HANA Real-time Analytics and Application Platform

Qi Su
Customer Innovation & Enterprise Platform
Agenda

Data Architecture
- Data virtualization (Smart Data Access)
- HANA-Hadoop integration
- Data tiering strategy
- Vora

Advanced Analytics
- Multi-modal Advanced Analytics
- Predictive / Machine Learning
- Natural Language Processing (NLP) & Text Analytics
- Geospatial
- Graph
Data Architecture
SAP Big Data Platform Evolution

HANA

In-memory

HANA Dynamic Tiering

Smart Data Access/ vUDF

Altiscale

SAP Vora
Virtualize data access to access data where it is located and reduce data movement

- Enables access to remote data just like a “local” table
- Smart federated query processing
- Supports data location agnostic development
- No special syntax to access heterogeneous data sources
- Benefit from HANA functionality without moving all data to HANA
- Rapid deployment of high-performance, data intensive transactional and analytical applications
HANA-Hadoop Integration

HANA & Hadoop Integration
• SQL on Hadoop via SDA (virtual tables) – Hive (SPS06), SparkSQL (SPS10)
• Remote caching with Hive (SPS07), SparkSQL (HANA2 SPS 00)
• Execution of MapReduce (SPS 09) and Spark Scala (HANA2 SPS 00) jobs via HANA (Virtual Functions), direct access to HDFS (SPS 09)

Key Benefits
• Deep Integration for storage & processing
• Optimized data access between HANA & Hadoop
• Data tiering to Hadoop for cold storage
Data Lifecycle Manager (DLM) for Hadoop

**Define a data aging strategy with DLM**
Leverage SAP HANA Dynamic Tiering (Warm-Store), in Hadoop native use cases with a tool based approach to model aging rules on tables to displace ‘aged’ data to HANA extended tables to optimize the memory footprint of data in SAP HANA.
Vora
SAP Vora

SAP Vora is an enterprise-ready, easy-to-use in-memory distributed computing solution to help organizations uncover actionable insights from big data.

- Builds upon Apache Spark
- Runs on Hadoop
- Seamless Integration with SAP HANA
Distributed Computing for the Digital Enterprise

SAP HANA Vora

Data Modeler

- OLAP
- Time Series
- Graph
- Doc Store

- Disk-to-Memory Accelerator
- Distributed Transaction Log

Spark

Hadoop

Insights from one single solution

In-memory distributed computing engines: OLAP, Time Series, Graph, JSON/Doc

Disk-to-memory accelerator

Enterprise-ready

Production-ready, integrated solution

Integration with SAP HANA

Easier to use

Intuitive web interface

One SQL entry point

Open consumption
Vora Value in the Hadoop world

Data access methods
- Security
- Data model
- Cypher, SQL browser

Functions & data types & tools
- Graph function
- Time Series function
- JSON Docu Store
- OLAP functions / relational modeling

Infrastructure
- Dedicated infrastructure

SAP HANA Vora
- Vora Tools
  (Data Browser, SQL Editor, OLAP Modeler)
- Security

Integrated Engines
- Graph
- Time Series
- Relational
- JSON Doc Store
- Future etc.
- Data Exchange

- Disk-to-Memory Accelerator
- Distributed Transaction Log

Shared Hadoop Infrastructure
# SAP HANA Vora Roadmap

**Product road map overview - key themes and capabilities**

**Current Release V1.3**
- Enhanced Security for Vora tables (Kerberos and role based authorization)
- HANA integration: Query optimization, auto discovery of HANA tables and views
- More OLAP features like Currency conversion
- Engine support for TimeSeries, Graph, DocStore and Disk store
- HANA integration: voraojdbc connection between HANA and VORA

**Planned Q1/2017**
- Completely distributed VORA (including next generation relational engine)
- EIM integration / Data quality tools
- Additional primary distributed persistence layer (DLOG)
- Cloud support (AWS, SCP)
- Insert/delete support for DocStore
- Optimizations for Time Series, Graph engines
- Support for SCPBDS fka. Altiscale

**Planned Q2/2017**
- Support for Apache Sentry / Ranger
- Integrated Machine learning
- Vora integration with IoT Platform
- Text Search
- Messaging/streaming on top of DLOG
- Extended cloud support (Azure, Google)
- Data pipeline and integration to Kafka
- Extend HANA Data Lifecycle Management to Hadoop through Vora integration

**Planned 2017**
- Video processing
- Multitenancy support
- Audio processing
- Image engine integration
- Distributed Text processing
- Reduce complexity of communication infrastructure (i.e. own JDBC endpoint)
- Spatial data processing

---

This is the current state of planning and may be changed by SAP at any time.
SAP Cloud Platform Big Data Services (Altiscale)
Big Data is complex
It gets more complicated as you scale
About SAP Cloud Platform Big Data Services (Altiscale)

Altiscale Data Cloud: GA in 2014

Led by experienced, renowned Hadoop team from Yahoo!

