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	Crystal Reports; Nina Bao / Donald Guo Web Intelligence Gregory Botticchio	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	Lumira; Francois Imberton Pred. Analytics, Pierpaolo Vezzosi	Apr 20

SAP Analytics Portfolio - BI Solutions 2/2

7	**BI Cloud**		Ty Miller	Cloud for Analytics BI in the Cloud	Apr 21
8	**Office Integration**		Alexander Peter	BI for Office /Analysis for Office & Live Office	Apr 27
9	**Dashboards & Apps**	 	Ian Mayor / David Stocker	Design Studio & Dashboards	Apr 28
10	**BI platform**		Maheshwar Singh		May 3
12	**Mobile**		Reena Sethy		May 12
11	**Big Data**		**Angela Harvey**		**June 2**

SAP Analytics & Big Data

June 2016

Angela Harvey, Director, Solution Management
Big Data & BI

Agenda

The Big Data Landscape

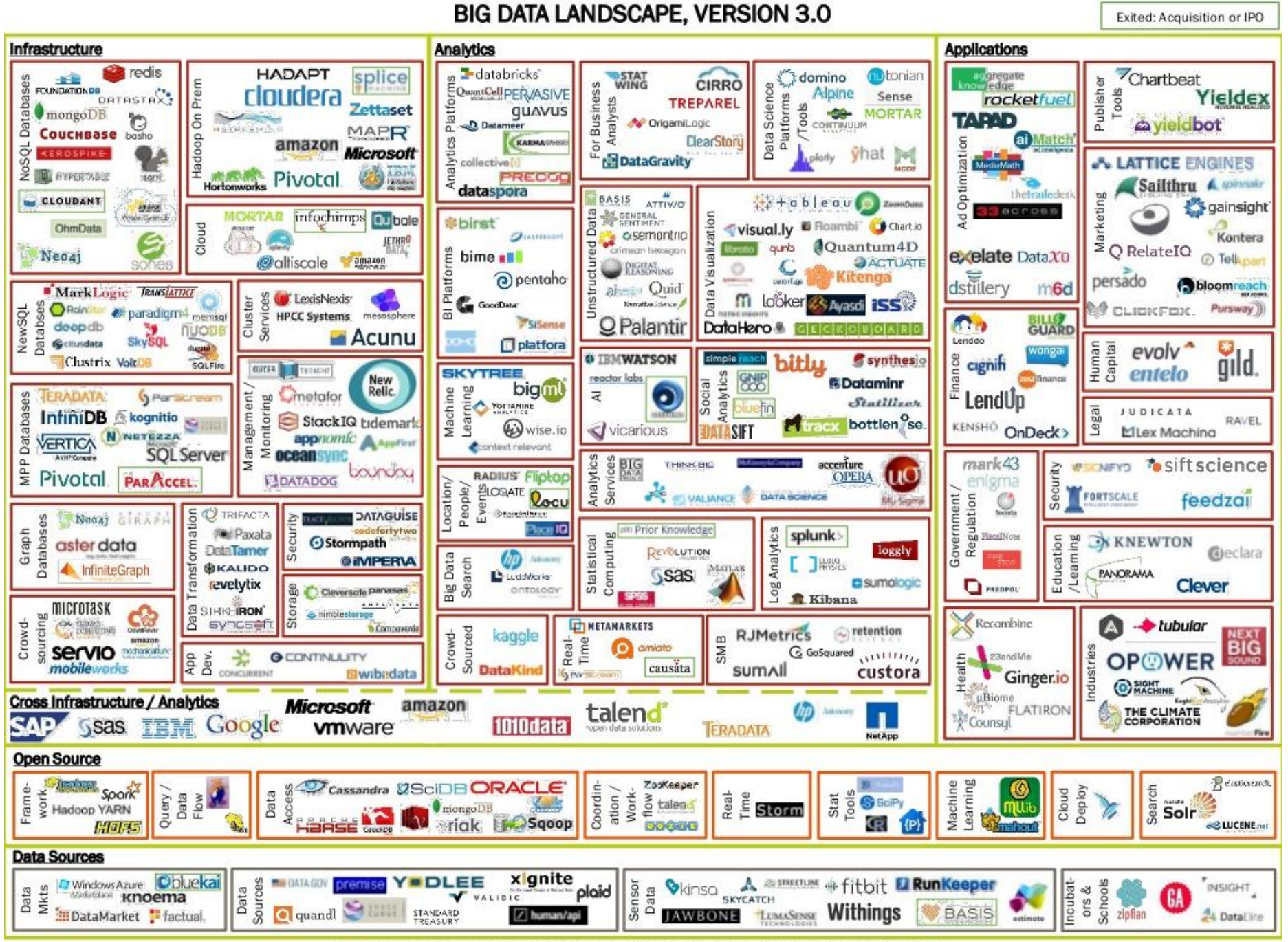
SAP Lumira & Big Data Wrangling

Demo

SAP HANA Vora & SAP Lumira

A Practical Big data Use Case

Big Data Landscape



© Matt Turck (@mattturck), Sutian Dong (@sutiandong) & FirstMark Capital (@firstmarkcap)


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…

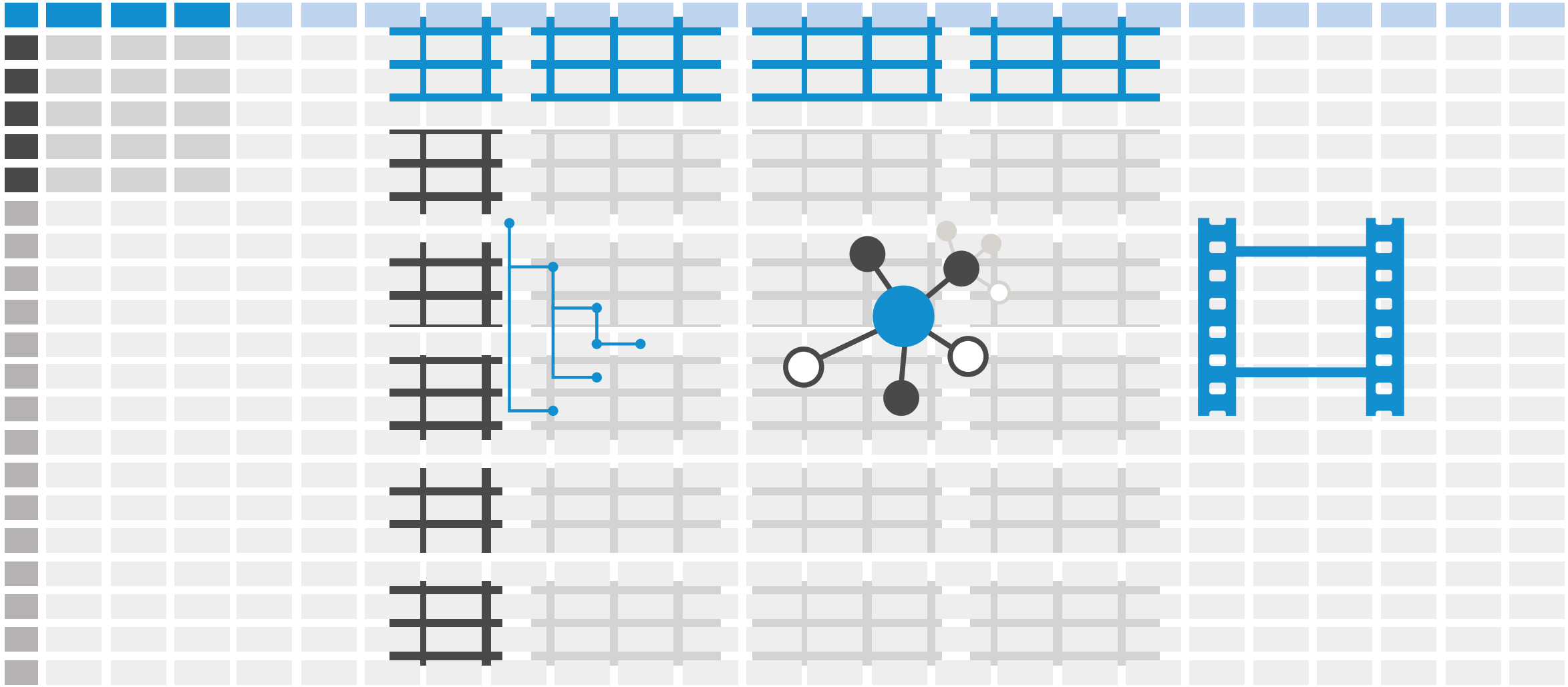
Why is Big Data Growing?

