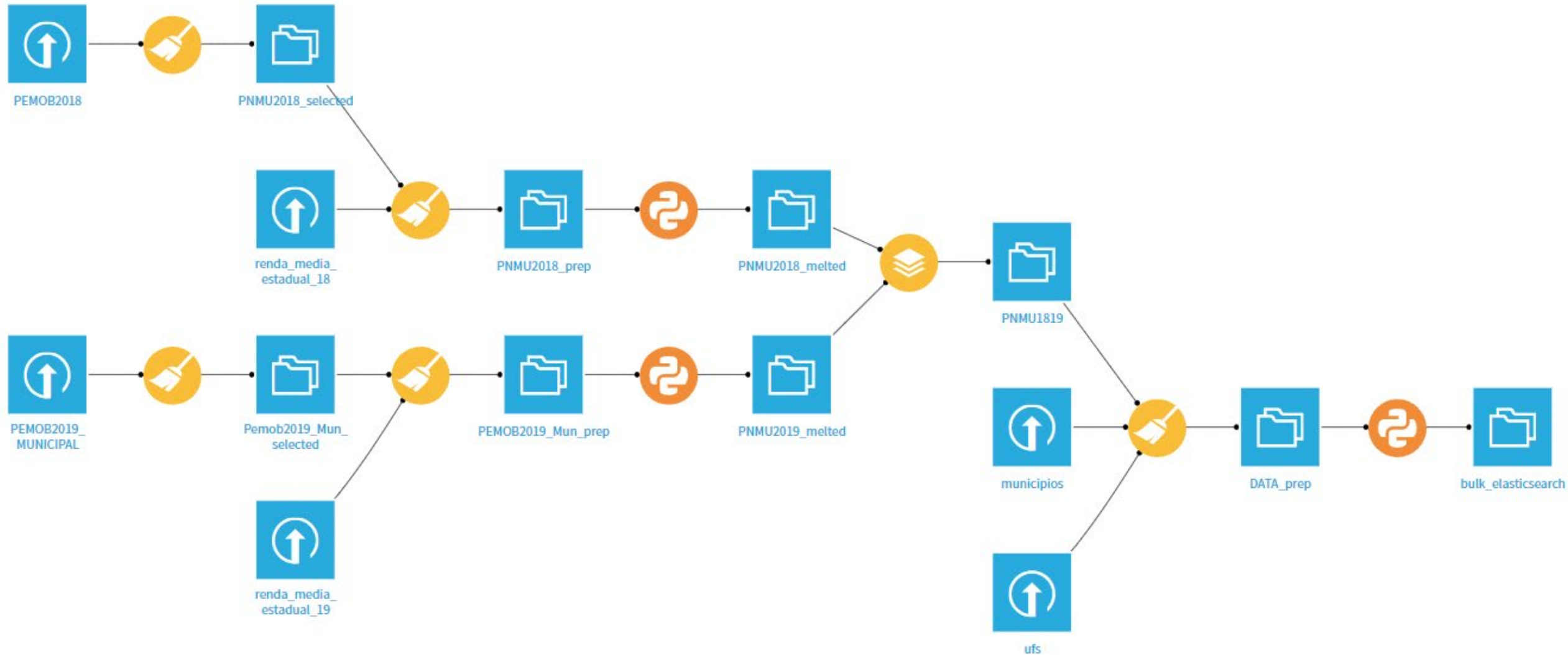


ETL: Extract, Transform and Load



ETL: Extract, Transform and Load

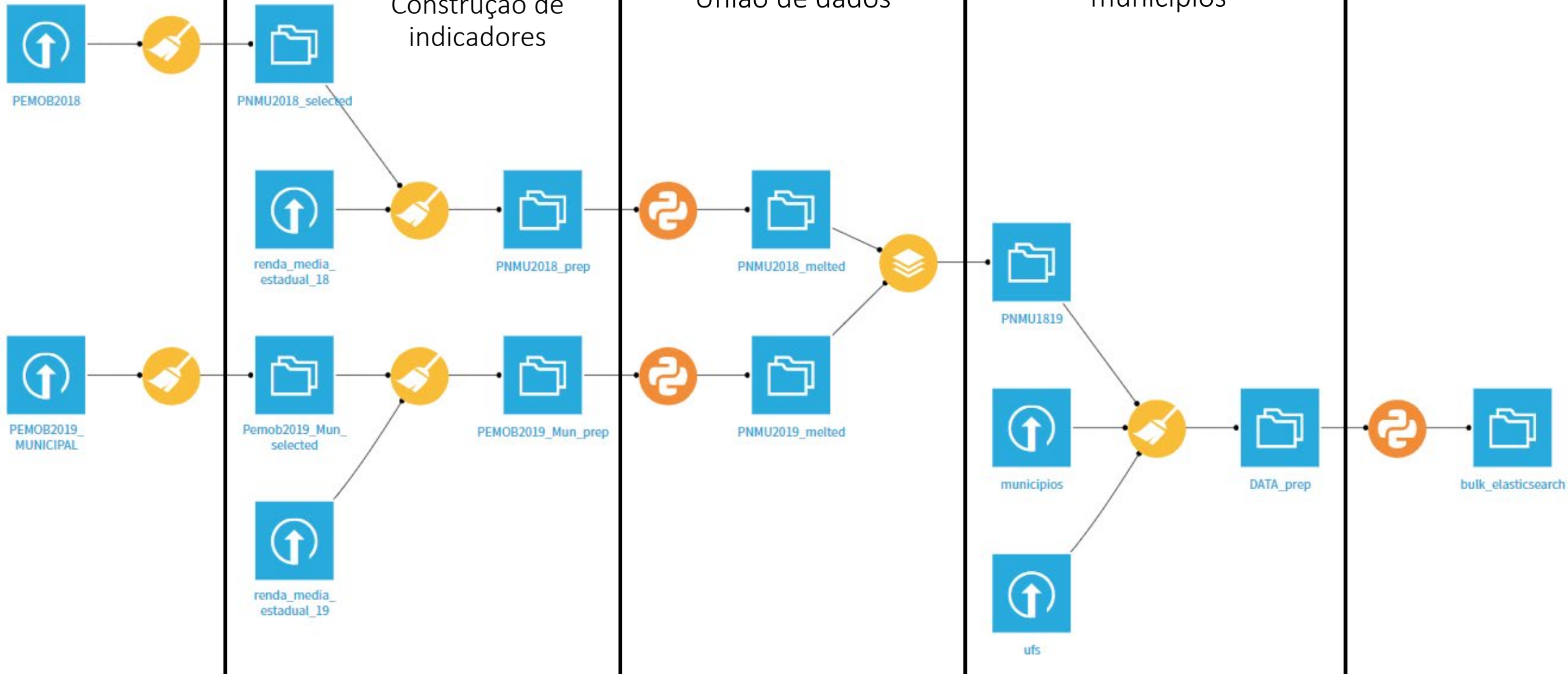
Seleção de variáveis

Limpeza de dados
+
Construção de indicadores

Formatação Long
+
União de dados

Enriquecimento com dados de municípios

Indexação Elasticsearch



ETL: Extract, Transform and Load

Seleção de variáveis

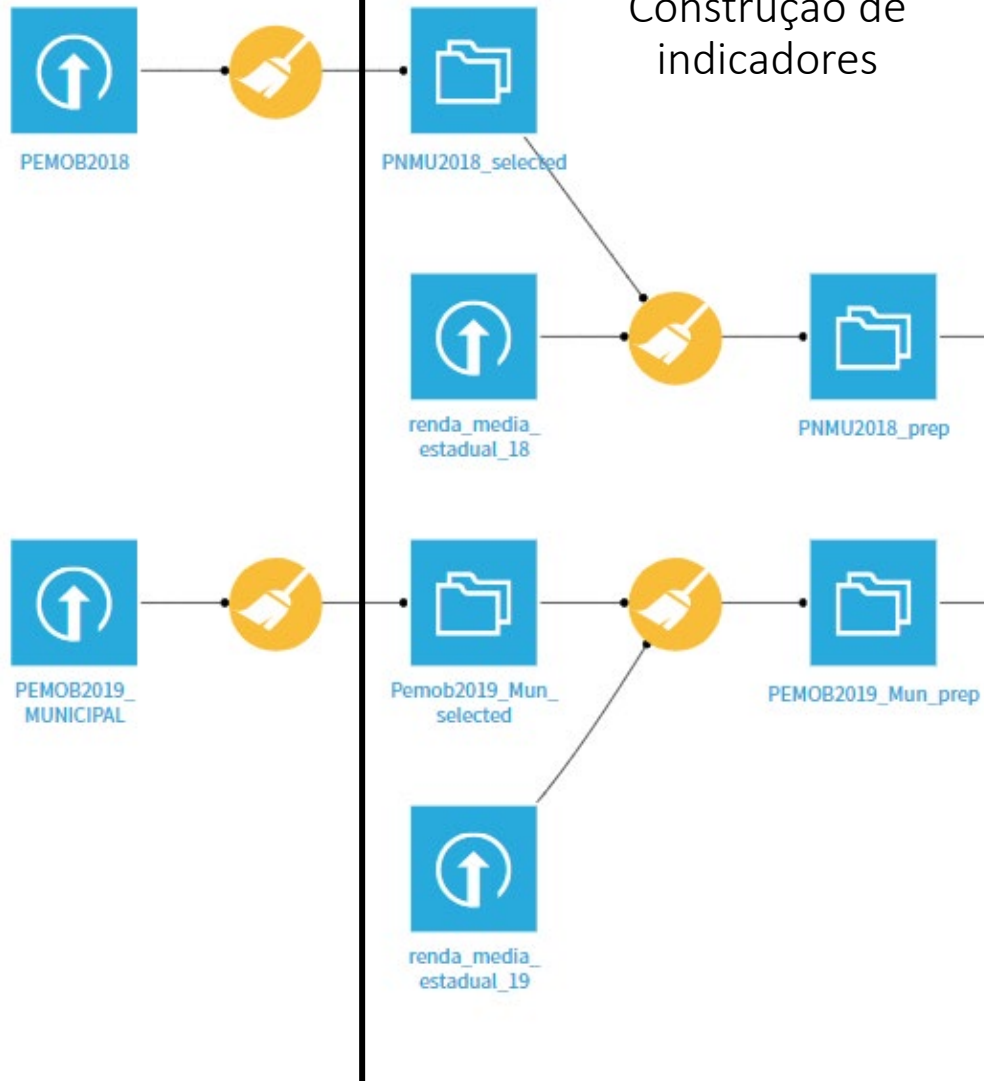


Id	Indicador	Variáveis 2018	Variáveis 2019
1	Idade média da frota	8.3.A	2.3.1.A
