Amazon Lookout for Metrics ワークショップ

2021/05/21

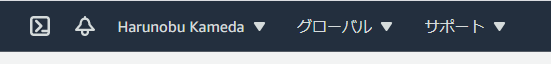
シニアエバンジェリスト 亀田

[S3バケットの準備とデータのアップロード]

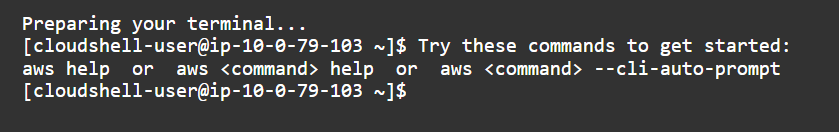
* 適当な名前でS3バケットを作成し、その名前を手元のメモにコピーしておきます。設定は全てデフォルトで問題ありません。



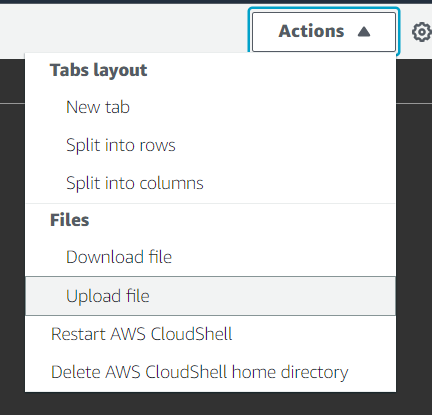
* Githubのレポジトリから[ecommerce.zip]をダウンロードします
* CloudShellを起動します（AWSマネージメントコンソール右上です）