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

Scale out, not up

Run on commodity clusters

Fault tolerant

Agile (semi & unstructured data)

DATASTAX

Couchbase

Cassandra

mongoDB

HBASE

Fast read/write

"OLTP" type use-cases

Examples, sensor data, clickstream...

Large-scale processing

"datawarehouse"

Recommendation engines, predictive, etc

Hortonworks

MAPR

cloudera

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

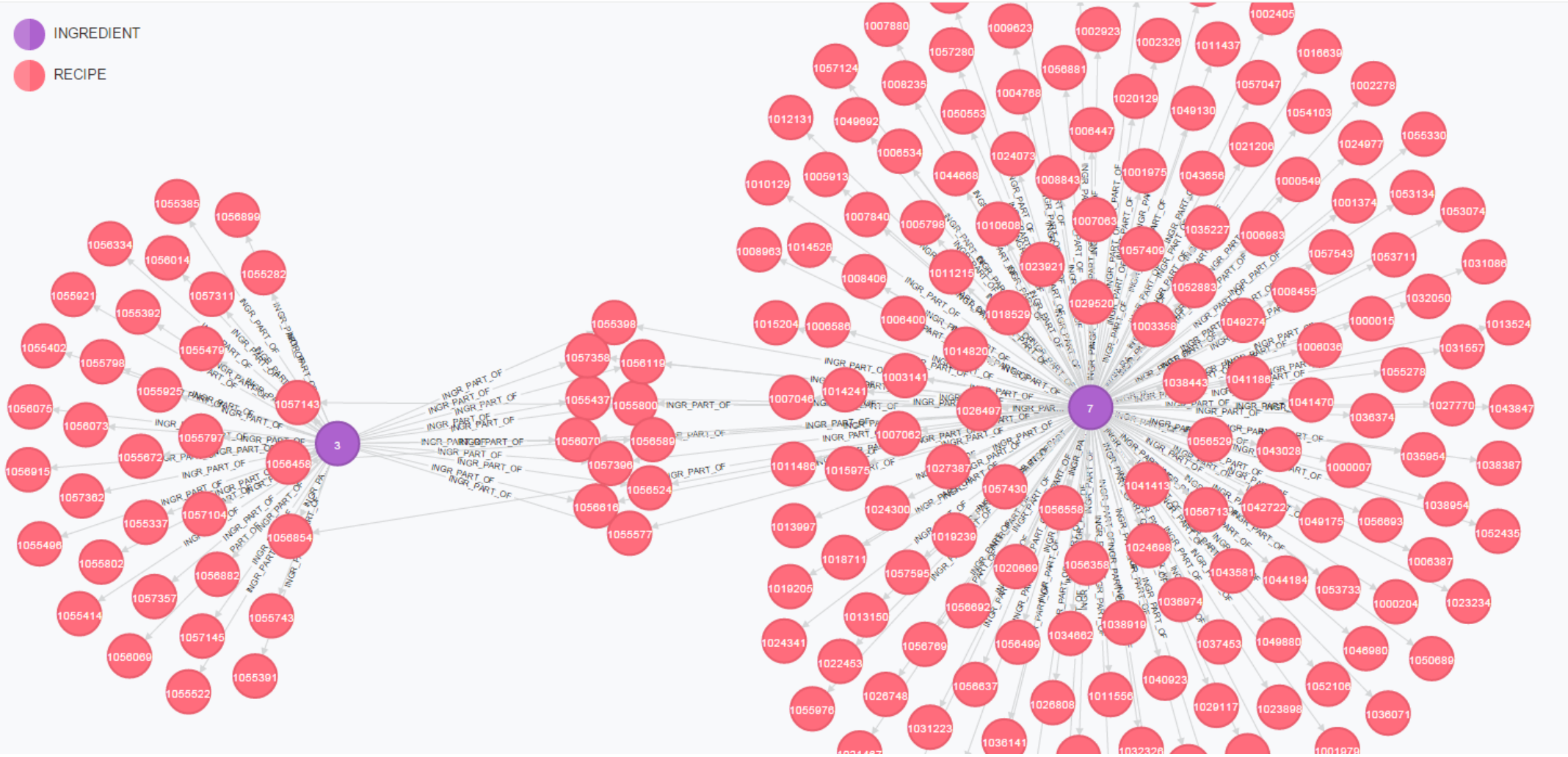
- SAP Analytics tools view Hadoop as just another data source, but needs special considerations
- Complement your existing data infrastructure with Hadoop or NoSQL and derive value with familiar SAP tools
- Use SAP Analytics directly against Big Data sources, or with HANA for real-time analytical capabilities

