TECHNOLOGY
LIKE NEVER BEFORE

EBDC Dresden
7. October 2014

The Need for BIG DATA Processing

Petra Streng
Solution Manager SAP SE
Industry Business Unit Life Sciences

Markus Tempel
Global Lead
Big Data Analytics Services Practice

Agenda

The Need

For Big Data Architectures

SAP Use Cases

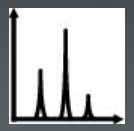
- Customer Projects
- Partner Projects

Big Data Platform

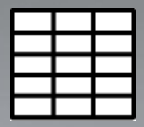
Insight into suitable IT Architecture and innovation platform

Growing Data Volumes in Diverse Healthcare Systems

Human genome/biological data
800 MB per full genome
15 PB+ in databases of leading institutes



Human proteome
160 Mil. data points (2.4 GB) per sample
3.7 TB raw proteome data on ProteomicsDB.org



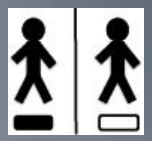
Clinical information management systems
Often more than 50 GB

PubMed biomedical article database
23+ Mil. articles

Cancer patient records
160,000 at
NCT Heidelberg

Medical imaging data
Scan of a single organ in 1s creates 10GB of raw data

Prescription data
1.5 Bil. records from 10,000 doctors and 10 Mil. Patients (100 GB)



Clinical trials
Currently more than 30,000 recruiting on ClinicalTrials.gov

Big Data Challenges – Oncology as an example

1 million human genomes fully sequenced by end of 2013, and 5 million by 2014

As many as 20 driver mutations possible for a single cancer cell

9 days for the analysis of a patient's sequencing data

15 million
new cancer patients each year

Causal mutations are different from patient to patient, and evolve over time

About 400 protein-coding genes showed somatic mutations driving tumor growth

Human genome:
800 MB per full genome,
15 PB+ in databases of leading institutes

Embrace Complexity

BIG DATA

Business Development – What Does It Take

For Big Data Analysis in Life Sciences?

Money

¥$€

Dietmar Hopp,
SAP Co-Founder,
invested 1 bil Euro
in biotech

Innovation

New players in R&D:

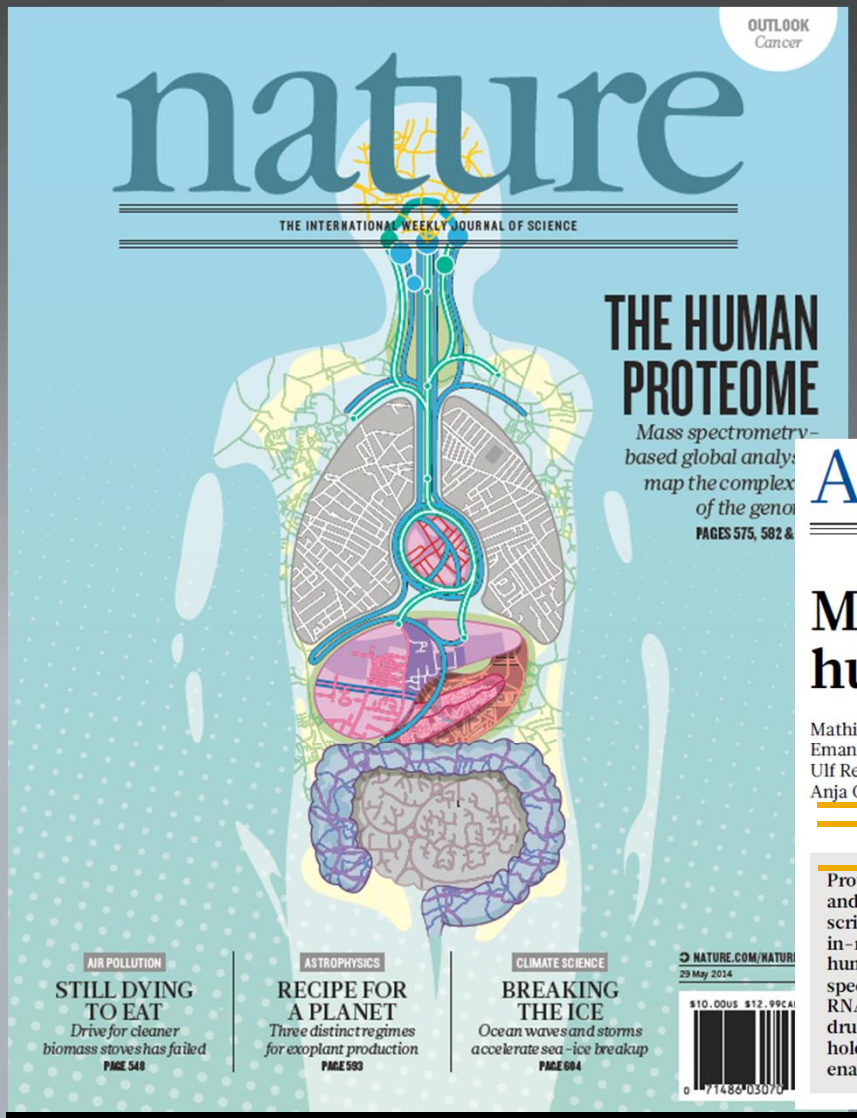
Hasso Plattner,
SAP Co-Founder,
invented a revolutionary
platform geared for Big Data