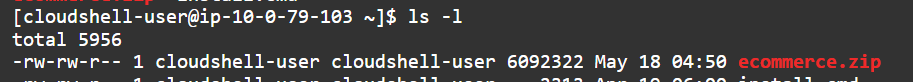
* 起動にしばらく時間がかかるので待つと以下が表示され使えるようになります。



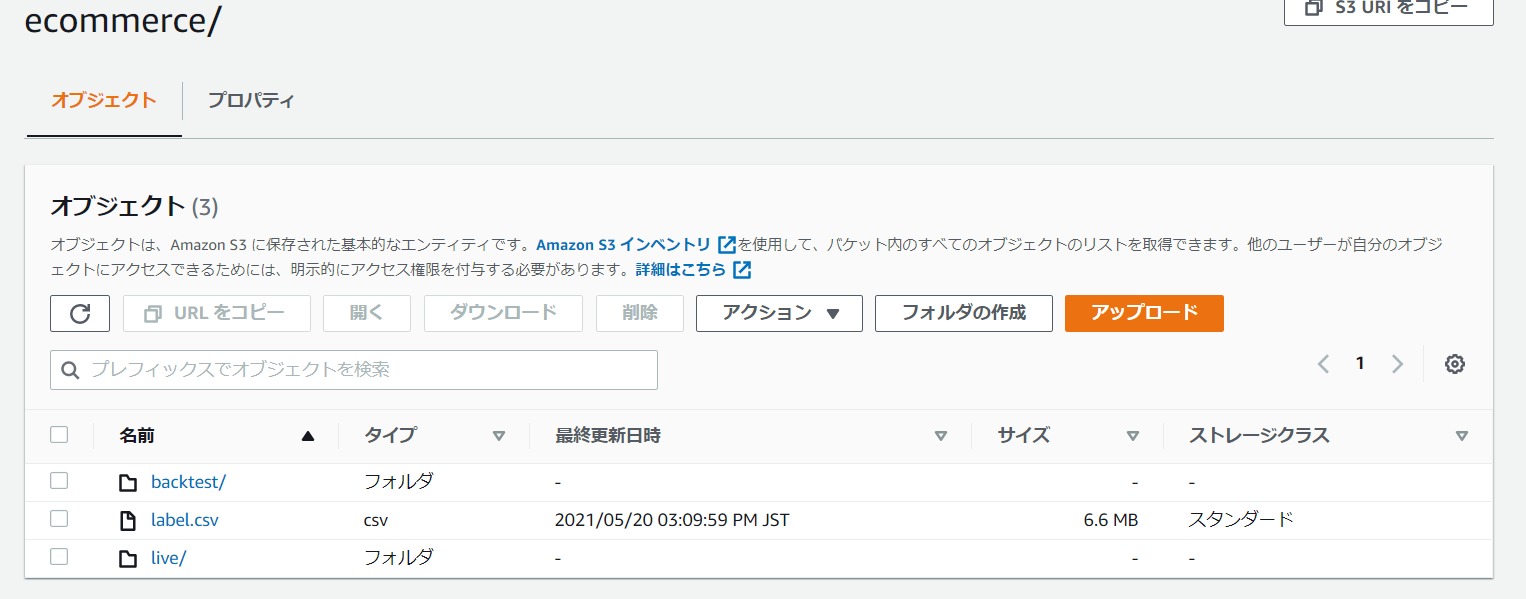
* 画面右上の[Actions]から[Upload File]を選びます



* [ecommerce.zip]をアップロードします。念のためls –l コマンドで確認しておきます。

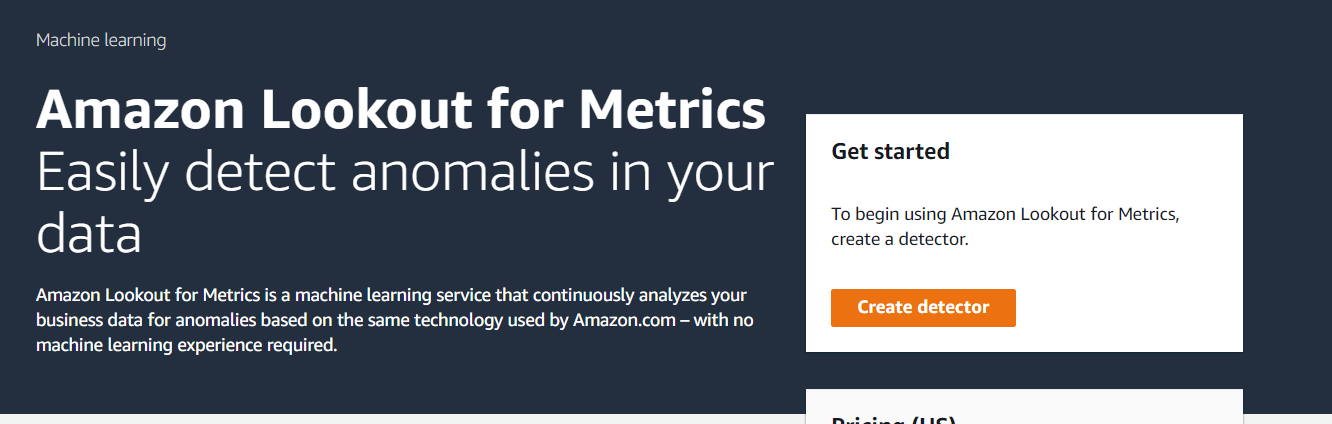


* [unzip ecommerce.zip]で解凍します
* [aws s3 sync ./ecommerce/ s3://{s3\_bucket}/ecommerce/]を実行し、作成されたディレクトリを纏めてS3にアップロードします。（S3マネージメントコンソールでもアップロード可能ですが数十倍高速です）。{s3\_bucket}は先程作成したS3バケット名に置き換えます
* アップロードが完了したらS3バケットを見てみます。Ecommerceファルダの下に1つのファイルと2つのフォルダができていれば成功です。

