



# SAP Analytics & Big Data

---

June 2016

Tania Dinnendahl

# Legal disclaimer

---

The information in this presentation is confidential and proprietary to SAP and may not be disclosed without the permission of SAP. This presentation 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 in 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's willful misconduct or gross negligence.

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.

# SAP Analytics – Overview Webinars

SAP Analytics – Overview



Jayne Landry

Mar 16

SAP BI – Overview & Roadmap



Olivier Duvelleroy /  
Saurabh Abhyankar

Mar 23

SAP EPM – Overview & Roadmap



Kirk Anderson

Mar 24

SAP EIM – Overview & Roadmap



Paul Medaille

Mar 30

SAP Predictive Analytics – Overview &  
Roadmap



Sven Bauszus

Mar 31









SAP GRC – Overview & Roadmap



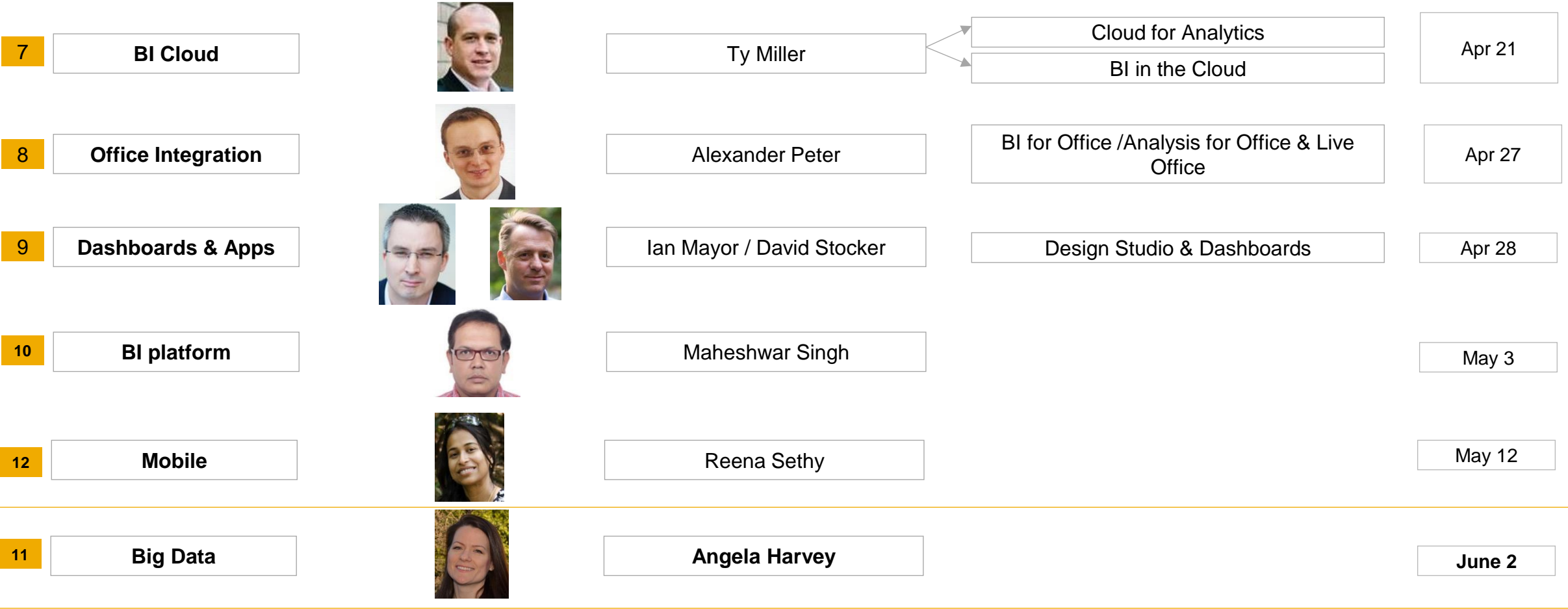
Kevin McCollom

Apr 13

# SAP Analytics Portfolio - BI Solutions 1/2

1	Reporting		Intro by Merlijn Ekkel	<ul style="list-style-type: none"> <li>Crystal Reports; Nina Bao / Donald Guo</li> <li>Web Intelligence Gregory Botticchio</li> </ul>	Mar 29
2	Adopting BI 4.1 /4.2		Merlijn Ekkel		Apr 5
3	What's new in BI 4.2		Merlijn Ekkel		Apr 7
4	Analytics Services Portfolio		Markus Schunter		Apr 12
5	Self Service BI		Ina Felsheim		Apr 14
6	Discovery & Analysis	  	Intro by Saurabh Abhyankar	<ul style="list-style-type: none"> <li>Lumira; Francois Imberton</li> <li>Pred. Analytics, Pierpaolo Vezzosi</li> </ul>	Apr 20

# SAP Analytics Portfolio - BI Solutions 2/2





# SAP Analytics & Big Data

---

June 2016

Angela Harvey, Director, Solution Management  
Big Data & BI

# Agenda

---

The Big Data Landscape

SAP Lumira & Big Data Wrangling

Demo

SAP HANA Vora & SAP Lumira

A Practical Big data Use Case

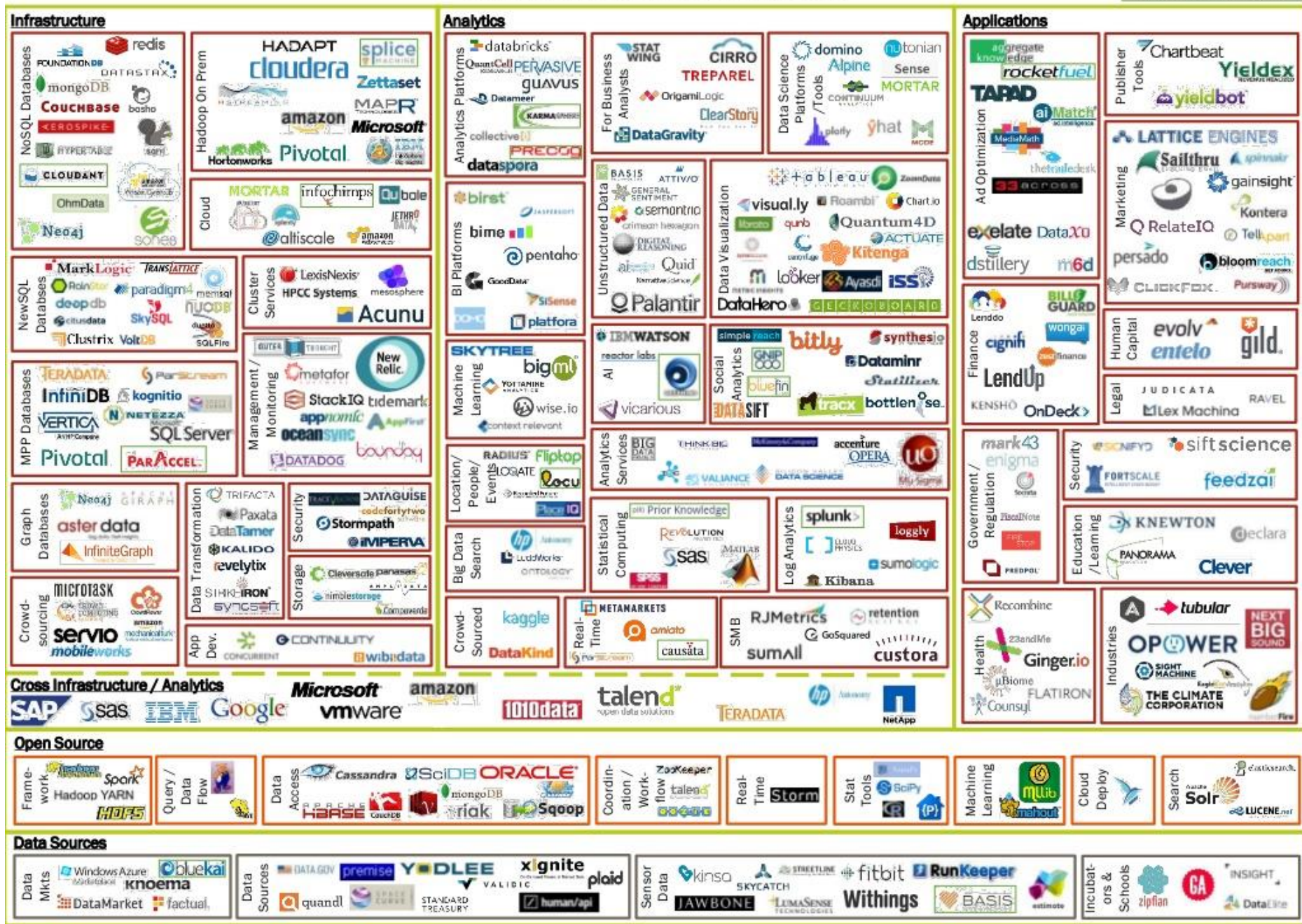
# Big Data Landscape





# BIG DATA LANDSCAPE, VERSION 3.0

Exited: Acquisition or IPO





**LOST**

**CONFUSED**

**UNSURE**

**UNCLEAR**

**PERPLEXED**

**DISORIENTED**

**BEWILDERED**

# Big Data is going places – want to go with it?

- Data volume in the Enterprise will **grow 50x year over year** until 2020—Hadoop Summit, 2014
- According to **Forrester**, most companies use **only 12%** of their data
- **IDC** predicted **Big Data** spending will grow to **\$125 billion in 2015**
- Do a search for **Data Scientist** roles...



