

SAP Solution Brief

SAP Solutions for Information Management

SAP Data Services

Objectives

Solution

Benefits

Quick Facts

Unlock the Value of Text Data with SAP® Data Services



The Best-Run Businesses Run SAP®

Combine structured and unstructured data for new insight

Combine structured and unstructured data for new insight

In today's world of Big Data, companies are grappling with volumes of raw data that are growing exponentially. Unstructured and semistructured data from sources such as documents, Web logs, e-commerce sites, social media, sensor devices, and multimedia are contributing to that growth. What if you could **gain new insight from this text data** to compete more effectively?

Successful businesses need unprecedented insight from Big Data. That means extracting useful intelligence from unstructured content and combining it with structured content for new contextual insight. By leveraging unstructured text data alongside regular data, you can gain significant insights into your business, identify emerging trends, and proactively respond to opportunities or potential risks.

With SAP® Data Services software, you can process large volumes of unstructured data to identify and extract relevant text. What's more, you can rapidly load that text into a high-performing database, where it can be combined with structured data for deeper analysis to gain exceptional business intelligence.



Process text data to extract relevant content

Process text data to extract relevant content

Extract valuable text data from Apache Hadoop without coding

SAP Data Services enables you to identify, transform, extract, and load meaningful information from text data sources into a high-performance database like SAP HANA® or SAP Sybase® IQ software. The result? You can perform contextual analysis of unstructured and structured data in real time.

Using the designer tool in SAP Data Services, you can identify the data sources (such as text messages, documents, PDFs, spreadsheets, presentations, Web pages, and XML files). You can also designate the targets (such as databases, business intelligence tools, and applications) into which the relevant text will be loaded.

Next, you configure the software to extract relevant text from the data sources. Optionally, you can configure an extraction dictionary or rule to customize how the text is pro-

cessed. To find customer sentiment, for example, you can generate a list of variant product names in a dictionary and use rules based on voice-of-the-customer data.

[Continued on next page](#) ►

USE CASE: IMPROVE CUSTOMER SATISFACTION AND REDUCE COSTS

An airline wants to automate how it responds to customer e-mails. Using SAP® Data Services software, the firm can classify incoming e-mails in various languages; extract relevant text about flights, times, and locations; eliminate duplicates; and automate accurate text responses. This boosts customer service and cuts call-center costs.



Objectives

Solution

Benefits

Quick Facts

Process text data to extract relevant content

Extract valuable text data from Apache Hadoop without coding

When executing text data processing, SAP Data Services applies linguistic techniques to identify the input text language. Using this linguistic information, the software extracts core entities (such as people, companies, or things) and domain facts (such as topics, requests, or sentiment) from the text.

Optionally, the software can apply data quality techniques such as matching, data cleansing, address cleansing, and geocoding. This enables you to load clean and accurate data into your database to help ensure confidence in decision making.

SAP Data Services supports 31 languages for linguistic markup and extraction, 13 languages for predefined core extractions, and 5 languages for sentiment analysis.

USE CASE: REDUCE WARRANTY AND MAINTENANCE COSTS

To understand the drivers behind rising warranty and maintenance costs, a manufacturer needs insight into employee maintenance logs. With SAP® Data Services software, staff members can extract and cleanse relevant text from the logs – categorizing it by part number, sentiment, or other variable. They can store the text in a high-performing database for deep analysis – reducing warranty costs by identifying and solving common problems.



Extract valuable text data from Apache Hadoop without coding

Process text data to extract relevant content

Extract valuable text data from Apache Hadoop without coding

When data is petabyte in scale, moving it to the point of processing is extremely inefficient. Using SAP Data Services to process text data at the source is an efficient way to extract only relevant data that you can store in a high-performing database for deeper analysis with structured data.

SAP Data Services delivers Apache Hadoop connectors that push down text data processing for semantic extraction to the source files. This lets you leverage the power of Hadoop MapReduce for parallel processing to increase performance, without writing code.

