

# SAS Fórum Portugal 2016

# Analytics Everywhere

27 de Outubro  
Centro de Congressos de Lisboa



# SPAUTORES

SOCIEDADE PORTUGUESA DE AUTORES

*Cláudia Moreira*

Coordenadora no DINSD - Departamento de Informática / SPA Digital

[claudia.moreira@spautores.pt](mailto:claudia.moreira@spautores.pt)



Diariamente são ouvidas milhões de músicas online.

**Desafio:** Como processar os direitos de autor?

YouTube



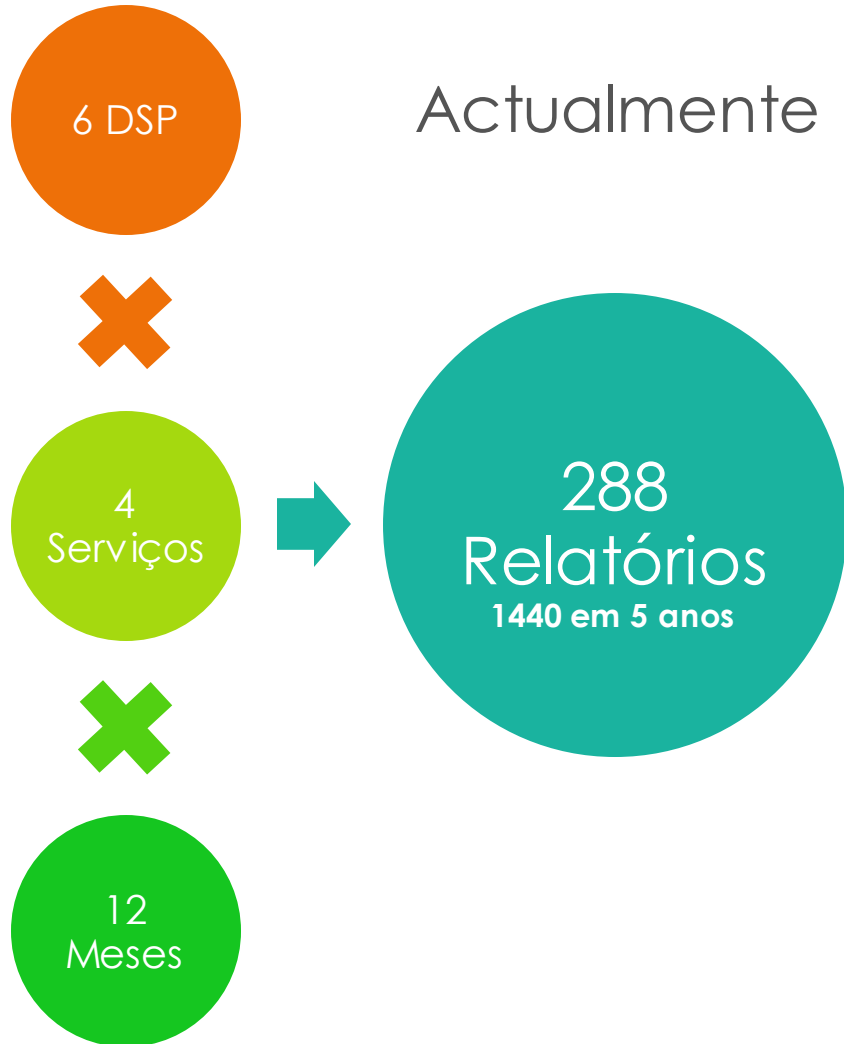
**SPAUTORES**

SOCIEDADE PORTUGUESA DE AUTORES

# DESAFIO

- Devido ao aumento do uso de plataformas *online* de *streaming* a nível mundial, torna-se difícil a recolha dos direitos de autor;
- O número de relatórios de vendas gerados por cada prestador de serviço digital (DSP - Digital Service Provider) está a aumentar exponencialmente, tomando-se difícil processá-los nos sistemas tradicionais;
- Os dados chegam em diferentes formatos e tamanhos extra-grandes (> 10GB);
- Dificuldades em processar a elevada quantidade de dados, fazendo com que os autores não recebam os seus direitos no devido tempo;
- Todos os dias surgem novas plataformas, com mais relatórios de vendas, sendo necessário ter uma solução aberta que possa processar todos os dados de forma rápida e eficiente.