# Why is Big Data Growing?

More data is accessible from a variety of sources (weblogs, sensors, social media)



New data stores cheaper in terms of hardware, software, and ETL

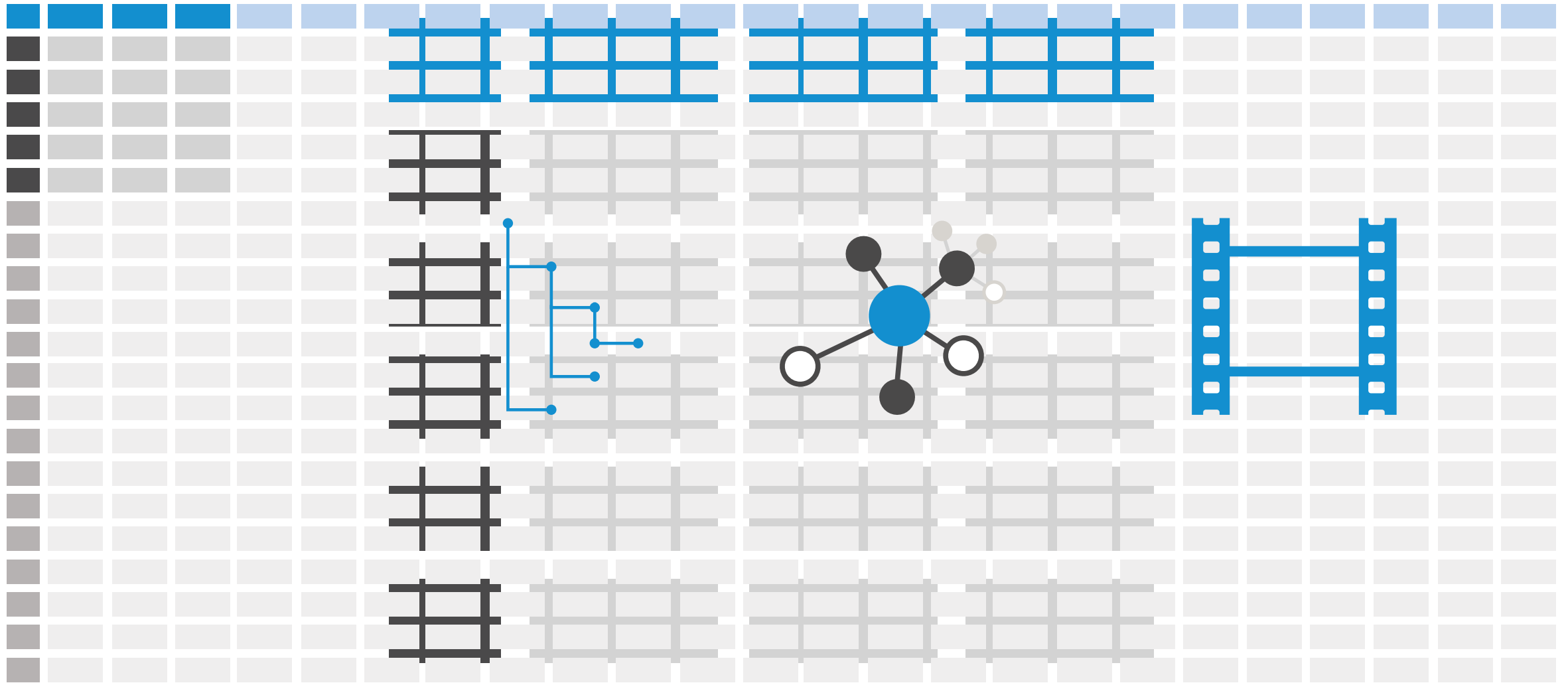


Organizations are recognizing the value of previously discarded data and new data types



# But what is Big Data?





# NoSQL and Hadoop

Not only SQL OR ~~SQL~~



DATASTAX



Fast read/write  
“OLTP” type use-cases  
Examples, sensor data,  
clickstream...

Scale out, not up  
Run on commodity clusters  
Fault tolerant  
Agile (semi & unstructured data)

Large-scale processing  
“datawarehouse”  
Recommendation engines,  
predictive, etc



# Analytics and Big Data

## SAP Lumira

- Connect to Hadoop thru Hive, EMR, Impala, Spark, Redshift. Additional sources with our API
- Customized data acquisition for MongoDB (POC) & Neo4j
- Leverage our Big Data visualizations or build your own

## SAP BI Suite

- Connect universes directly to Hadoop thru EMR, Hive, Impala then report using any client tool (Web Intelligence, Crystal Reports, Dashboards)
- Stream data into Design Studio dashboards for realtime analytics

## SAP Predictive Analysis

- Beyond knowing what happened understand why, or model what could
- Tease more information out of Big Data sources, creating more attributes for better modeling
- Fast—pushing the predictive calculations to Hadoop removes the need to bring data to the desktop



### Data Sources

- SAP Analytics tools view Hadoop as just another data source, but needs special considerations
- Complement your existing data infrastructure with Hadoop or NoSQL and derive value with familiar SAP tools
- Use SAP Analytics directly against Big Data sources, or with HANA for real-time analytical capabilities

