

Usos de IA/Machine Learning na Construção de Catálogos Inteligentes

Marcel Gomes de Carvalho, BSc, OCP
Diretor Lustrabits
marcel.carvalho@lustrabits.com.br



LUSTRABITS®



SEFIN

Governo do Estado de Rondônia
Secretaria de Estado de Finanças

Objetivo

- Criação de Catálogos inteligentes a partir de qualquer fonte de dados
- Com baixa supervisão humana.
- Primeiro passo no mundo analítico.

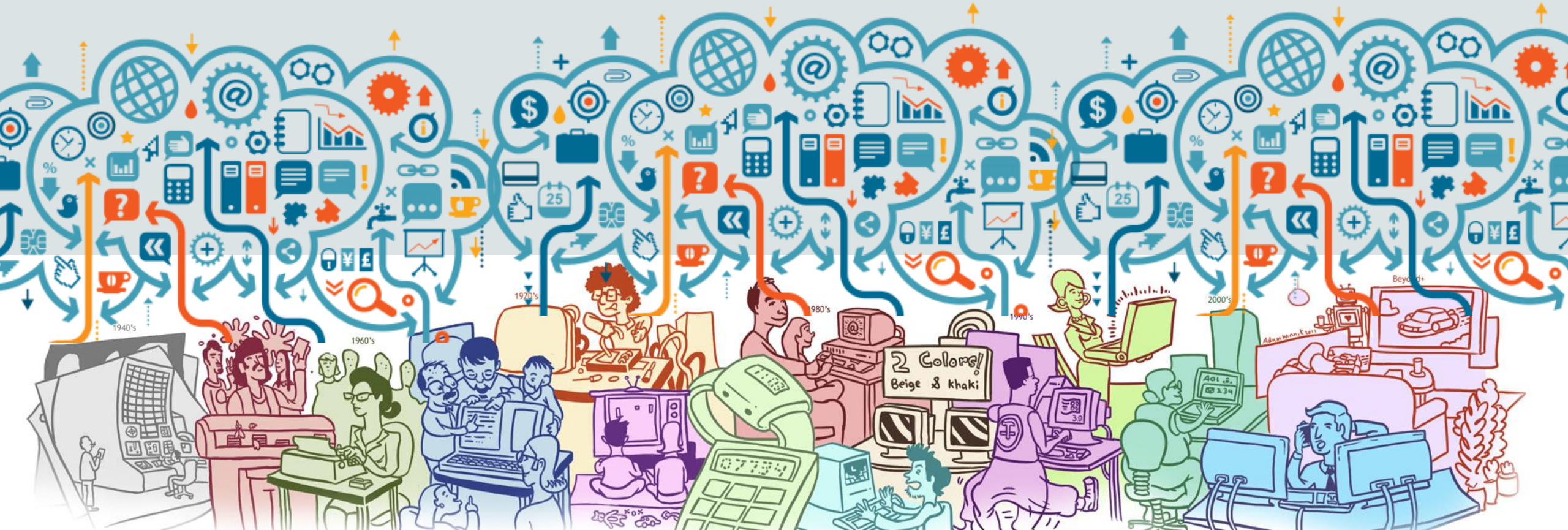


Alguns Fatos e Conceitos Envolvidos

- Volume de dados em crescimento exponencial;
- Em muitos casos a qualidade dos dados é baixa;
- BIG DATA versus Formas de Armazenamento de Dados Convencionais;
- Arquitetura Tradicional versus Arquitetura Orientada a Dados;
- Mineração de Dados;