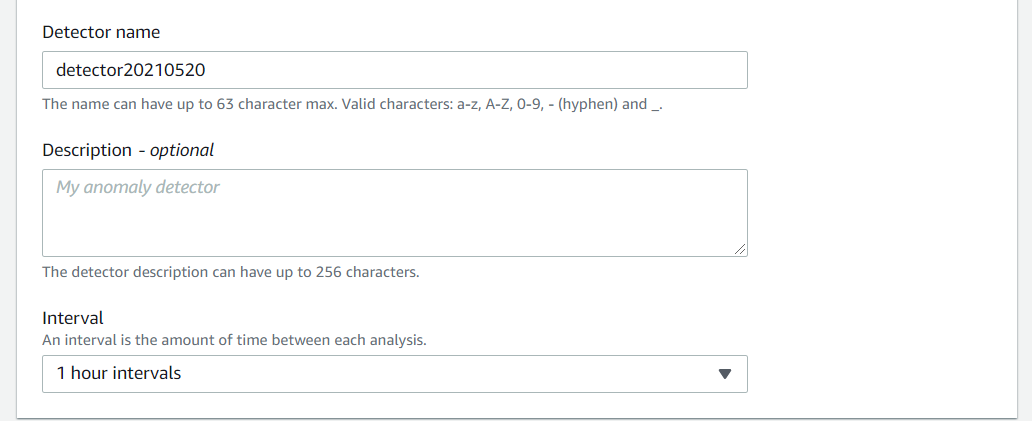


**[Lookout for MetricsでDetector（検出器）を作成]**

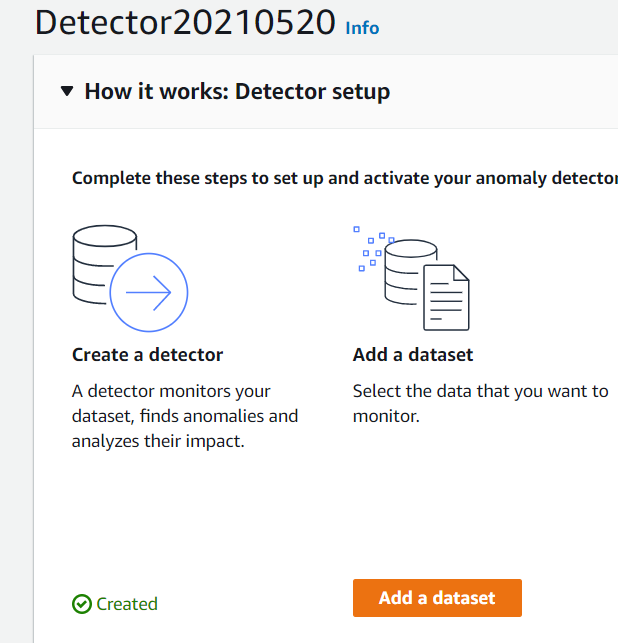
* Management ConsoleでLookout for Metircsの画面に行き、[Create detector]を押します



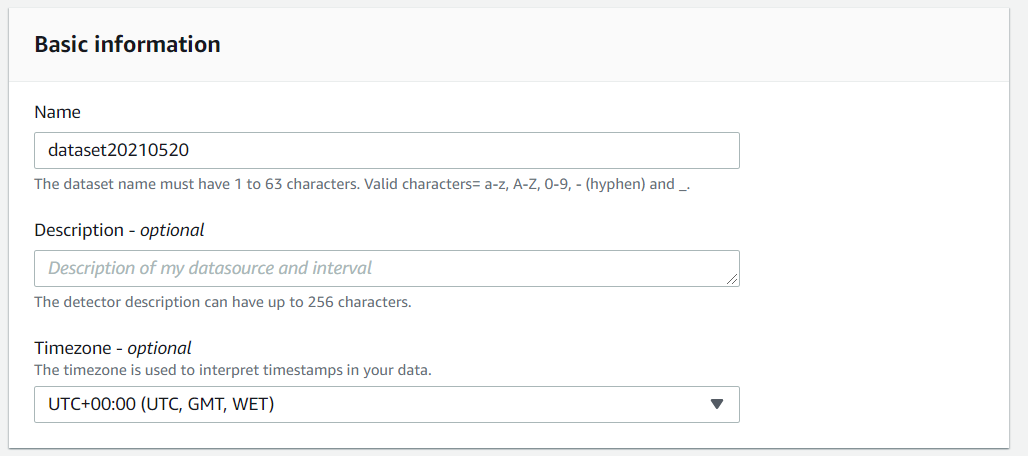
* [Detector Name]に適当な名前を付け、[Interval]を[1 hour]に設定します。このIntervalは皆さんが用意したデータ間隔と一致している必要があります



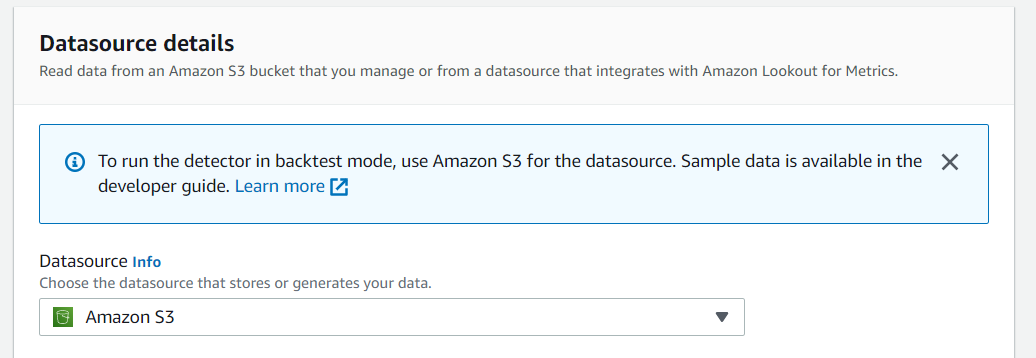
* [Create]を押します。



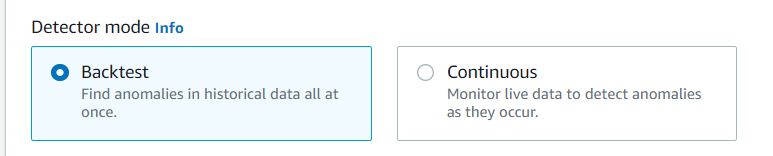
* Detecorができましたので、続いて[Add a dataset]を押して、学習準備を設定します
* 適度な名前を[Name]に設定します



