Dataikuフロー整理術

2024/10

アジェンダ

なぜ整理術を身につけることは重要か？

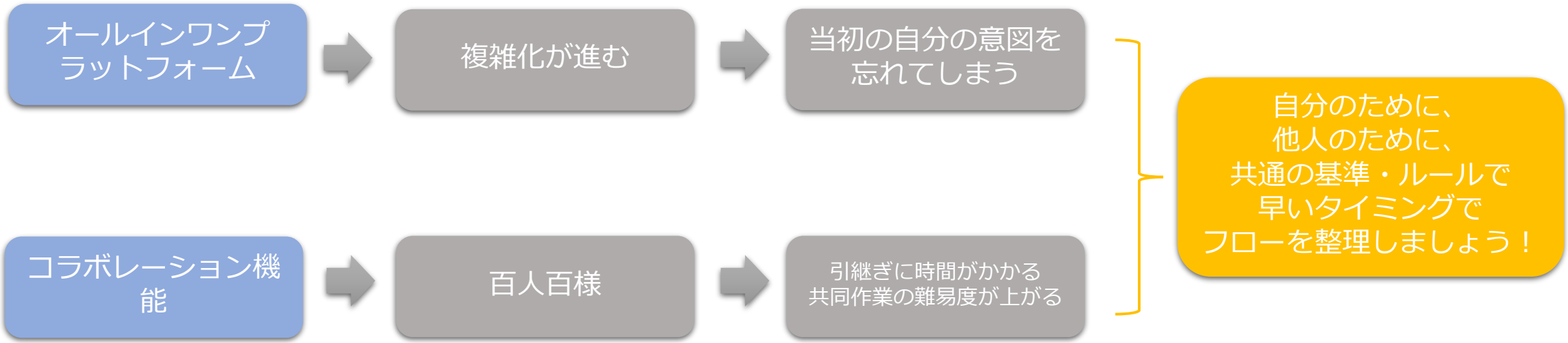
フローの整理術：その1~6

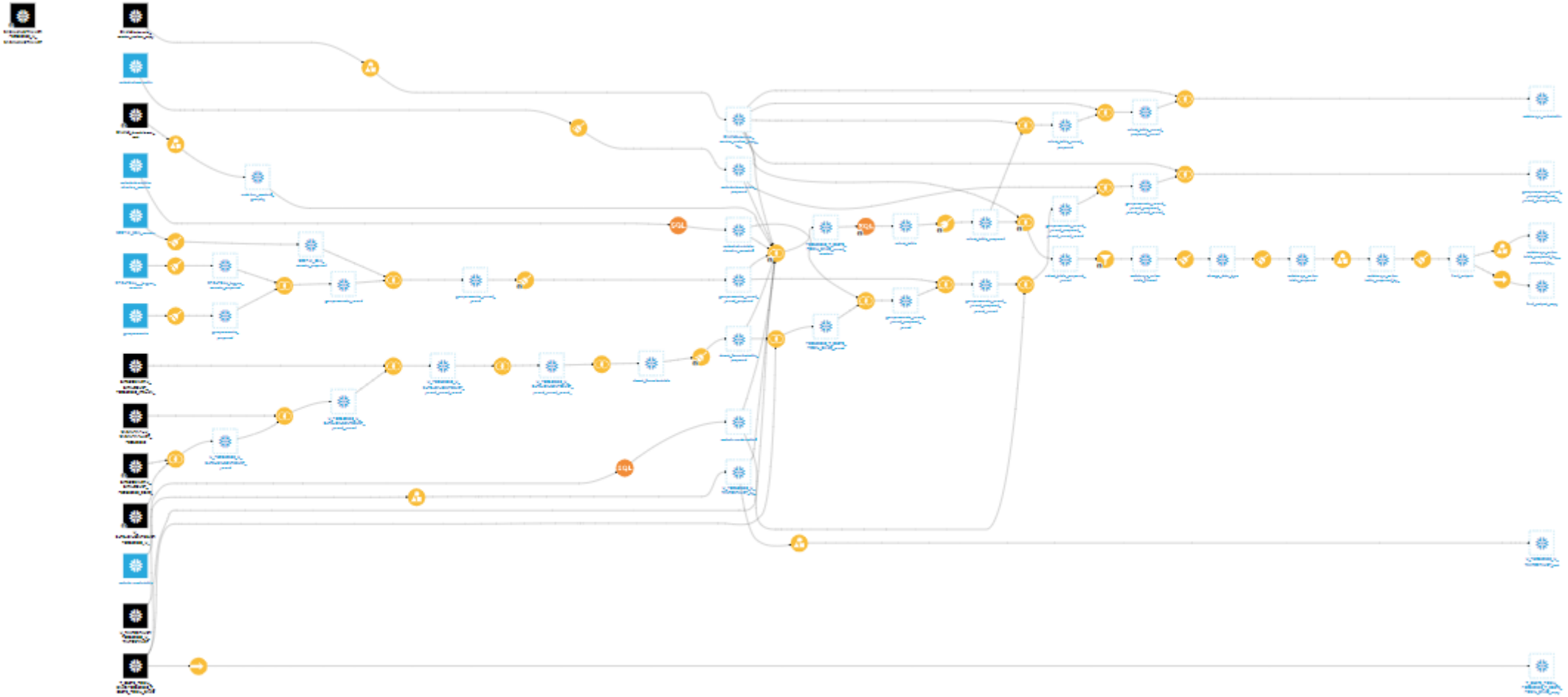
フロー整理に関連した「あるといいな」

Dataikuフロー整理術