IDC's Digital Universe Study

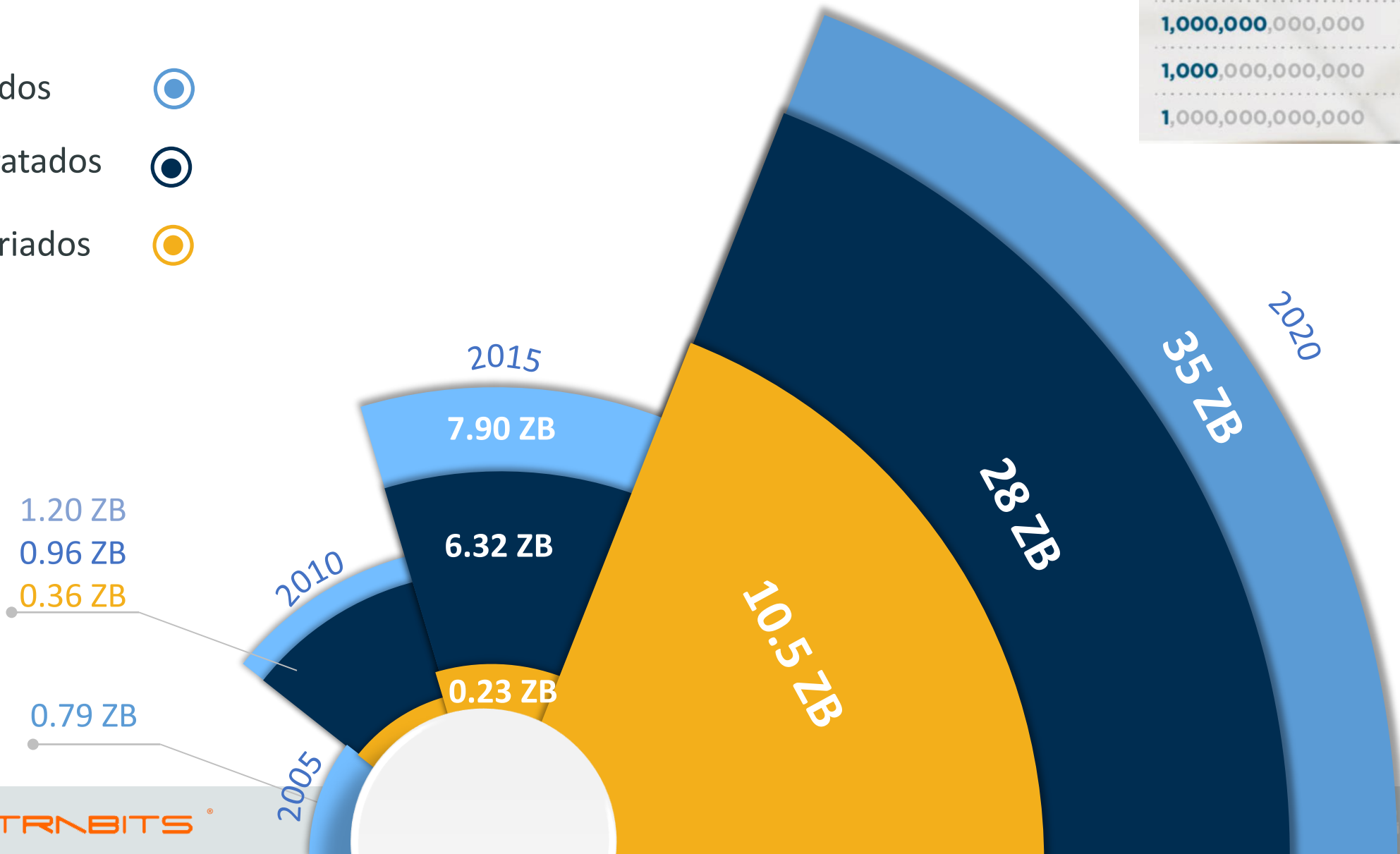


IDC's Digital Universe Study

- Total Dados ◎
- Dados tratados ◎
- Dados Criados ◎

WHAT IS A ZETTABYTE?

1,000,000,000,000	gigabytes
1,000,000,000,000	terabytes
1,000,000,000,000	petabytes
1,000,000,000,000	exabytes
1,000,000,000,000	zettabyte