Thinking about Big Data Differently

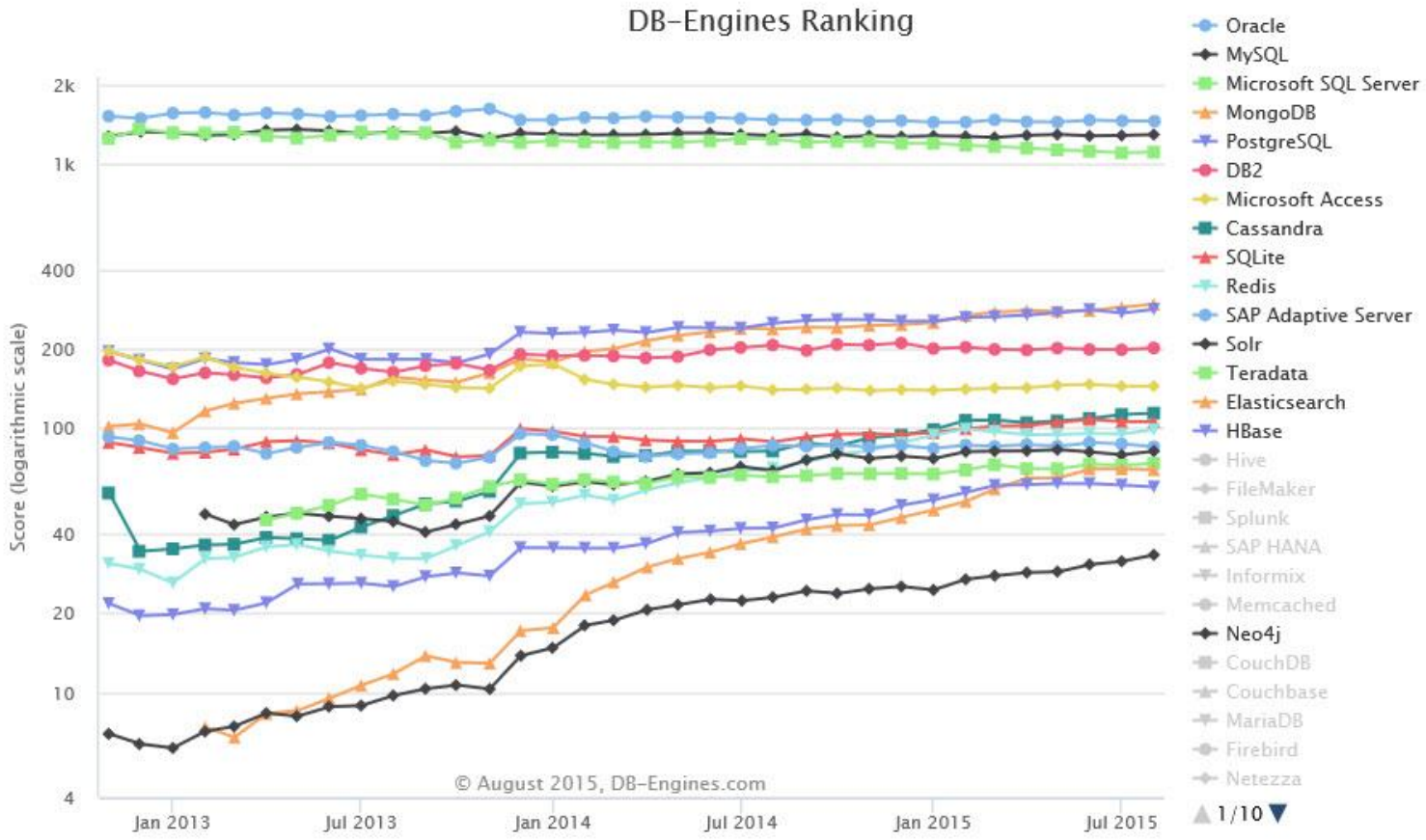
MongoDB


```
{
        "_id" : ObjectId("5447f3328ab4cb34af6e0987"),
        "hash" : "000000000671b1e2b5c146ae4a256ab6ae673eb09c8bae09500372fda89a9010f",
        "confirmations" : 324656,
        "size" : 216,
        "height" : 1282,
        "version" : 1,
        "merkleroot" : "64c7c3e4ae3ac1f104307389c1a83cecc60dcf4d810e1cf3860329fc554601e7",
        "tx" : [
                "64c7c3e4ae3ac1f104307389c1a83cecc60dcf4d810e1cf3860329fc554601e7"
        ],
        "time" : 1232561773,
        "nonce" : NumberLong("2388031237"),
        "bits" : "1d00ffff",
        "difficulty" : 1,
        "chainwork" : "0000000000000000000000000000000000000000000000000000000503050305 03",
        "previousblockhash" : "000000005bcc4c9b9f0fd22ea09985abd17896a063be8ac81390e3d56b5eec71",
        "nextblockhash" : "000000005bc20e95e988500042863abf7ccfdc188b69762108dc653947281c58",
        "date" : "1/21/2009",
        "timestamp" : "10:16:13",
        "tcount" : 1,
        "BTCtransactions" : [
                {
                        "tx" : "64c7c3e4ae3ac1f104307389c1a83cecc60dcf4d810e1cf3860329fc554601e7",
                        "block_hash" : "000000000671b1e2b5c146ae4a256ab6ae673eb09c8bae09500372fda89a9010f",
                        "date" : "1/21/2009",
                        "timestamp" : "10:16:13",
                        "decodedTransaction" : {
                                "txid" : "64c7c3e4ae3ac1f104307389c1a83cecc60dcf4d810e1cf3860329fc554601e7",
                                "version" : 1,
                                "locktime" : 0,
                                "vin" : [
                                        {
                                                "coinbase" : "04ffff001d02030a",
                                                "sequence" : NumberLong("4294967295")
                                        }
                                ],
                                "vout" : [
                                        {
                                                "value" : 50,
                                                "n" : 0,
                                                "scriptPubKey" : {
                                                        "asm" : "046e0f9086d0a9202bff6be49d059d0830335b88cd00f6d04b29171bda8618e26b82466effcf56c5939870fd9d55930e3
                                                        "hex" : "41046e0f9086d0a9202bff6be49d059d0830335b88cd00f6d04b29171bda8618e26b82466effcf56c5939870fd9d55930
                                                        "reqSigs" : 1,
                                                        "type" : "pubkey",
                                                        "addresses" : [
                                                                "1Fx9WBMHRBXTQhUUS3cuuCQWJbLfkzKg4U"
                                                        ]
                                                }
                                        }
                                ]
                        }
                }
        ]
}
```

Graphs



How popular are they?



