timestamp:biw

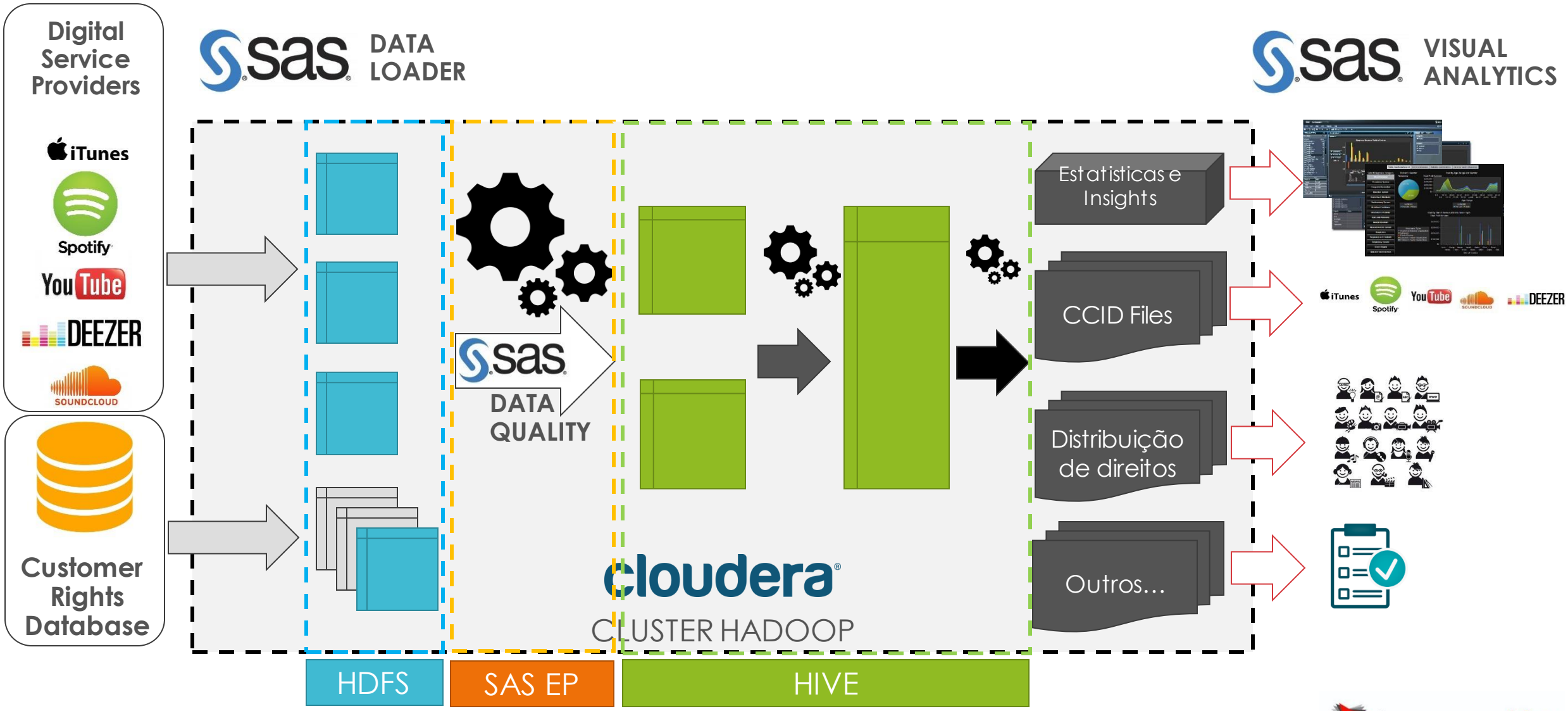
turning information into knowledge

COLLECTING COPYRIGHTS MASSIVELY WITH BIG DATA

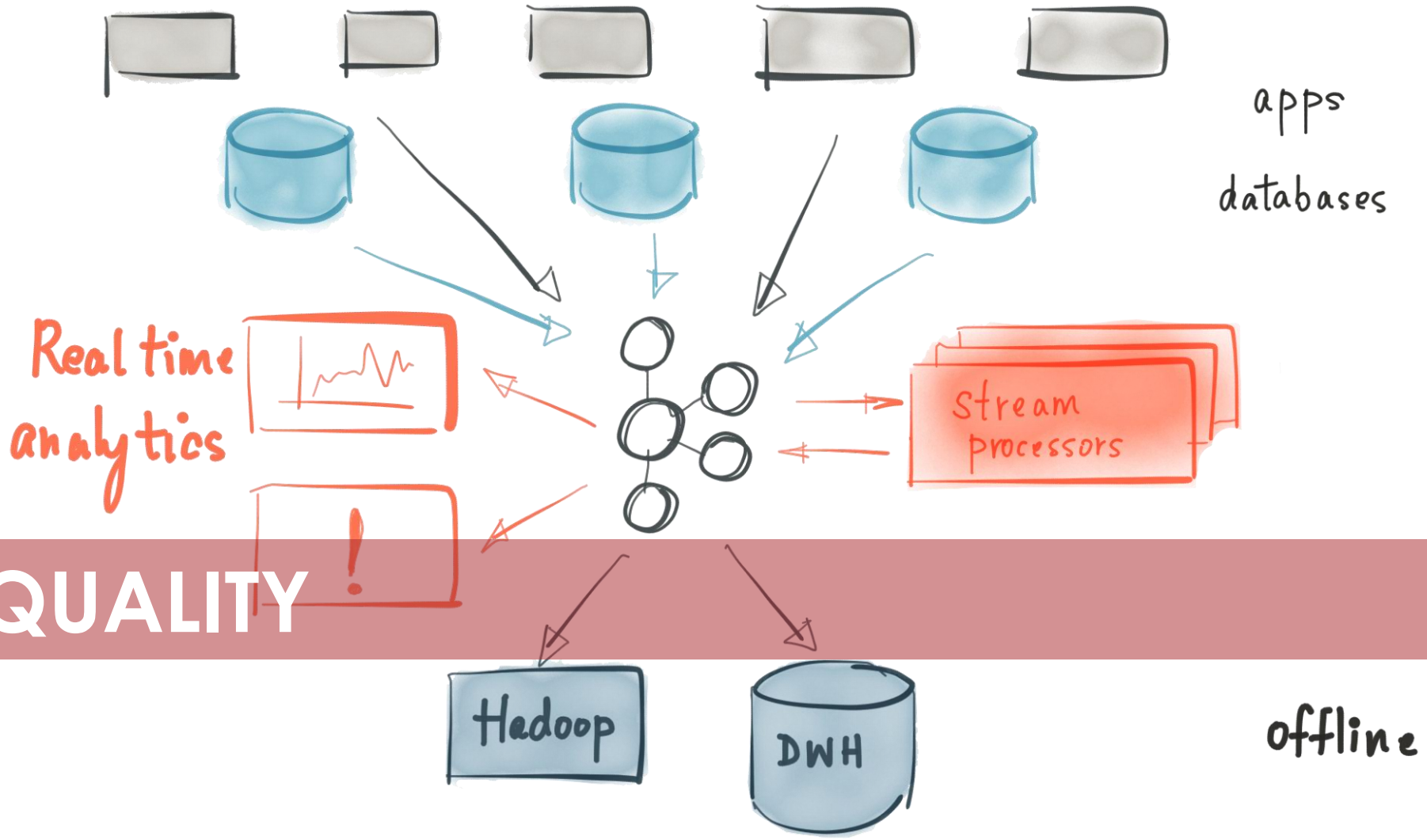
BIG DATA & SAS

- **O que é Big Data?**
  - Big Data e Hadoop são sinónimos.
  - É um conjunto de aplicações/serviços aplicacionais distribuídos.
- **Para que serve e quais são as vantagens face ao que já existe?**
  - Para processar volumes de informações, nos seus formatos nativos, uma vez que a sua principal vantagem é a capacidade de leitura e escrita (face a um sistema tradicional)
- **Como se usa de uma forma prática?**
  - Através de ferramentas visuais como o **SAS Data Loader** (para carregamento e processamento), **SAS Data Quality** (para melhorar a informação) e o **SAS Visual Analytics** (para reporting e análise).