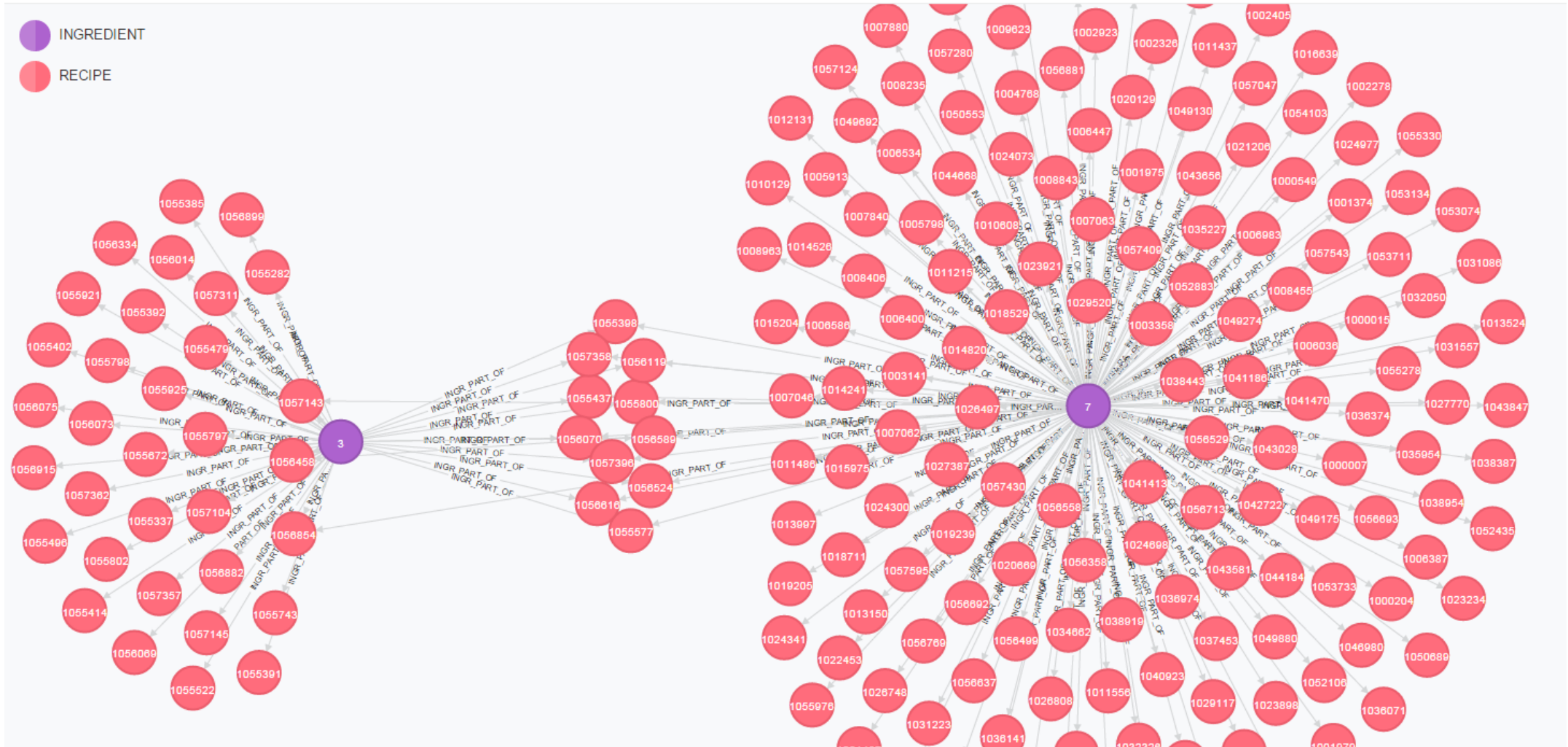


# Thinking about Big Data Differently

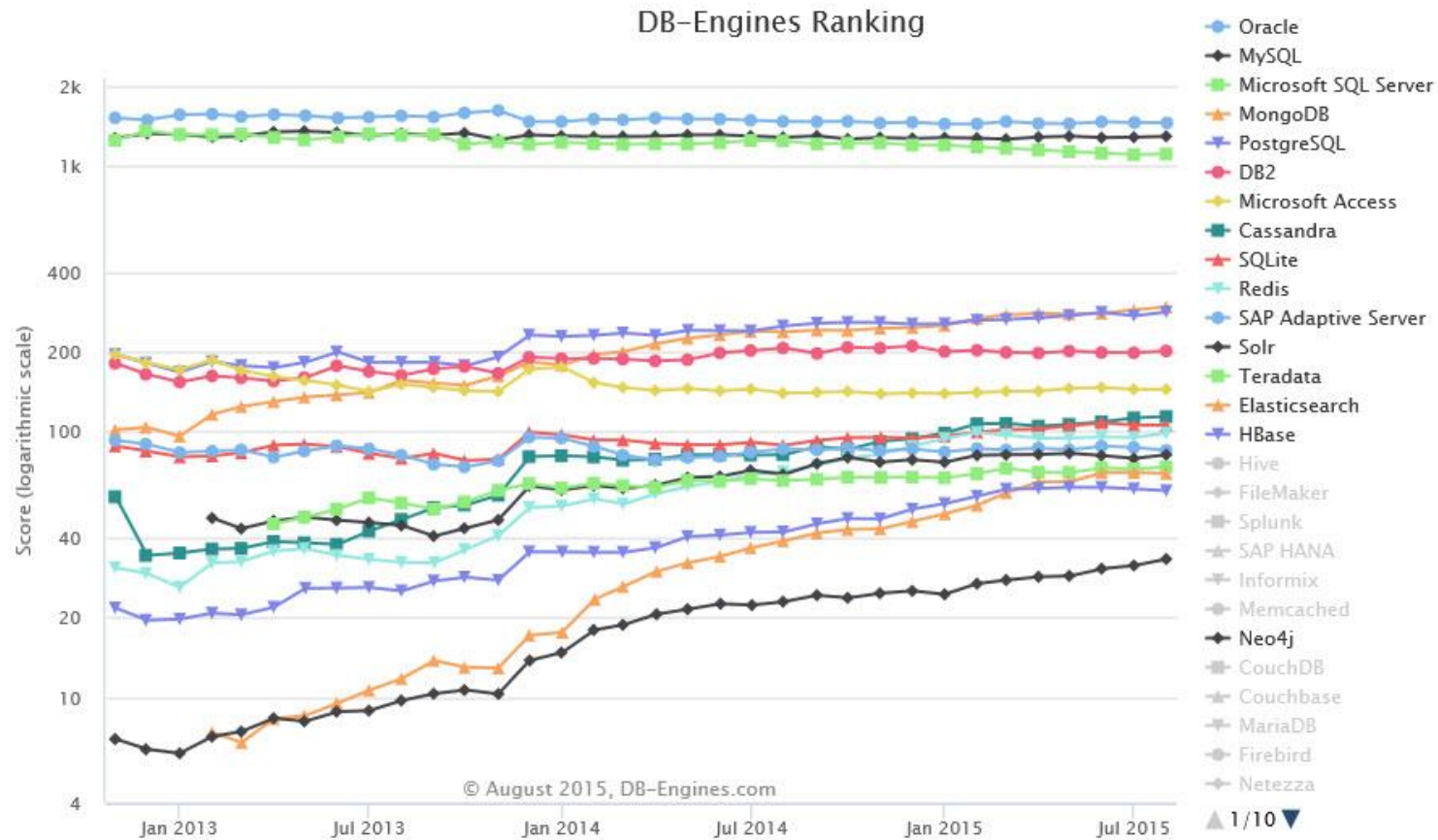




# Graphs



# How popular are they?



[http://db-engines.com/en/ranking\\_trend](http://db-engines.com/en/ranking_trend)

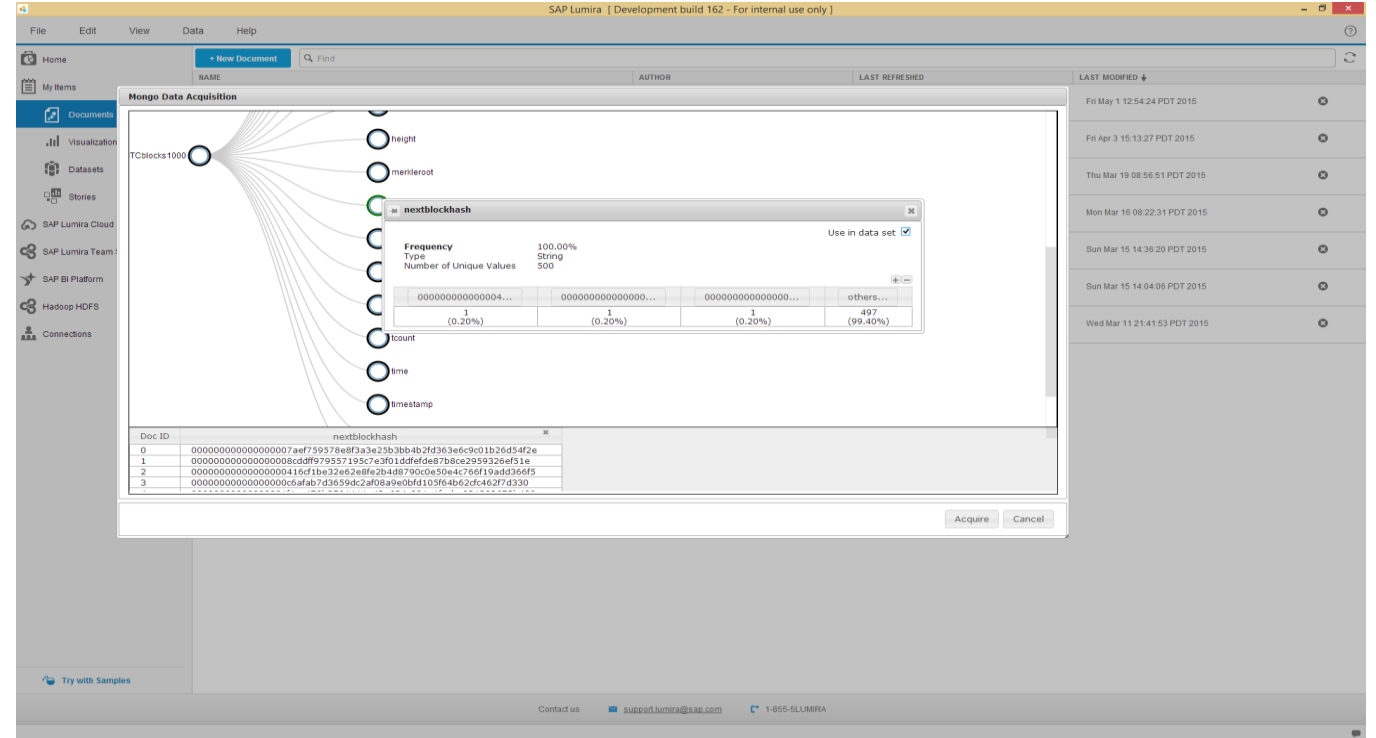
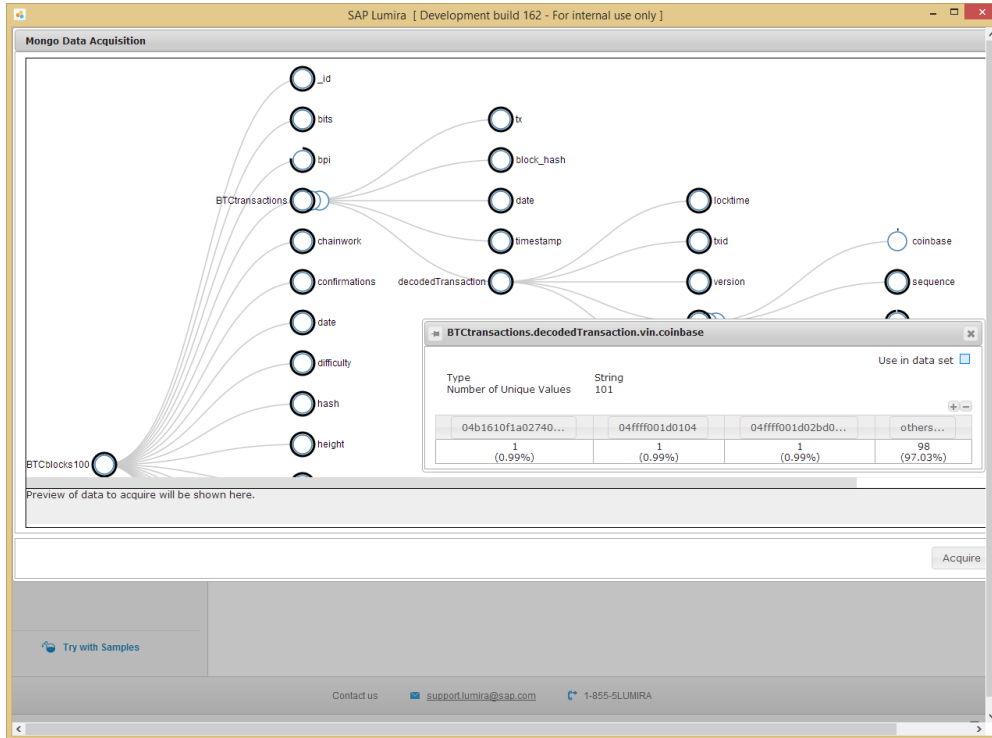
# Adjust to these new types of sources