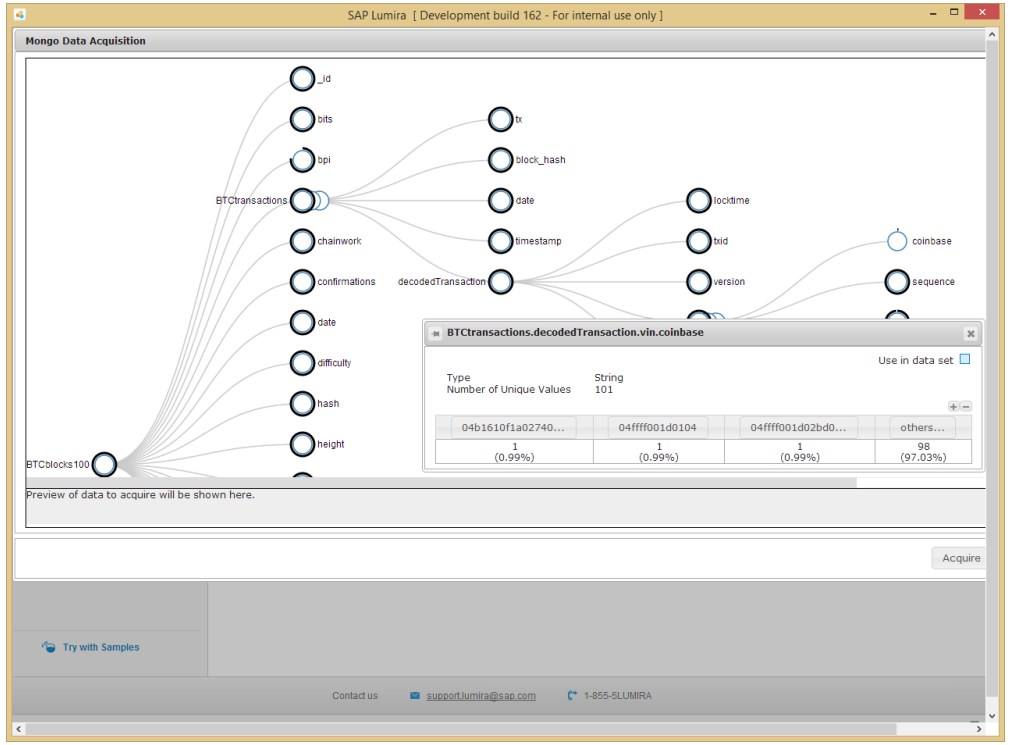
http://db-engines.com/en/ranking_trend

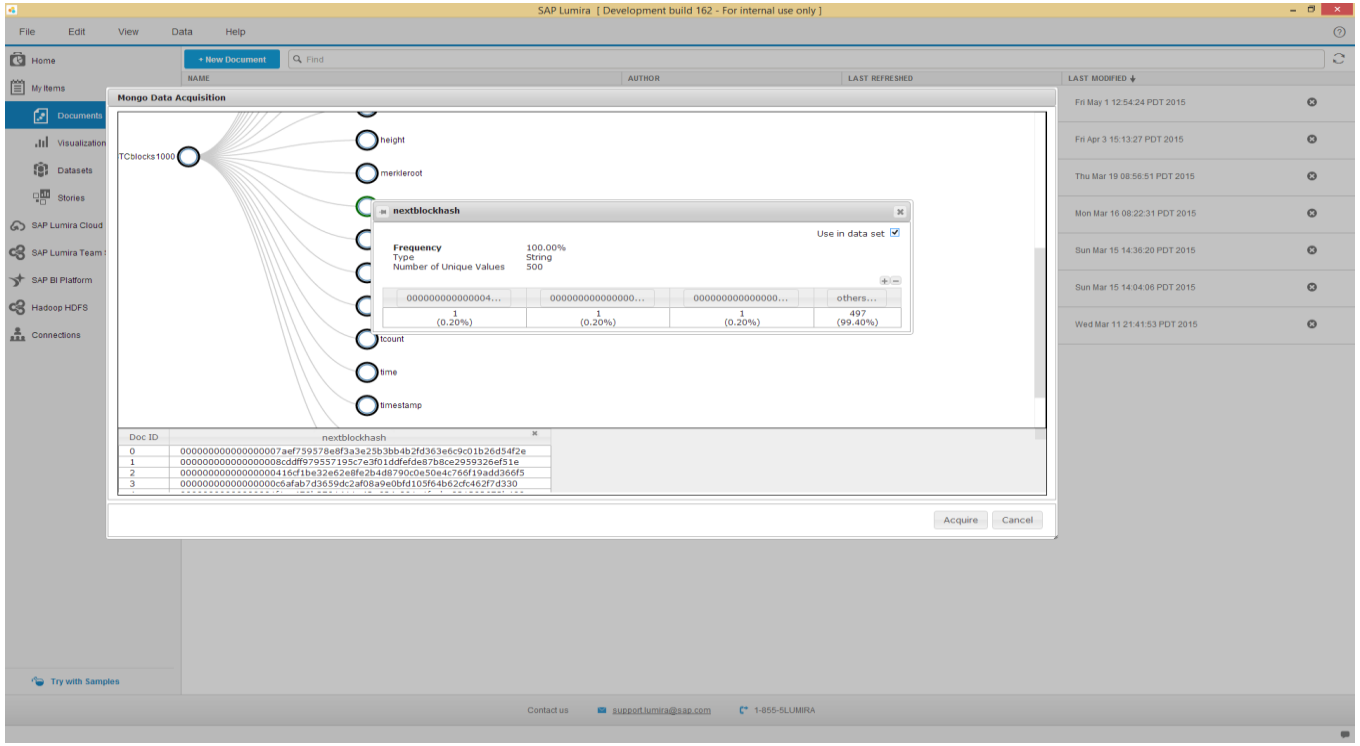
Adjust to these new types of sources

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

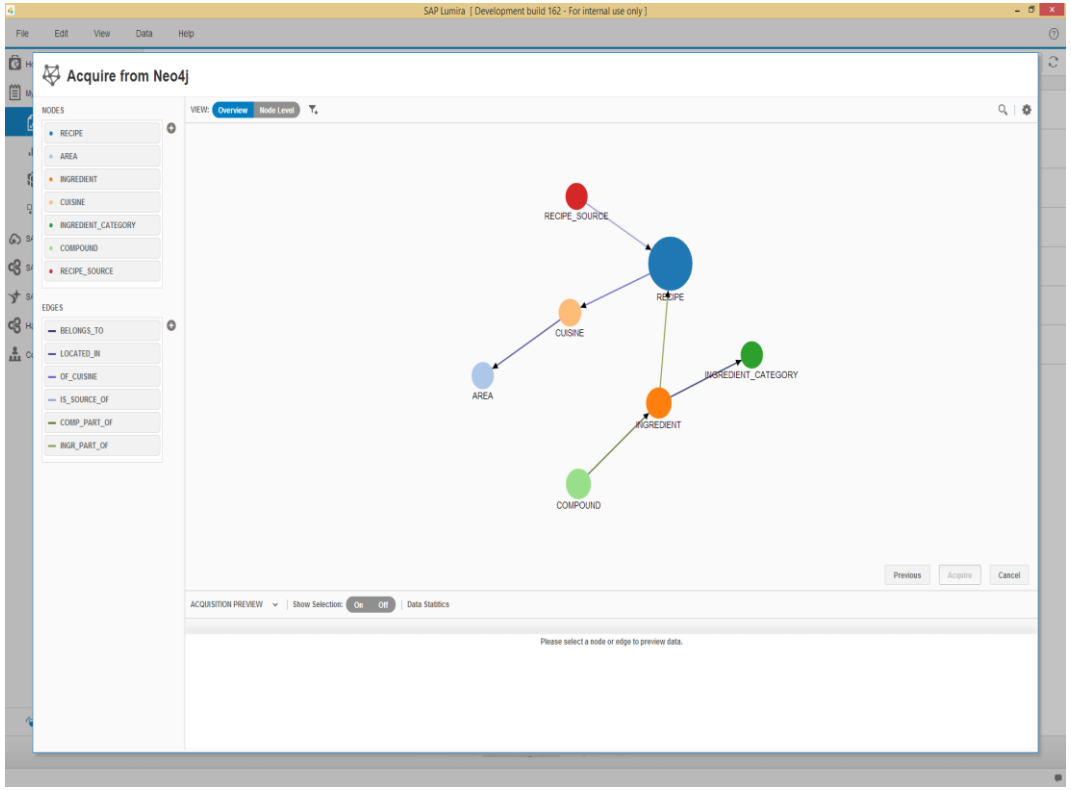
- Flattening the data is the end goal, not the beginning
- Support schema-less sources
- Dedicated acquisition experience