ARQUITETURA



STREAM DATA PLATFORM



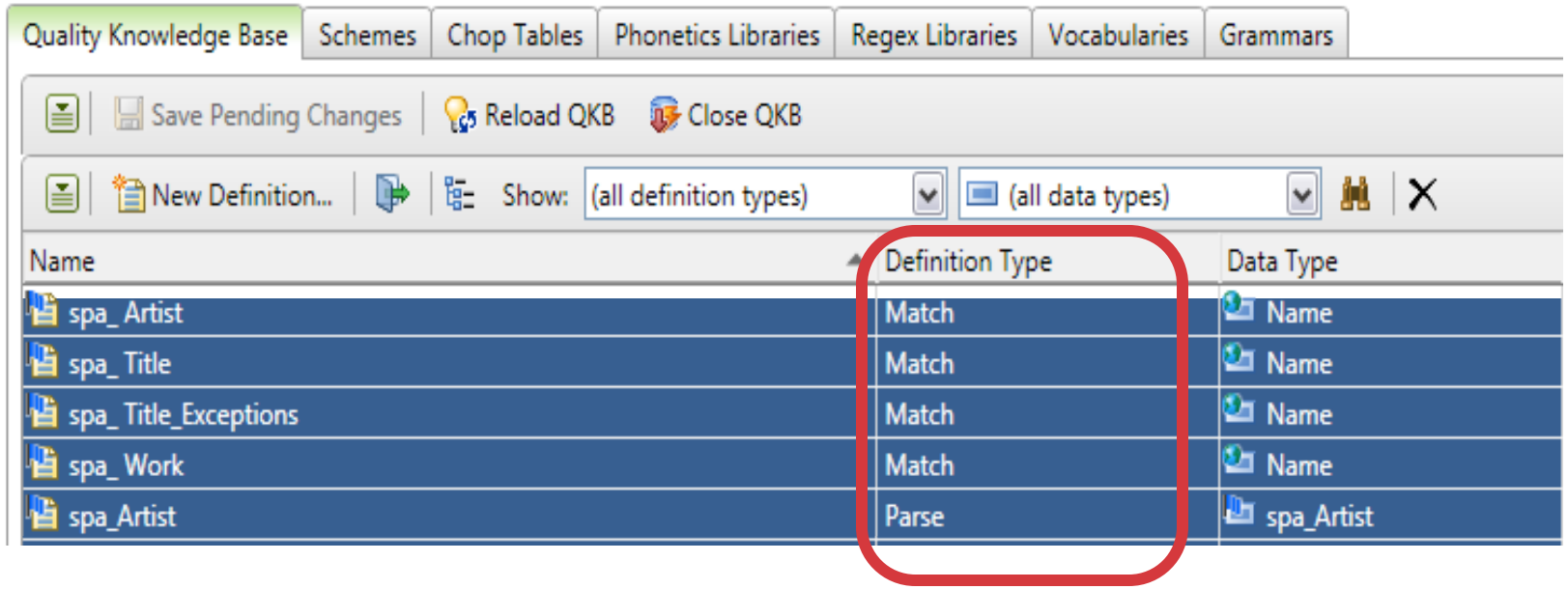
DATA QUALITY

DATA QUALITY - DESAFIO

- Matching de *Obras* e *Artistas* dos Digital Service Providers na base de dados de referência