* [Datasource]にAmazon S3を選択します



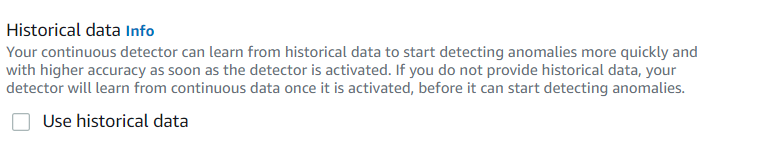
* [Detector Mode]をBacksetに設定します



解説：Continuousモードはライブデータをもとに学習、そして異常検知を行います。先程S3バケットにアップしたデータのliveフォルダには以下のように1時間ごとに売り上げデータが命名規則に従い別ファイルとして生成されています。



Continuousモードでは1時間間隔でこの中身を用いて、学習＆推論を行います。皆さんのapplicationは命名法則にしたがい設定したInterval間隔でログを吐き出すようにしておく必要があります。命名法則にしたがう限り、設定不要で自動でファイルを読み込み続けます。この際精度を向上させるために、[Historical Data]として過去のデータを学習させることも可能です。勿論この場合過去のデータと今のapplicationへのリクエストにおいて、変更が加わってないことが前提になります。



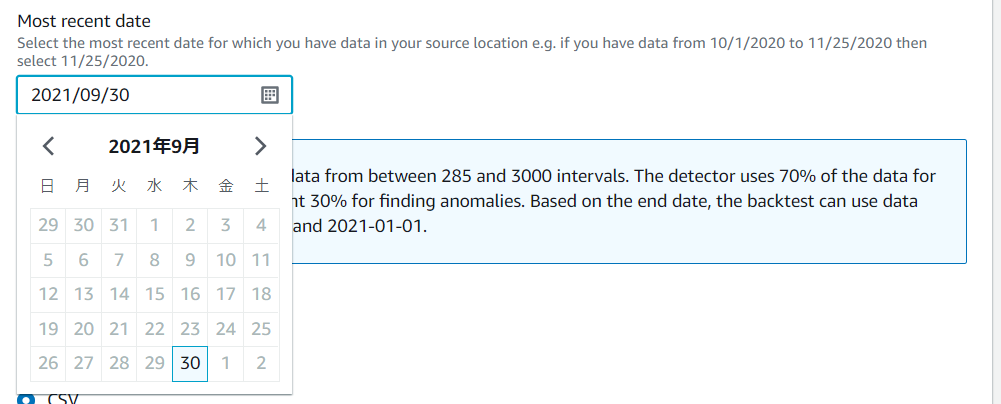
このハンズオンではContinuousモードの場合、最低でも数時間待たないと結果が出力されないため、[Backtest]モードを使用します。

