### データセット

#### 概要

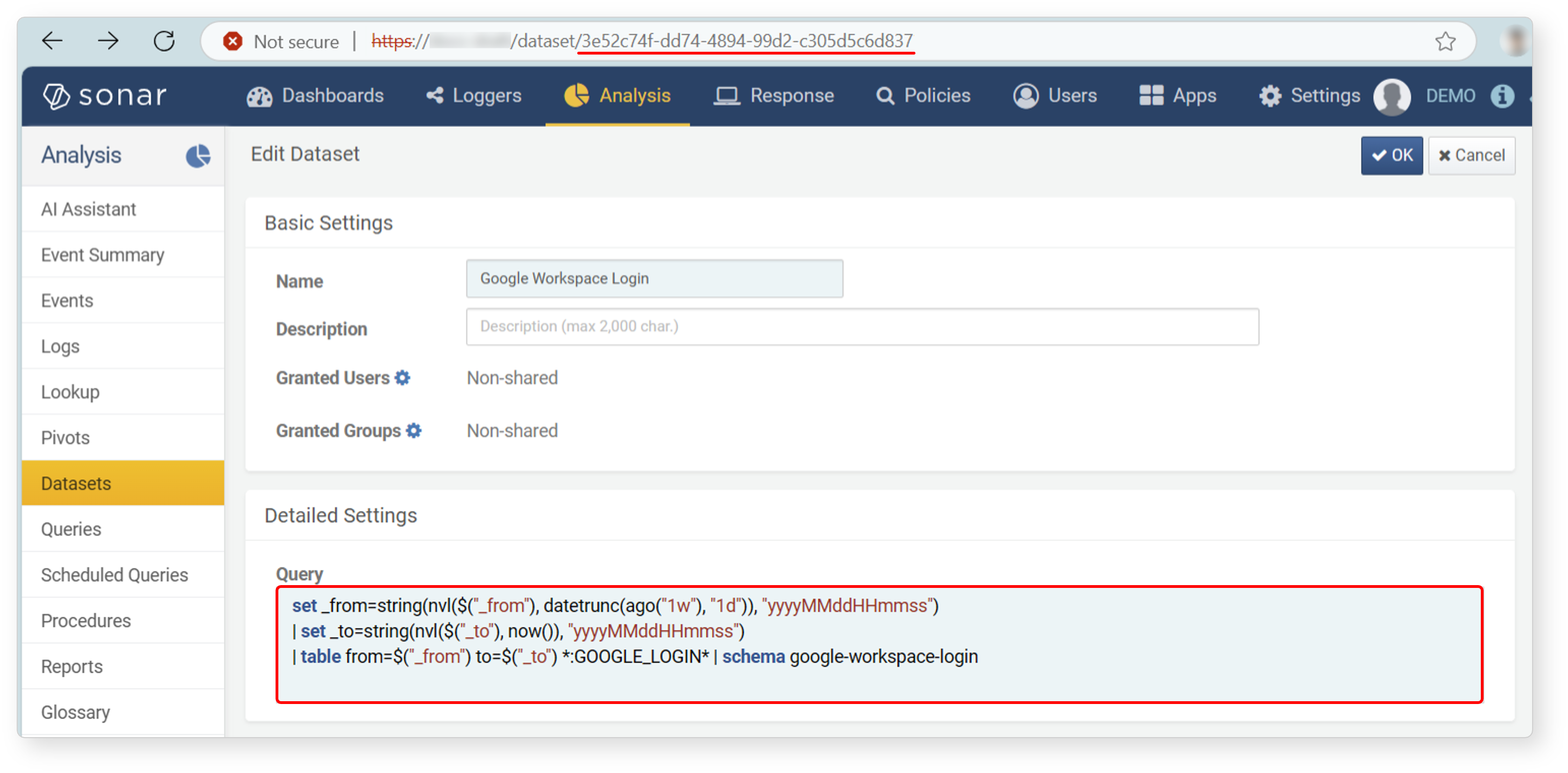
データセットは、[ピボット](https://docs.logpresso.comnull)や[ダッシュボードウィジェット](https://docs.logpresso.comnull)など、定期的なデータ更新が必要な分析機能において基礎データソースとして使用されます。

データセットの作成

データセットの中核は、データの取得と処理に使用される[ログプレッソクエリ](https://docs.logpresso.comnull)です。データセットは通常、時間範囲を指定してデータを取得するため、クエリは[table](https://docs.logpresso.comnull)または[fulltext](https://docs.logpresso.comnull)コマンドのdurationオプションで始まり、その後[スキーマ](https://docs.logpresso.comnull)コマンドでログスキーマを適用する形が一般的です。

以下は、[Google Workspace](https://logpresso.store/ja/apps/google-workspace)アプリ（バージョン1.7.2504.0）に含まれる「Google Workspace ログイン」データセットで使用されているクエリ例です。

set \_from=string(nvl($("\_from"), datetrunc(ago("1w"), "1d")), "yyyyMMddHHmmss")| set \_to=string(nvl($("\_to"), now()), "yyyyMMddHHmmss")| table from=$("\_from") to=$("\_to") \*:GOOGLE\_LOGIN\* | schema google-workspace-login



すべてのデータセットは、Webブラウザのアドレスバーに表示されるGUIDで識別できます。

以下は、[FortiGate](https://logpresso.store/ja/apps/fortigate)アプリ（バージョン1.3.2502.0）に含まれるFortiGate SSL VPNデータセットのクエリ例です。

fulltext tt=t duration=1d subtype == "vpn" and log\_desc == "SSL VPN\*" from \*:FW\_FORTIGATE\* | schema fortigate-sslvpn

データセットクエリコマンド

保存済みデータセットを利用するには、[dataset](https://docs.logpresso.comnull)コマンドを使用します。datasetコマンドは、guidオプションにデータセットのGUIDを指定してデータを取得します。このコマンドは、データセット出力を利用するダッシュボードウィジェットのクエリでよく使用されます。以下は、[Google Workspace](https://logpresso.store/ja/apps/google-workspace)アプリに含まれる**Google Workspace ログインタスク統計**ウィジェットの例です。

dataset guid="3e52c74f-dd74-4894-99d2-c305d5c6d837" | pivot count() rows Task | limit 10000

クエリロジックによっては、実行のたびに結果が変動する場合があります。例えば、過去1時間に収集されたログを基に統計を算出するよう定義されたデータセットは、異なるタイミングでクエリを実行すると、常に最新の1時間分のデータを反映した異なる結果が得られます。