**We need to be able to bring their data into our analytical experience**

- Flattening the data is the end goal, not the beginning
- Support schema-less sources
- Dedicated acquisition experience



# MongoDB and Lumira



# Neo4j and Customized Acquisition Experience

The screenshot shows the SAP Lumira 'Acquire from Neo4j' interface. On the left, there are two panels: 'NODES' and 'EDGES'. The 'NODES' panel lists: RECIPE (blue), AREA (light blue), INGREDIENT (orange), CUISINE (light orange), INGREDIENT\_CATEGORY (green), COMPOUND (light green), and RECIPE\_SOURCE (red). The 'EDGES' panel lists: BELONGS\_TO, LOCATED\_IN, OF\_CUISINE, IS\_SOURCE\_OF, COMP\_PART\_OF, and INGR\_PART\_OF. The main graph area displays a network with a central 'RECIPE' node (blue) connected to 'AREA' (light blue), 'CUISINE' (light orange), 'INGREDIENT' (orange), and 'INGREDIENT\_CATEGORY' (green). 'RECIPE\_SOURCE' (red) is also connected to 'RECIPE'. 'INGREDIENT' is further connected to 'COMPOUND' (light green). At the bottom, there are 'Previous', 'Acquire', and 'Cancel' buttons, and an 'ACQUISITION PREVIEW' section with 'Show Selection' set to 'On' and 'OR' selected. A message at the bottom reads 'Please select a node or edge to preview data.'

The screenshot shows the same SAP Lumira 'Acquire from Neo4j' interface, but with the 'ACQUISITION PREVIEW' table visible. The table has columns for 'ID', 'name', and 'name'. The data is as follows:

ID	name	name
2888	2125	Turkey
2889	2126	American
2890	2127	Cajun_Creole
2891	2128	Canada
2902	2129	Southern_SoulFood
2903	2130	Southwestern
2904	2131	Scandinavia
2905	2132	Scandinavian

At the bottom, there are 'Previous', 'Acquire', and 'Cancel' buttons, and an 'ACQUISITION PREVIEW' section with 'Show Selection' set to 'On' and 'OR' selected. The table has checkboxes for 'ID', 'name', and 'name' columns.

# Why Hadoop matters to Analytics

Companies transform entire business model by using the new architecture to keep **all its business data for longer** and to use for a wider variety of analytics.





# Hadoop – Practical Use Cases



Archive



Data of Unknown Value



Data Lakes for Exploratory Analytics

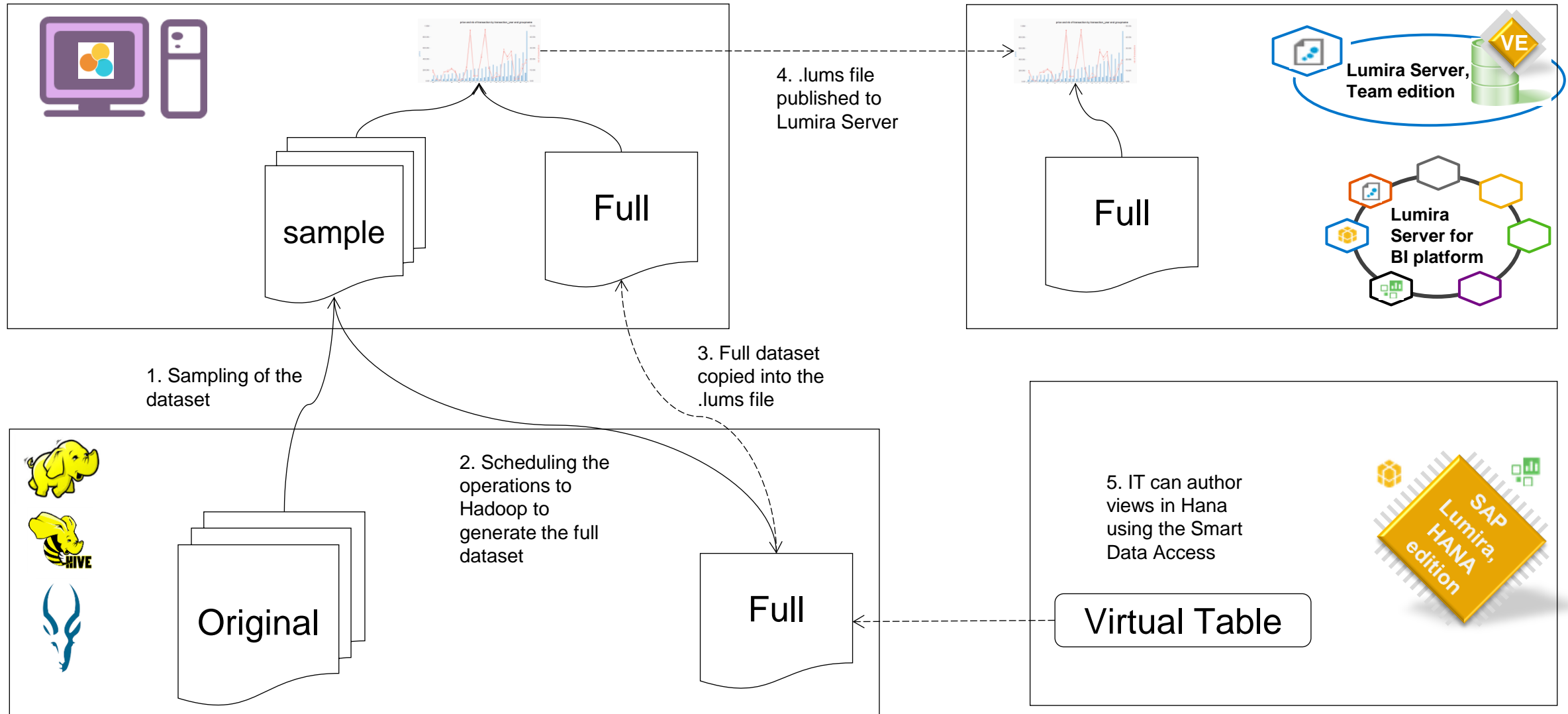
# More data more problems...

- Visualizations on big data...
- Connecting to more data sources...
- Performance expectations of end users & and of course transferring big data is not a good idea...