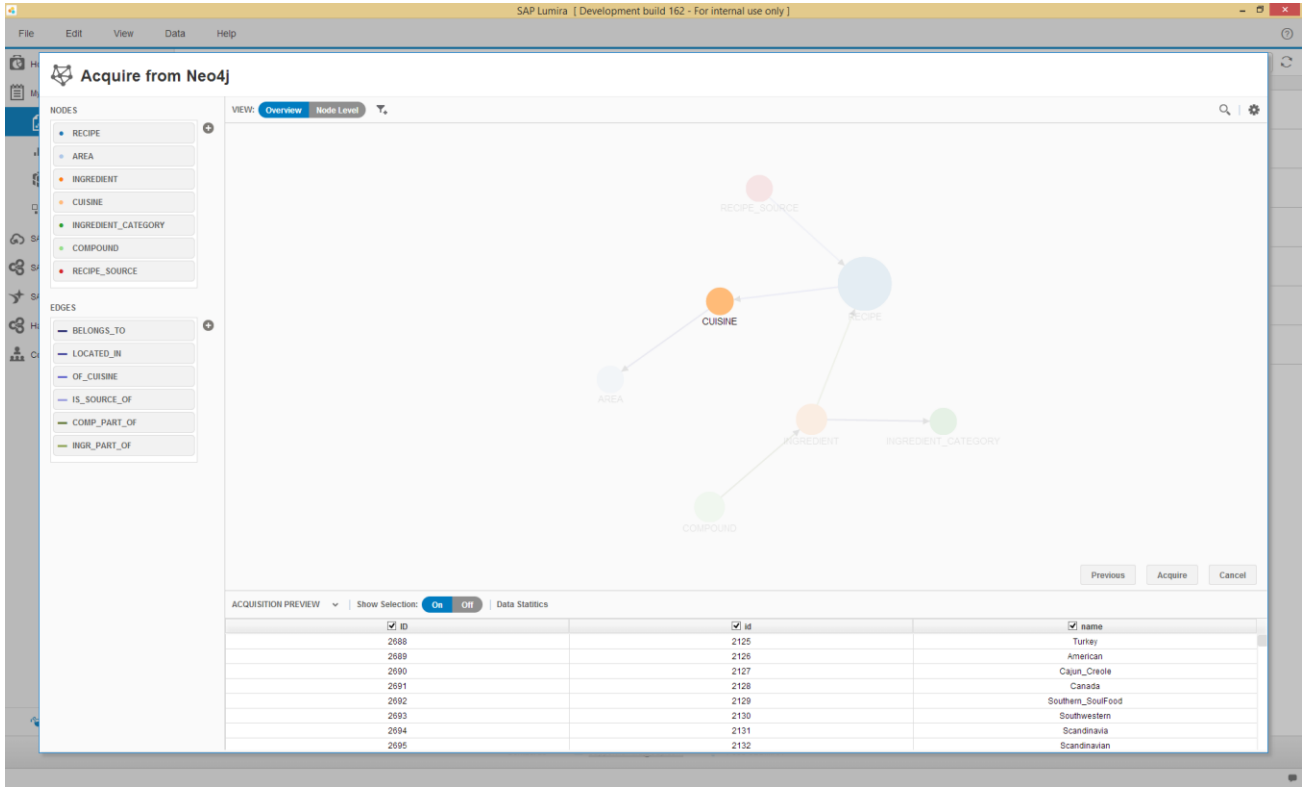
MongoDB and Lumira





Neo4j and Customized Acquisition Experience





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

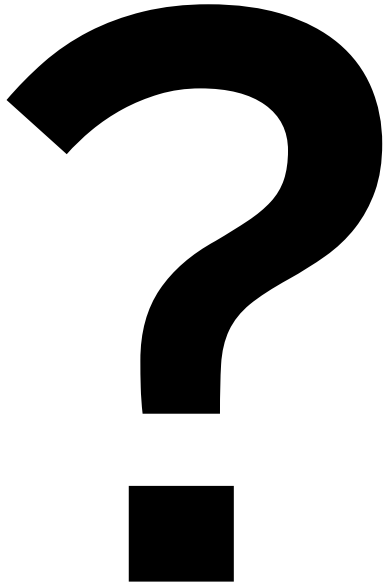
Data of Unknown Value

Data Lakes for Exploratory Analytics

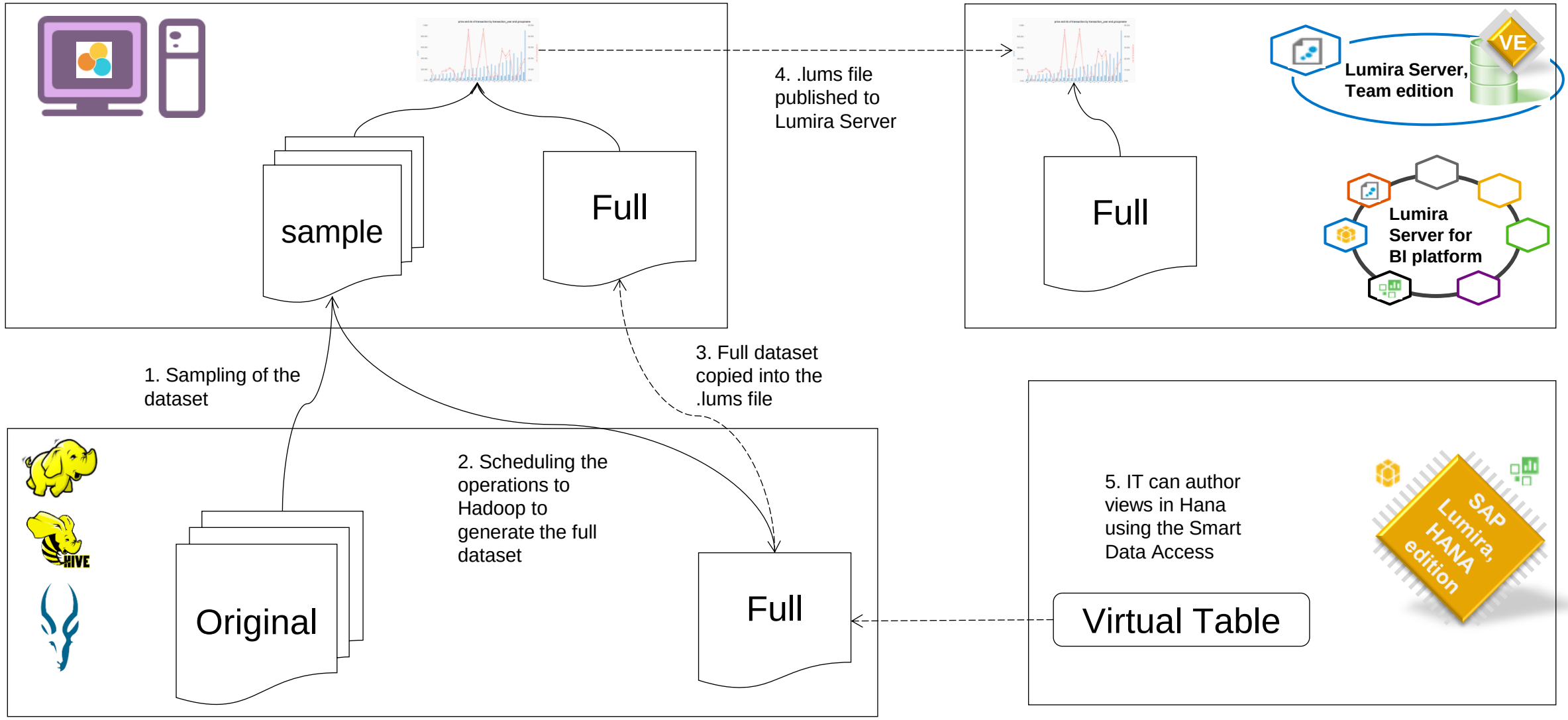