• Raymie Stata, CEO. Former Yahoo! CTO, well-known advocate of Apache Software Foundation
• David Chaiken, CTO. Former Yahoo! Chief Architect

Built and managed by veterans of Big Data, SaaS, and enterprise software

• Google, LinkedIn, Netflix, VMware, Oracle, and Yahoo!

Recognized innovator in Hadoop-as-a-Service

40,000 nodes
500 PB
1,000 users
$ billions at stake
SAP Cloud Platform Big Data Services: Big Data That Just Works

- **Fast time to value** - days not months
- **Easier, faster scalability** - with elastic scaling
- **Operations support** - so your jobs get done
- **Lower TCO** - for fast investment payback
- **Enterprise ready** - for business-critical applications
Advanced Analytics
Multi-modal Advanced Analytics
Use Case: 30 Day Re-admission Propensity

Requirements:

- Predictive / machine learning models on structured clinical patient data sources
- Geospatial – distance to clinics, hospitals, service providers; mash up census block information on demographics with patient data
- Text extraction, analysis, classification – physician and nurse notes
- Incorporate results of text & geospatial analysis into predictive analysis
Traditional solution architecture: integrate multiple purpose-built analytical application silos

- Predictive/ML (SAS, R)
- GIS (ESRI)
- NLP (OpenNLP, Lucene, Solr)
- Database/Data Lake
Use Case: Public Health Surveillance - Real Time UI

Keyword Search

Geospatial Heatmap

Timeline of Occurrences

Graph of Relationships
Typical solution architecture for advanced analytics applications

Database / data lake

Adhoc data exploration tool

Predictive analytics / machine learning library

Geographical Information System (GIS) & geospatial visualization

Natural language processing & text analysis

Graph processing engine

Real time data stream engine

Application & data integration effort

Data quality & data governance
SAP HANA data application platform

SAP HANA PLATFORM
ON-PREMISE | CLOUD | HYBRID

Application Services
- Web Server
- JavaScript
- Fiori UX
- Graphic Modeler
- Application Lifecycle Management

Processing Services
- Spatial
- Graph
- Predictive
- Search
- Text Analytics
- Streaming Analytics
- Series Data
- Business Functions

Integration & Quality Services
- Data Virtualization
- ELT & Replication
- Data Quality
- Hadoop & Spark Integration
- Remote Data Sync

Database Services
- Columnar OLTP+OLAP
- Multi-Core & Parallelization
- Advanced Compression
- Multi-tenancy
- Multi-Tier Storage
- Data Modeling
- Openness
- Admin & Security
- High Availability & Disaster Recovery
Predictive / Machine Learning
SAP Strategy for Artificial Intelligence (AI) & Machine Learning (ML)

Intelligent Enterprise powered by Application Intelligence

Leverage SAP business network

Data Science expertise

Platform

Machine Learning Services

cloud | onprem

© 2017 SAP SE. All rights reserved.
Apps: Service Ticket Intelligence, Customer Retention Insight, Brand Intelligence, Resume Matching, Cash Application, SAP Fraud Management, S/4HANA Procurement predict contract consumption

Business Services: CV matching, recommender, etc

Functional Services: image classification, image feature extraction, topic detection, time series changepoint detection, etc

Lifecycle Mgmt.: training, scoring, monitoring, model mgmt., etc
SAP Cloud Platform (SCP) Predictive Services

- **REST web services**
  - Time series forecast
  - Key influencer detection
  - Outlier detection
  - Model based what-if simulation
  - Model scoring
- Easily embed predictive functionality in any HCP application
- No complex modeling or data science required
- Language agnostic
SAP HANA Predictive Ecosystem

SAP Data Science Organization

SAP Predictive Analytics
SAS, 3rd Party Statistical Tools
SAP and Custom Applications
Business Intelligence Clients