Environment

What is the required legal and socio-economic framework to foster collaboration and adoption of new innovations?

What is SAP doing in nature?

How Life Sciences Research and IT Technology can benefit each other



92 % coverage of the human proteome

www.proteomicsdb.org

ARTICLE

doi:10.1038/nature13319

Mass-spectrometry-based draft of the human proteome

Mathias Wilhelm[1,2]*, Judith Schlegl[2]*, Hannes Hahne[1]*, Amin Moghaddas Gholami[1]*, Marcus Lieberenz[2], Mikhail M. Savitski[3], Emanuel Ziegler[2], Lars Butzmann[2], Siegfried Gessulat[2], Harald Marx[1], Toby Mathieson[3], Simone Lemeer[1], Karsten Schnatbaum[4], Ulf Reimer[4], Holger Wenschuh[4], Martin Mollenhauer[5], Julia Slotta-Huspenina[5], Joos-Hendrik Boese[2], Marcus Bantscheff[3], Anja Gerstmair[2], Franz Faerber[2] & Bernhard Kuster[1,6]

Proteomes are characterized by large protein-abundance differences, cell-type- and time-dependent expression patterns and post-translational modifications, all of which carry biological information that is not accessible by genomics or transcriptomics. Here we present a mass-spectrometry-based draft of the human proteome and a public, high-performance, in-memory database for real-time analysis of terabytes of big data, called ProteomicsDB. The information assembled from human tissues, cell lines and body fluids enabled estimation of the size of the protein-coding genome, and identified organ-specific proteins and a large number of translated lincRNAs (long intergenic non-coding RNAs). Analysis of messenger RNA and protein-expression profiles of human tissues revealed conserved control of protein abundance, and integration of drug-sensitivity data enabled the identification of proteins predicting resistance or sensitivity. The proteome profiles also hold considerable promise for analysing the composition and stoichiometry of protein complexes. ProteomicsDB thus enables navigation of proteomes, provides biological insight and fosters the development of proteomic technology.

Proteomics Database

Dedicated to expedite the identification of the human proteome and its use across the scientific community

Proteome-based Cancer Research

Dedicated to identify early cancer signals ("fingerprints") to derive at diagnostic tests via protein mass spectroscopy signals

Award Winning

Personalized Medicine

Forbes | New Posts +3 posts this hour | Most Popular 5 LinkedIn Strategies | Lists The Most Powerful People | Video One Worl

BUSINESS | 11/14/2013 @ 12:56AM | 758 views

The White House Honors SAP, Stanford and NCT

Jacqueline Vanacek, SAP

+ Comment Now + Follow Comments

In 2013 we celebrated two magnificent achievements in biology: the 60th Anniversary of Watson and Crick's DNA double helix and the 10th Anniversary of the completion of the Human Genome Project.