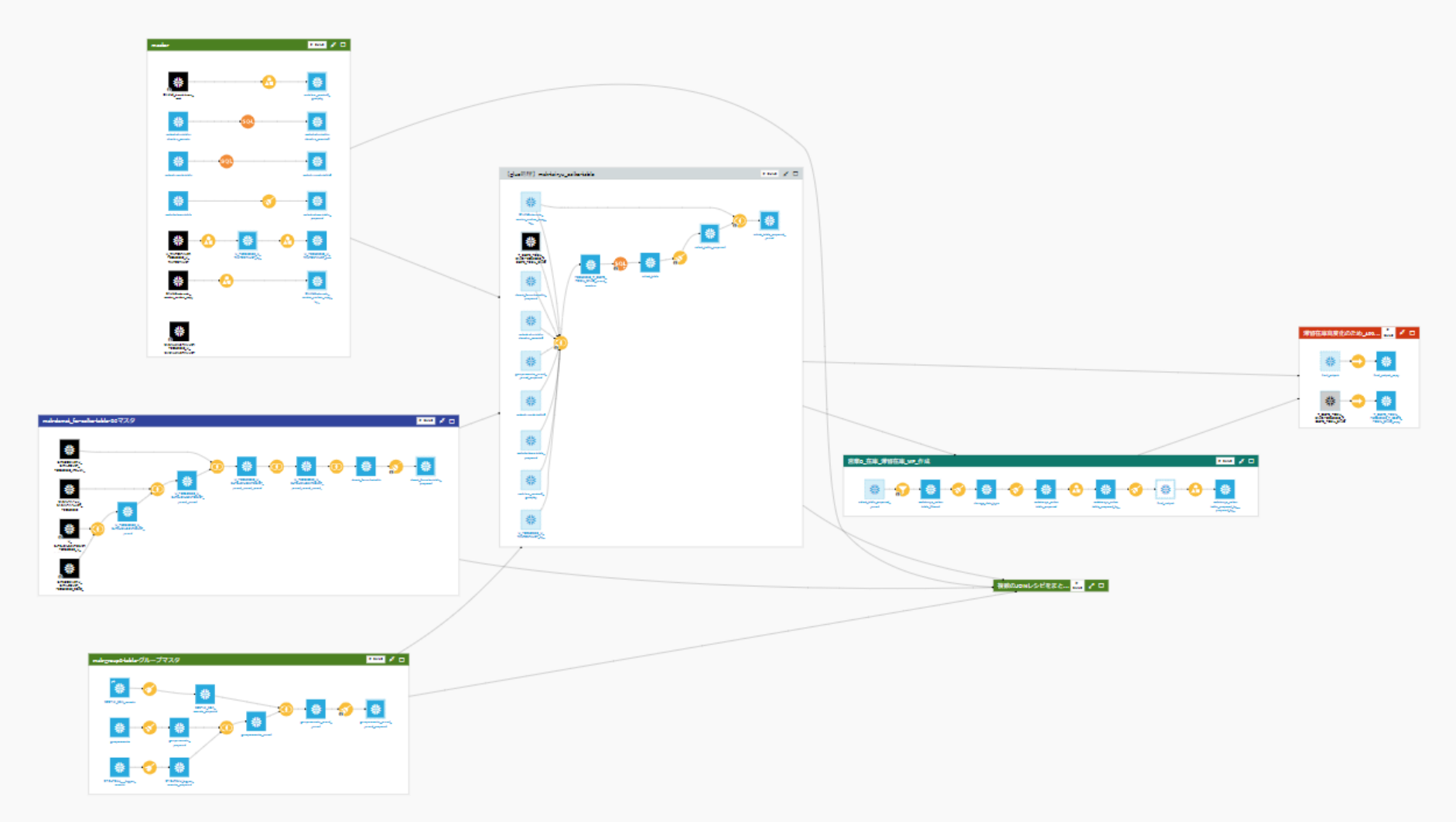
なぜ整理術を身につけることは重要か？

DATAIKUの特徴





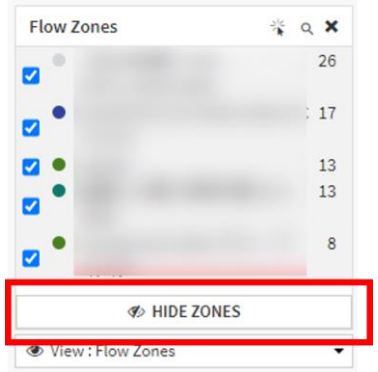
整理術その1：ZONE



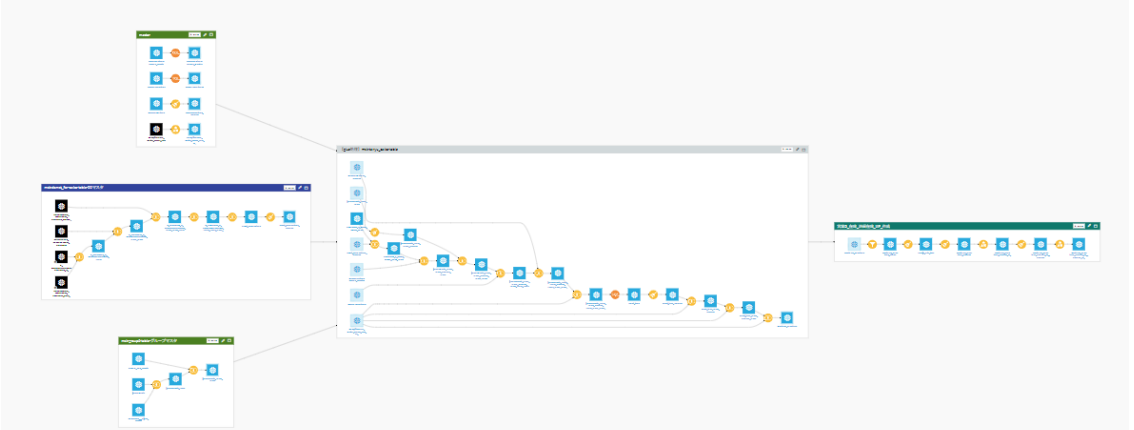
整理術その1：ZONE

ZONEを活用することで、フローの可読性を高められるのと、使用難易度を低くできる。