ユーザー権限

ログプレッソ・ソナーのユーザーアカウントを持つすべてのユーザーは、自身のデータセットを作成・利用できます。ただし、意味のあるデータセットを構築するには、基礎となるログテーブルへのアクセス権限が必要です。

ダッシュボードウィジェットでデータセット結果を表示する場合、ダッシュボードの共有設定でアクセス権限を管理できます。そのため、データセット単位で個別に権限を設定する必要はありません。

アプリデータセット

[アプリ](https://docs.logpresso.comnull)をインストールすると、そのダッシュボードウィジェットをサポートするために必要なデータセットが同梱されます。これらのデータセットは、アプリの再インストールやアップデート時に自動的に更新されます。

アプリインストール時に提供されるデータセットの所有者は、アプリをインストールした管理者アカウントとなります。標準ユーザーアカウントがこれらのデータセットを利用（例：カスタムウィジェットの作成やクエリでの利用）できるようにするには、該当ユーザーまたはユーザーグループとデータセットを共有する必要があります。

アプリに同梱されているデータセットのクエリ文を確認すると、以下のようなテーブルパターンがよく見られます。

\*:TABLE\_NAME\*

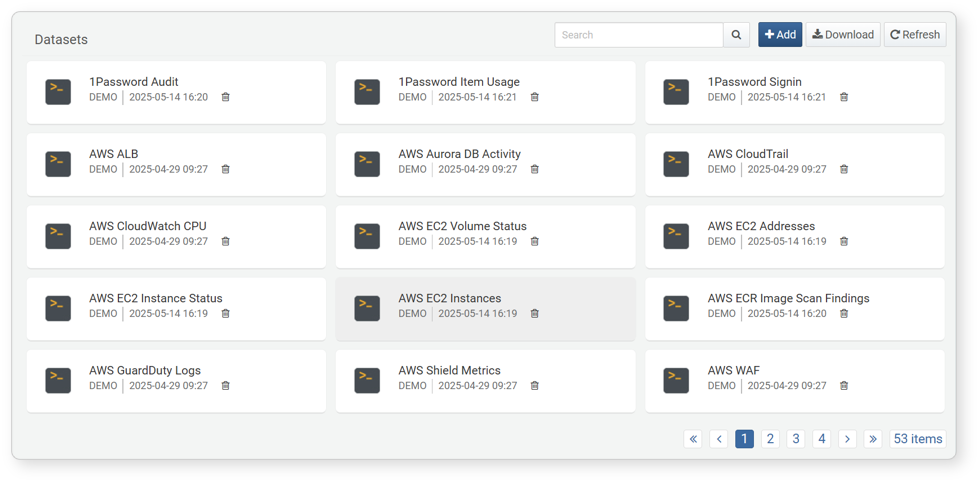
\*はすべてのデータノードからのデータを示します。テーブル名の末尾の\*はワイルドカードで、任意のサフィックスにマッチします。テーブル名は通常、FW（ファイアウォール）、NAC（ネットワークアクセス制御）、WEB（Webサーバー）、NDR（ネットワーク検知＆レスポンス）、EDR（エンドポイント検知＆レスポンス）など、データソースの分類で始まります。

ログプレッソでは、テーブル名にアプリ識別子や追加サフィックスを付与してデータソースを明示することを推奨しています。例えば、冗長構成のPalo Altoファイアウォールが2台ある場合、FW\_PALOALTO\_ACTIVEとFW\_PALOALTO\_STANDBYのように命名します。

ロガー追加時は、必ずアプリのデータセットクエリを確認し、正しいテーブル命名規則を使用しているか確認してください。

#### データセットの検索

**分析 > データセット**で、データセット一覧の閲覧や検索が可能です。



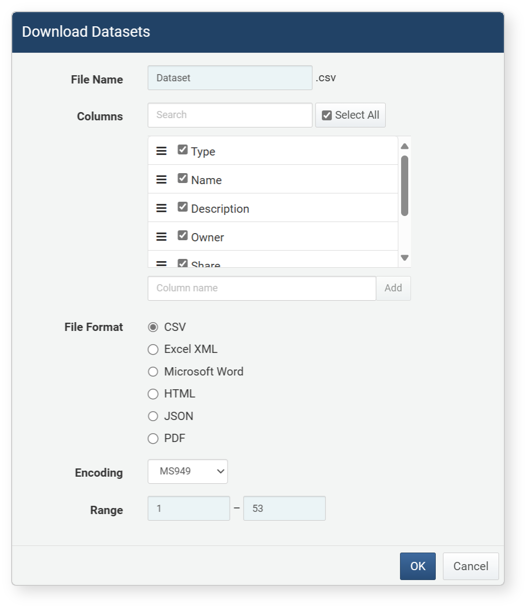
データセット一覧から特定のデータセットを探す場合は、ツールバーの検索ツールを利用します。検索ツールは、入力した単語を**名前**に含むデータセットを検索します。大文字・小文字は区別されません。

データセット一覧のダウンロード

データセット一覧をダウンロードするには、以下の手順を実行します。

ツールバーの**ダウンロード**をクリックします。

**データセット一覧ダウンロード**ダイアログで、ファイルに含めるデータセットプロパティを入力または選択し、**ダウンロード**をクリックします。



* **ファイル名**：ダウンロードするデータセット一覧ファイル名（デフォルト：Ticket）
* **カラム**：ファイルに含めるデータセットプロパティ。すべて選択をクリックすると、利用可能なすべてのプロパティが含まれます。
* **ファイル形式**：ダウンロードするファイルの形式（デフォルト：CSV）
* **CSV**：CSVファイル
* **Excel XML**：Microsoft Excelで閲覧可能なXMLファイル
* **Microsoft Word**：DOCXファイル
* **HTML**：HTMLファイル
* **JSON**：JSONファイル
* **PDF**：PDFファイル
* **エンコーディング**：ファイルのエンコーディング（UTF-8、UTF-18 BE、MS949；デフォルト：MS949）
* **範囲**：ファイルに含めるデータセット数。指定した件数分、最新登録順で出力されます。