Photo: Shutterstock

The White House Honors SAP, Stanford and NCT

SAP received special recognition from the US White House together with the **Stanford School of Medicine** and the **National Center for Tumor Diseases (NCT Heidelberg)** to help accelerate the Human Genome Project's therapeutic promise of personalized medicine

http://www.forbes.com/sites/sap/2013/11/14/the-white-house-honors-sap-stanford-and-nct/

National Center for Tumor Diseases (NCT)

Gain Insight into Cancer Research

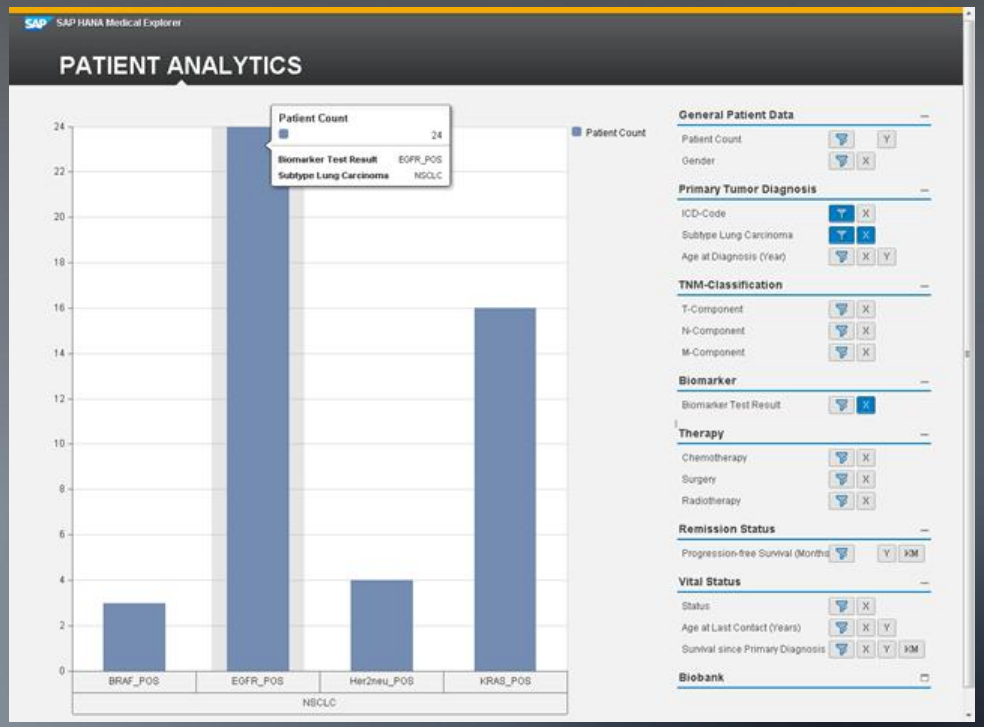
NCT NATIONAL CENTER FOR TUMOR DISEASES HEIDELBERG

supported by
German Cancer Research Center (DKFZ)
Heidelberg University Medical Center
Hospital for Thoracic Diseases
German Cancer Aid

- Interface for **real-time analyses of clinical data**: Various sources of **structured** (tumor documentation, medical records, clinical trials) and **unstructured** (doctor letters, treatment guidelines, trial reports, publications) nature.