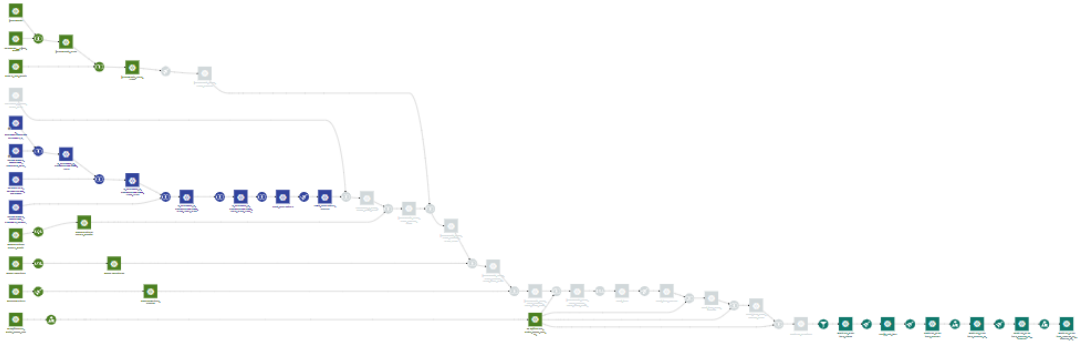
・処理の群れをまとめる
　→同じ目的を持つ処理を１つのゾーンに集約
・特定のZONEのみ実行する
　→目的によって指定されたZONEのみ実行することもできる
　　(シナリオ設定も)
・ZONEの色
　→色に意味を持たせよう。
・ZONEビューの表示/非表示
　→ZONEビューを非表示にすることもできる