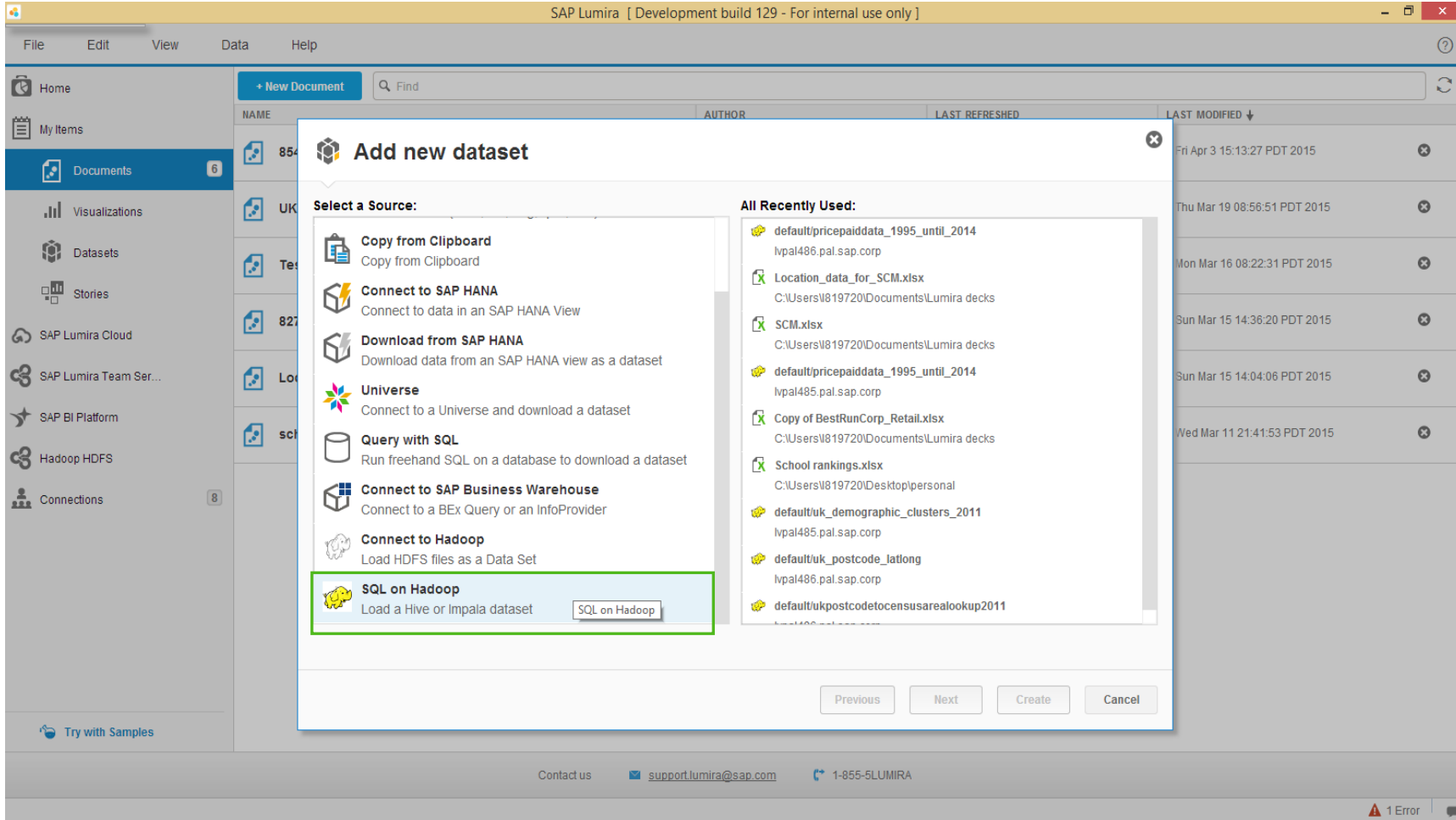
# How do we bring End-User Self-Service to Hadoop environments



# Data Wrangling on any data size in Hadoop!



# Data Wrangling on Hadoop



# Data Wrangling on Hadoop

**Add new dataset**

Dataset Name:  Sample Dataset?  Yes  No Cell Sampling Rate % (Up to 50M cells):

**pricepaiddata\_1995\_until\_2014**

<input checked="" type="checkbox"/>	Column Name	Preview of Column Values
<input checked="" type="checkbox"/>	ABC id	{72E1329E-4D04-45E3-8640-0A964713E17F},{51E78E19-2EE8-41C1-...
<input checked="" type="checkbox"/>	123 price	250000,46500,145000,90000,66500
<input checked="" type="checkbox"/>	ABC transaction_date	2006-10-16 00:00,1995-11-19 00:00,2006-08-03 00:00,1996-08-16 00:...
<input checked="" type="checkbox"/>	ABC postcode	"E17 8NQ","SR7 0LE","CW12 4QT","CF62 3EP","PE28 3JX"
<input checked="" type="checkbox"/>	ABC property_type	"T","S","D","F"
<input checked="" type="checkbox"/>	ABC old_new	"N","Y"
<input checked="" type="checkbox"/>	ABC duration	"F","L"
<input checked="" type="checkbox"/>	123 paon	
<input checked="" type="checkbox"/>	ABC saon	"",26 - 32","FLAT 3","12",1"
<input checked="" type="checkbox"/>	ABC street	"BOUNDARY ROAD","SHARPLEY DRIVE","LONGDOWN ROAD","RHOO...
<input checked="" type="checkbox"/>	ABC locality	"LONDON","SEAHAM","CONGLETON","RHOOSE","SOMERSHAM"

Total Number of Columns: 15  Show Only Selected Columns **15 Columns Selected = Fetching up to 50M number of cells**



**Add new dataset**

Dataset Name:  Sample Dataset?  Yes  No Cell Sampling Rate % (Up to 50M cells):

**pricepaiddata\_1995\_until\_2014**

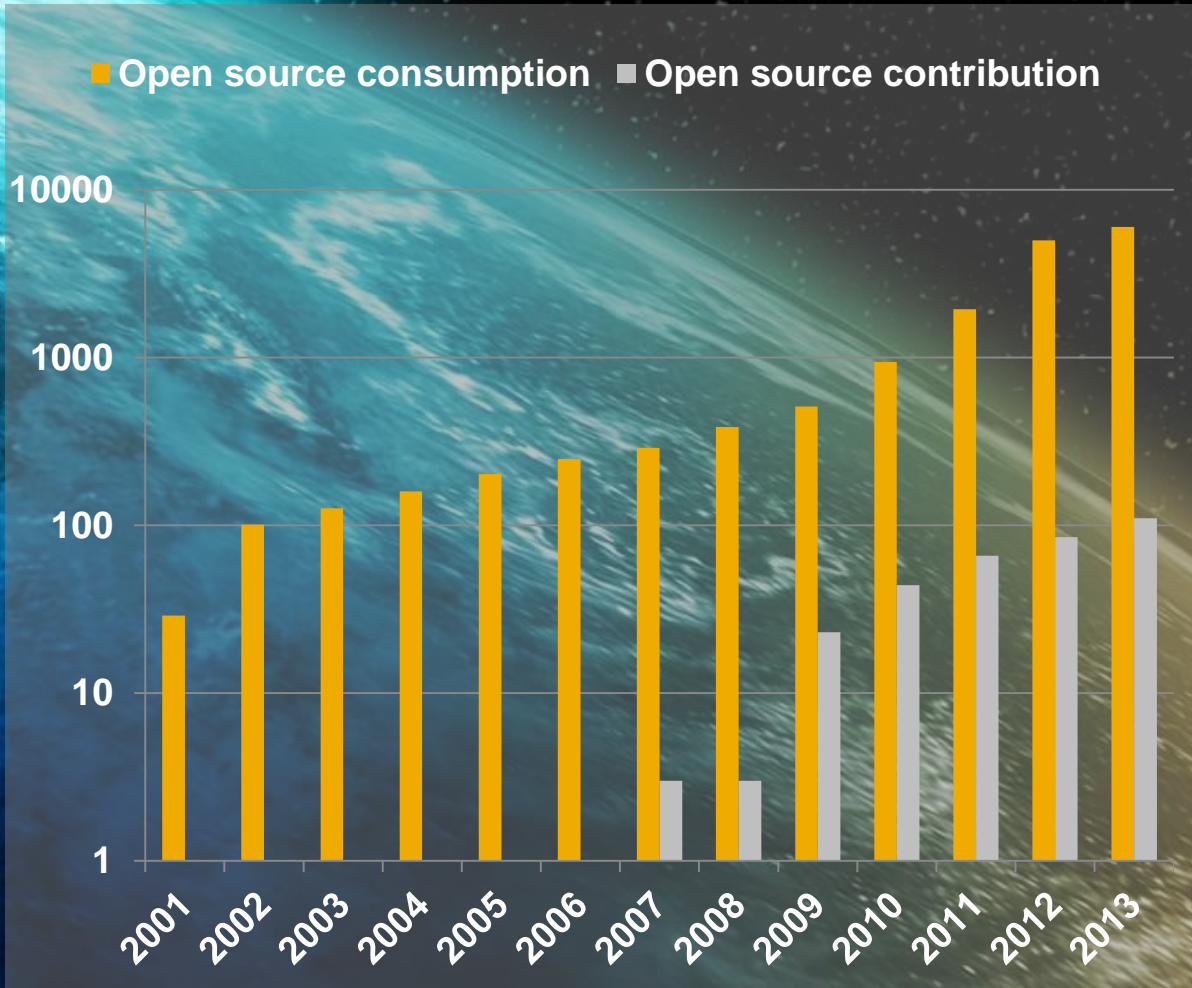
<input checked="" type="checkbox"/>	Column Name	Preview of Column Values
<input checked="" type="checkbox"/>	ABC id	{72E1329E-4D04-45E3-8640-0A964713E17F},{51E78E19-2EE8-41C1-...
<input checked="" type="checkbox"/>	123 price	250000,46500,145000,90000,66500
<input checked="" type="checkbox"/>	ABC transaction_date	2006-10-16 00:00,1995-11-19 00:00,2006-08-03 00:00,1996-08-16 00:...
<input checked="" type="checkbox"/>	ABC postcode	"E17 8NQ","SR7 0LE","CW12 4QT","CF62 3EP","PE28 3JX"
<input checked="" type="checkbox"/>	ABC property_type	"T","S","D","F"
<input checked="" type="checkbox"/>	ABC old_new	"N","Y"
<input checked="" type="checkbox"/>	ABC duration	"F","L"
<input checked="" type="checkbox"/>	123 paon	
<input checked="" type="checkbox"/>	ABC saon	"",26 - 32","FLAT 3","12",1"
<input checked="" type="checkbox"/>	ABC street	"BOUNDARY ROAD","SHARPLEY DRIVE","LONGDOWN ROAD","RHOO...
<input checked="" type="checkbox"/>	ABC locality	"LONDON","SEAHAM","CONGLETON","RHOOSE","SOMERSHAM"