* [Historical data]に以下の値を入力します。{bucketname}は今日作成のものに置換してください

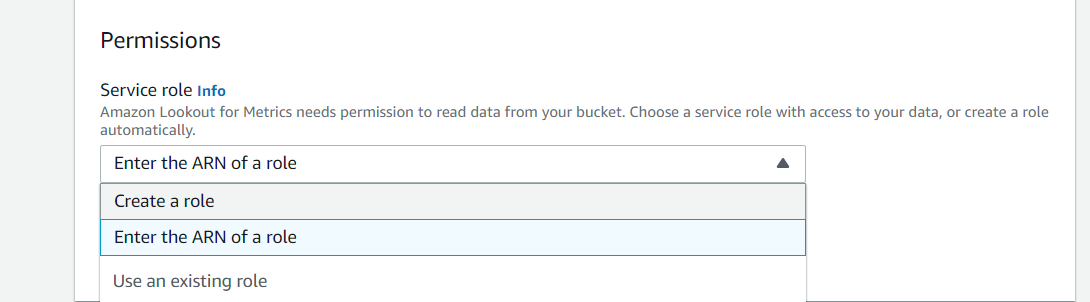
s3://{bucketname}/ecommerce/backtest/

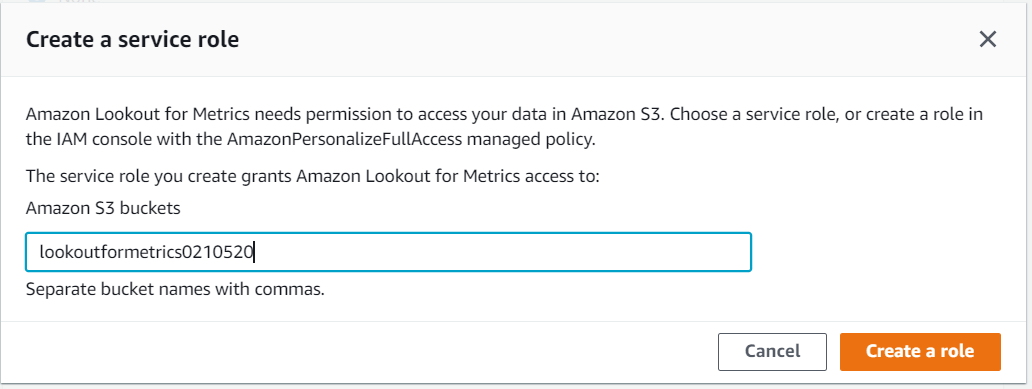
ここで指定されたデータの70%が学習に使用され、残り30%が推論に使用されます。また設定したIntervalの285回分のデータ（今回で言うと285時間分）が最低必要です。

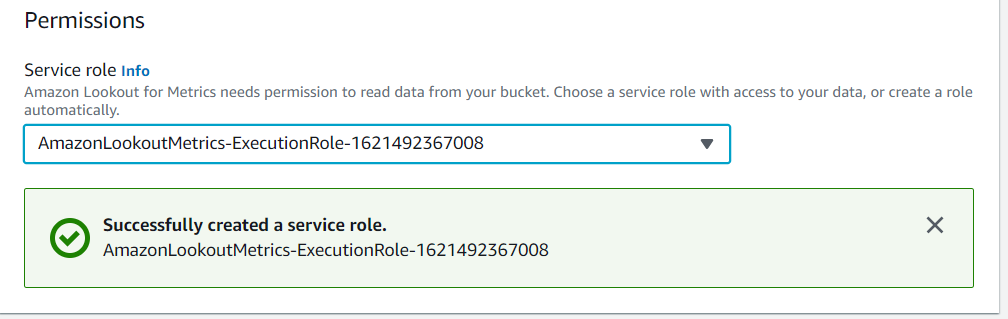
* [Most recent date]に[2021/09/30]と設定します



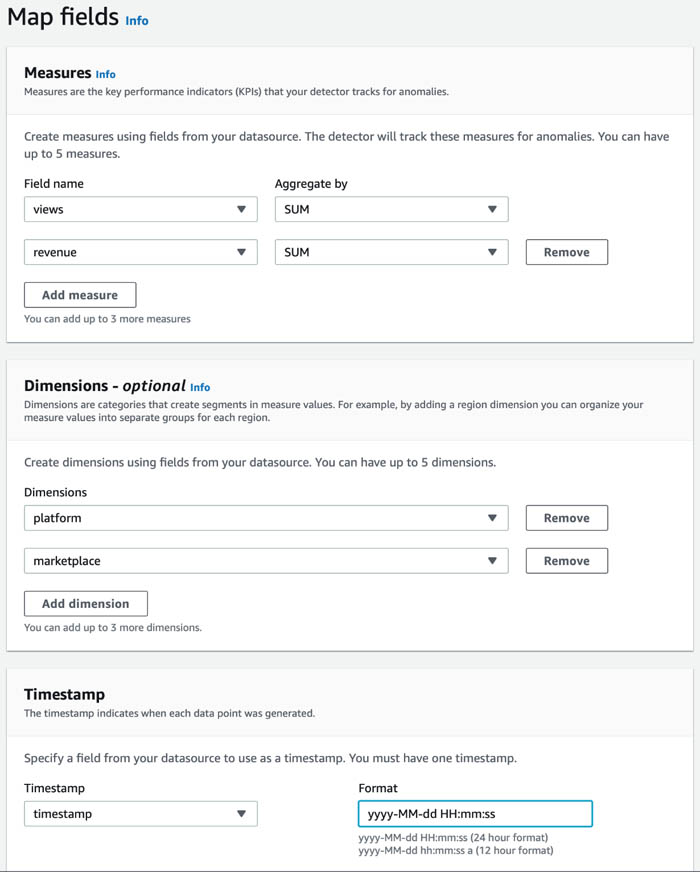
* [Permission]では、[Create a role]を選びます。現れたダイアログにS3バケット名を入力し、[Create a role]を押します