データセット一覧の更新

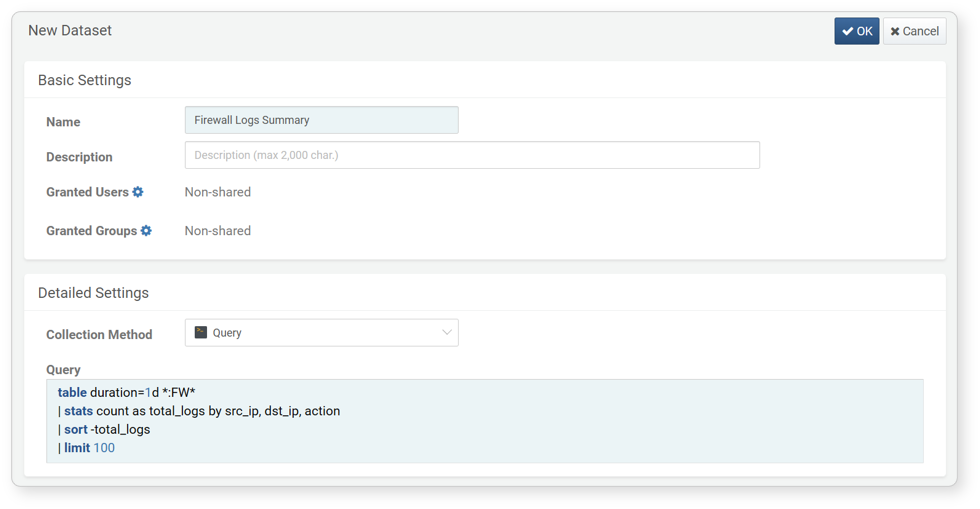
最新情報に更新するには、ツールバーの**更新**をクリックします。

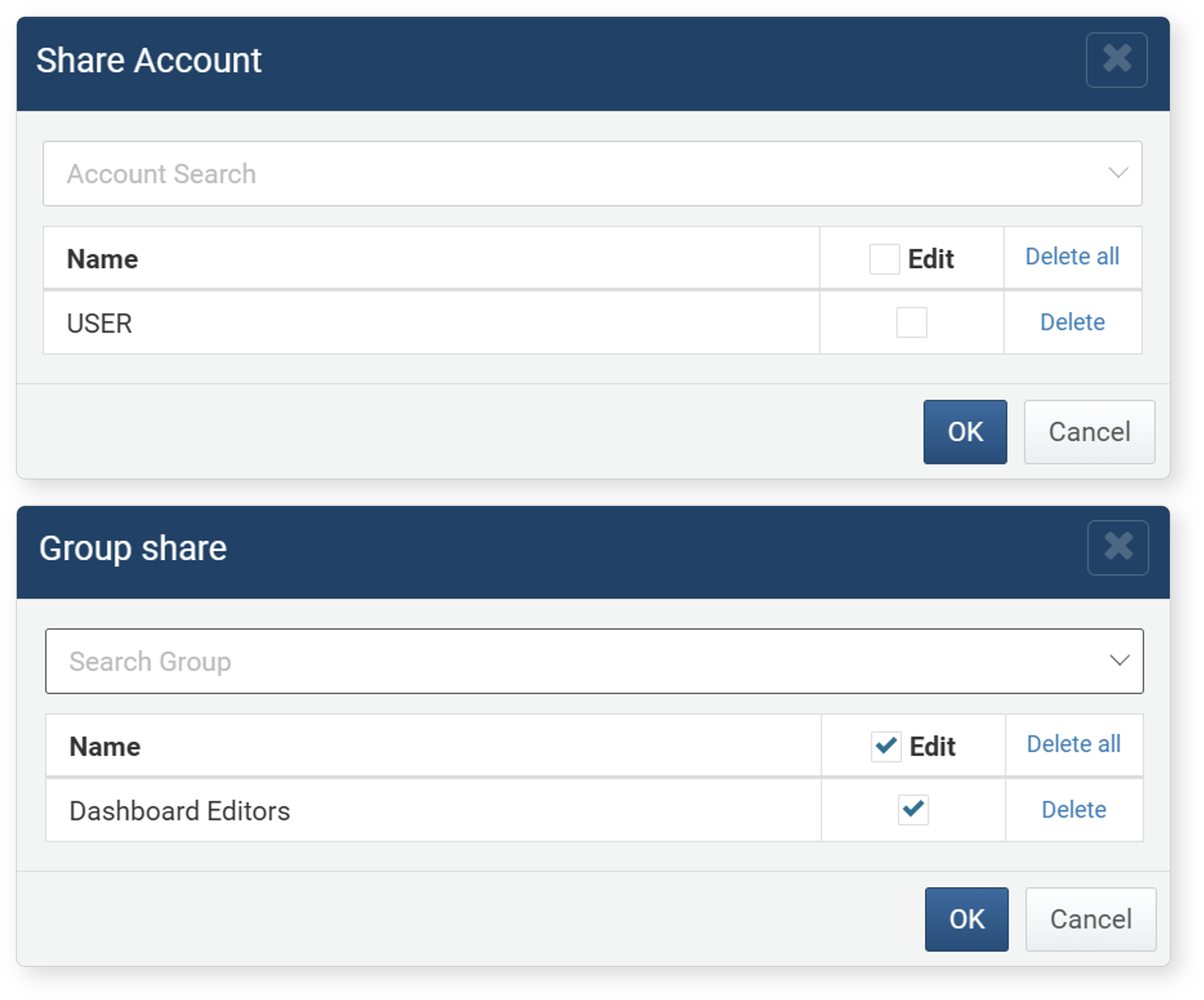
#### データセットの追加（クエリ）

データセットを追加するには、以下の手順を実行します。

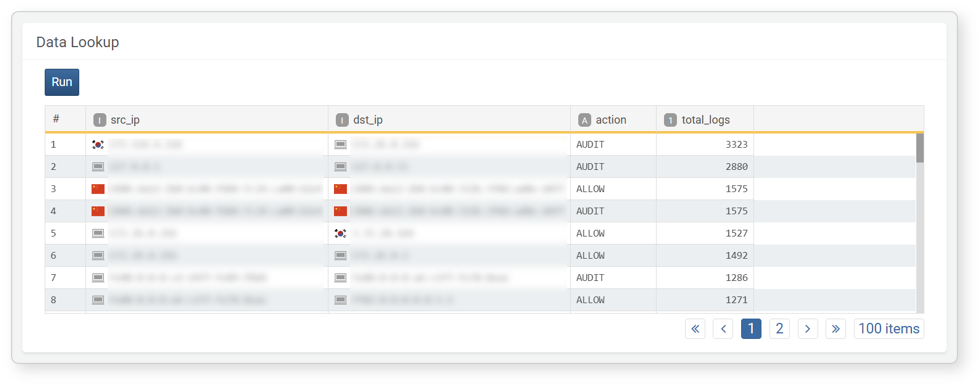
[データセット一覧](https://docs.logpresso.comnull)画面上部のツールバーで**追加**をクリックします。