# DESAFIO

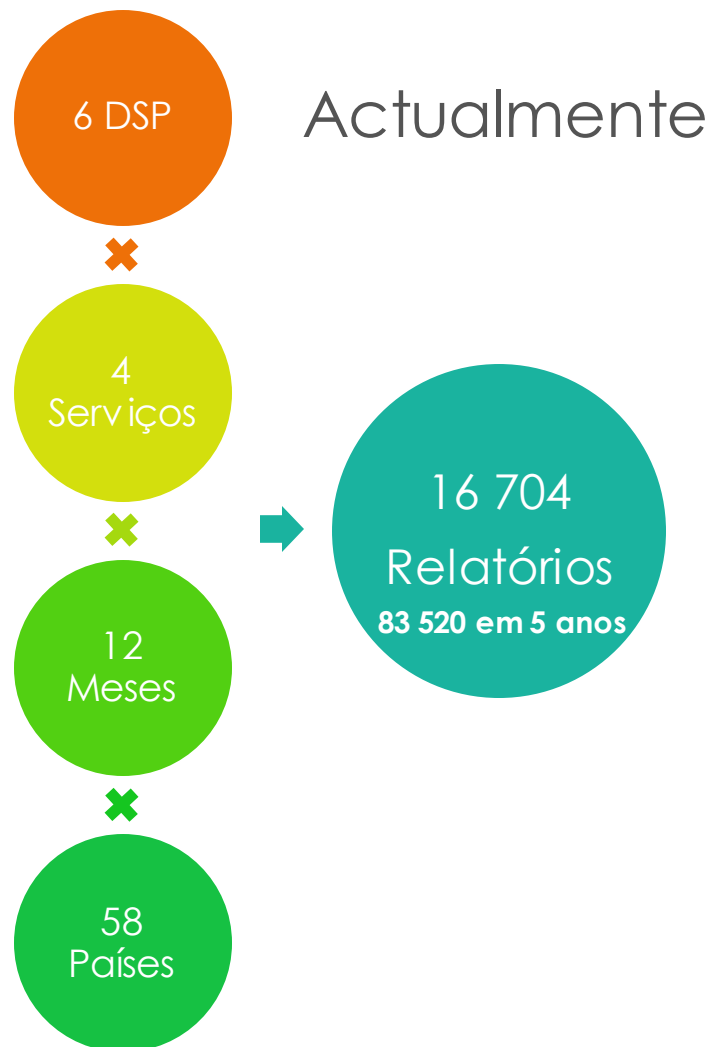


# DESAFIO

Número de relatórios de vendas para o Território PT.

Qual o número de relatórios de vendas para o mundo?

# DESAFIO



A green road sign with a white border and four mounting bolts. The sign features the text "Big Data" in a large, white, sans-serif font at the top. Below it, the text "Straight Ahead" is written in a smaller white font, followed by two white arrows pointing upwards. The sign is set against a background of a blue sky with light clouds.