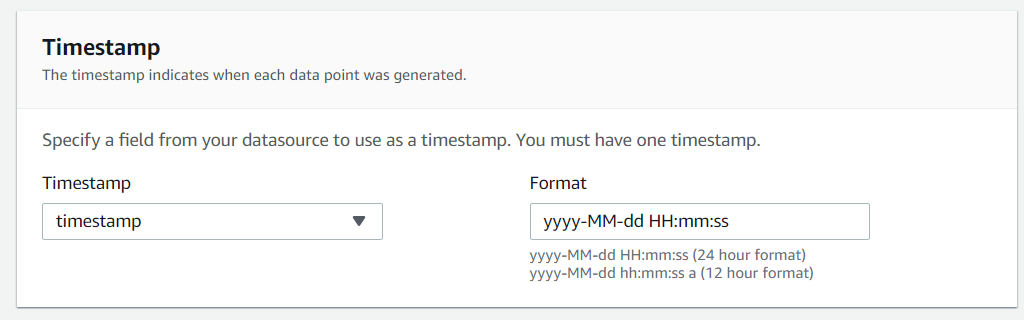


* [Next]を押します。（表示されるダイアログでエラーとなる場合、IAM作成が間に合っていない可能性がありますので、もう一度Nextボタンをおしてください）
* 次の画面の[Measures]と[Dimensions]を以下の画像のように設定します

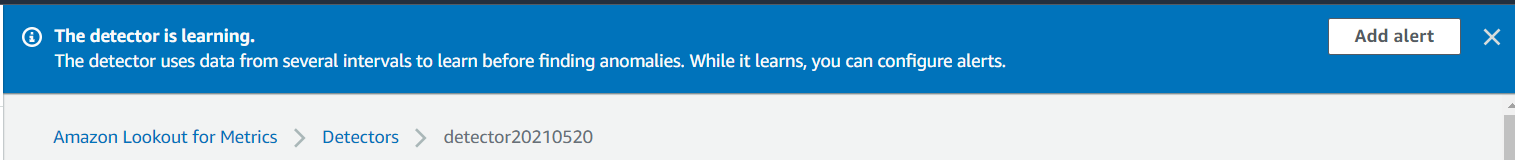


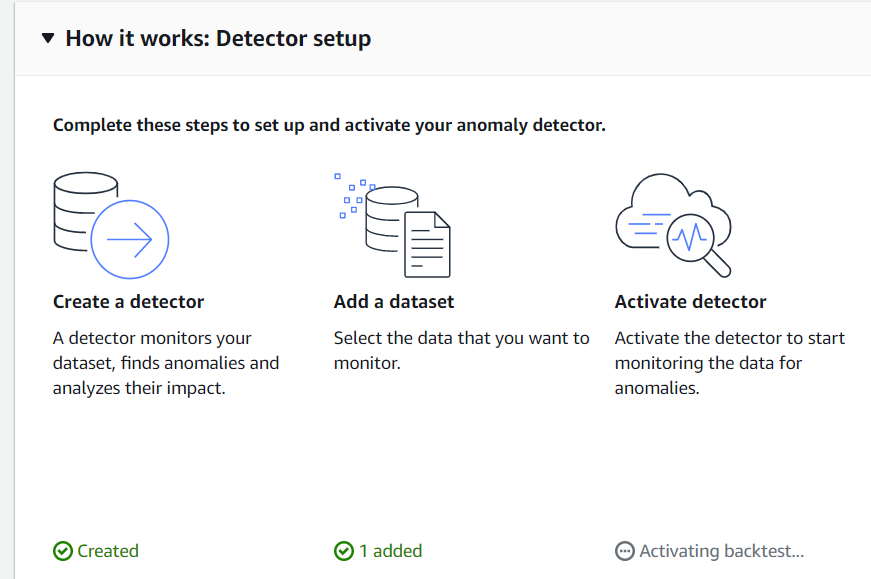
解説：Measureは異常検知を行う対象となる値です。Dimensionsはそれに影響を与える可能性のある項目です。例えば対象データがGlobalで展開されているECサイトの売り上げ及びページビューとした場合、Dimensionsには国やプラットフォーム（PC向けやモバイル向け等）をセットすることで、検知された異常に影響を与えているパラメータを知ることができます