**データセット追加**画面で必要な値を設定します。



* **名前**：データセットの一意な名称（最大255文字）
* **説明**：データセットの詳細説明（最大2000文字）
* **許可ユーザー**/**許可グループ**：データセット権限を付与するアカウントまたはアカウントグループ。データセット権限を付与するユーザーまたはユーザーグループを指定します。
* 
* **データセットタイプ**：データセットのタイプ（デフォルト：クエリ）
* **クエリ**：クエリをデータセットとして登録
* **相関分析**：相関分析用データセット。詳細は[データセット追加（相関分析）](https://docs.logpresso.comnull)を参照してください。

データセットクエリの検証には、画面下部のデータセットクエリ欄で**実行**をクリックします。



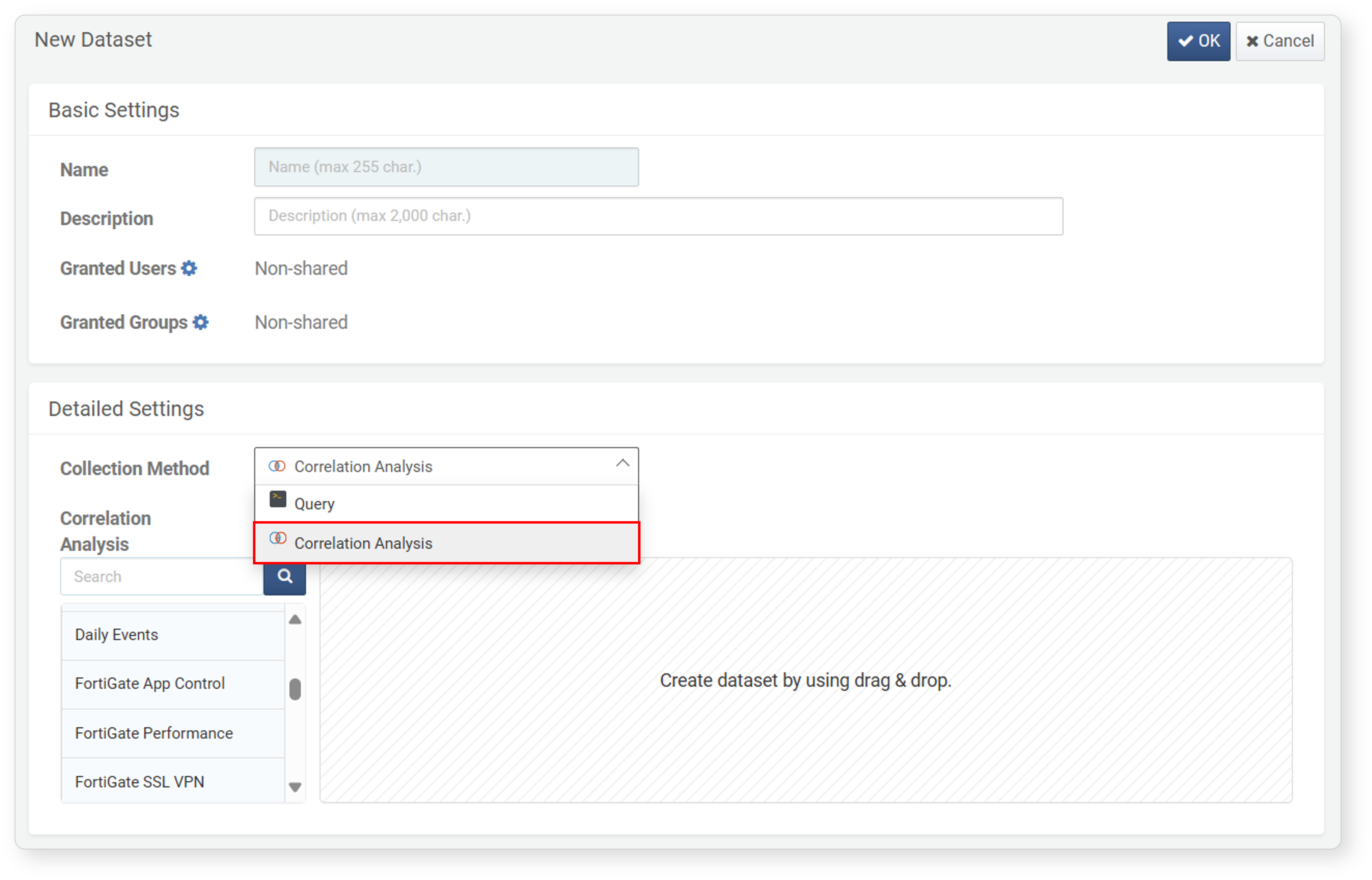
設定が完了したら、画面上部の**OK**をクリックします。

共有設定で付与されたアカウントは「ユーザー権限アカウント」と呼ばれます。これに対し、管理者アカウント（クラスター管理者、管理者）は、共有設定に関係なくデータセットの追加・変更・削除が可能です。データセットを共有していても、ユーザー権限アカウントが基礎データソース（例：テーブル、行動プロファイル、イベント）への読み書き権限を持たない場合、「ピボットエンティティの編集」はできず、データセットの編集もできません。データセット共有権限を付与する前に、必ずソースへのアクセス権限を確認してください。