ZONEを表示：



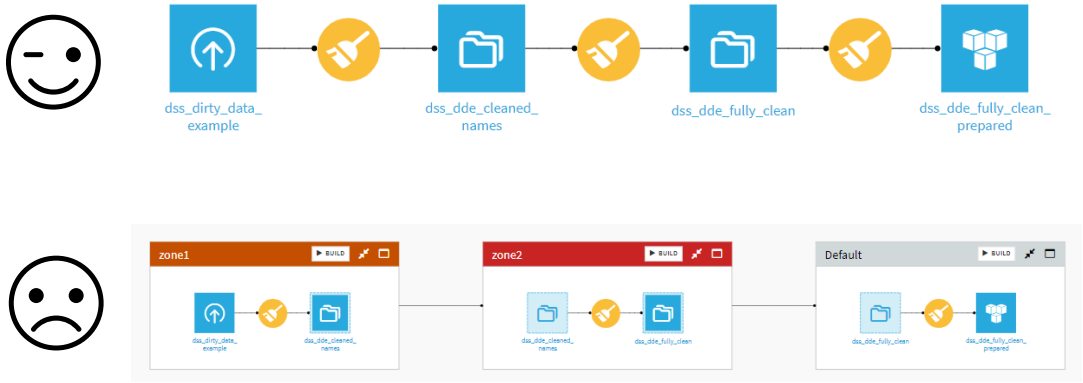
ZONEを非表示：



ZONE：どの粒度でZONEで分けるか

〇同じ目的を持つ処理を１つのゾーンにまとめる

✕細かすぎる



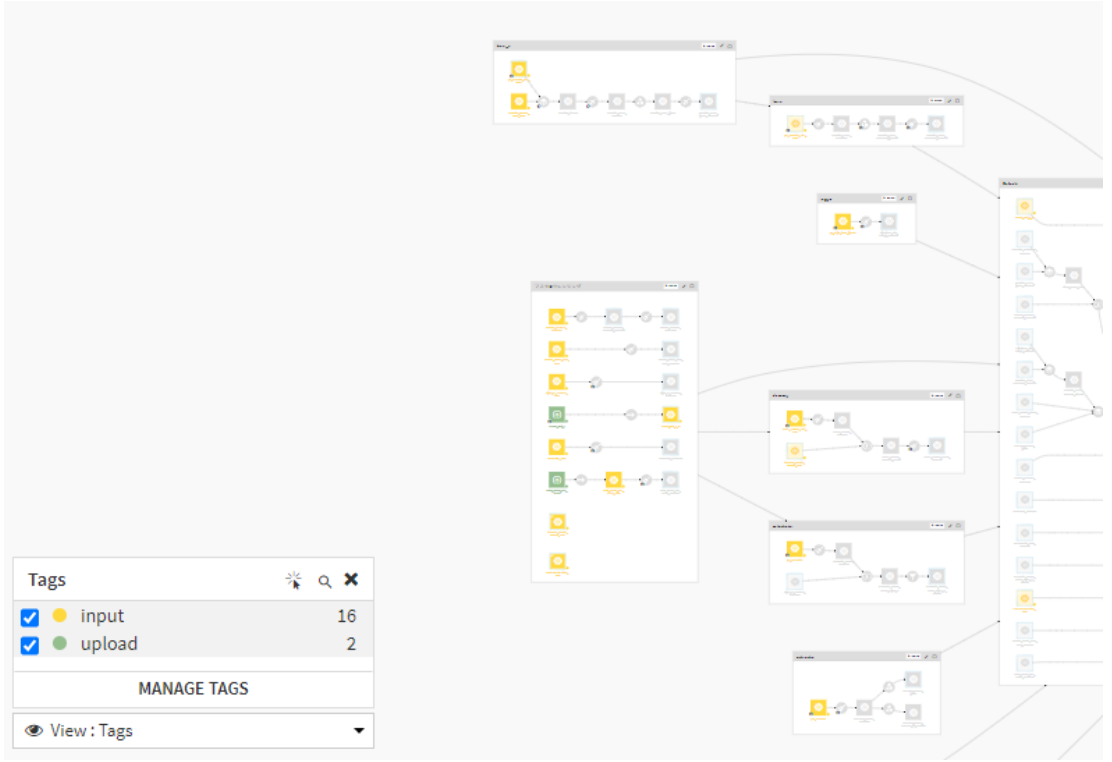
整理術その2：タグ

ZONEと組み合わせて使用することで、フローの可読性をワンランクアップ

よく使うタグ：

Input/Output/test(開発)/マスタ

・データセットやレシピを分類
　→検索しやすくするため
・タグビューの表示/非表示
　→タグを俯瞰



ZONEとタグの使い分け

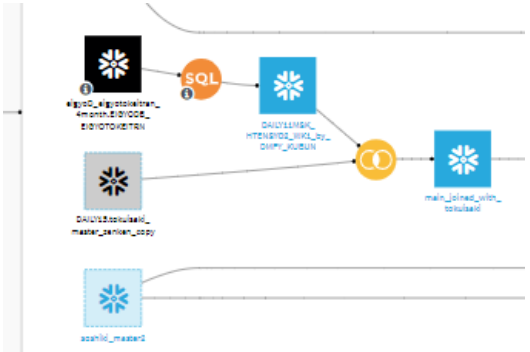
	ZONE	タグ
目的	同じ意味を持つレシピやデータセットの「まとまり」を作る	後から検索しやすくするために、特殊なレシピやデータセットをマーク
レシピやデータセットの数	多め	少なめ
使用頻度	高い	必要に応じて
その他	ZONEにタグをつけることは可能	