Dados Estruturados



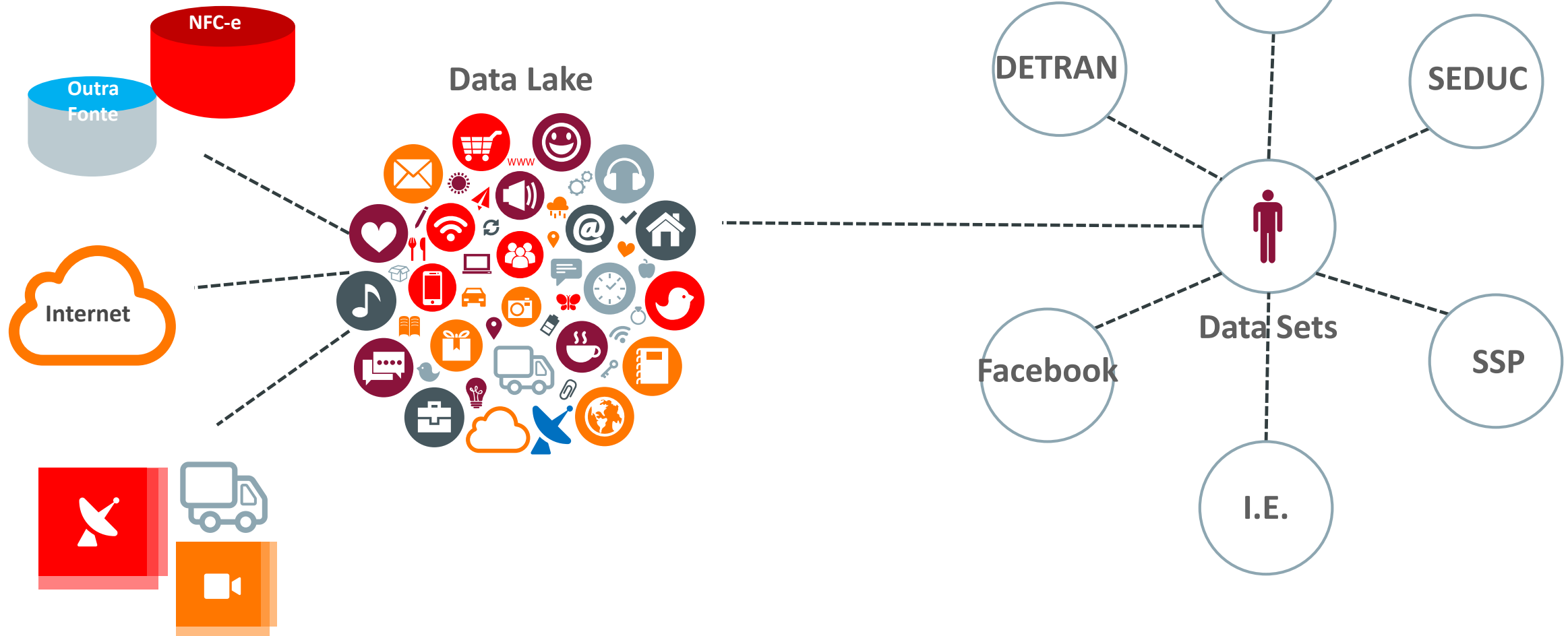
SEFIN

Governo do Estado de Rondônia
Secretaria de Estado de Finanças

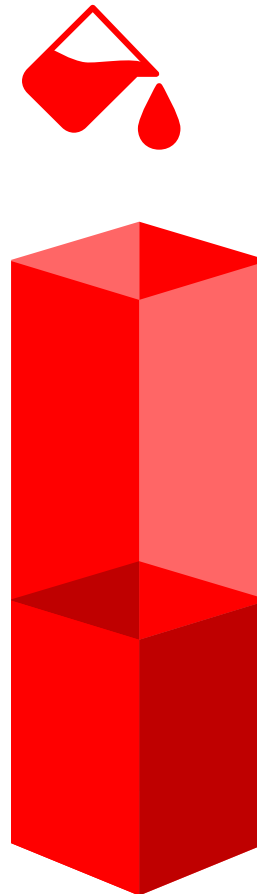
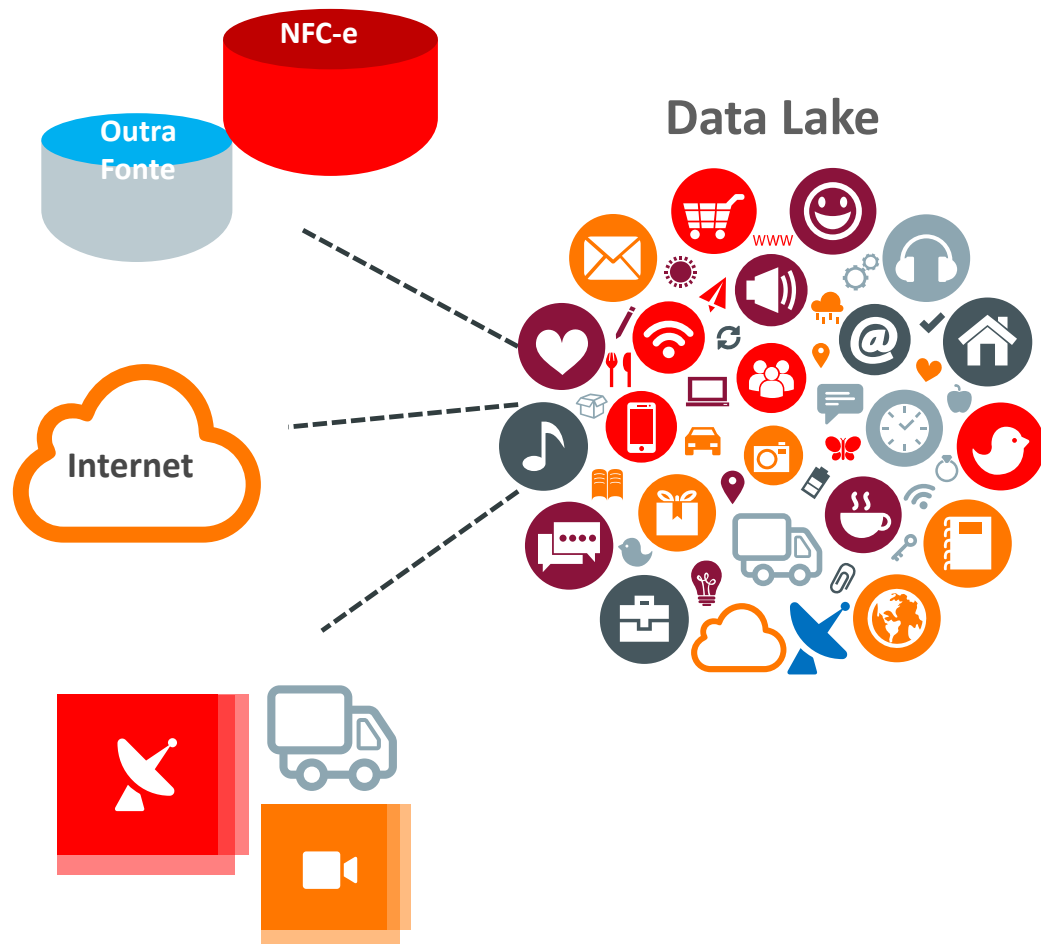
Arquitetura Orientada à Dados