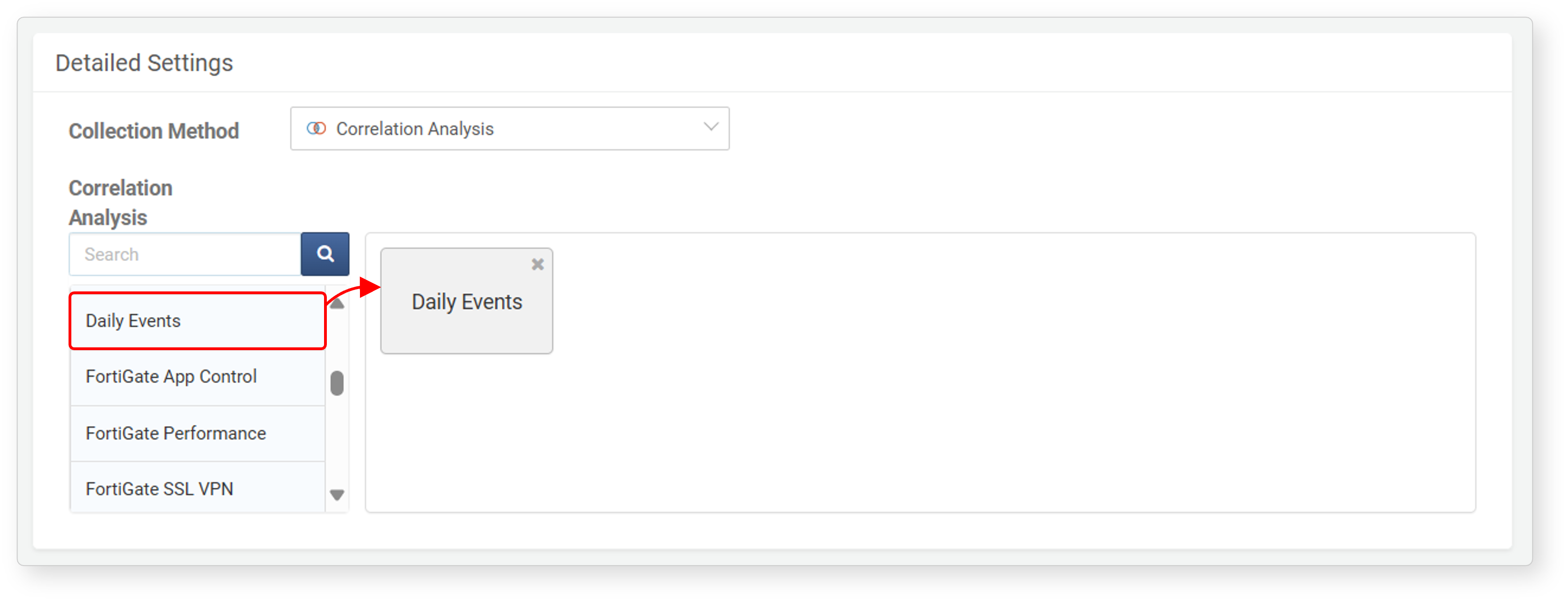
#### データセットの追加（相関分析）

既存のデータセットを利用して相関分析用データセットを作成できます。

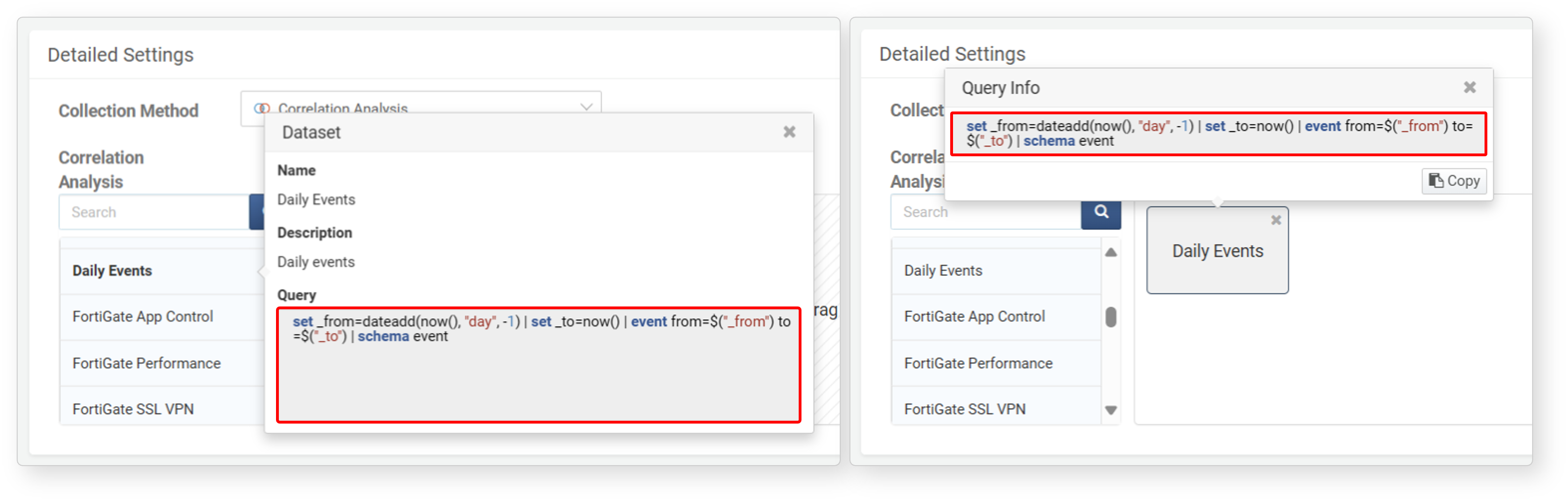
[データセット追加](https://docs.logpresso.comnull)画面で、**データセットタイプ**を**相関分析**に設定します。