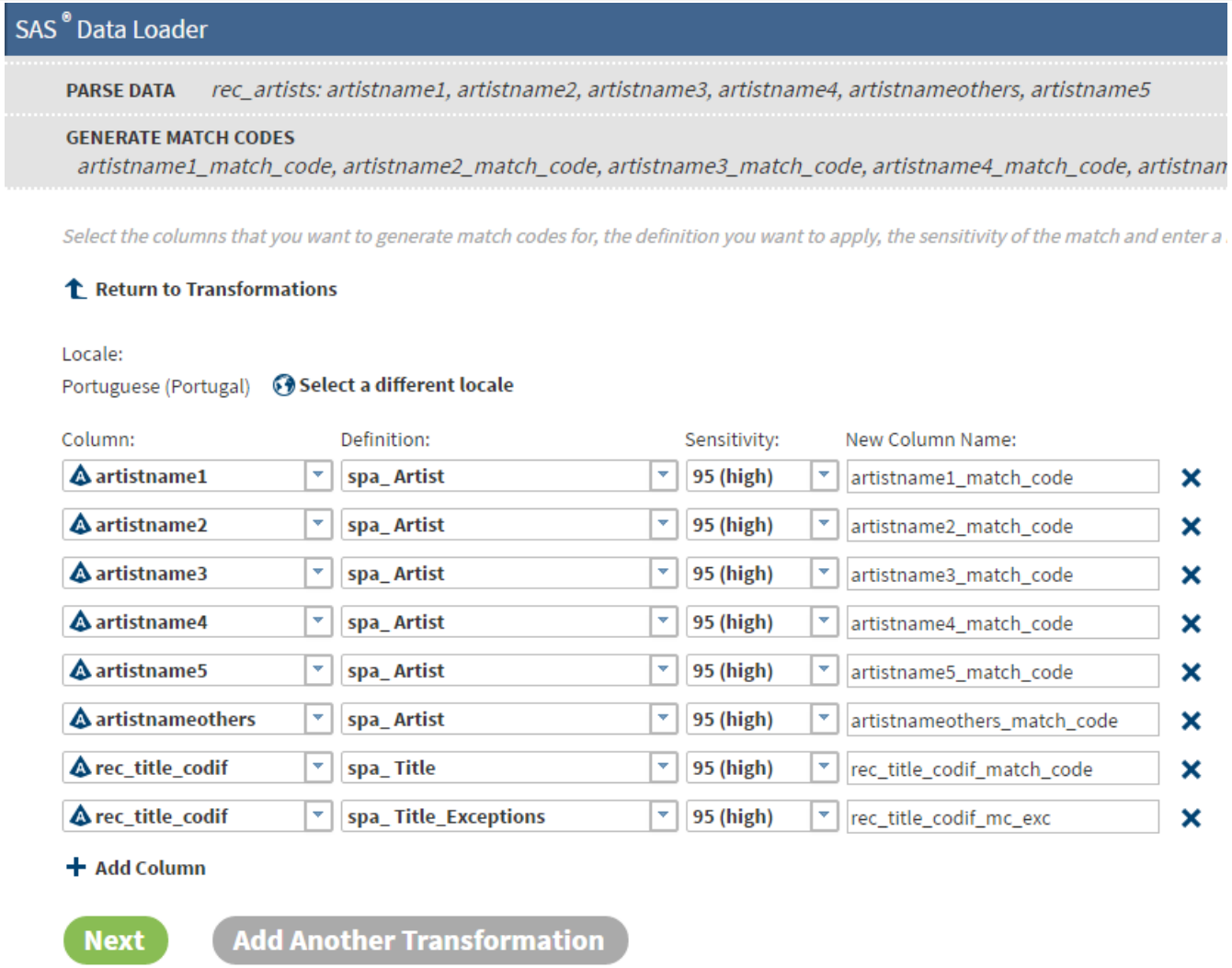
DATA QUALITY - ABORDAGEM

- Configuração SAS Data Quality Knowledge Base
- Definições de
  - ***Parsers***
  - ***Match Codes***
- Aplicado a
  - ***Autores***
  - ***Artistas***
  - ***Obras***
  - ***Títulos***