Arquitetura Orientada à Dados



Laboratório de Dados



Construir Hipóteses

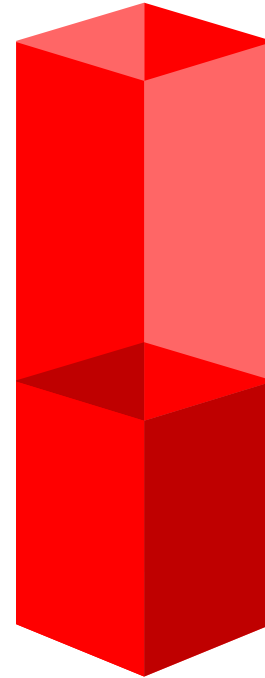
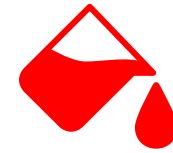
Explorar Dados

Entender os Dados

Preparar Dados



Laboratório de Dados



Construir Hipóteses

Explorar Dados

Entender os Dados

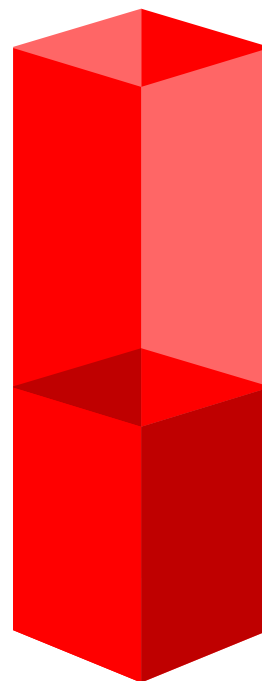
Preparar Dados



Construção Modelo de Aprendizagem



- NCM
- GTIN
- Descrição
 - Marca
 - Complemento
 - Volume
 - Embalagem
- Unidade
- Qtde
- Valor Unitário
- Valor Total



Construir Hipóteses
Explorar Dados
Entender os Dados
Preparar Dados

Desafios 1 – Volume e Qualidade

- Grande volume de dados:
 - Mais de 70 milhões de NFC-es no período;
 - Aproximadamente 350 milhões de itens de NFC-e;
- Baixa qualidade das informações. Muitas NFC-es preenchidas sem muito critério pelos emitentes;
- Solução desenvolvida deve ser abrangente o suficiente para aplicação em outros segmentos da economia sem a necessidade de retrabalho;