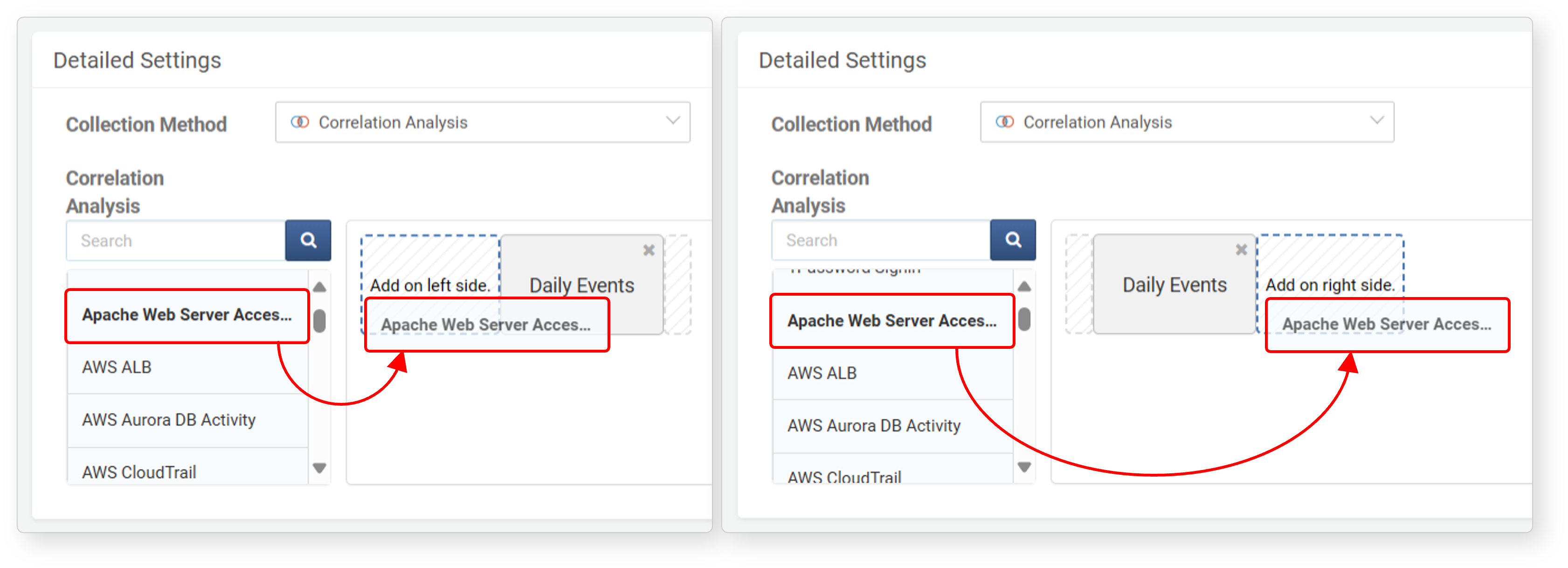
左側のパネルからデータセットを選択し、右側のボックスにドラッグします。



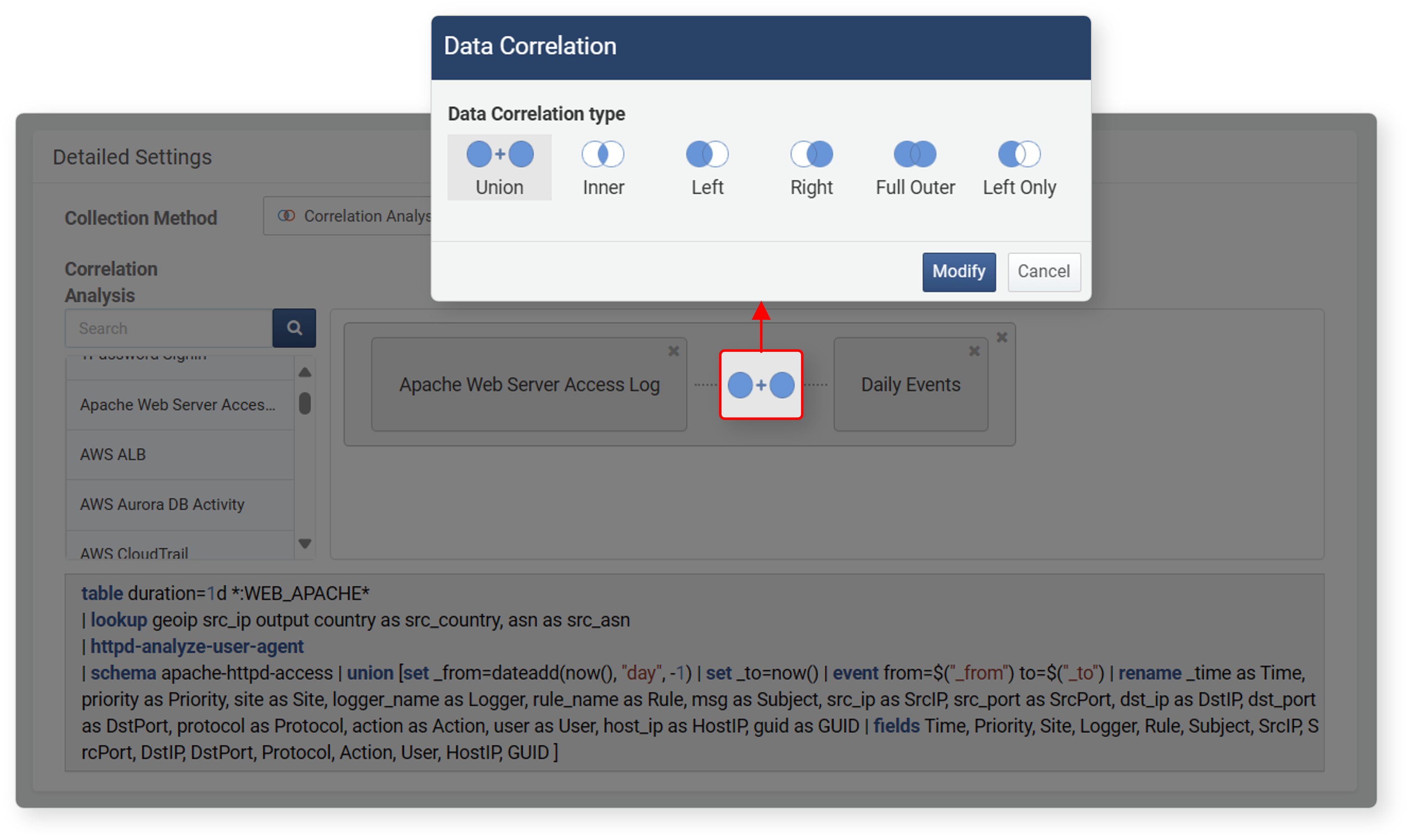
一覧のデータセットをクリックすると、参照用にクエリが表示されます。



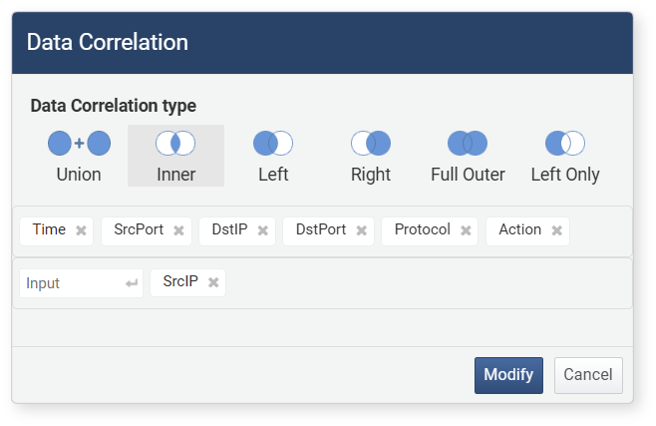
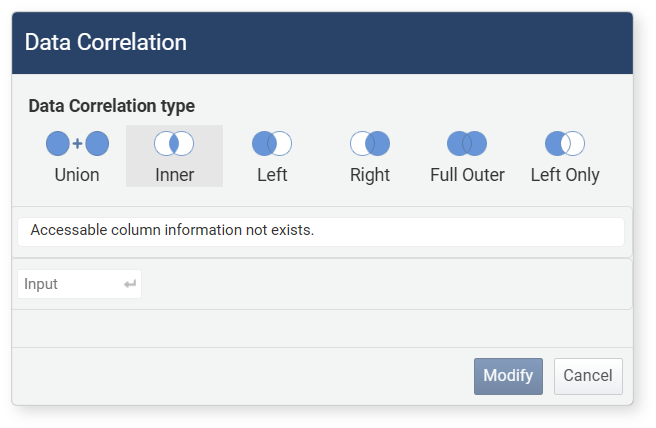
既に配置済みのデータセットの左右にドラッグすると、「右に追加」「左に追加」のツールチップが表示されます。希望の位置にドロップして相関分析の順序を設定します。



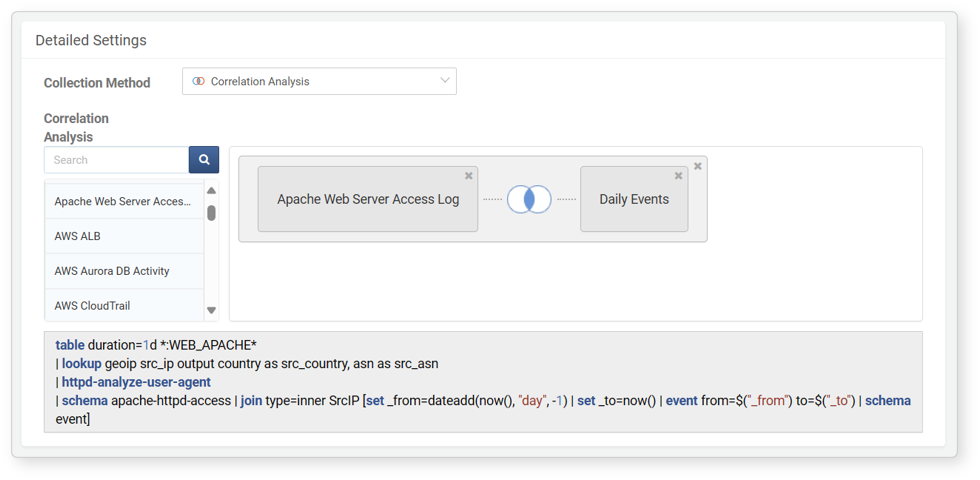
相関タイプラベルをクリックして**相関**タイプを変更できます。対応タイプは[union](https://docs.logpresso.comnull)および[join](https://docs.logpresso.comnull)コマンドのドキュメントを参照してください。