- Medical records from **150,000 patients** and 3,600,000 interactions, and a selection of doctor letters from 120 doctors

"In future we want everyone coming for a diagnosis to go through an SAP HANA scan just as they have an MRI/Ultrasound scan today."
Professor von Kalle, NCT Heidelberg



Screenshot: A search for lung carcinoma patients

Genomics: SAP and Stanford

Genome analysis in real-time of 1000 genome data

STANFORD SCHOOL OF MEDICINE

Dr. Carlos D. Bustamante

Analyze genome wide patterns of variation within and between species to address fundamental questions in biology, anthropology, and medicine: **implications for global health and disease**

"We have been thrilled to work with SAP and HPI on a collaboration to accelerate DNA sequence analysis. In our pilot projects, we are seeing **dramatic speedups** in **computing on human genome variation data** from many samples. We are dreaming of what will soon be possible as we integrate phenotype, genomics, proteomics, and exposome data to empower complex trait mapping using millions of health records."

- Professor Carlos D. Bustamante at the Stanford University School of Medicine

Plant Genomics: SAP & University of British Columbia

Match geno- and phenotypes

Dr. Loren Riesberg

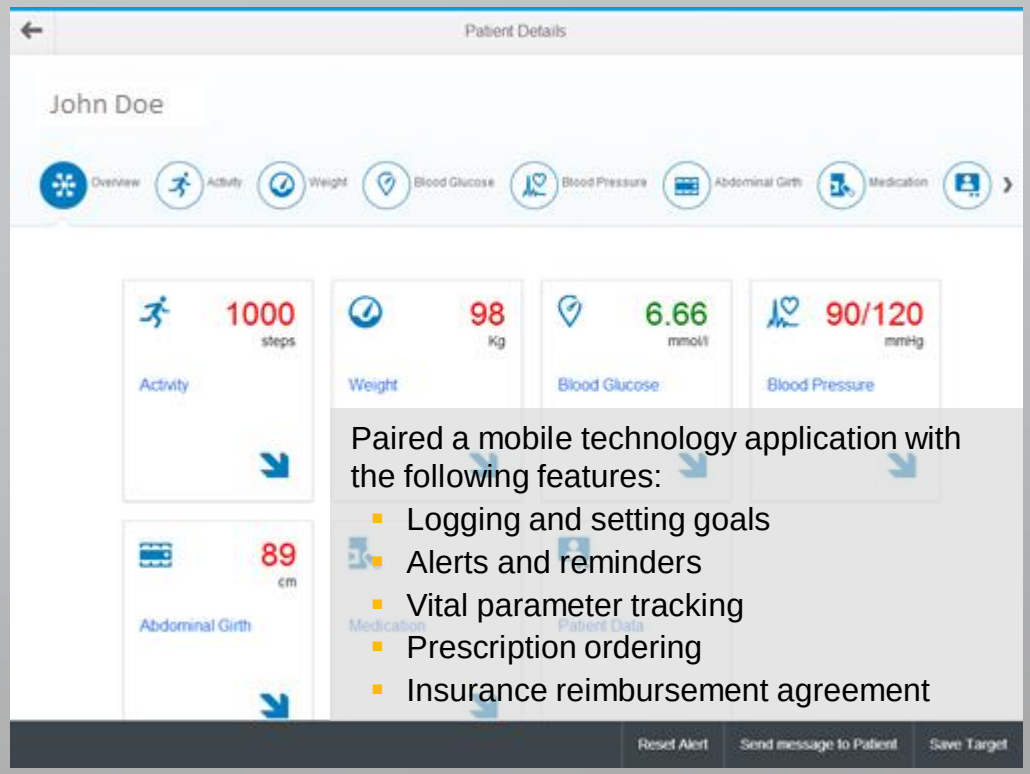
Integrates high-throughput genomic methods, bioinformatics, ecological experiments & evolutionary theory to **study the origin & evolution of species, domesticated plants & weeds**.