SAP HANA Platform

Predictive Analysis Library (PAL)
Automated Predictive Library (APL)
R Integration for SAP HANA

Data Pre-Processing and Loading
SAP Data Services, SLT, SAP Replication Server, 3rd party data movement tools
## PAL Algorithm Roadmap

### Algorithms

<table>
<thead>
<tr>
<th>Clustering</th>
<th>Association</th>
<th>Social Network Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-means/median/medoid</td>
<td>A-priori</td>
<td>Link prediction</td>
</tr>
<tr>
<td>ABC Classification</td>
<td>KORD</td>
<td></td>
</tr>
<tr>
<td>Kohonen SOM</td>
<td>FP-growth</td>
<td></td>
</tr>
<tr>
<td>DBSCAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agglomerate</td>
<td>Linear Regression</td>
<td></td>
</tr>
<tr>
<td>Hierarchical</td>
<td>Exponential/Logarithmic/Polynomial</td>
<td></td>
</tr>
<tr>
<td>Gaussian Mixture</td>
<td>Ridge/Lasso Regression</td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latent Dirichlet</td>
<td>Inter-Quartile Range test</td>
<td></td>
</tr>
<tr>
<td>Allocation (LDA)</td>
<td>Grubbs Outlier Test</td>
<td></td>
</tr>
<tr>
<td>DenStream</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression</th>
<th></th>
<th>Preprocessing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td></td>
<td>Anomaly Detection</td>
</tr>
<tr>
<td>Exponential/Logarithmic/Polynomial/Ridge/Lasso Regression</td>
<td></td>
<td>Variance Test</td>
</tr>
<tr>
<td>Outlier</td>
<td></td>
<td>Chi squared test</td>
</tr>
<tr>
<td>Inter-Quartile Range test</td>
<td></td>
<td>Random Distribution Sampling</td>
</tr>
<tr>
<td>Grubbs Outlier Test</td>
<td></td>
<td>Distribution fitting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Binning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Univariate and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>multivariate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>statistical functions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dataset partitioning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weibull failure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AUC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kaplan-Meier</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Survival Analysis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time Series</th>
<th></th>
<th>PMML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single/Double/Triple/Brown Exponential Smoothing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Croston’s method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIMA, seasonal,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>auto</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forecast smoothing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### HANA2 sp0/sp1

<table>
<thead>
<tr>
<th>Classification</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient Boost Decision Tree</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear Discriminant Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision tree scoring in Smart Data Streaming</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regression</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cox Proportional Hazard Model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Association</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential pattern mining</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factorization machine</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time Series</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Preprocessing</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast Fourier Transform</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collinearity test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANOVA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
R Integration for SAP HANA

- Embedding R scripts within the SAP HANA database execution
- Enhancements are made to the SAP HANA database to allow R code (RLANG) to be processed as part of the overall query execution plan
- This scenario is suitable when the modeling and consumption environment sits on HANA and the R environment is used for specific statistical functions

Sample Code in SAP HANA SQLScript

```sql
CREATE FUNCTION LR( IN input1
SUCC_PREC_TYPE, OUT output0
R_COEF_TYPE)
LANGUAGE RLANG AS '''
    CHANGE_FREQ<-input1$CHANGE_FREQ;
    SUCC_PREC<-input1$SUCC_PREC;
    coefs<-coef(glm(SUCC_PREC ~
        CHANGE_FREQ, family = poisson ));
    INTERCEPT<-coefs["(Intercept)"];
    CHANGEFREQ<-coefs["CHANGE_FREQ"];
    result<-as.data.frame(cbind(INTERCEPT,CHANGEFREQ ))
''';
TRUNCATE TABLE r_coef_tab;
CALL LR(SUCC_PREC_tab,r_coef_tab );
SELECT * FROM r_coef_tab;
```
Provide Data Scientists and Business Analysts with sophisticated algorithms to take the next step in understanding their business and modeling outcomes

• Holistic solution for predictive analytics and advanced data visualization
• Build models and apply to scenarios using R, automated, SAP PAL, and SAP APL algorithms to forecast future outcomes
• Open to most major data sources for a comprehensive analysis
• Library of advanced visualizations within the modelling tool
• Optimized for SAP HANA to support huge data volumes and in-memory processing
Automated Analytics: SAP Predictive Analytics
Data Scientist in a Box

Provide Business Analysts and Data Scientists with a fully automated process

Data preparation
- Create 1000s of derived attributes
- Define metadata once
- Builds analytic dataset automatically

Predictive modeling/Data mining
- Regression/Classification
- Segmentation
- Forecasting
- Association rules
- Social Network Analysis

Advanced model deployment and management
Natural Language Processing (NLP) / Text Analytics
Search, Text Analysis, and Mining

Search like….