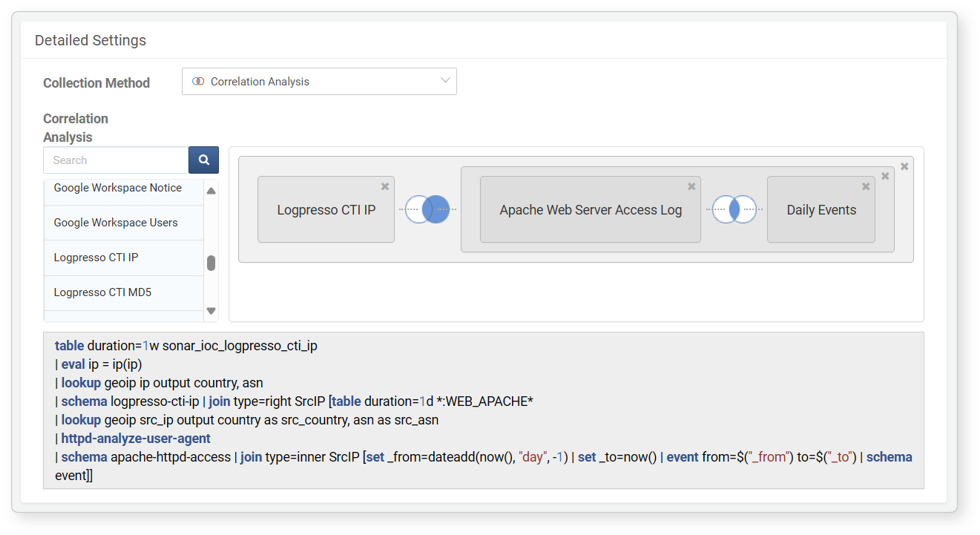
**Union**以外の相関分析タイプ（**Inner**、**Left**、**Right**、**Fully Outer**、**Left Only**）では、両データセットに存在する相関フィールドを定義する必要があります。

* フィールド横の「**+**」ボタンをクリックして相関キーとして追加します。
* 相関キーは両データセットに共通するフィールドでなければなりません。
* 
* 結合可能なフィールドがない場合、「**結合可能なカラム情報が存在しません。**」というメッセージが表示されます。
* 

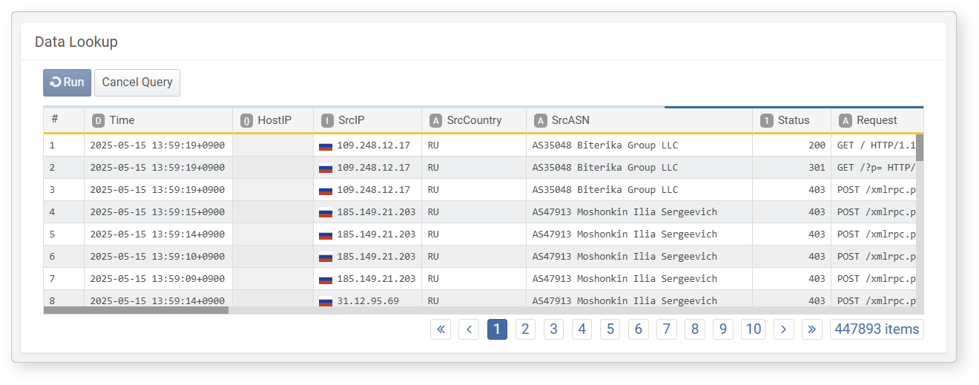
**相関分析**ダイアログで設定が完了したら、**修正**をクリックします。ビジュアルダイアグラムと対応するクエリが表示されます。