More data more problems…

- 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…

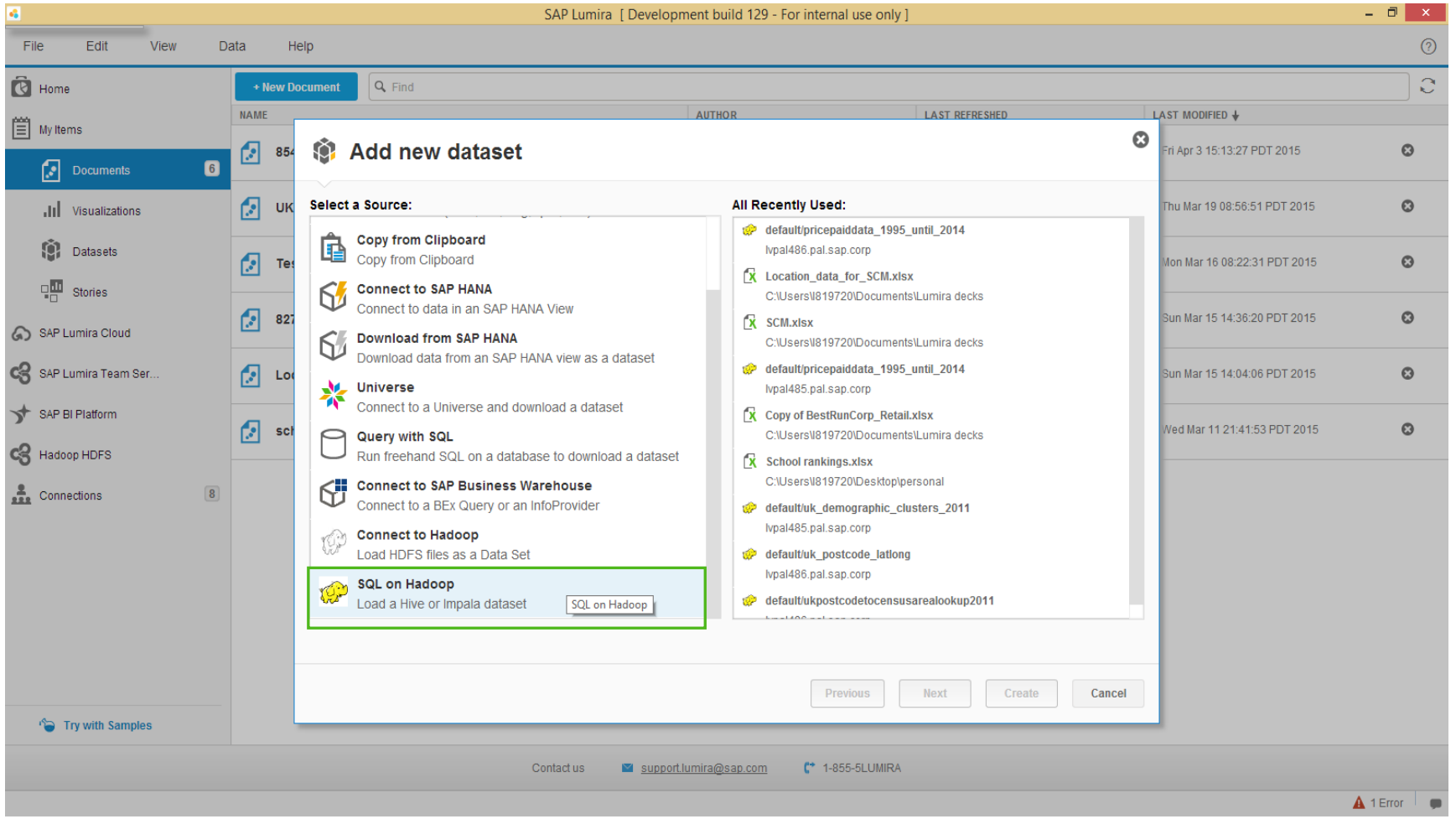
How do we bring End-User Self-Service to Hadoop environments



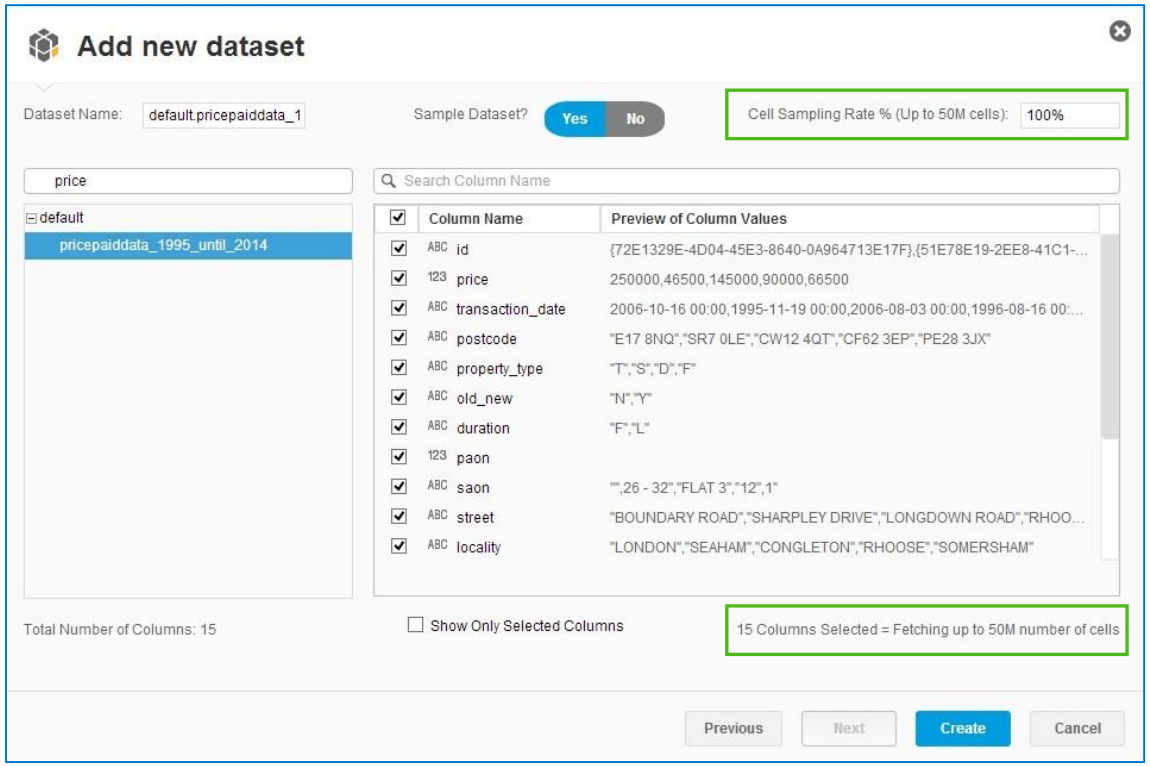
Data Wrangling on any data size in Hadoop!