* Timestampに[yyyy-MM-dd HH:mm:ss]を設定します。



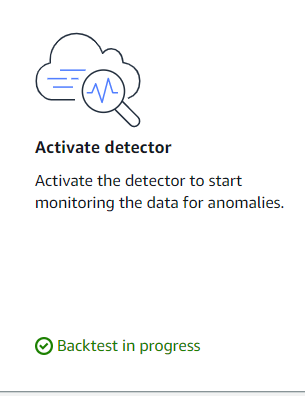
* [Next]をおします。次に表示される確認画面で[Save and activate]を押します。さらに表示されるダイアログで[Active]を押すと、学習が開始されます



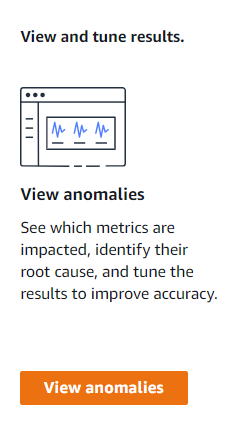


数十分（25分程度ですが、環境、タイミングにより変動します）かかるので待ちます

* しばらく待つと学習が完了しステータスが変更になります。



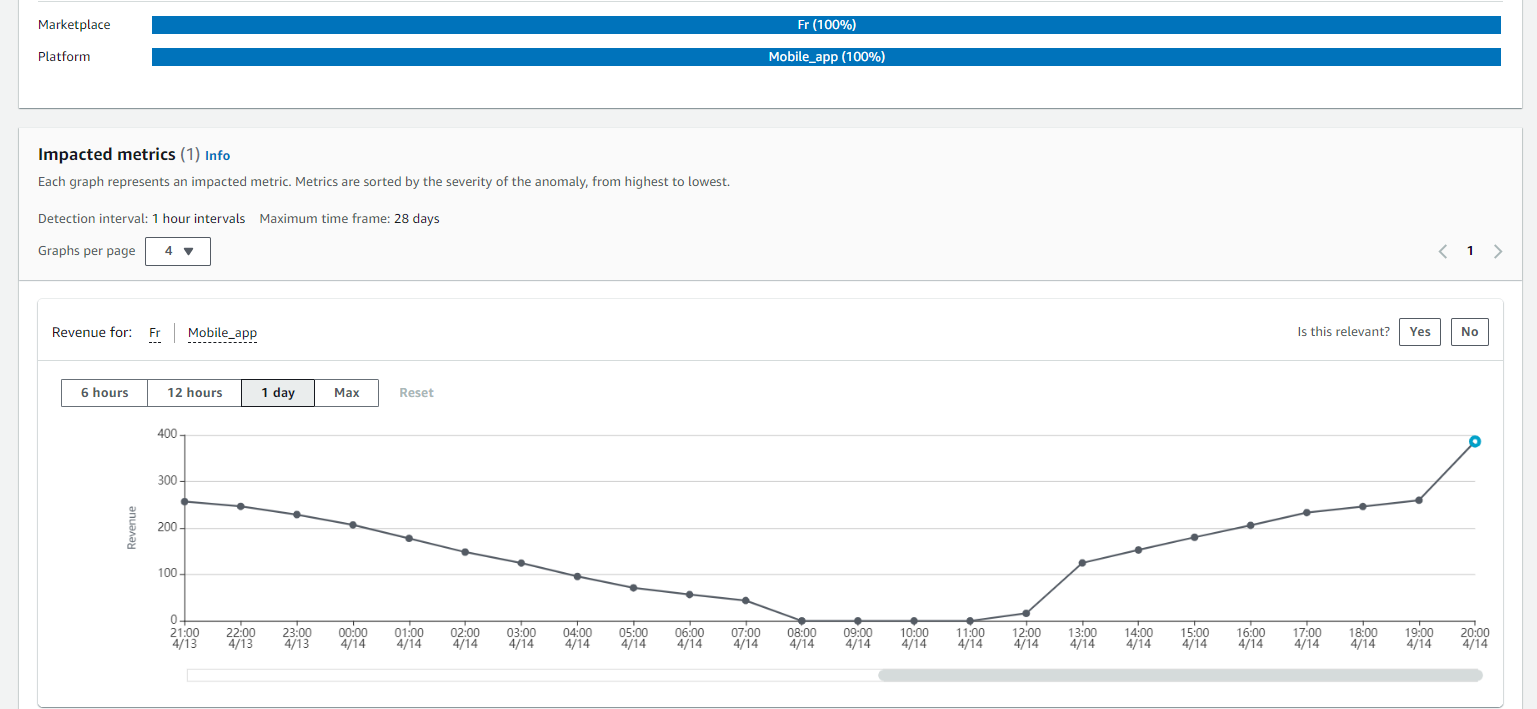
推論が開始され始めていますので、[View anomalies]を押すと検知された異常が表示されています。