手順2～5は、2つ以上のデータセットを相関させる場合も繰り返し実行できます。



画面下部の**実行**をクリックして相関を検証します。



完了したら、画面上部の**OK**をクリックしてデータセットを保存します。

#### データセットの編集（クエリ）

データセットを編集するには、以下の手順を実行します。

[データセット一覧](https://docs.logpresso.comnull)から編集したいデータセットカードをクリックします。

**データセット編集**画面でデータセットプロパティを修正し、OKをクリックします。編集可能なプロパティの詳細は[データセット追加（クエリ）](https://docs.logpresso.comnull)を参照してください。

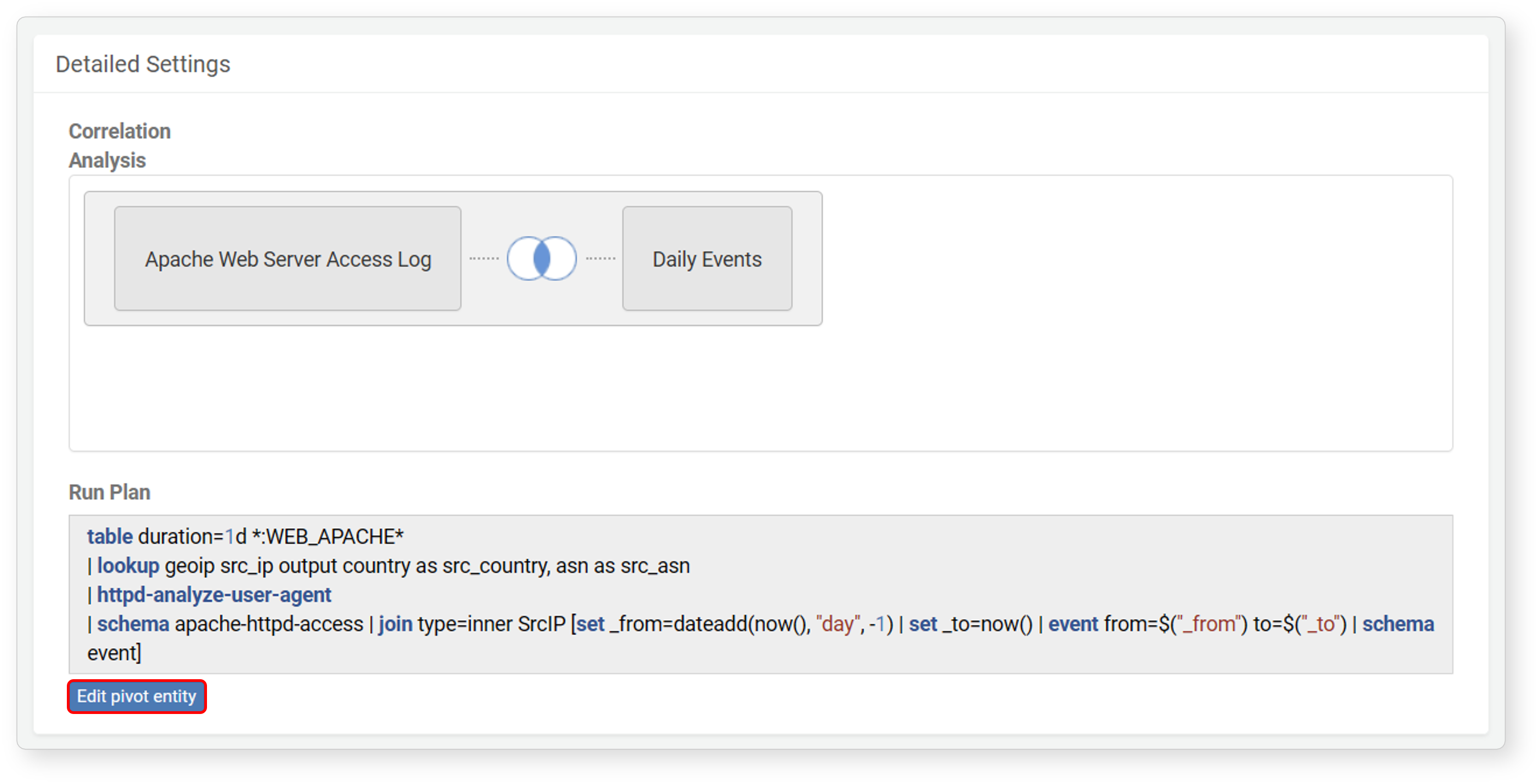
**OK**をクリックして変更を適用します。

#### データセットの編集（相関分析）

相関分析データセットを編集するには、以下の手順を実行します。

[データセット一覧](https://docs.logpresso.comnull)から編集したいデータセットカードをクリックします。

**データセット編集**画面で、実行計画下部の**ピボットエンティティの編集**をクリックします。これにより、データセットの[ピボット](https://docs.logpresso.comnull)編集画面に遷移します。



ピボット画面で分析条件を必要に応じて修正し、ツールバーの**データセット保存**をクリックします。

**データセット保存**ダイアログで以下を選択します。

* **保存**：既存のデータセットを上書き保存
* **名前を付けて保存**：新しい名前で保存
* **キャンセル**：保存しない場合

#### データセットの削除

データセットを削除するには、以下の手順を実行します。

削除したいデータセットカードのアイコンをクリックします。

**データセット削除**ダイアログで削除対象のデータセット名を確認し、**削除**をクリックします。削除しない場合は**キャンセル**をクリックします。

バッチルールシナリオの分析ロジックに適用されているデータセットは削除できません。ただし、データセットに含まれる個別データセットは削除可能です。これらを削除しても、作成時に登録された元のデータセットクエリは保持され、基礎データソースが利用可能な限り分析は正常に動作します。

#### データセットの活用

データセットは、以下の機能で活用できます。

登録済みデータセットを相関分析データセット作成時に適用

バッチルールシナリオの分析ロジックでデータセットを利用

**ピボット**、**クエリ**、**ダッシュボード**で**ロード**機能を使い、データセットレコードを読み込み・適用

[dataset](https://docs.logpresso.comnull)コマンドでクエリ文中からデータセットを参照