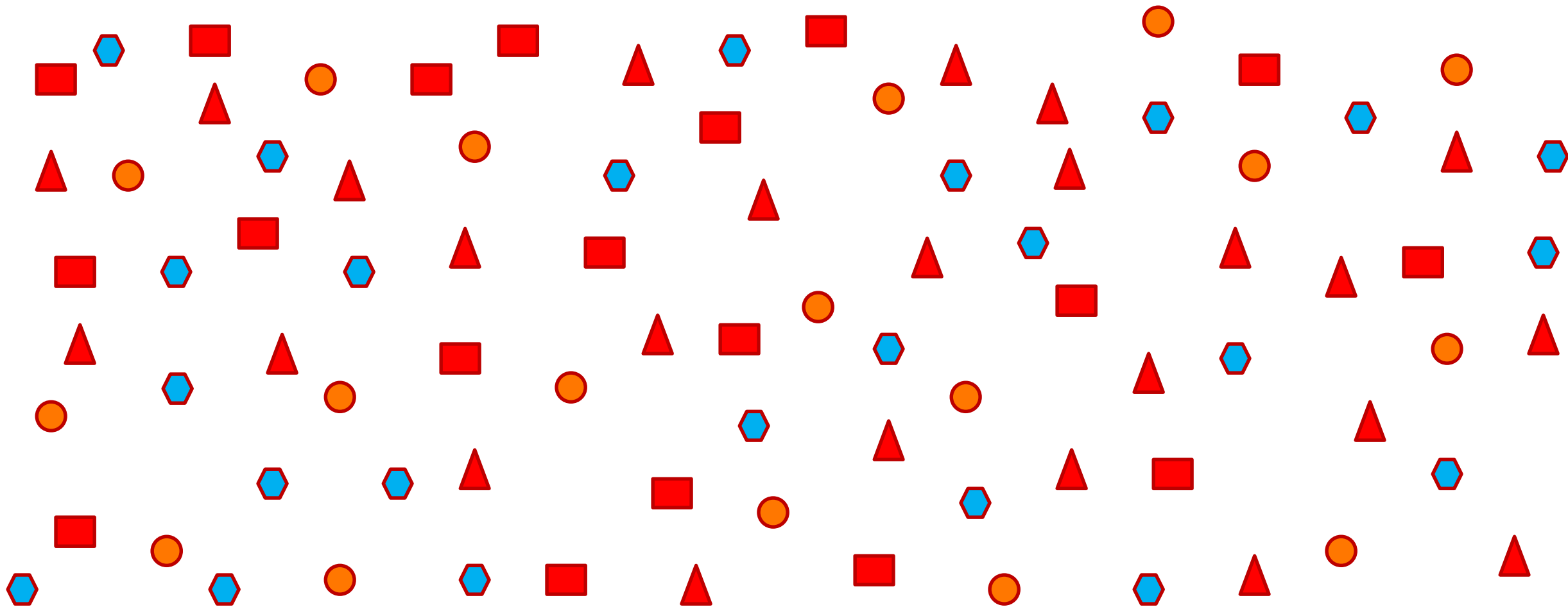
Desafios 2 - Aprendizagem

- NCM e GTIN devem ser usados como parâmetros “fortes” na mineração de dados? Deve-se descartá-los e usar somente as descrições dos itens?
- PMPF pressupõe preço unitário. Como efetuar as conversões de unidades entre as diferentes apresentações do item do Objeto de Estudo?
 - ✓ Caixas;
 - ✓ Qtde de Capsulas
- Como evitar que operações pontuais realizadas em grande volume contaminem o valor do PMPF? Preços muito baixos ou demasiadamente altos devem ser descartados;



A solução

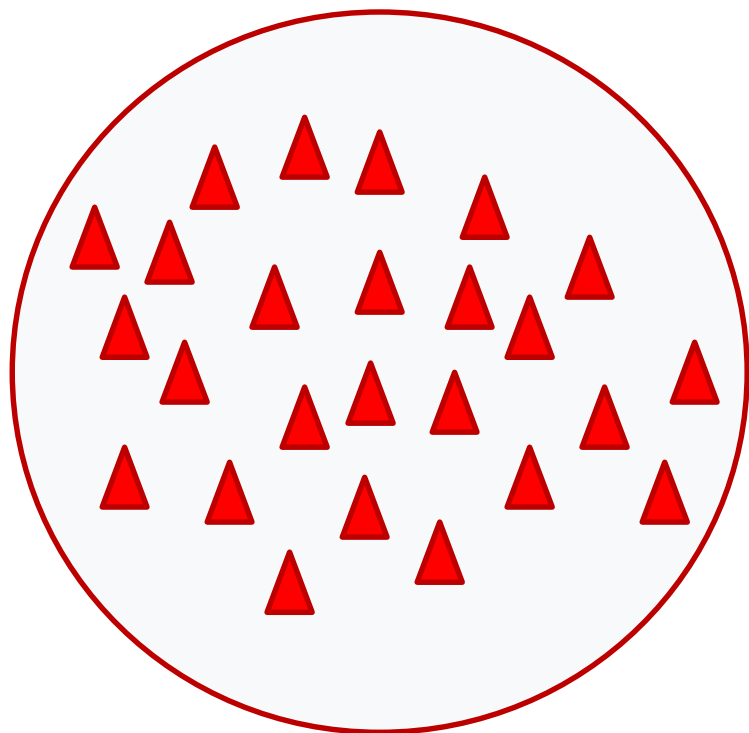
- *Clusterização* – ilustração:



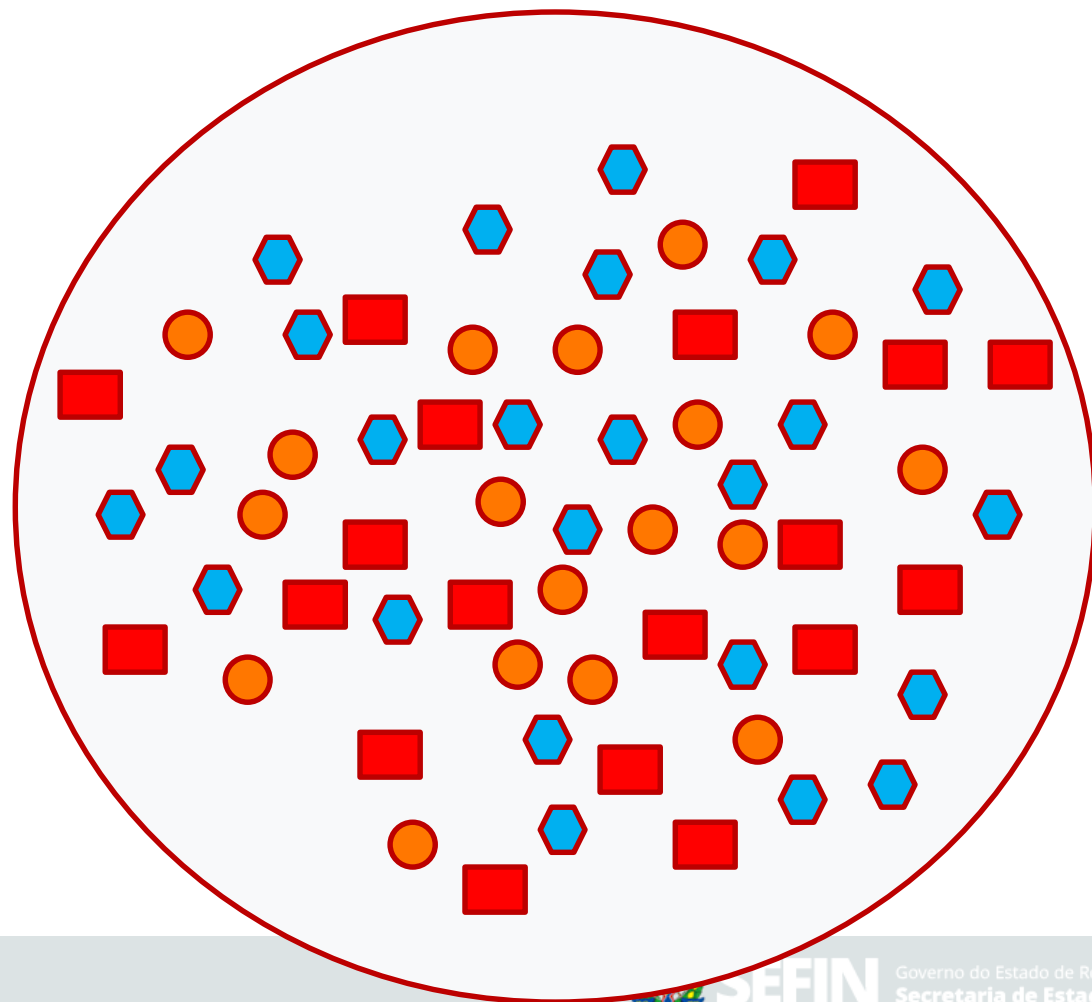
A solução

- *Clusterização* – ilustração:

Cluster 1 – “Medicamentos”



Cluster 2 – “Não Medicamentos”



A solução

- Primeiro Aprendizado – identificação dos itens de “Medicamentos”
 - Segundo Aprendizado – Objetos categóricos a partir do conjunto original:
 - ✓ Nome Comercial;
 - ✓ Princípio Ativo;
 - ✓ Complemento
 - ✓ Fabricante;
 - ✓ Tipo;
 - ✓ MG;
- A definição adequada dos Objetos Categóricos é determinante para o sucesso da mineração de dados / Aprendizado.