Text Analysis: Entity Extraction & Sentiment Analysis

<PERSON>Jim</PERSON> bought <QUANTITY>300</QUANTITY> shares of <ORGANIZATION>Acme Corp.</ORGANIZATION> in <DATE>2006</DATE>

Text Mining:

Related terms, top relevant terms in a document, related documents, document categorization
SAP HANA Text Search

- Full text search is enabled by a **full text index**
- A full text index is created “implicitly” or explicitly
  - Implicit by usage of column data type “TEXT” or “SHORTTEXT”
  - Explicit via CREATE FULLTEXT INDEX statement
- When a FULLTEXT INDEX is created these steps are executed
  - File filtering
    - Converting binary file types like pdf, doc, ppt, xls, rtf, eml, msg into plain text
  - Linguistic Analysis
    - Tokenization – decompose word sequence, e.g. “the quick brown fox” → “the” “quick” “brown” “fox”
    - Stemming – reduction of tokens to linguistic base form, e.g. houses → house; ran → run
    - Part-of-Speech Identification, e.g. quick: Adj; houses: Nn-Pl
- Full text index is “attached” to the table column
- Additional configuration options available (e.g. language, fuzzy search)
SAP HANA Text Search

- The `CONTAINS()` predicate is used in SQL to run a full text search against a table or a model.
  - **Basic Search**
    ```sql
    SELECT CAMP_NO, DEFECT FROM RECALLS WHERE CONTAINS(DEFECT, 'throttle engine');
    ```
  - **Freestyle Search**
    ```sql
    SELECT CAMP_NO, DEFECT, CONSEQUENCE FROM RECALLS
    WHERE CONTAINS(*, 'throttle engine');
    ```
  - **Fuzzy Search**
    ```sql
    SELECT CAMP_NO, DEFECT FROM RECALLS WHERE CONTAINS(DEFECT, 'thrtle', FUZZY(0.7));
    ```
  - **Relevance Ranking**
    ```sql
    SELECT SCORE() AS S, CAMP_NO, DEFECT FROM RECALLS
    WHERE CONTAINS(DEFECT, 'throttle engine') ORDER BY S DESC;
    ```
  - **Highlighting / Snippets**
    ```sql
    SELECT SNIPPETS(DEFECT), CAMP_NO FROM "_SYS_BIC"."nhtsa/J_RECALLS_S"
    WHERE CONTAINS(DEFECT, 'throttle');
    ```
SAP HANA Text Analysis

- HANA supports in-database *Text Analysis*
- Text analysis is an optional process “on top of” full text indexing
- Text analysis results are stored in a table
  - Multiple text analysis options
    - Linguistic markup, e.g. tokens, stems, parts of speech tags
    - Core extraction, e.g. persons, locations, companies, dates, etc.
    - “Voice of Customer” fact extraction, e.g. sentiments, requests, topics, emoticons, profanity, etc.
    - Enterprise: membership info, management change, product release, M&A, organizational info
    - Public Sector: action & travel events, military units, person alias, relationship, spatial references, etc.
  - Language support
    - Up to 32 languages for linguistic markup
    - 15 languages for predefined core extraction
    - 10 languages for voice of customer extraction
- The text analysis process runs inside the HANA database – feeds from and writes to HANA tables
- Technically, the text analysis is performed by the TA SDK – same NLP technology from Data Services TDP and formerly “Inxight”
## SAP HANA Text Analysis Supported Languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Linguistic Analysis</th>
<th>Entity Extraction</th>
<th>Fact Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sentiment Analysis</td>
</tr>
<tr>
<td>Arabic</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catalan</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese (Simplified)</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Chinese (Traditional)</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Croatian</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danish</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>X</td>
<td>X</td>
<td>X**</td>
</tr>
<tr>
<td>English</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Farsi</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>German</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Greek</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hebrew</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hungarian</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indonesian</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Japanese</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Korean</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Norwegian (Bokmal)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norwegian (Nynorsk)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polish</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portuguese</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Romanian</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Serbian (Cyrillic)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serbian (Latin)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slovak</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slovenian</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Swedish</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thai</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Predefined Core Extraction Examples