Project goals:

- Complete pre-variant analysis pipeline for sunflower genome data & benchmark against existing tools
- Develop post-variant analysis algorithms & visualizations
- Predictive analysis – using HANA-R integrations & SAP's existing predictive tools

Real Life Evidence through Remote Patient Engagement

Capture medical device data to drive lifestyle change through insight



Paired a mobile technology application with the following features:

- Logging and setting goals
- Alerts and reminders
- Vital parameter tracking
- Prescription ordering
- Insurance reimbursement agreement

Patient stratification → Remote monitoring → Analyze outcome

Transforming Healthcare with SAP Technology
National Rural Health Mission in India

Planned to enroll

270 million

school children across India

Currently

> 60 000

children enrolled

A pilot program for 270 million school children **to start a lifetime of data collection, using mobile tablets for data entry and cloud storage for all health data.**

Goal: Determine the need for medical support, prevent epidemics, and provide **analysis capabilities to aid in the understanding of health trends across the population**.

Partner Engagement for Genome Analysis

216x

faster by reducing genome analysis from several days to only 20 minutes

408,000x

faster than traditional disk-based systems in a technical PoC

Initially it took MKI two days to find differences in genome data between cancer patients and healthy people

.

"Our solution is to incorporate SAP HANA along with Hadoop and R to create a single real-time big data platform. Data mining will be handled by R and assisted by HANA. Data pre-processing prior to data analysis and high-speed storage will be managed by Hadoop. With this we have found a way to shorten the genome analysis time from several days down to only 20 minutes."

Yukihisa Kato, CTO and Director of MITSUI KNOWLEDGE INDUSTRY

Partner Engagement for Treatment Decision Support

300x faster

An exome analysis now only requires 3 minutes!

75%

of the time drug therapies are ineffective, because every patient's cancer is different, unique to his/her genetic makeup.

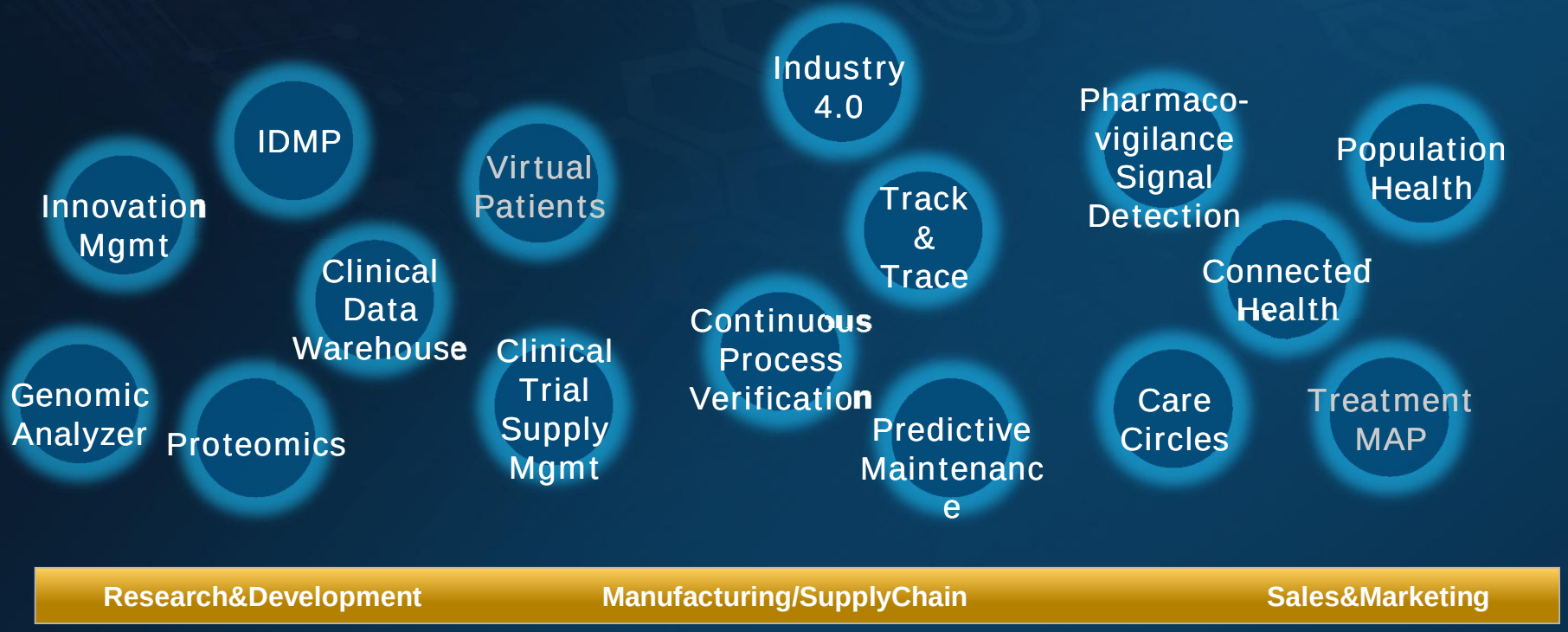
With TreatmentMAP™ MolecularHealth has created a state of the art result report providing