Amoxicilina - Legrand
500mg, caixa com 21 cápsulas

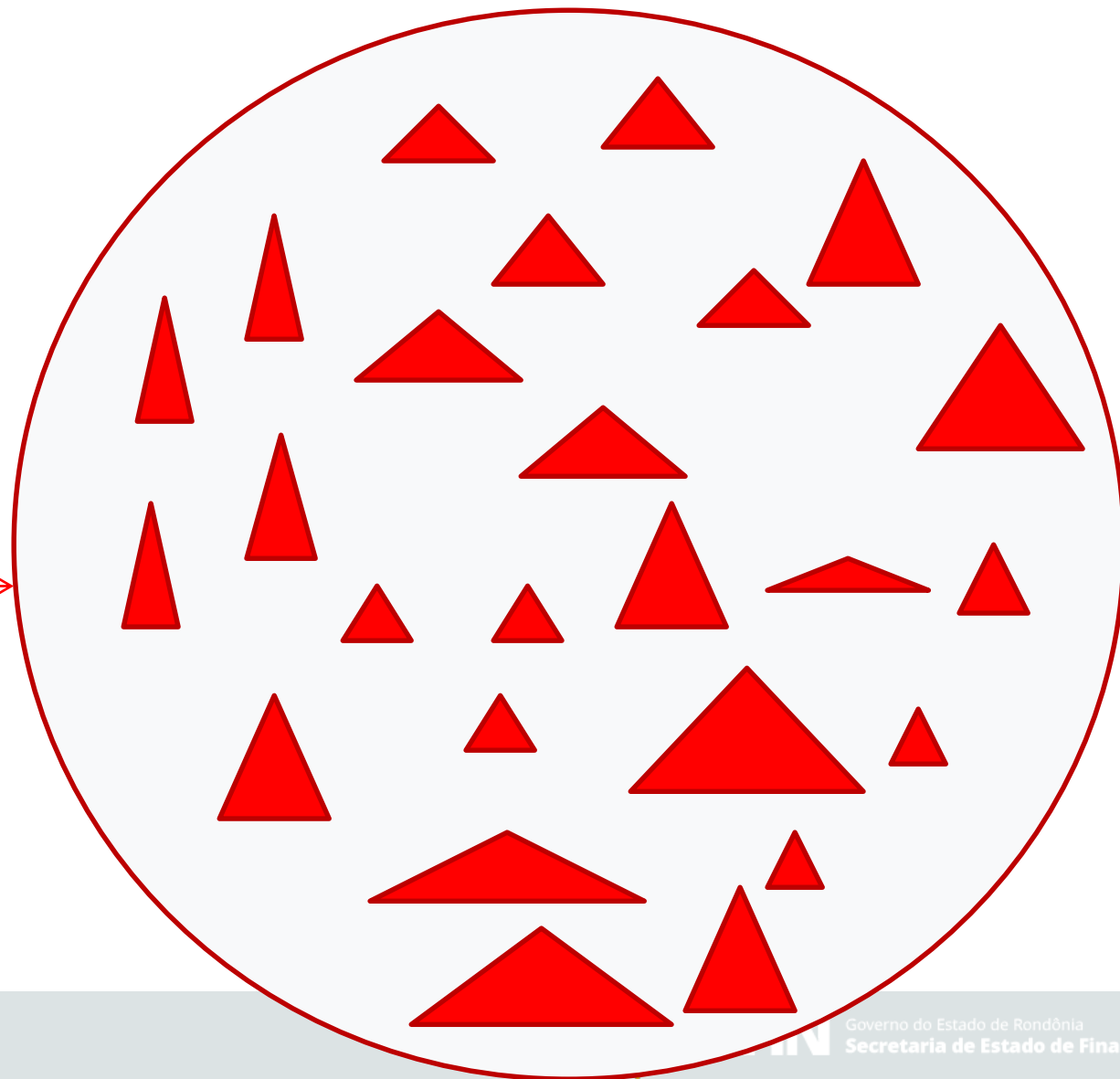
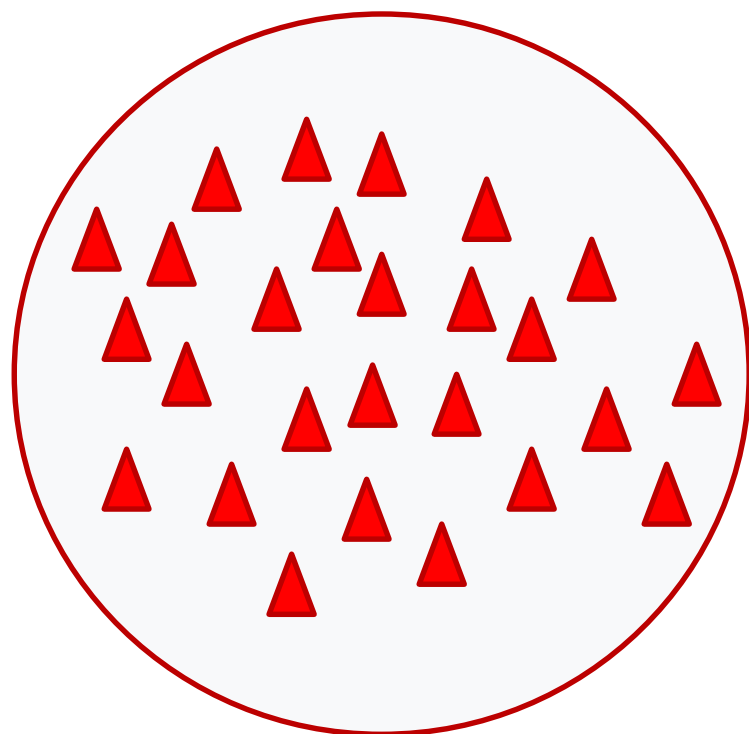
MS 1356904250024 ? PMC/PA R\$ 35,60 ?