整理術その3： Share Dataset

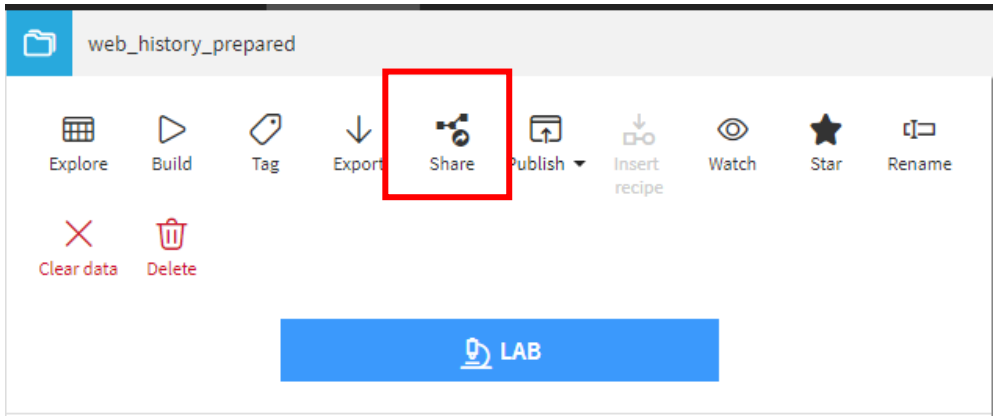
別プロジェクトで作られたデータセットを引用する場合、
Share Dataset機能でデータセットをそのまま参照することがおすすめ。

- ✓ Inputのデータセットが、どのフローによって作られたか追うのにとても便利になる。
- ✓ Data QualityやChartsなどの設定も継承されるので、データセットが管理しやすくなる。