Total Number of Columns: 15  Show Only Selected Columns **1% of 15 Selected Columns = 2M Cells will be fetched**

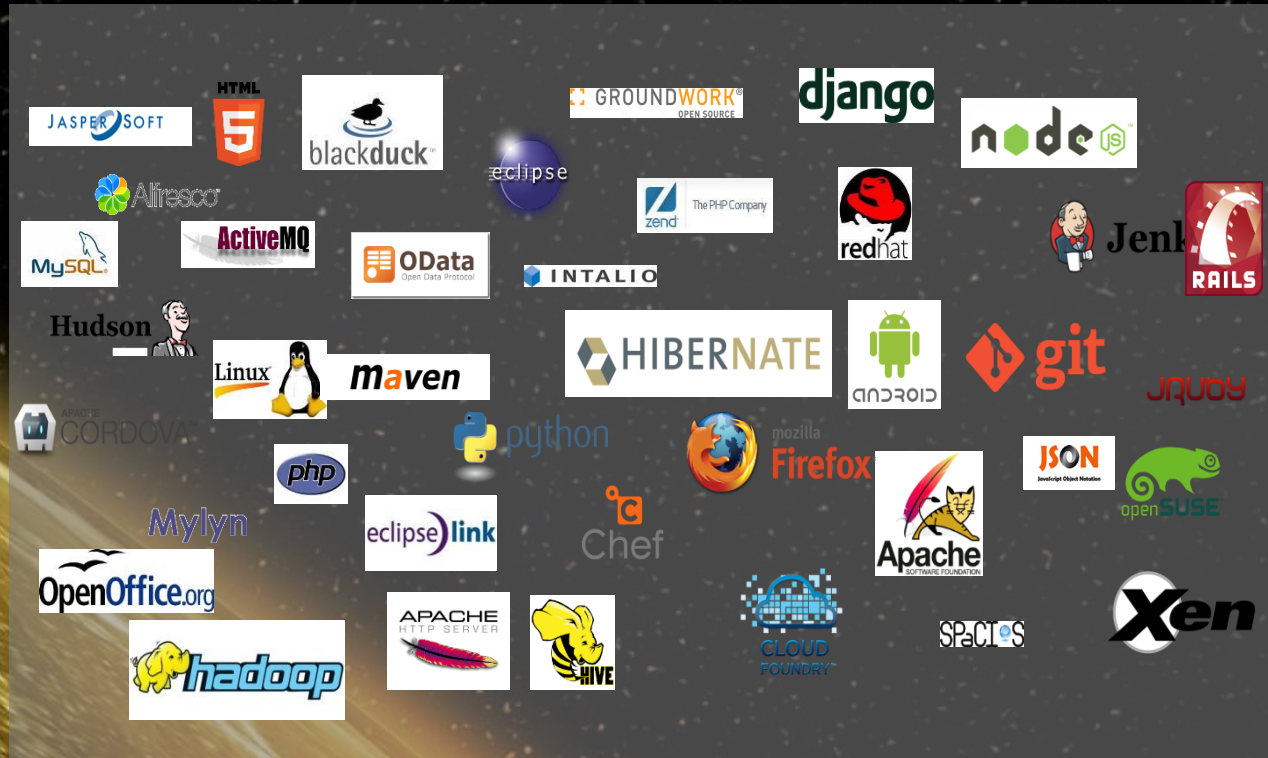
# Data Wrangling on Hadoop

The screenshot displays the SAP Lumira interface in the 'Visualize' tab. The main window shows a horizontal bar chart with 10 categories labeled abc1 through abc10 on the Y-axis and a 'Measure' on the X-axis ranging from 0.00 to 50.00. A tooltip titled 'Sampling Information:' is overlaid on the chart, containing the text '2941815/294181500' and 'Sample cells / original cell count', along with a 'Generate Full Dataset' button. The interface includes a left sidebar with 'MEASURES' (paon, price) and 'DIMENSIONS' (city, county, district, duration, id, locality, old\_new, paon, postcode, price, property\_type, record\_status). The top menu bar includes 'File', 'Edit', 'View', 'Data', 'Help', 'Prepare', 'Visualize', 'Compose', and 'Share'. The status bar at the bottom indicates 'Sample of default.pricepaiddata\_1995\_until\_2014', 'Showing: 196121/196121 Rows - 15/15 Columns', and 'Never Refreshed'.

# HADOOP is key Part of SAP's Open Source Development usage



## SAP Contributes to over 100 Open Source Projects





# The SAP focus: End-to-end value chain

## Mobile applications and BI

CONSUME

- High Performance Applications
- Business Planning & Forecasting
- Reporting & Dashboards
- Adhoc & OLAP Analytics
- Data Exploration & Visualization
- Predictive Analytics
- Lumira / BI

COMPUTE

### SAP HANA Platform

Application Development Environment

STREAM PROCESSING

ANALYTICS, TEXT, GRAPH, PREDICTIVE ENGINES

SPATIAL PROCESSING

STORAGE

In-Memory Calculation engine Column Storage Series Data Storage Dynamic Tiering

Data model & data Fast computing High performance analytics Store time-series data Aged data in Disk

INGEST

Smart Data Streaming Smart Data Access Smart Data Integration Smart Data Quality

Stream Processing Virtual Tables User Defined Functions Transformations & Cleansing

SOURCE

- ERP
- OLTP
- Geo
- Store & forward
- Text
- Social
- Logs
- Machine
- Sensor

### Hadoop / NoSQL

MapReduce

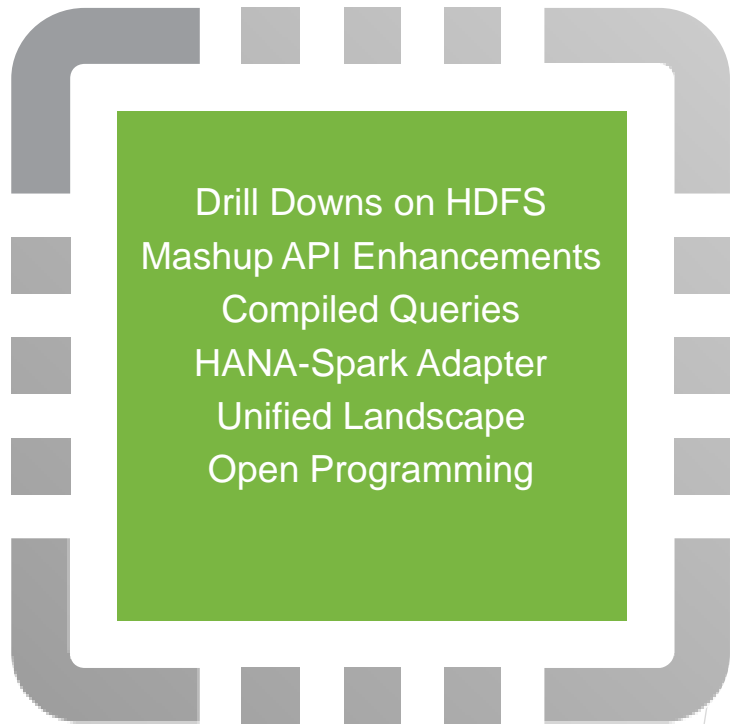
YARN

HDFS

Not Only SQL

# SAP HANA Vora

## What's Inside and What Does It Do?



SAP HANA Vora is an in-memory query engine which leverages and extends the Apache Spark execution framework to provide enriched interactive analytics on Hadoop.