- Actionable treatment options,
- Details on each of the treatment options as well as
- Drug-drug interactions to ensure effective treatment

Molecular Health handles the complete process:

- Sample collection
- Sequencing genetic tests,
- Deep-level analysis run against the current medical research
- Summary recommendations and result report

Uncover value.
Create breakthroughs.
Experience simplicity.



Examples of SAP and Partner Solutions and Customer Projects based on HANA

SAP HANA Database Technology

as basis for our bioinformatics work

Bulk load
Fast insertion of large genomic datasets or other relevant datasets

Text Retrieval and Extraction
Search doctor's notes, diagnoses, etc.(unstructured data)

Lightweight Compression
Fit big data in main memory while allowing fast retrieval

Multi-core/ parallelization
Speedup of relevant queries across many nodes

SQL
SQL interface on columns & rows
Easily connect with other tools (e.g. Rstudio)

On-the-fly extensibility
Adapting to new format requirements without going offline (e.g. changing VCF files)

BIGDATA Portfolio

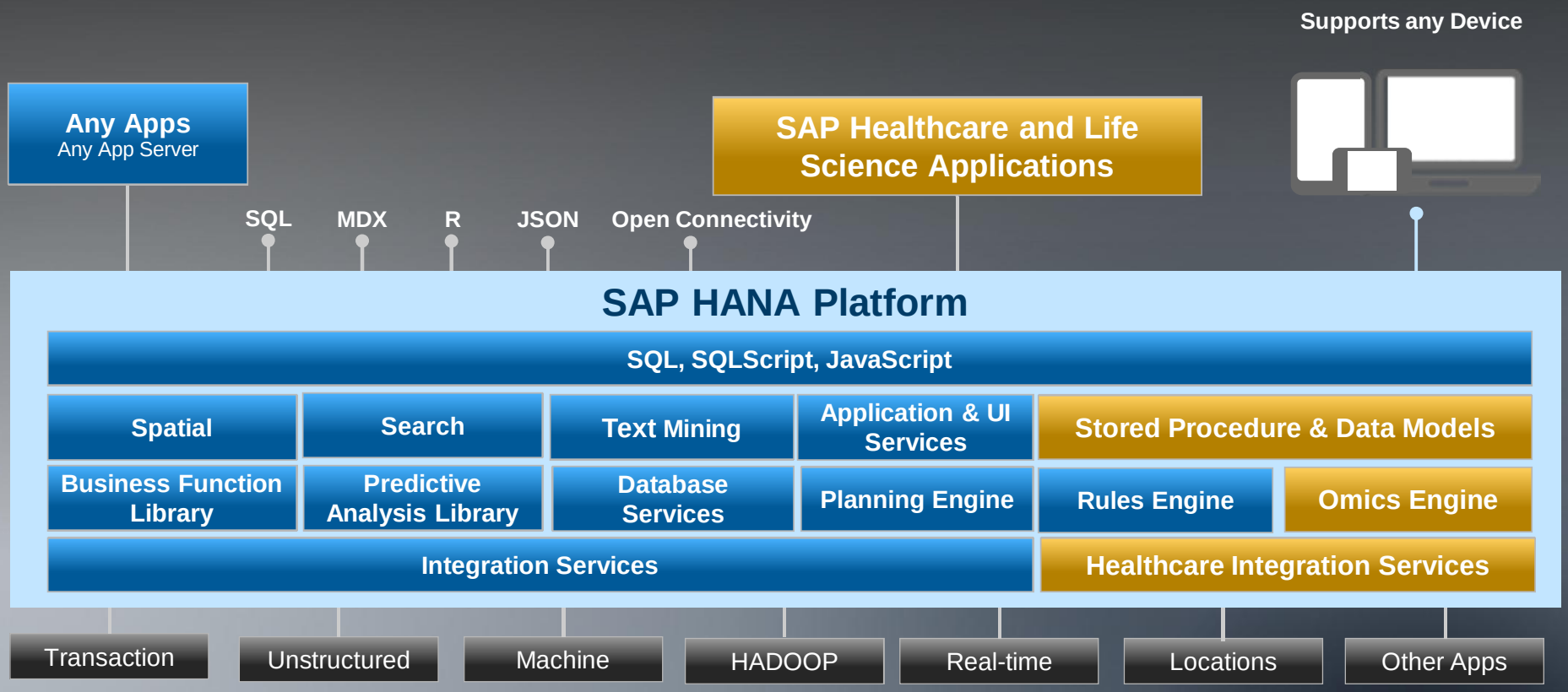
Key Components of a World Class Big Data Portfolio