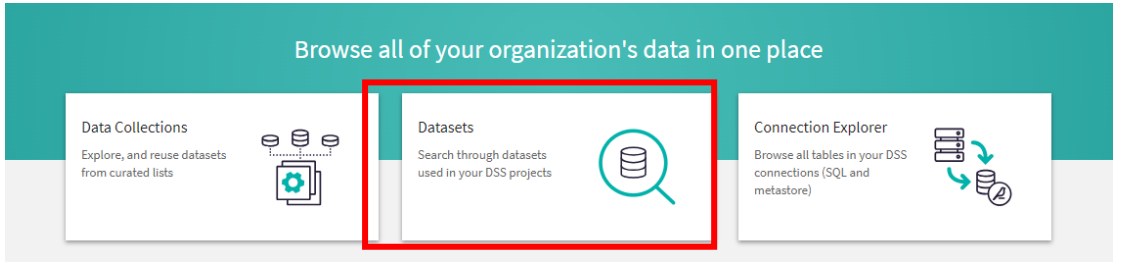
※シェアされたデータセットは下記のように黒いアイコンで表示される。



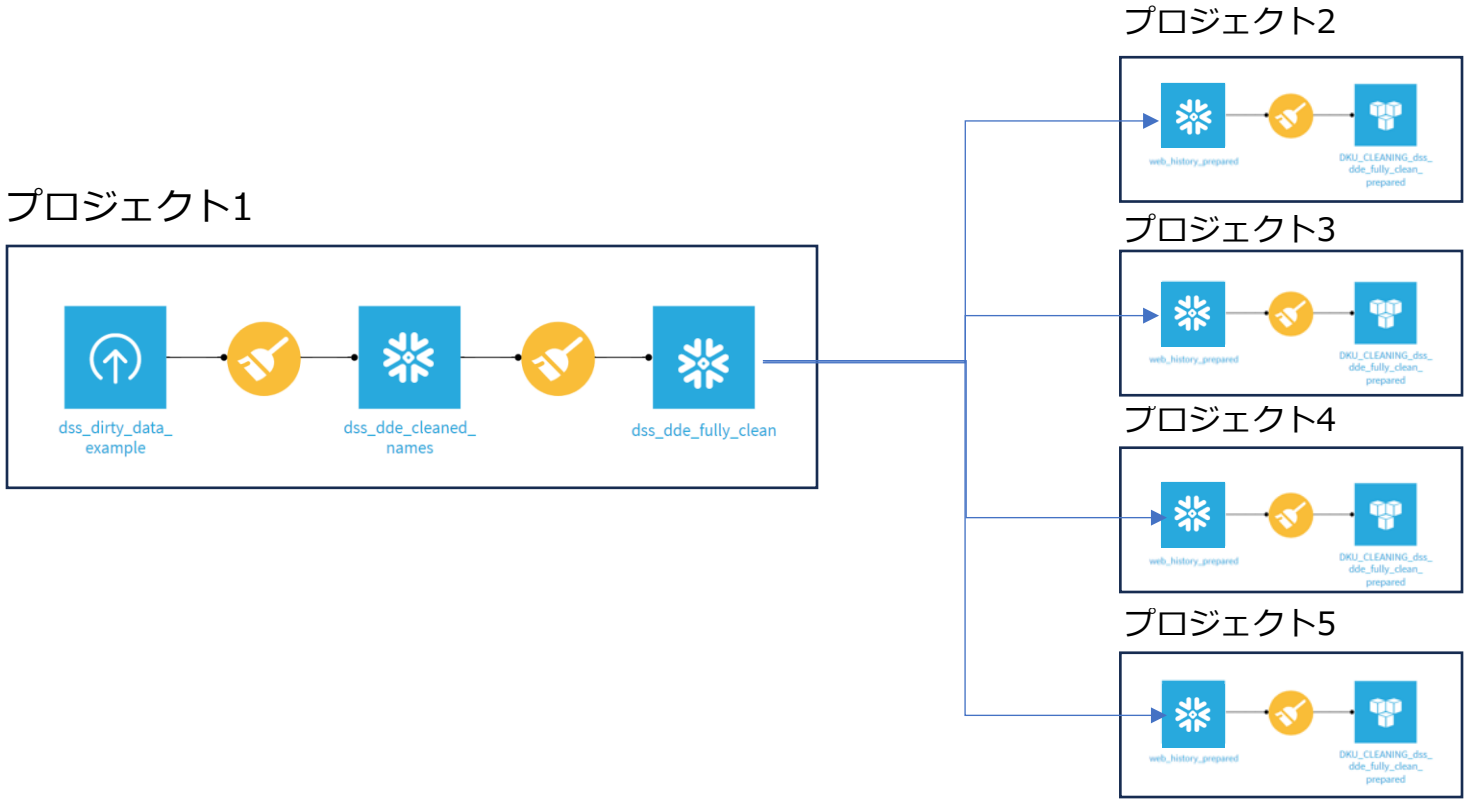
データセットをシェアする方法①



データセットをシェアする方法②



整理術その3： Share Dataset



Share Dataset

Share Datasetを使うメリット：

・Data QualityやChartsなどの設定が継承される

・別プロジェクトによって作られたデータであることがすぐわかる
ソースプロジェクトへの移動も便利

・サンプルの更新は一回行えばすべて更新される

・(データセットの格納先がデータベースの場合)テーブルの差し替えは楽に済む

整理術その4：注釈(フロー)

あらゆる処理に注釈をつけよう。注釈をつける理由はたくさんある。
・理解を助ける
・メンテナンスが容易になる
・将来の自分を助ける
など

Dataikuでは、いろいろなところに注釈をつけることができる。
・データセットやレシピ
・ZONE
・PrepareレシピのStep

データセットやレシピ、ZONEは右側のサイドバーから注釈を付けられる

整理術その4：注釈(Prepareレシピ)