A solução

- *Clusterização* – segunda fase:

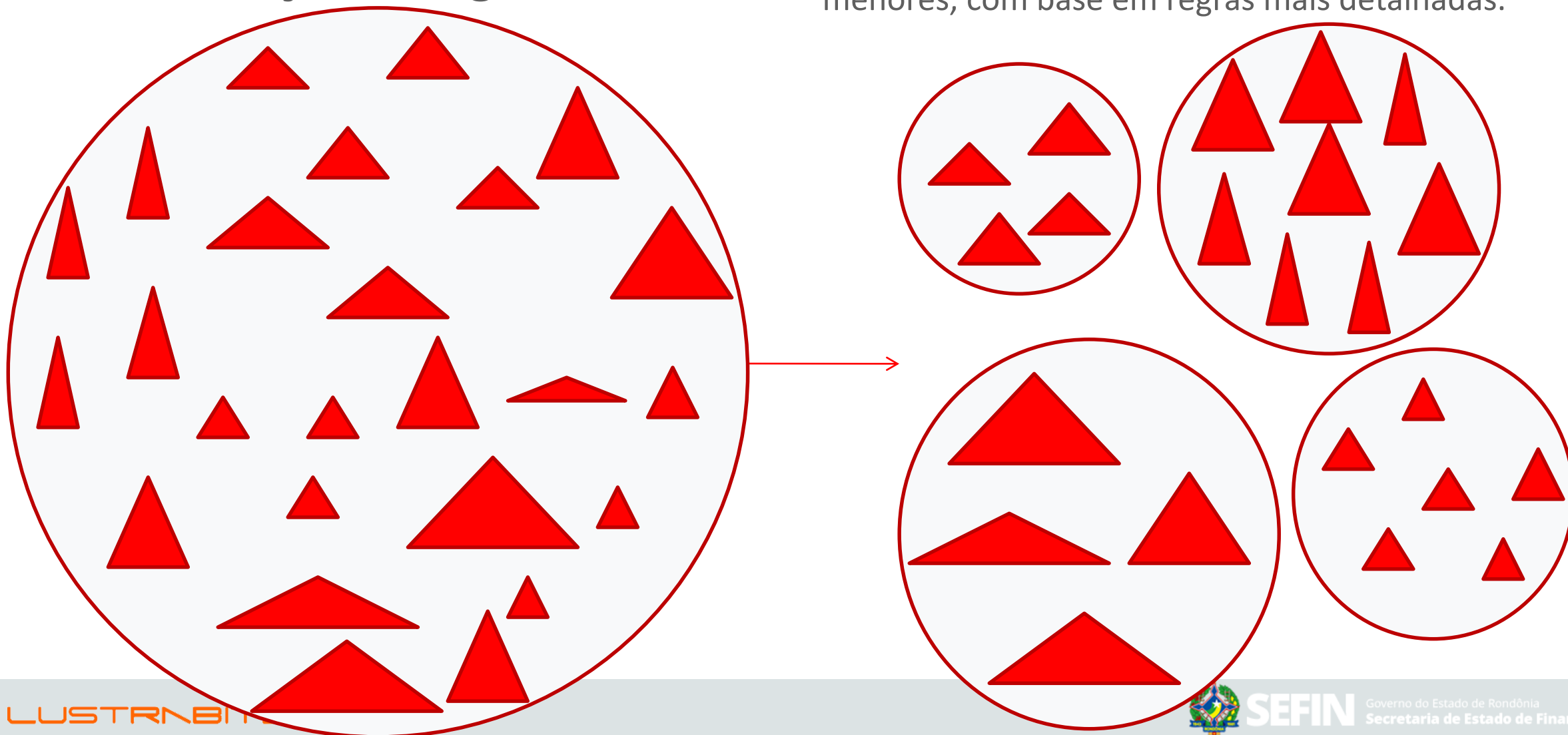
Cluster 1 – “Medicamentos”



A solução

- *Clusterização* – segunda fase:

Divisão do cluster resultante da fase 1 em clusters menores, com base em regras mais detalhadas.



Usos e Aplicações

- **Detecção automática de não conformidade**
- **Cálculo automático preço médio produtos**
- **Utilização do preço médio para compras Governo**
- **Utilização do Preço médio cálculo inflação**
- **Planejamento de Incentivo Tributário Estadual**
- **Estoque Virtual**
- **Desonerações Governamentais na Prática**



Uso 1 - Calculo PMPF

- ✓ Após o aprendizado, o cálculo do PMPF – Preço médio ponderado ao consumidor final – pode ser calculado por período, localização geográfica para cada um dos itens;

**Objetivo:
Publicação do PMPF**



Uso 2 – Aferir o Imposto

- ✓ Com a lista de preços médios, a mineração sobre a base de itens de NF-e, modelo 55, pode ser realizada.
- ✓ Objetivo desta fase é alimentar a base de dados usada na cobrança do ICMS.
 - I. Detecção de não conformidade NCM/MVA
 - II. Avaliação de o PMPF utilizado está correto.;

**Objetivo:
Identificação do
Subfaturamento / NCM
divergente**





Obrigado!



SEFIN

Governo do Estado de Rondônia
Secretaria de Estado de Finanças