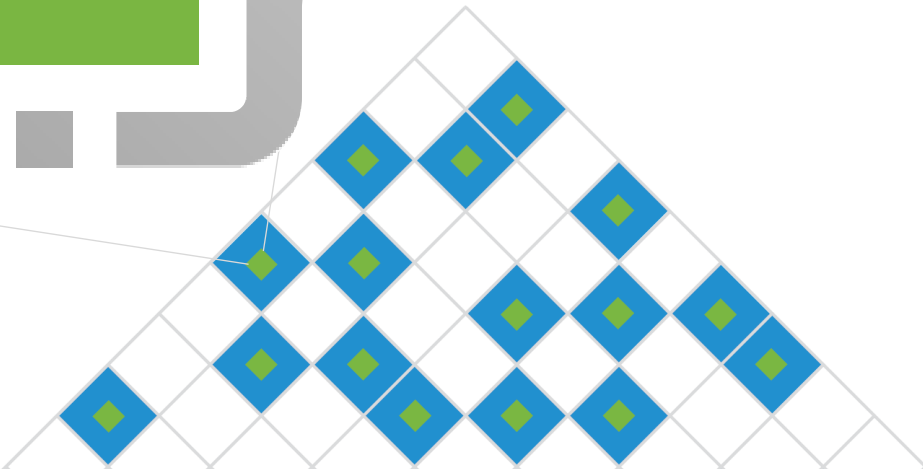
Make  
Precision  
Decisions



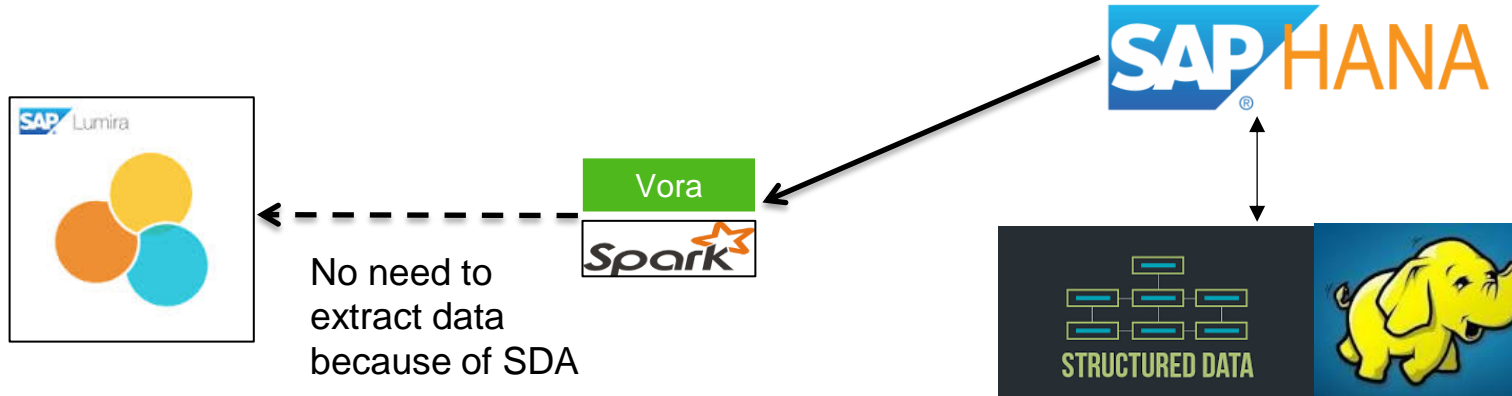
Democratize  
Data  
Access



Simplify  
Big Data  
Ownership

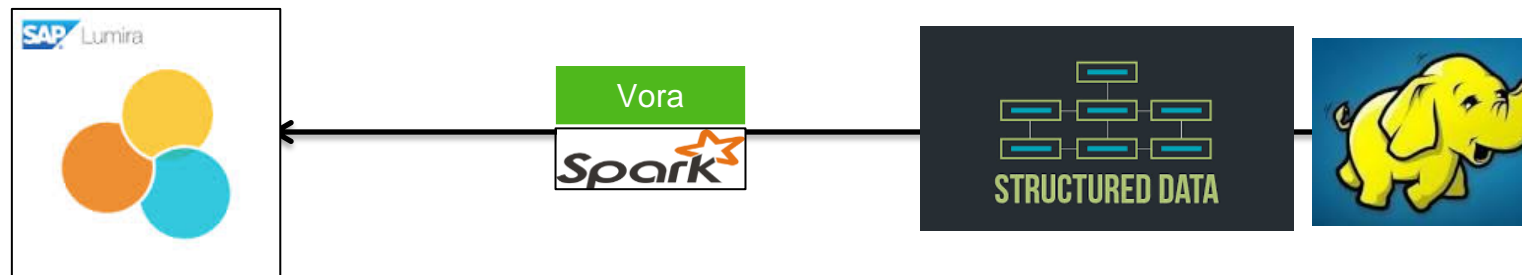


# SAP Lumira and SAP HANA Vora—you've got options



- **With HANA**

- Online connectivity
- Take advantage of Vora's ability to integrate with corporate data like HANA
- Improved query time (vs Spark) thru compiled queries



- **Without HANA**

- Get started quickly
- Go directly against Hadoop
- Improved query time thru compiled queries

# **A Practical Big Data Use Case**

## **SAP IT's Need for IT Operations Analytics**

# Snapshot of SAP's landscape

## SAP Cloud Infrastructure Services

- Over **42 data centers** in **28 locations** and **12 countries**
- Over **87,000 servers** on **20,000+** physical hosts
- **5,000+ TB RAM**
- **100,000+** CPU core computing power
- **Cloud storage** capacity of over **87 petabytes**
- Virtualization rate of **70%**
- Daily **backup volume** of over **1 petabyte**