BIG DATA **PLATFORM**	BIG DATA **ANALYTICS**	BIG DATA **APPLICATIONS**	BIG DATA **SERVICES**
Accelerate how you acquire, analyze and act on Big Data insights	Unleash the power of Big Data with collective insight across your business	Adopt new business models and revenue streams with applications that deliver Big Data insights	Achieve tangible results from your Big Data initiatives with services that apply advanced data science to your business
→ In-Memory Platform → Analytics Database → Hadoop → Event Processing → Data Services	→ Predictive Analytics → Visualization → Text Analysis → Business Intelligence	→ Business Applications → Custom Applications	→ Full Offering: Discover, Plan, Realize, Run → Data Science Experts: Predicting your business better than you can

SAP is making a significant investment to lead the market and is going **BIG** with BIGDATA:
www.sapbigdata.com

SAP HANA Platform

More than just a database



SAP HANA Platform for Healthcare – industry specific extensions of SAP HANA.
Providing **breakthrough capabilities** for **healthcare** and **life sciences** applications from SAP and its partners, while **reducing time-to-value and TCO**.

Acquire, Analyze and Act on BIGDATA

Acquire

Effectively, rapidly and efficiently acquire and consolidate massive amounts of diverse and arbitrary data

Analyze

Achieve real results to get the insights you need with a variety of means alone or in combination

Act

Real-time response time regardless of data volume or data location manage and integrate massive volumes of data

SAP HANA: Unprecedented **hyper-performance** achieved through deep synergies between software and hardware innovations

BIGDATA augments traditional analytics …

Adding BIGDATA technologies enables analytics solutions…

- **to include more types of data**
- **to consume higher volumes of information.**
- **to process more information in less time**
- **to maintain a higher quality of analysis**
- **to offer an improved user experience.**
- **to allow users to explore all data beyond DWH models**