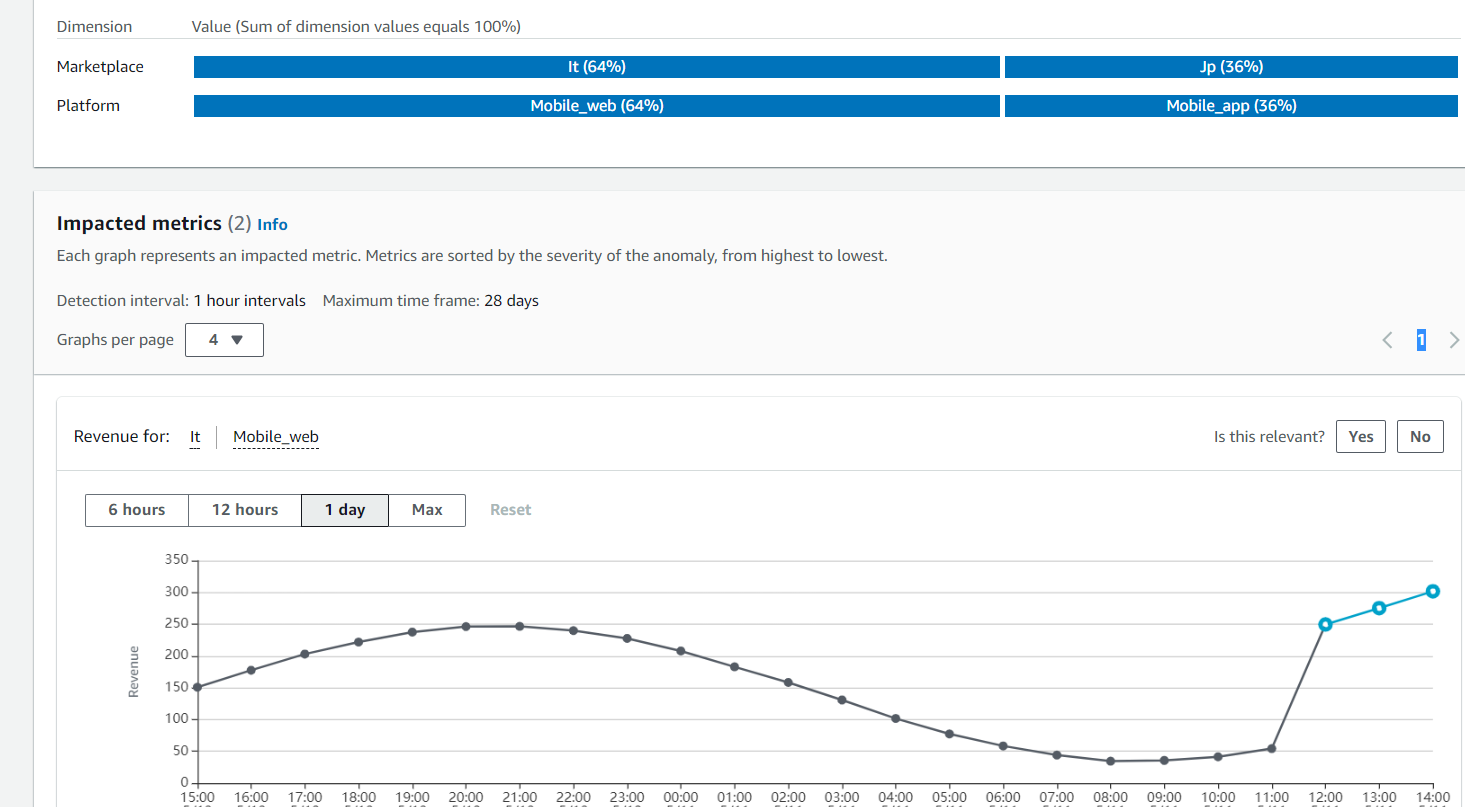
* [Severity threshold]のつまみを調整すると、異常値をもとに絞り込むことができます



* 異常は、中央値からの乖離であり、良い物悪いもの両方が含まれています。例えば以下の場合、なにがしかの理由でフランス売り上げが伸びたことがわかります。



多くの異常は単一Dimensionsにより構成されていますが、中には以下のように複数要因が絡んでいるものも存在しています。



これにより更なる調査を行うための絞り込みなどが容易になります。

おつかれさまでした！削除は以下を行ってください

* Detector
* S3 バケット
* IAM Role