわしたい

加工整形手順(レシピ)の共有
/ 再利用

ユースケース 2: 加工整形手順(レシピ)の共有



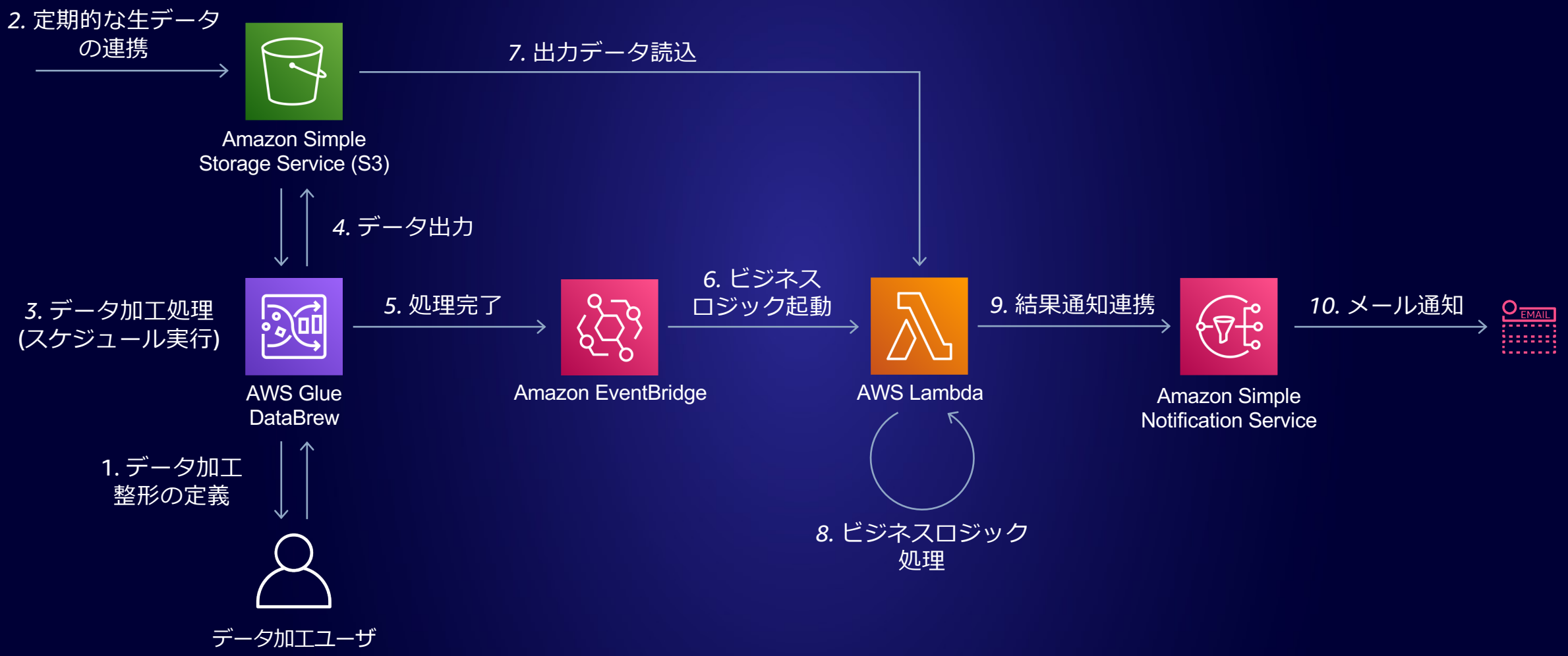
ユースケース 3

- 手作業でレポートを作っており、毎週半日くらいがレポート作成に費やされている
- 月次でレポートしていた情報をデータ更新に応じてタイムリーにお客様に届けたい

データ加工整形の自動化

ユースケース 3: データ加工の自動化

ユースケース 3: データ加工の自動化



料金

AWS Glue DataBrew の料金

- DataBrew インタラクティブセッション
  - ビジュアルツール利用時間に基づきセッション(30分)毎に請求される
  - 利用料金は 1.00 USD / セッション
  - DataBrew プロジェクトを開くとセッションが開始され、未操作の時間が続いた場合、自動的にサスペンドする

(料金の例)

プロジェクトを開いてから 90 分間ビジュアルツールを利用した場合、請求額は 3 USD = 1.00 (USD) x 90 / 30 (セッション)

https://aws.amazon.com/jp/glue/pricing/

AWS Glue DataBrew の料金

- DataBrew ジョブ
  - ❖ ジョブの実行に使用された AWS Glue DataBrew ノードの数に基づき、1 分毎に請求される
  - ❖ 利用料金は 0.48 USD / DataBrew ノード時間
  - ❖ デフォルトは各ジョブに 5 ノード 割り当てられる

(料金の例)

初期設定でジョブが 12 分で完了した場合、
請求額は 0.48 USD = 5 (ノード) x 0.48 (USD) x 12 / 60 (時間)

https://aws.amazon.com/jp/glue/pricing/

すぐに始められます

- AWS Glue DataBrew は、東京リージョンで利用可能
- 初めてのご利用の場合、ビジュアルツールを無償で 40セッション(20時間) 利用可能

https://aws.amazon.com/jp/glue/pricing/

まとめ

まとめ

- AWS Glue DataBrew は、ビジネスユーザの方も簡単にデータを読み込み、加工整形をすぐに始められるビジュアルツール
- 実施したデータ加工手順をレシピとして保存し、共有可能
- AWS Glue DataBrew が提供する機能を利用することでデータ加工整形を効率化し、ビジネス価値創出により多くのリソースを注力可能に

※ 気象情報について

気象庁「世界の天候データツール(ClimatView 日別値)」(https://www.data.jma.go.jp/gmd/cpd/monitor/dailyview/index.php) より
2020年のニューヨークのデータを加工して作成・利用しております

Thank you!

川村 誠

kawamako@amazon.co.jp

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Please complete the session survey

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.