SAP Predictive Analysis

Advanced Visualization

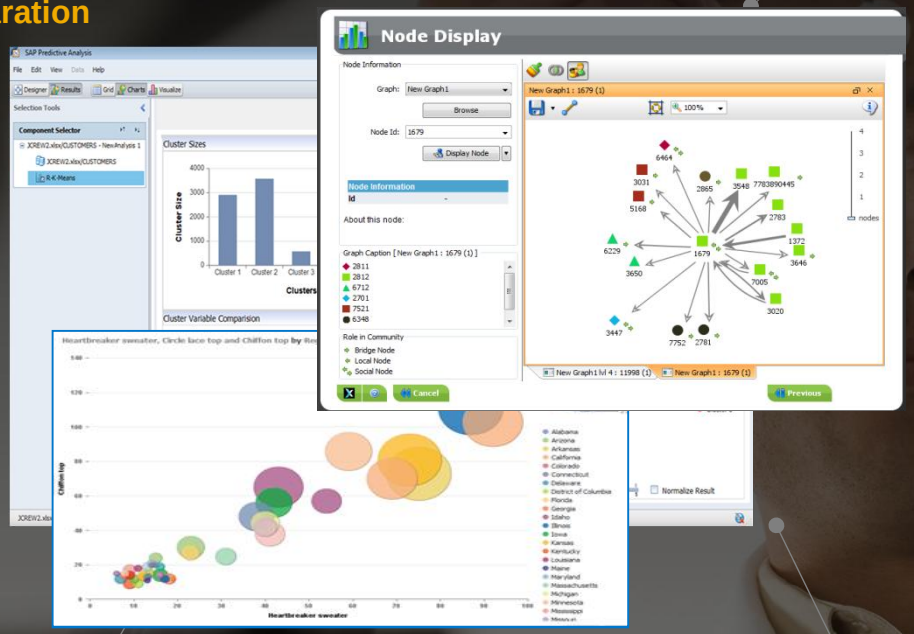
- Direct access to Advanced Visualizations
- Superset solution includes SAP Visual Intelligence library
- Stunning visualizations

Ease of Use

- Drag & drop data selection, preparation, processing
- Easy sharing /collaboration of findings
- Built for business analysts

Rich Pre-Built Modelling Functionality

- Automatic Data Preparation
- Classification
- Regression
- Anomaly Detection
- Attribute Importance
- Association Rules
- Clustering
- Feature extraction



Multi Language Support

- EN, DE, ES, PO, FR, JP, IT

Automated Approach to solving predictive problems

- Rapid development of predictive models
- Focused on predictive functions not algorithms
- Ability to deliver large amounts of models quickly

R Data Mining Language Support

- Native installer included
- ~12 R algorithms included
- 3,500+ R Model library and growing
- Custom R, JAVA, etc.

Integration

- Native integration with SAP HANA
- Leverage existing BOBJ universes
- Publish actionable results to mobile & BI clients

BENEFITTING
FROM SAP HANA

COMBINE
massive amount of data from all available sources

SEGMENTATION
of very large data sets

PREDICTIVE
power to identify risks for patient's health or company's financial performance early on

REAL-TIME
analysis of very large data set of genome or proteome data, orders, invoices, financial transitions, medical records, doctor letters or medical publications

Contacts

Petra Streng
Solution Manager
SAP for Life Sciences

SAP AG
Dietmar-Hopp-Allee 16
69190 Walldorf, Germany

+ 49 170 8555 664
petra.streng@sap.com

Legal Disclaimer

The information in this document is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. This document is not subject to your license agreement or any other service or subscription agreement with SAP. SAP has no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation and SAP's strategy and possible future developments, products and or platforms directions and functionality are all subject to change and may be changed by SAP at any time for any reason without notice. The information on this document is not a commitment, promise or legal obligation to deliver any material, code or functionality. This document is provided without a warranty of any kind, either express or implied, including but not limited to, the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This document is for informational purposes and may not be incorporated into a contract. SAP assumes no responsibility for errors or omissions in this document, except if such damages were caused by SAP intentionally or grossly negligent.

All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations.

Readers are cautioned not to place undue reliance on these forward-looking statements, which speak only as of their dates, and they should not be relied upon in making purchasing decisions.

© 2014 SAP AG or an SAP affiliate company. All rights reserved.