NAME_DESIGNATOR  c/o, attn
TITLE  President
PERSON  Barak Obama
PEOPLE  Greeks
LANGUAGE  Greek
ADDRESS1  245 First Street Floor 16
          Cambridge, MA 02142
ADDRESS2
LOCALITY  Cambridge
REGION@MINOR  Napa County
REGION@MAJOR  Connecticut
COUNTRY  Brazil
CONTINENT  South America
GEO_FEATURE  Mount Fuji
GEO_AREA  Scandinavia
ORGANIZATION@COMMERCIAL  AT&T
ORGANIZATION@EDUCATIONAL  University of Washington
ORGANIZATION@OTHER  FBI
PRODUCT  iPhone
TICKER  NYSE:SAP
SOCIAL_MEDIA@TWITTER_ID  @SAP
SOCIAL_MEDIA@TWITTER_TOPIC  #HANA
DATE  2/14/2011
DAY  Monday
MONTH  June
YEAR  2011
TIME  3:47pm
TIME_PERIOD  3 days, from 9 to 5pm
HOLIDAY  Memorial Day
CURRENCY  17 euros
MEASURE  217 meters
PERCENT  4%
PHONE  617-677-2030
NIN@NL_BSN  510889517
URI@EMAIL  john.smith@sap.com
URI@IP  165.14.2.0
URI@URL  http://sap.com
Syntactic Entities:
NOUN_GROUP  big umbrella
PROP_MISC  Cup o’ Soup
Major **fact types** classified:

- **Sentiments**: expression of a customer’s feelings about something
- **Problems**: a statement about something which impedes a customer’s work
- **Requests**: expression of a customer’s desire for an enhancement/change

Any rule may have an associated **TOPIC** sub entity which describes the person, service, product, etc. which the sentiment, problem, or request is about.

*I hate *this book*.  
*I never received *the book*.  
*Please send me *a new book*.

Additional **fact types** classified:

- **Profanity**: defines a set of pejorative vocabulary
  - Ambiguous: words and phrases that are pejorative only in certain contexts
    - *Those hooligans threw toilet paper on my lawn.*
  - Unambiguous: words and phrases that are always pejorative
    - *I cannot express how angry I am with this asshole.*

- **Emoticons**: expression of someone’s feelings about the whole sentence or situation
  - Weak Positive: *Loving my new BlackBerry Q10!* 😊 *No iPhone needed over here.*
  - Strong Positive: *The show was hilarious :D*
  - Weak Negative: *I hate this phone I'm using :-(*
  - Strong Negative: *The Dow Jones fell 200 points :-((((

**Additional social slang coverage**

**GR8, LOL, WTF, I Loooooooove blah, *smiles*, luv, yikes, hooray, etc.**
Text Mining

Text mining works at the document level – making semantic determinations about the overall content of documents relative to other documents. Whereas text analysis does linguistic analysis and extracts information embedded within each document.

Functions based on Vector Space Model

- Identify similar documents
- Identify key terms of a document
- Identify related terms
- Categorize new documents based on a training corpus

Scenarios

- Highlight the key terms when viewing a patent document
- Identify similar incidents for faster problem solving
- Categorize new scientific papers along a hierarchy of topics
Geospatial
SAP HANA Platform for Spatial Processing

Geographical Enablement Framework
Geocoding & Integration Services
Ability to geocode, cleanse, merge, and provision data

Spatial Data Types
Natively store 2D and 3D vector data types

Spatial Functions
Over 80 native SQL based geospatial functions

Spatial Content
Geospatial content & starter content

Mapping Services
Mapping API delivered; open to any mapping service

Application Services
Quickly develop and deploy custom geospatial solutions

Geocoding & Integration Services
Ability to geocode, cleanse, merge, and provision data

Native support of geospatial data types, storage, processing, analysis, and services
Graph Analysis
The Property Graph Model

The Property Graph Model provides directed, attributed (vertices and edges) multi-relational graphs as the central data structure. (BOM, social-, chemical-, biological-, and other networks.)
Graph Capabilities

Built-in Graph Functions

• Neighborhood Search – Breadth-first search
• Single source shortest path
• Strongly connected components
• Pattern matching subgraph

Custom Algorithms

• Cypher language for pattern matching
• GraphScript scripting language
Thank You!

Qi Su
QI.SU@SAP.COM