Once SAP Data Services extracts semantic entities and patterns, the software can rapidly load relevant data into SAP HANA, where it can be combined with structured data for

contextual analysis in real time. Deep integration between SAP HANA and the SAP BusinessObjects™ Business Intelligence suite lets you deliver results quickly to business users for competitive advantage.

Continued on next page ►

USE CASE: RESPOND QUICKLY TO CUSTOMER SENTIMENT

A company wants to better understand how customers view its products by analyzing sentiment posted on Facebook. With SAP® Data Services software, the firm can extract the social media data from its Apache Hadoop system without coding. Using text data processing, it can analyze sentiment, identify negative perceptions, and quickly respond to improve customer satisfaction.



Objectives

Solution

Benefits

Quick Facts

Process text data to extract relevant content

Extract valuable text data from Apache Hadoop without coding

SAP Data Services can interact with Hadoop in multiple ways, including the following:

- Apache Hive support – To query data in Hive database tables, SAP Data Services generates Hive Query Language (HQL) scripts. These scripts can support simple joins, order data, filter data, and apply functions (such as average, minimum, maximum, and so on). Hive converts the query into a MapReduce job, and the resulting data files are stored in the Hadoop Distributed File System (HDFS).
- HDFS direct access – To directly access files within a distributed file system, SAP Data Services leverages the libhdfs application programming interface for HDFS and reads those files into memory for processing.
- HDFS access via Apache Pig – To perform standard query operations or text data processing, SAP Data Services generates Apache Pig scripts against HDFS files.

It pushes down text data processing for semantic extraction based on linguistic markup and pattern matching through distributed MapReduce processing within the Hadoop cluster. Results are then stored in the Hadoop Distributed File System.



Mine text from large volumes of unstructured data in Apache Hadoop with text data processing.



Gain value from structured and unstructured data

Gain value from structured and unstructured data

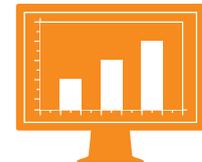
By converting unstructured data into structured information, SAP Data Services delivers new insights from text data. It processes a variety of data sources including e-mails, documents, and all kinds of free-form text stored in an enterprise resource planning (ERP) system or database. It then extracts relevant text and rapidly loads it into SAP HANA, SAP Sybase IQ, or another database for deeper analysis with structured data.

Now it's easier than ever to find nuggets of information from Big Data. That's because SAP Data Services lets you push down text data processing to Apache Hadoop to process text data at the source – without coding.

SAP Data Services also helps ensure data quality. You can cleanse, match, and deduplicate information before it is loaded into a data store. The software also enhances textual data by enabling address cleansing and geocoding to support spatially enabled applications.

And with support for 31 languages, SAP Data Services harnesses the value of text data across the world – enhancing global competitiveness.

SAP Data Services is designed to help you gain value from all your Big Data.



Objectives

Solution

Benefits

Quick Facts

Summary

SAP® Data Services software enables you to access and integrate structured and unstructured data from data sources across your enterprise for deeper insights.

Objectives

- Extract relevant text from unstructured documents at the source
- Access relevant text from large volumes of content in Apache Hadoop, without coding
- Create customized dictionaries to identify what text to extract
- Create custom rules to identify patterns between text extractions
- Ensure the quality of extracted data
- Load relevant text rapidly into a database, business intelligence tool, or application for deeper analysis

Solution

- Unified solution for extracting, transforming, cleansing, and loading relevant text data from unstructured data sources
- Intuitive interface for modeling flow between source and target systems
- Flexible data acquisition with file-format filtering as well as database, Hadoop, and application connectivity
- Enhanced data quality via matching, deduplication, address cleansing, and geocoding
- Support for 31 languages

Benefits

- Gain unprecedented insight from structured and unstructured text
- Increase developer productivity and lower cost
- Increase user satisfaction

Learn more

To find out more, call your SAP representative or visit www.sap.com/eim.



© 2013 SAP AG or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. The information contained herein may be changed without prior notice.

Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and other countries.

Please see <http://www.sap.com/corporate-en/legal/copyright/index.epx#trademark> for additional trademark information and notices.



The Best-Run Businesses Run SAP™