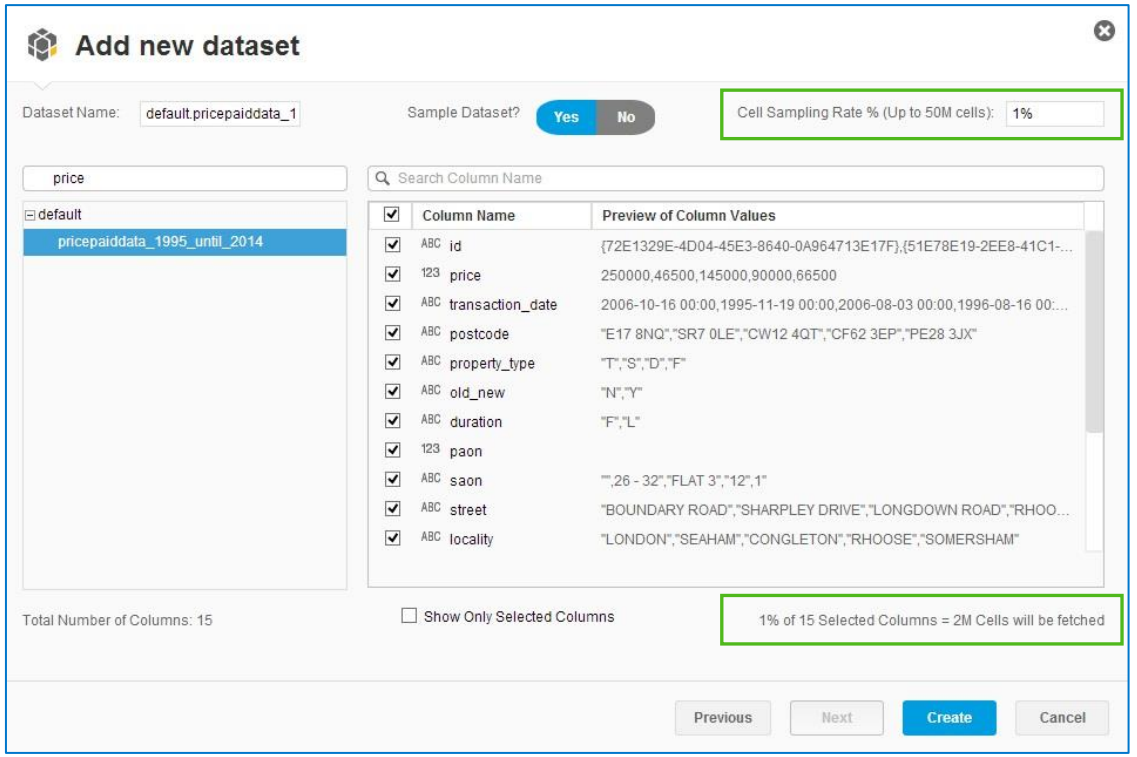


Data Wrangling on Hadoop

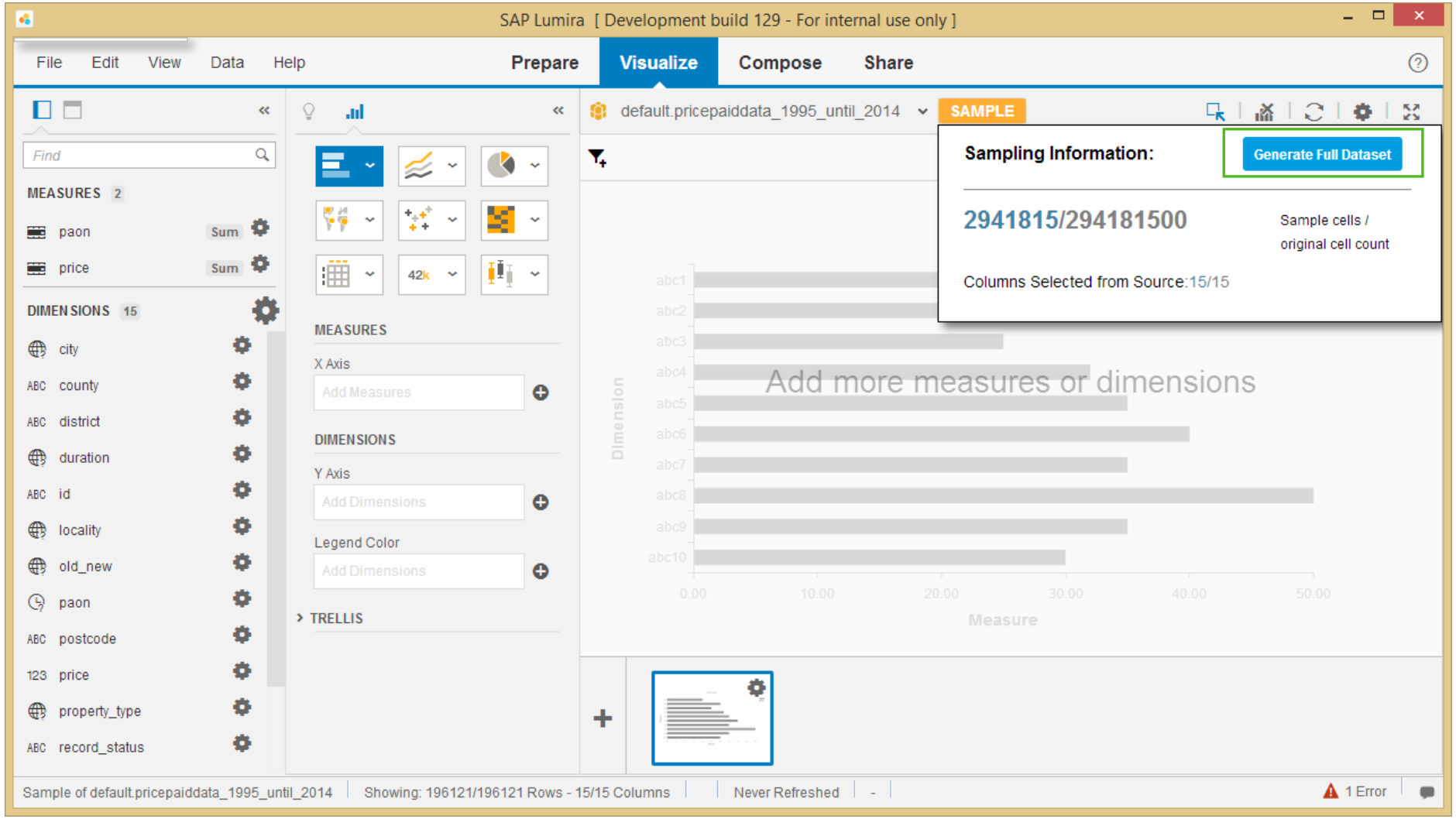


Data Wrangling on Hadoop

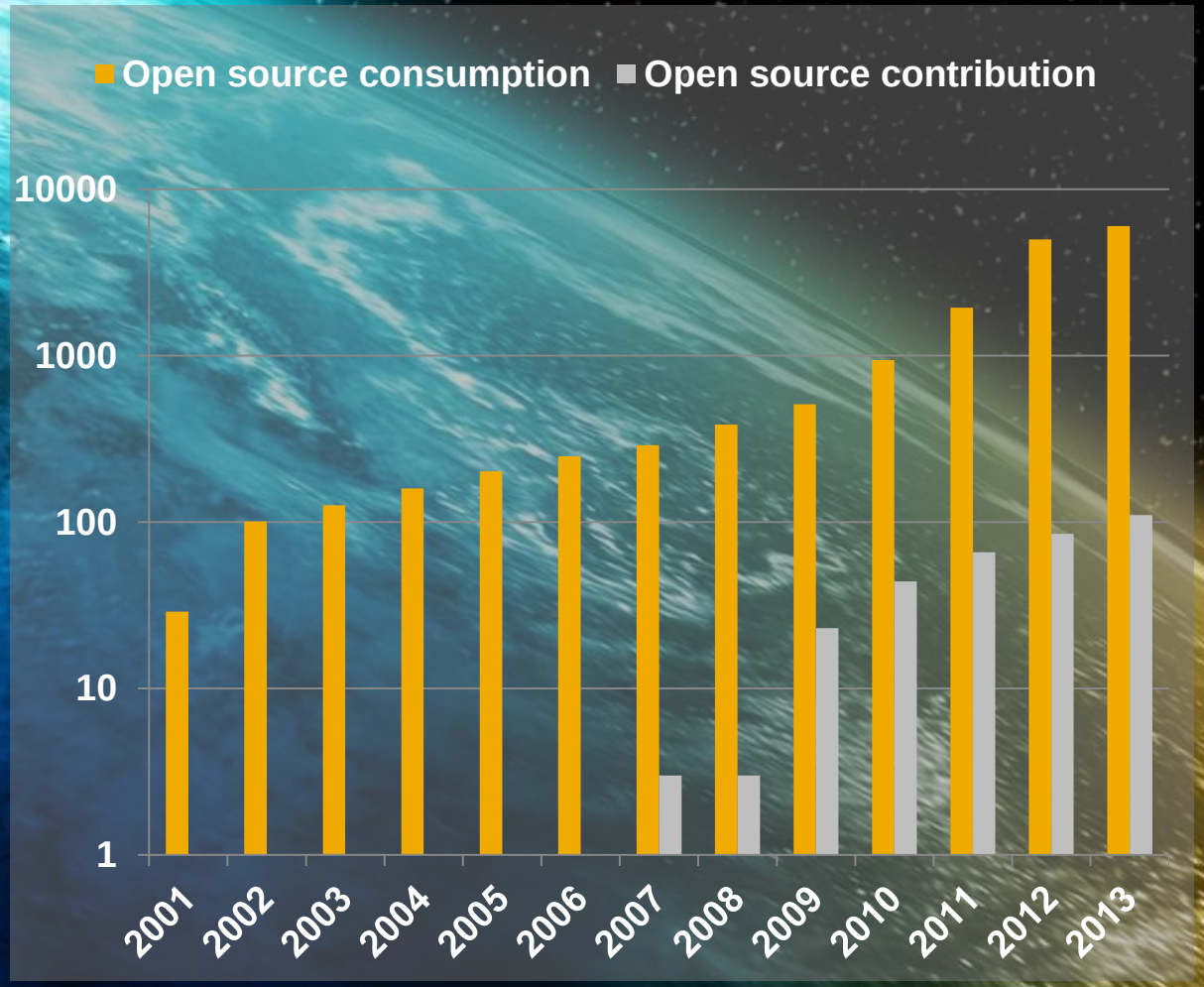




Data Wrangling on Hadoop

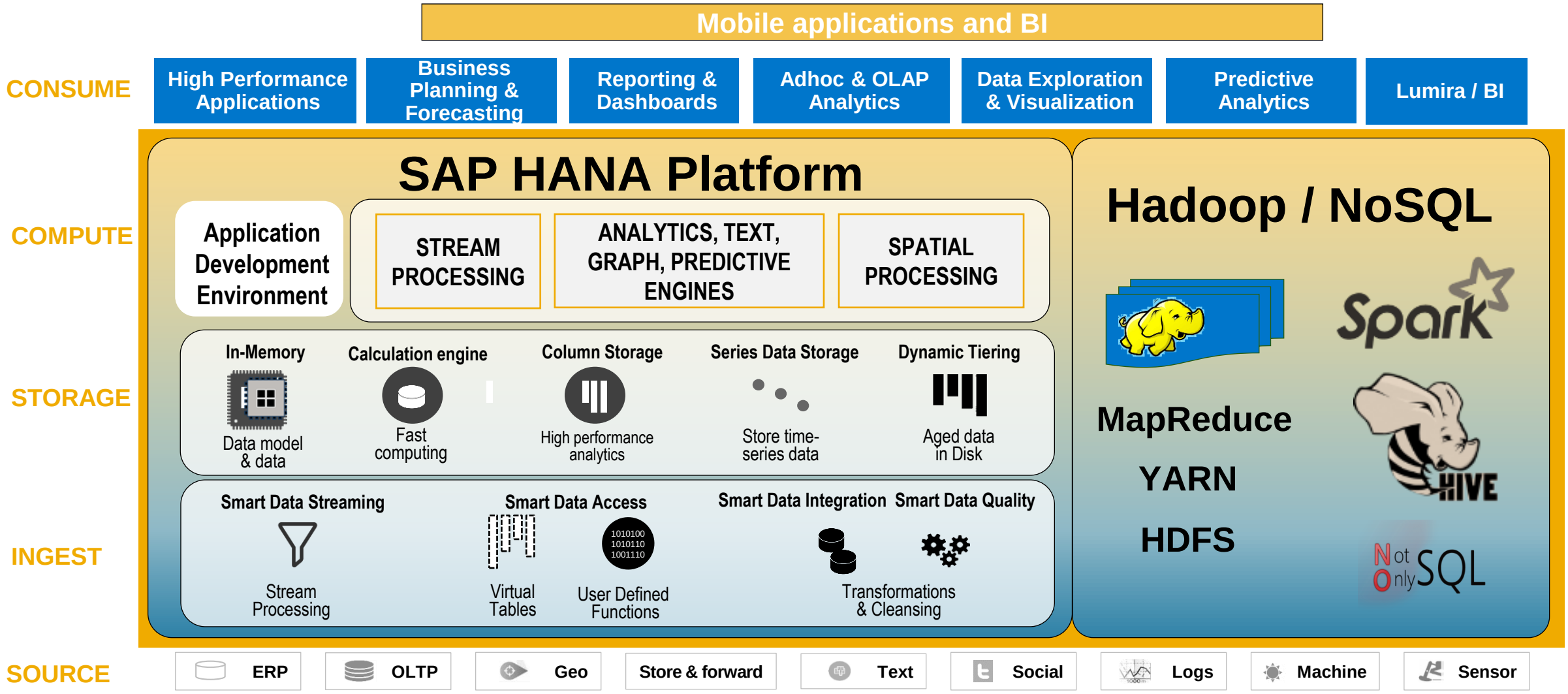


HADOOP is key Part of SAP's Open Source Development usage