DATA QUALITY - UTILIZAÇÃO

- Utilizado através de:
  - Reference Database
  - SAS Data Loader
- Os Match Codes obtidos são usados nos critérios de cruzamento de dados



timestamp:biw turning information into knowledge

DATA QUALITY - EXEMPLO

Reference Database

Rod Stewart	Tonight's the Night (Gonna Be Alright)	Rod Stewart Tom Dowd
4B~2$&G6F^	9GDT$$$VBS$$$$	4B~2$&G6F^ 7HB$$$D^~J$$$

Data Loader

Cleanse Data
Cleanse data in Hadoop by performing data quality transforms

Rod Steward	Tonight Night	Rod Stewart Tom Dowd
4B~2$&G6F^	9GDT$$$VBS$$$$	4B~2$&G6F^ 7HB$$$D^~J$$$

DSP Data

Rod Steward	Tonight Night	Rod Stewart feat Tom Dowd

Browse Tables
Browse tables or open a table to see its contents

Chain Directives
Run multiple directives in a specific order

Run Status
Show the status of current and previous directive executions

Saved Directives
Open a previously created directive to run, view or edit

Cluster-Survive Data
Define rules to cluster similar records into groups and optionally create a best record to ...

Delete Rows
Delete rows from a selected table. Requires Hive 14 or above.

Match-Merge Data
Match-merge rows from one or more source tables into a single row and output a single ...

Query or Join Data
Query a table, or join data from multiple tables

Run a Hadoop SQL Pro...
Run custom SQL code

Run a SAS Program
Run custom SAS code

Sort and De-Duplicate ...
Query, sort, or de-duplicate the data in an existing

Transform Data
Transform data from a Hadoop table

SAS DATA LOADER

Transpose data from a Hadoop table

Copy Data from Hadoop into a database

Copy data from a database into Hadoop

Import data from a file into Hadoop

Load Data to LASR
Copy data from a source and load it into LASR. Existing data in the target table will be re...

Cleanse Data
Cleanse data in Hadoop by performing data quality transforms

Profile Data
Generate a profile report of the data in a table

Saved Profile Reports
Explore previously generated profile reports

SAS DATA LOADER


graph
stats
PROCESS
ESS
ANALYTICS
timestamp:biw
turning information into knowledge
BUS
model
insight
OPPORTUNITIES
GRAPHICS
PREDICTION

VANTAGENS

- O Hadoop é uma plataforma de armazenamento altamente escalável, rápido e de baixo custo;
- SAS Data Loader permite que os utilizadores funcionais acedam facilmente ao Hadoop;
- SAS Data Quality potencia o *matching* dos dados provenientes dos DSPs na base de dados de referência;
- SAS Visual Analytics fornece *insights* e estatísticas sobre a informação;
- Os processos IT de negócio reduzem drásticamente o tempo de processamento;

PORQUÊ A TIMESTAMP:BIW?

- 11 anos dedicados a projetos em Data Warehouse, BI & Analytics e EPM
- Experiência em Big Data
- Metodologias de desenvolvimento para soluções Big Data
- Soluções Hadoop suportados em ambientes em Produção
- Consultoria especializada em soluções Big Data

POSICIONAMENTO FACE AO BIG DATA

Transformando
A informação armazenada permite gerar análises preditivas e novos fluxos de valor para o negócio

Otimizando
O desempenho operacional é otimizado com a utilização de Big Data, melhorando processos, tempos e resultados

Explorando
O potencial da utilização de Big Data leva a experiências Hadoop e resultados básicos, mas sem efeitos práticos e reais no negócio

Informado
Big Data é um termo conhecido, mas não se reflete em estratégias de negócios ou processos

ASK BIGGER QUESTIONS

E VOCÊS, EM QUE ESTÁGIO DE ADOÇÃO SE ENCONTRAM?

OBRIGADO

timestamp:biw
turning information into knowledge