Unified SAP Cloud Delivery

### Application Services



HANA Enterprise Cloud  
HANA Cloud Plattform  
S/4 HANA

SAP Ariba

SAP Fieldglass

SAP Hybris



CONCUR

SAP SuccessFactors

### Platform Services

SAP | non-SAP

### Infrastructure Services

Compute | Network | Storage | Backup

# Challenges from a provider point of view

Performance



Scalability



Zero Downtime



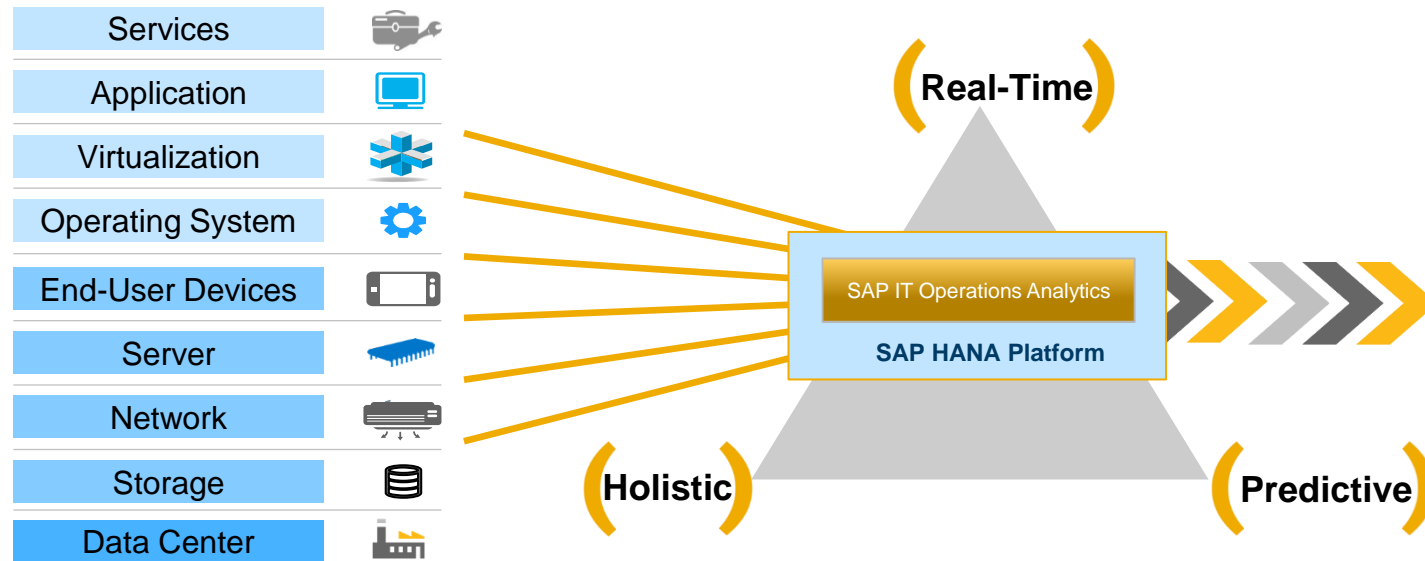
Security / Data Protection



# SAP IT Operations Analytics

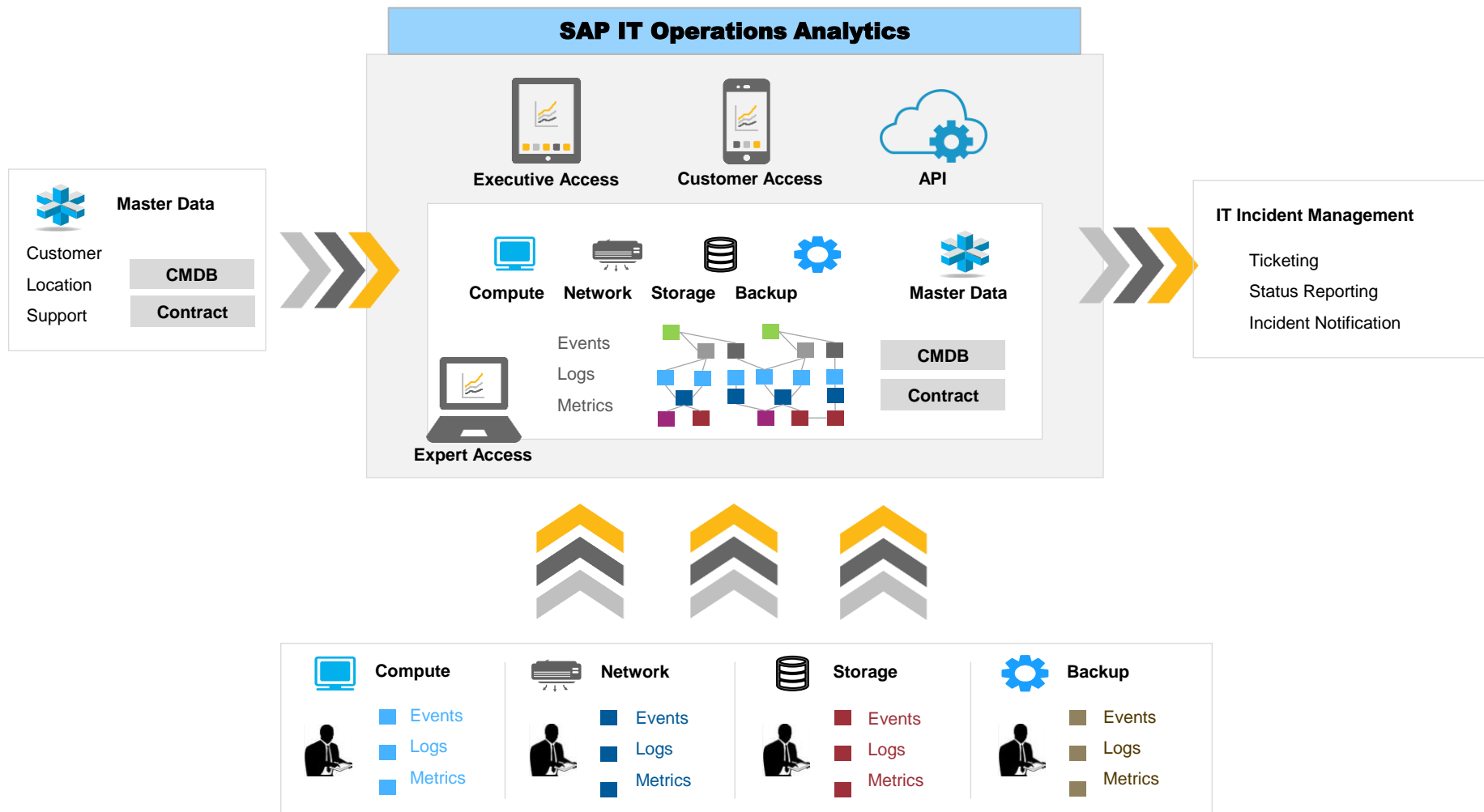
Now GA!

“ SAP IT Operations Analytics leverages SAP HANA to give you a **holistic** view of your data center in **real-time** with additional insights to **predict** critical events. ”



# Root Cause Analytics

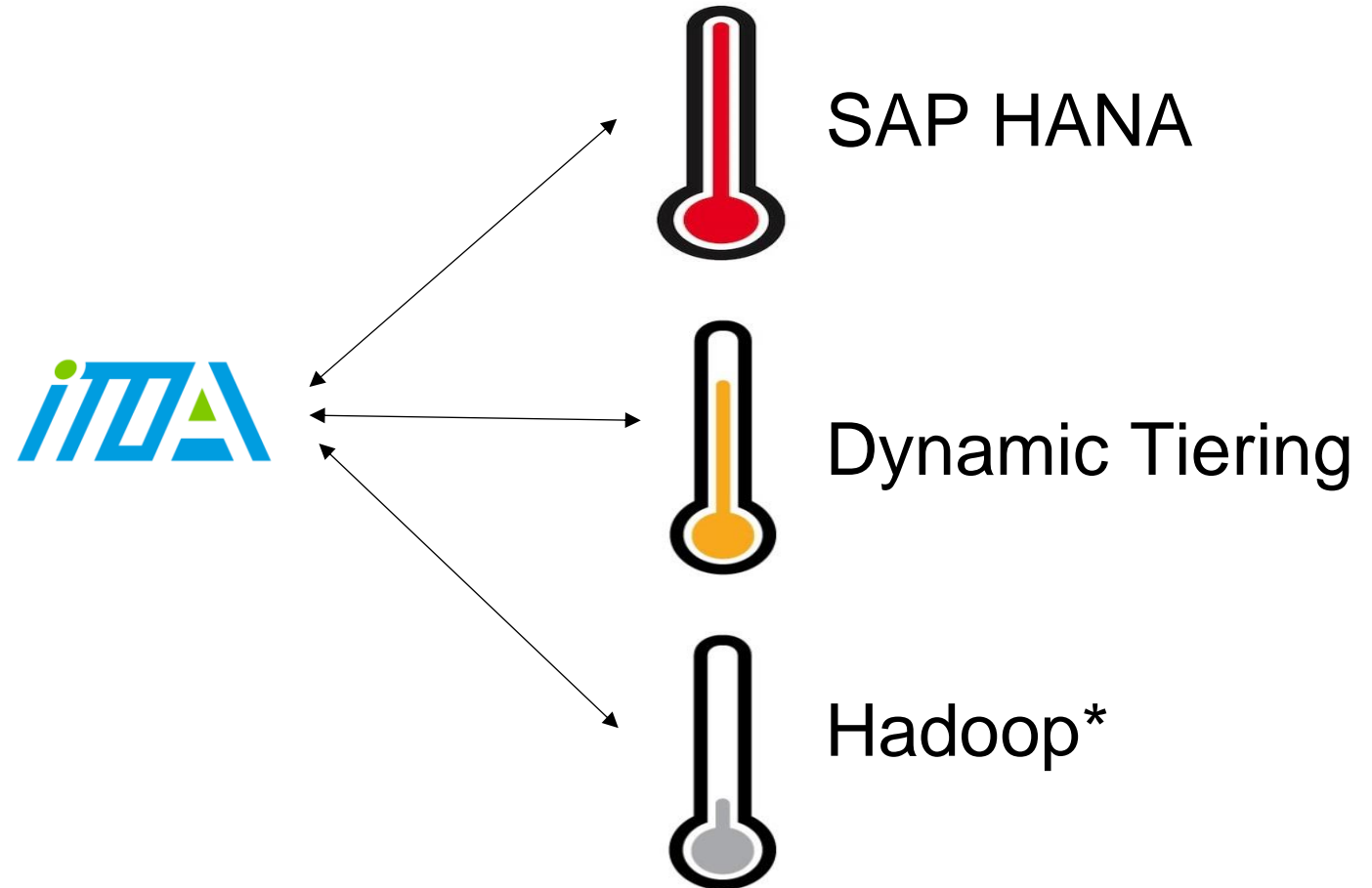
*“Holistic RCA in one single System”*





# Data Tiering

- Different data levels for aged out data, non-events etc.
- Cost effective
- Drive adoption through IT
- Mass storage for compliance or predictive
- Still accessible thru a single interface



**\*PLANNED INNOVATIONS—SUBJECT TO CHANGE**

# SAP IT Operations Analytics

## Realizing the impact on your data center

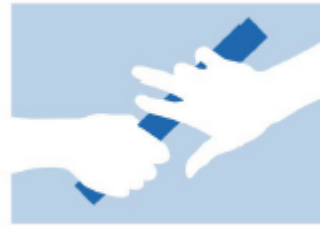
### Responsive IT Operations



#### Maximizing IT Responsiveness

by having a holistic view within and all over different hardware and software layers, facility sensors to avoid business impacting situations.

**Lower resolution times**



#### Optimizing IT operations

to drive cost-efficiency by acquiring, managing and acting on massive volumes of data to optimize resource and capacity management.

**Less operating costs**



#### Mitigate IT risks

to ensure business continuity by providing transparency on device connectivity in real-time to detect threats or predict outages.

**More uptime**



# Thank you!

Angela Harvey

[angela.harvey@sap.com](mailto:angela.harvey@sap.com)

@AngelaHarveySAP

# © 2016 SAP SE 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 SE or an SAP affiliate company.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. Please see <http://global12.sap.com/corporate-en/legal/copyright/index.epx> for additional trademark information and notices.

Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors.

National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP SE or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP SE or SAP affiliate company 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.

In particular, SAP SE or its affiliated companies have 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 SE's or its affiliated companies' strategy and possible future developments, products, and/or platform directions and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. 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.