2	Quilometragem de vias exclusivas dedicadas	5.3.A a 5.3.J	1.6.1.A a 1.6.1.M
3	Quantidade de pontos de embarque e desembarque	1.5.F	1.4.2
4	Número de passageiros transportados (ônibus, vans e microônibus)	3.4.A a 3.4.E, 3.4.G a 3.4.K	3.4.1.A a 3.4.1.E, 3.4.1.G a 3.4.1.K
5	Índice de Passageiros Transportados por Quilômetro (IPK)	3.4.F, 3.4.L, 1.7.A a 1.7.D	3.4.1.F, 3.4.1.L, 1.7.1.A, 1.7.1.B
6	Divisão Modal	1.8.G a 1.8.J	4.8.1.G a 4.8.1.J
7	Percentual de receita extratarifária do sistema de transporte coletivo	3.2.B, 3.2.H, 3.2.A, 3.2.B, 3.2.G, 3.2.H	3.2.1.B, 3.2.1D, 3.2.1.A, 3.2.1.B, 3.2.1.C, 3.2.1.D
8	Arrecadação por multas e utilização da infraestrutura de trânsito por habitante	5.2.B, 5.4.E, 9.5.F	5.1.1.A, 5.2.2.A, 9.1.4, 11.5.1.C
9	Composição da planilha de custos	3.7.B a 3.7.F	3.7.1.A a 3.7.1.E
10	Peso do custo do transporte público na renda média dos habitantes	3.1.A, renda média estadual (Pnad Contínua)	3.1.1.A, renda média estadual (Pnad Contínua)

ETL: Extract, Transform and Load

Seleção de
variáveis

Limpeza de dados
+
Construção de
indicadores



Limpeza de dados e tratamento de dados:

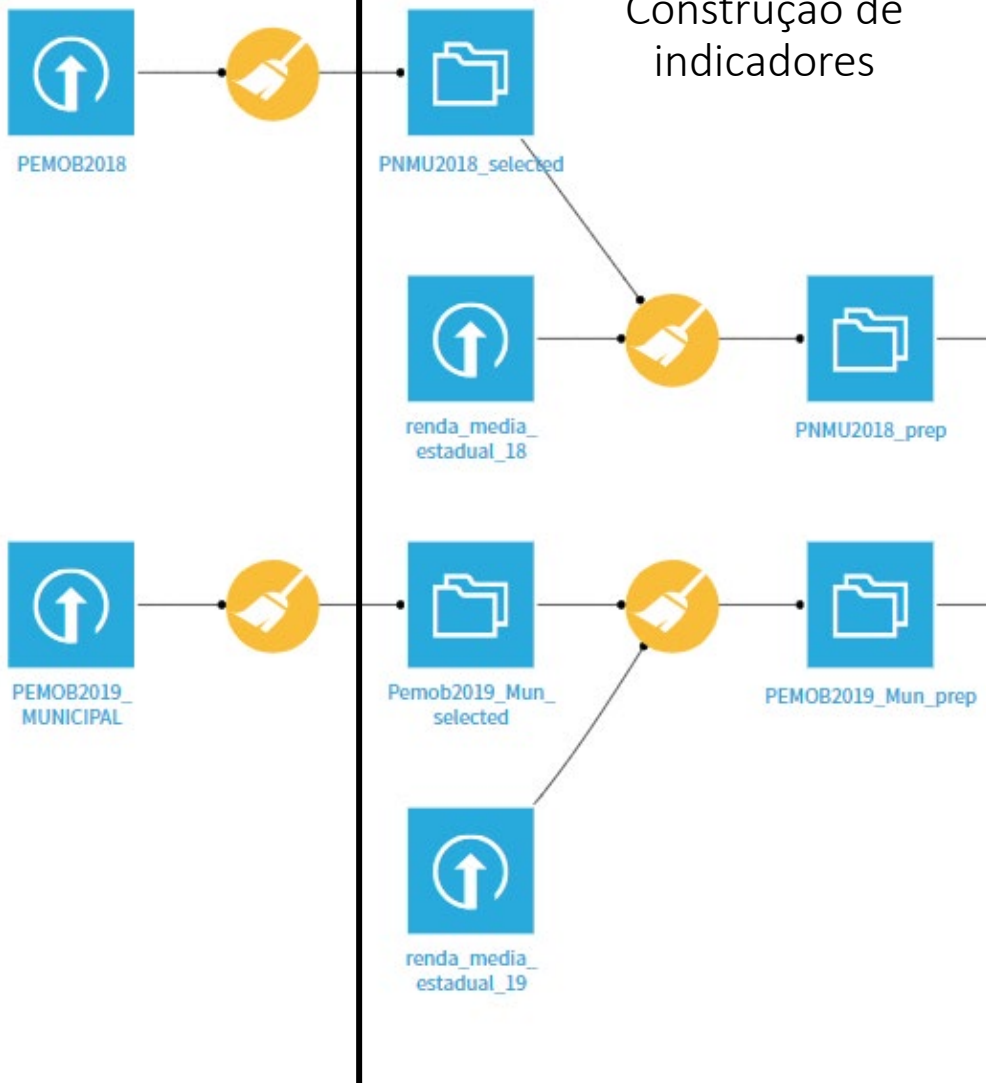
- **Dados faltantes como:** "", "ND", "NR", "NA", "-" ou "Em 2018 não houve. devido à falta de sistema"
- **Tratamento de valores numéricos:**
 - Remoção da expressão regular: "\(.*\)"
 - Remoção de espaços
 - Remoção de pontos como separadores de milhar
 - Remoção expressões como: "R\$", "RS", "%", ","
 - Padronização de ponto flutuante como "."
 - Padronização de porcentagens no intervalo [0,1]
- **Porcentagens – Divisão Modal e Composição da Planilha de Custos:**
 - Cálculo de complemento não informado
 - Cálculo de porcentagem excedente
- **Enriquecimento dos dados:**
 - Renda média estadual mensal

ETL: Extract, Transform and Load

Seleção de variáveis

Limpeza de dados +
Construção de indicadores

Construção de Indicadores:



Id	Cálculo 2018	Cálculo 2019
1	8.3.A	2.3.1.A
2	5.3.A a 5.3.J	1.6.1.A a 1.6.1.M
3	1.5.F	1.4.2
4	3.4.A a 3.4.E + 3.4.G a 3.4.K	3.4.1.A a 3.4.1.E + 3.4.1.G a 3.4.1.K
5	$(3.4.F + 3.4.L) / (1.7.A a 1.7.D)$	$(3.4.1.F + 3.4.1.L) / (1.7.1.A + 1.7.1.B)$
6	1.8.G a 1.8.J	4.8.1.G a 4.8.1.J
7	$(3.2.B + 3.2.H) / (3.2.A + 3.2.B + 3.2.G + 3.2.H)$	$(3.2.1.B + 3.2.1.D) / (3.2.1.A + 3.2.1.B + 3.2.1.C + 3.2.1.D)$
8	5.2.B + 5.4.E + 9.5.F	5.1.1.A + 5.2.2.A + 9.1.4 + 11.5.1.C
9	3.7.B a 3.7.F	3.7.1.A a 3.7.1.E
10	$(3.1.A \times 50) / \text{renda média estadual (Pnad Contínua)}$	$(3.1.1.A \times 50) / \text{renda média estadual (Pnad Contínua)}$

Provendo acesso conveniente aos dados