**Big Data**

**Straight Ahead**





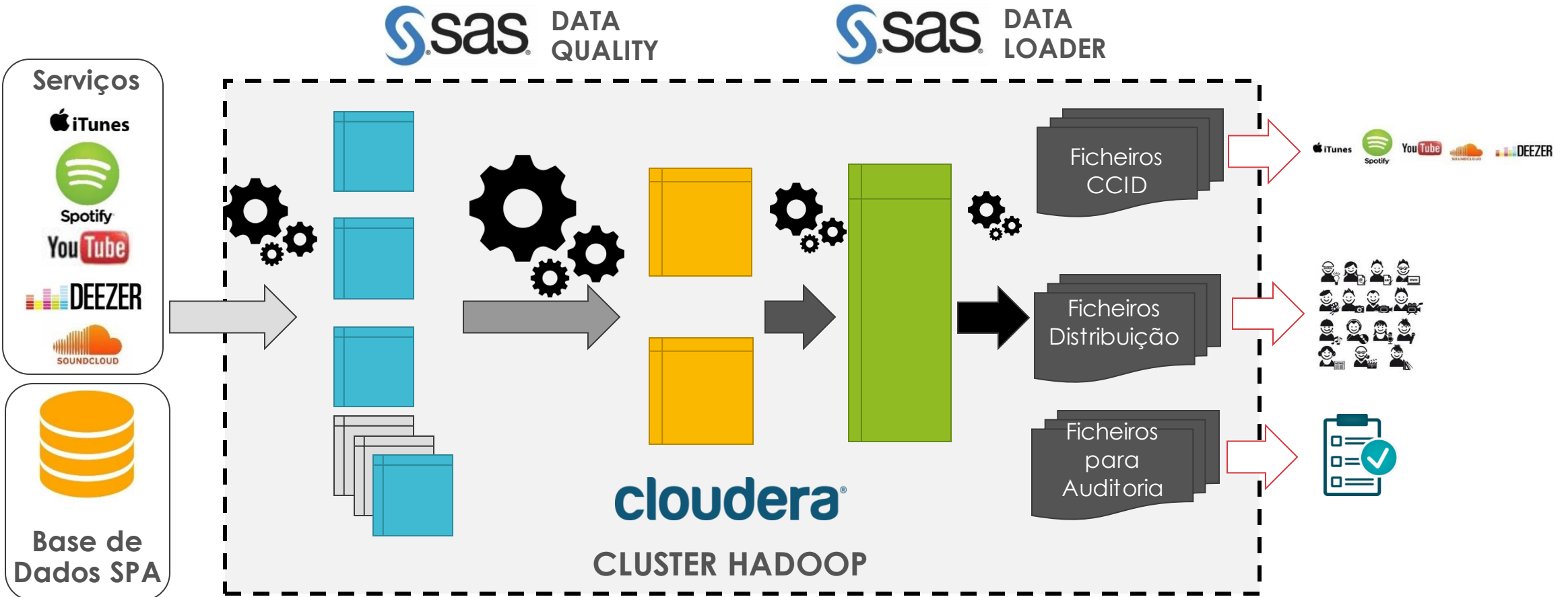
# SOLUÇÃO

- Implementar um *cluster Hadoop*, gerido pelo *software Cloudera*, que é escalável e flexível o suficiente para crescer tendo em consideração as necessidades de processamento;
- Utilização do SAS Data Loader (aplicação gráfica) para desenvolver processos de forma relativamente simples para:
  - Carregar e transformar grandes volumes de dados;
  - Procurar correspondência dos dados recebidos com os dados existentes na nossa base de dados;
  - Calcular a representatividade das gravações e o valor final a cobrar;
  - Gerar CCID (detalhe da factura), ficheiro de distribuição, auditoria dos dados e outros.



**THE POWER TO KNOW®**

# ARQUITECTURA



Cluster Hadoop - 2 horas

Sistema tradicional - 15 dias



# VANTAGENS

- *Hadoop* é:
  - uma plataforma de armazenamento altamente escalável e rápida;
  - permite aceder facilmente a novas fontes de dados e em diferentes tipos de dados.
- *SAS Data Loader* permite que os utilizadores funcionais utilizem o *Hadoop* de forma simplificada;
- Reduz o tempo de execução dos processos de dias para horas.

# VANTAGENS

## Visibilidade do negócio

Baixa capacidade para processamento dos relatórios



Alta capacidade de processamento a baixo custo

## Ficheiros processados

15 dias por relatório de vendas



2 horas por relatório de vendas

## Tempo de processamento

24 relatórios de vendas processados por ano



4368 relatórios de vendas previstos em 1 ano

# ASK BIGGER QUESTIONS



# Obrigada

 **SPAUTORES**  
SOCIEDADE PORTUGUESA DE AUTORES

[claudia.moreira@spautores.pt](mailto:claudia.moreira@spautores.pt)



Implementado pela  **timestamp:biw**  
turning information into knowledge