ⓘをクリックして、注釈を記入することができる

Prepareレシピの補足

同じ目的を持つ処理群をグループにまとめる

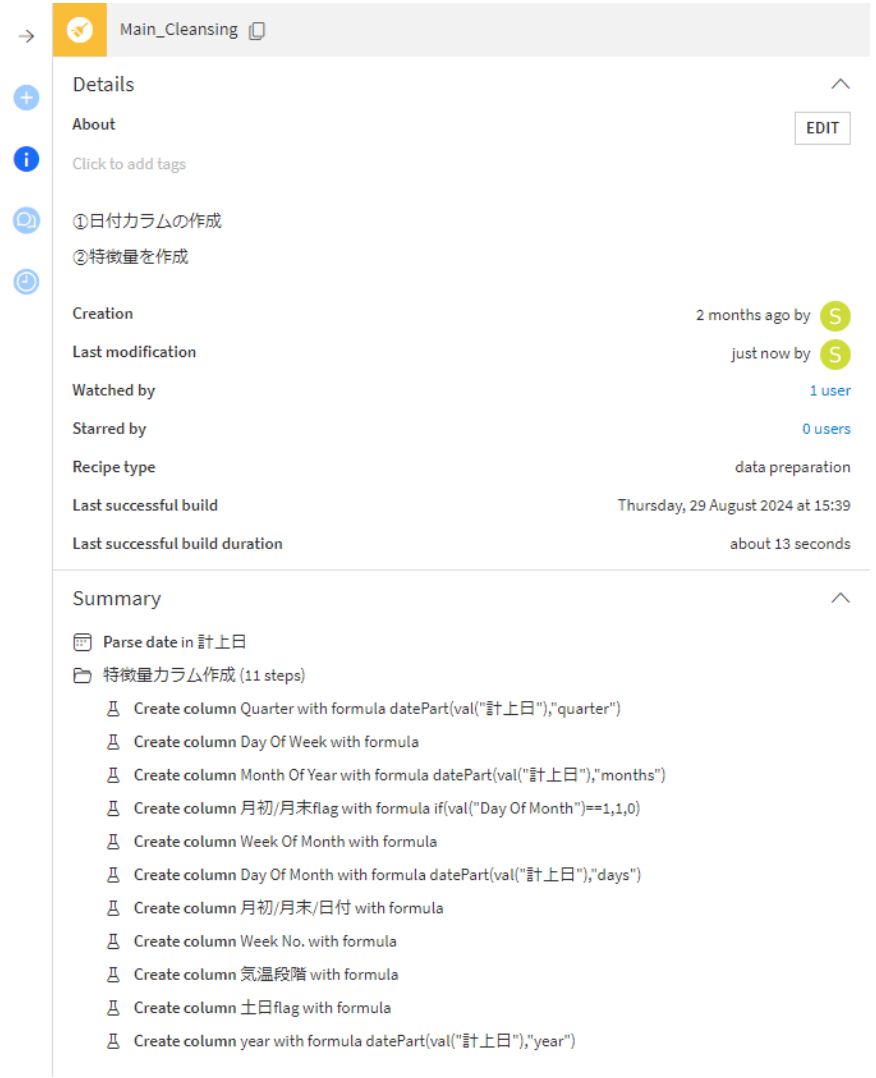
散乱した処理→

↑グループ化された処理群

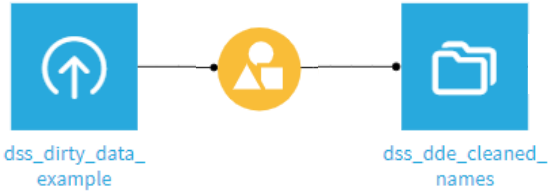
整理術その4：注釈

注釈は書きたければいくらでも書ける
どのぐらい注釈をつければよいか？

Detailsを見るだけでレシピの概要がわかるように

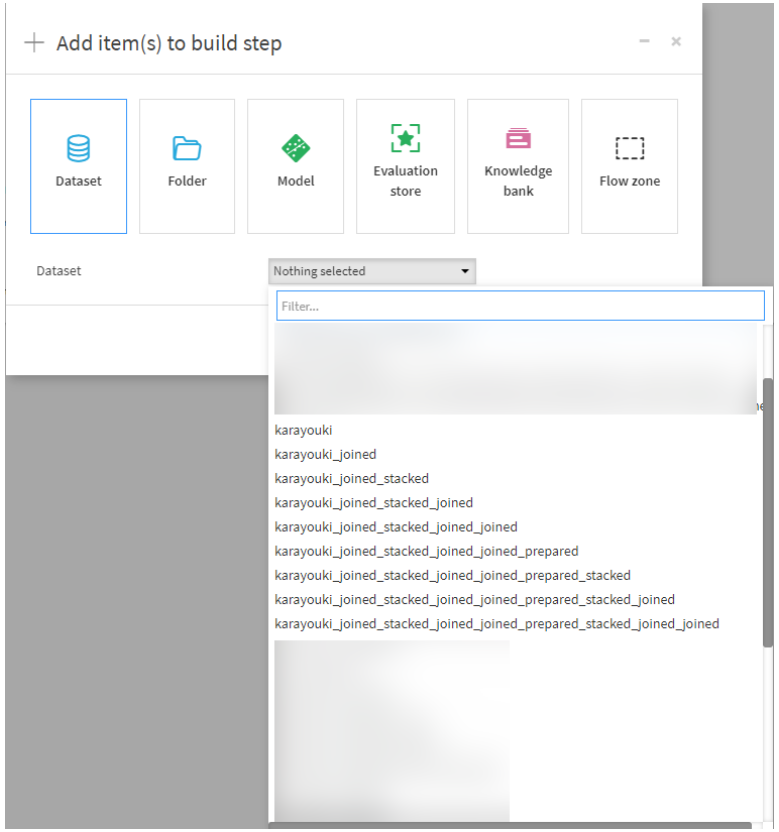


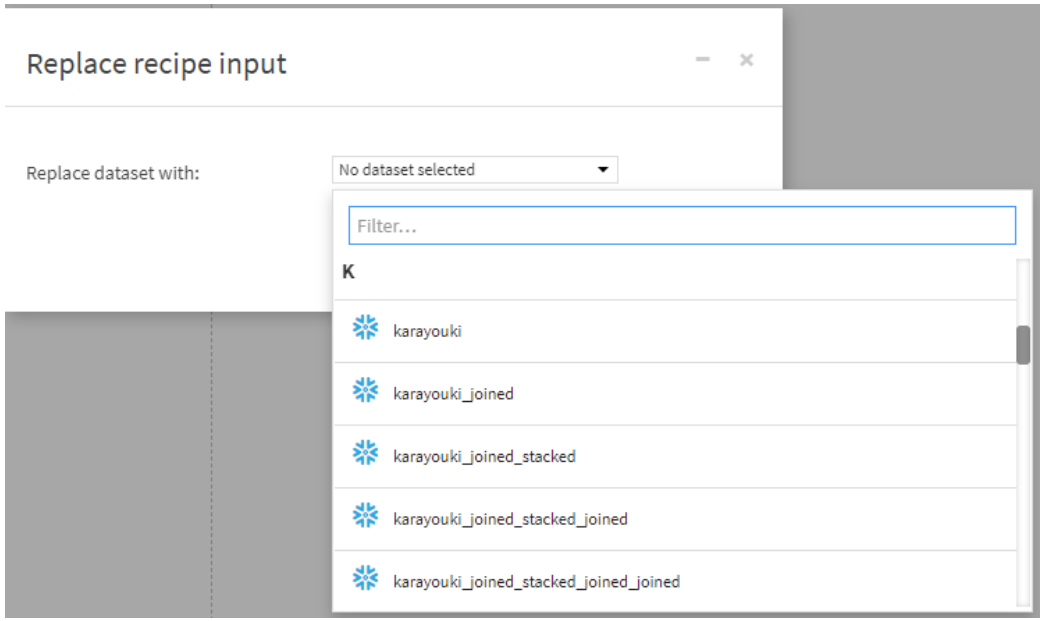
整理術その5：レシピの数を減らしましょう



※prepareレシピの処理内容はカラム作成もしくはフィルターの場合はGroupレシピに代替できる

整理術その6：ネイミング





まとめ

整理術その1：ZONE

整理術その2：タグ

整理術その3： Share Dataset

