How SAP HANA Leverages the
Cloud to Glean Business Insights
from Unstructured Data

SAP’s in-memory computing and cloud-based delivery model makes a strong business case for investing in more innovative ways of analyzing unstructured data.

Executive Summary

For most organizations, significant volumes of data are locked in the form of unstructured data generated by the broad sweep of social and mobile interactions that pervade our personal and professional lives. With SAP’s “in-memory” technology, called HANA, made available in the cloud, organizations can utilize the power of cloud computing to access tools required to effectively and efficiently analyze unstructured data.

Unstructured data is generated in various forms during formal and informal operations. Textual forms of unstructured data include e-mails, contracts (contained in PDF documents), medical records and claims and customer reviews and feedback shared via traditional Web sites and social media. Existing business intelligence (BI) and analytical tools are in our estimation less efficient at analyzing unstructured textual information as compared with structured data stored in database tables.

Research institutes such as PRIME Research, the Thomas J. Watson Research Center and the Fraunhofer Institute for Algorithms and Scientific Computing are said to be making use of open source frameworks like Apache UIMA for extracting information from textual unstructured data for research purposes. This approach is focused primarily around feature-based opinion mining, summarization, aspect-based sentiment analysis, detecting fake reviews, etc. However, a huge gap is observed when it comes to applications that can be used by business to drive decisions based on insights gained from textual unstructured data.

This paper describes a cloud-based solution that uses information extractors such as Apache UIMA for extracting information from textual unstructured data and SAP HANA for generating business insights that correlate with structured enterprise data contained in back-end systems. SAP HANA cloud platform provides a robust platform as a service to develop applications that can provide...
analytical scalability for in-memory data analysis via RESTful API, federated user authentication and integration with back-end enterprise systems.

**Architectural Design**

**Use of SAP HANA Cloud**

SAP HANA cloud is a platform that can host various components of the solution. An Information extractor such as Apache UIMA is first deployed on the cloud. Unstructured data is fed into the information extractor, either in real time or periodically. The information extractor uses industry-specific models as a reference for extracting information._extracted entities and structured enterprise data is saved into persistent data storage. The SAP HANA cloud provides HANA as an in-memory database storage device stored in industry-specific schemas.

Consider a case of a manufacturing company that is interested in obtaining insights from unstructured data generated in the form of official e-mail communications with its vendors and customers. Apache UIMA can be trained to identify and distinguish customers and vendors from their e-mail communication. Other allied information such as product, part, material, location, date, verbs and adjectives can be extracted and saved to the HANA database (containing manufacturing industry-specific schema) in the cloud for correlation analysis. This can help in sentiment analysis or sourcing marketing information.

**Industry-Specific Apps for Business Insights**

Apps specific to an industry are used to analyze data stored in HANA from which users can derive insights. These apps make use of HANA’s Predictive Analysis Library (PAL) to execute various algorithms on the data. Algorithms for cluster analysis, classification analysis and association analysis are currently available. New apps can be custom developed and deployed per the organization’s requirements.

**API for External Use**

The SAP HANA cloud’s application programming interface is exposed for external consumption. External applications that may be on a cloud or reside in on-premises systems can use these APIs to consume business insights. Developers can then make creative use of these insights to develop rich user interfaces using diverse technologies such as JavaScript or Adobe Flash or Microsoft Silverlight. APIs can be consumed by mobile apps as well by using a simple REST protocol.

**Data-Loading Mechanism/Use Cases**

- **Persistent storage scenario:** The user of this cloud service can opt to store data in the HANA system persistently. This provides continuously available analytics to the user, who can then develop his own external apps that make use of the insights provided by the cloud service. He can also access custom-built apps developed by third parties on SAP cloud that sit on top of HANA.

**An Integrated Approach to Meaning Making**

![Figure 1](image-url)
• **Temporary storage scenario:** This mechanism is suitable for one-time consumption of the analytic service (e.g., to generate insights for an advertisement campaign by analyzing consumer reviews). Once the campaign is completed, there’s no need to keep data on the cloud.

• **On-the-fly scenario:** This is purely in-memory analytics where data is not stored permanently within the HANA database. Unstructured data can be manually fed to the information extractor via a user interface or programmatically via API services. Extracted information is stored as temporary tables (in-memory) in HANA; SAP’s PAL is used to analyze the entities. Once the output is calculated, the temporary table is deleted so that there’s no permanent storage. This will be of use to other systems’ software and applications that require analytics to complete a task or transaction. For example, a vendor may want to query a partner’s HANA database by feeding in some of its work-in-progress materials to receive insights about inventory requirements contained in unstructured data stored in HANA.

**Cloud-Powered Analytics Empowers the Masses**

SAP HANA cloud, with its inherent in-memory analytics, provides the right mix of cloud computing’s economic advantages with fast and potent computing power. HANA’s PAL simplifies the process for converting raw data into insights that help inform smarter and more timely business decisions.

The REST API brings this solution into the hands of all the developers who are interested in developing their applications around the insights sourced from unstructured data. As such, the solution bridges unstructured data analytics with enterprise functions.

### Footnote

1 HANA is an abbreviation that stands for high-performance analytical appliance.

### About the Author

*Rahul Aware is a Senior Associate within Cognizant’s Enterprise Application Software Business Unit. He has nine-plus years of experience as a SAP technical consultant and is a certified SAP professional with a flair for new technologies and their applications. He can be reached at Rahul.Aware@cognizant.com.*

### About Cognizant

Cognizant (NASDAQ: CTSH) is a leading provider of information technology, consulting, and business process outsourcing services, dedicated to helping the world’s leading companies build stronger businesses. Headquartered in Teaneck, New Jersey (U.S.), Cognizant combines a passion for client satisfaction, technology innovation, deep industry and business process expertise, and a global, collaborative workforce that embodies the future of work. With over 50 delivery centers worldwide and approximately 162,700 employees as of March 31, 2013, Cognizant is a member of the NASDAQ-100, the S&P 500, the Forbes Global 2000, and the Fortune 500 and is ranked among the top performing and fastest growing companies in the world.

Visit us online at [www.cognizant.com](http://www.cognizant.com) for more information.

---

© Copyright 2013, Cognizant. All rights reserved. No part of this document may be reproduced, stored in a retrieval system, transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the express written permission from Cognizant. The information contained herein is subject to change without notice. All other trademarks mentioned herein are the property of their respective owners.