整理術その4：注釈

整理術その5：レシピの数を減らしましょう

整理術その6：ネイミング

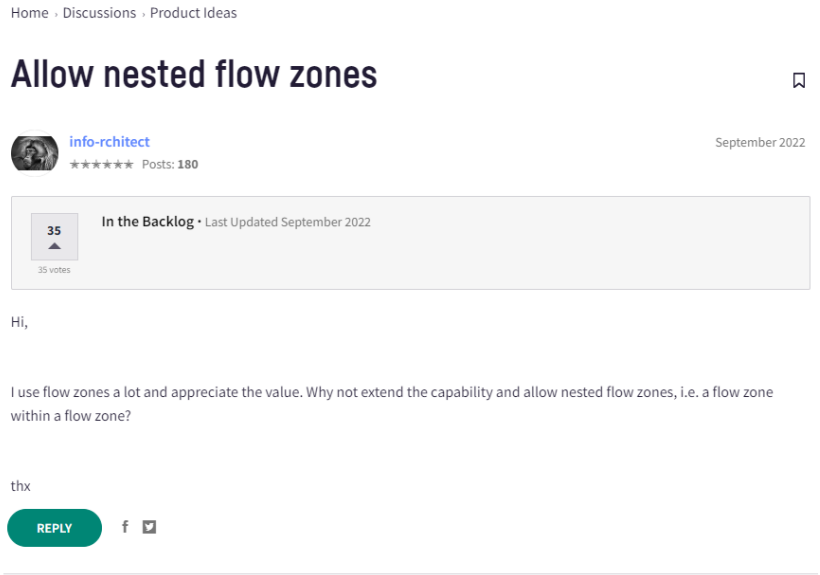
フロー整理に関連した「あるといいな」

・ネストされたZONE

・アイコンの自由配置、もしくはZONEの自由配置

・データセットネーミングの日本語サポート

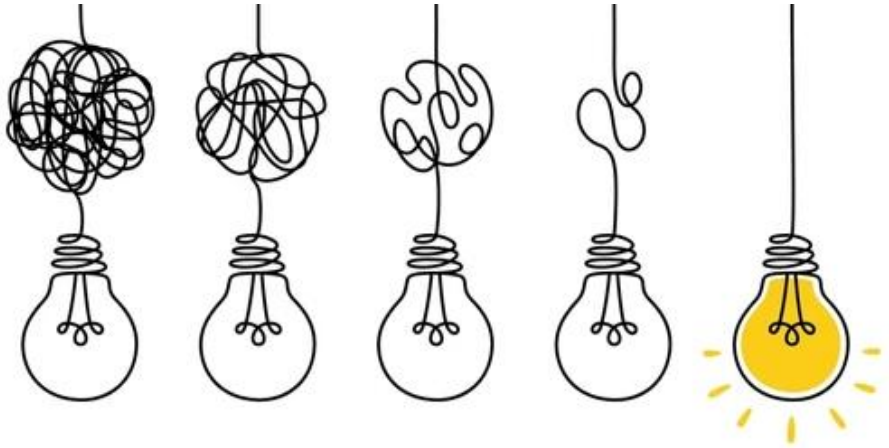
・データセットの選択のUIから行えるように



最後に

「工、其の事を善くせんと欲すれば
必らず先ず其の器を利くす。」

---孔子の『論語』



https://www.truestar.co.jp/

Shibuya Center Place 8F | 1-16-3 Dogenzaka | Shibuya-ku | Tokyo 150-0043 | Japan
Tel: 03 5422 6561 | Fax: 03 5422 6562 | e-mail: info@truestar.co.jp