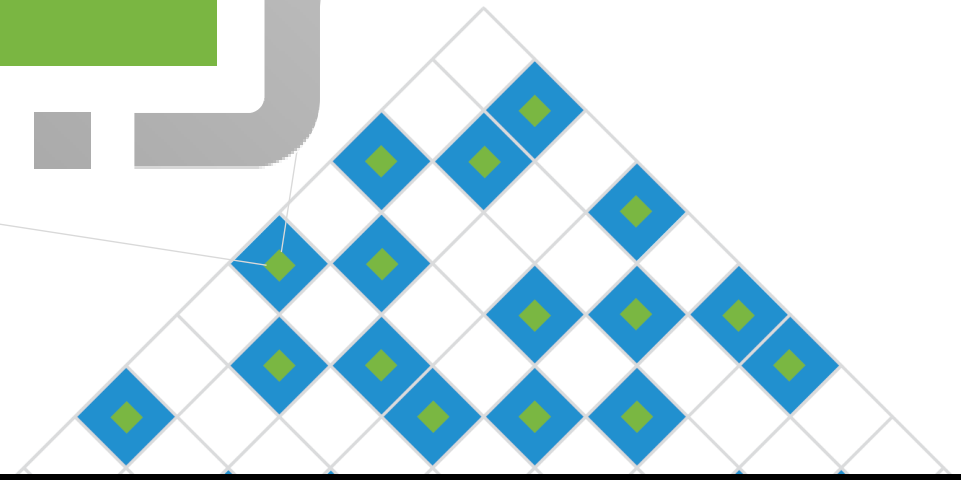
SAP Contributes to over 100 Open Source Projects

The SAP focus: End-to-end value chain



SAP HANA Vora
What's Inside and What Does It Do?

Drill Downs on HDFS
Mashup API Enhancements
Compiled Queries
HANA-Spark Adapter
Unified Landscape
Open Programming



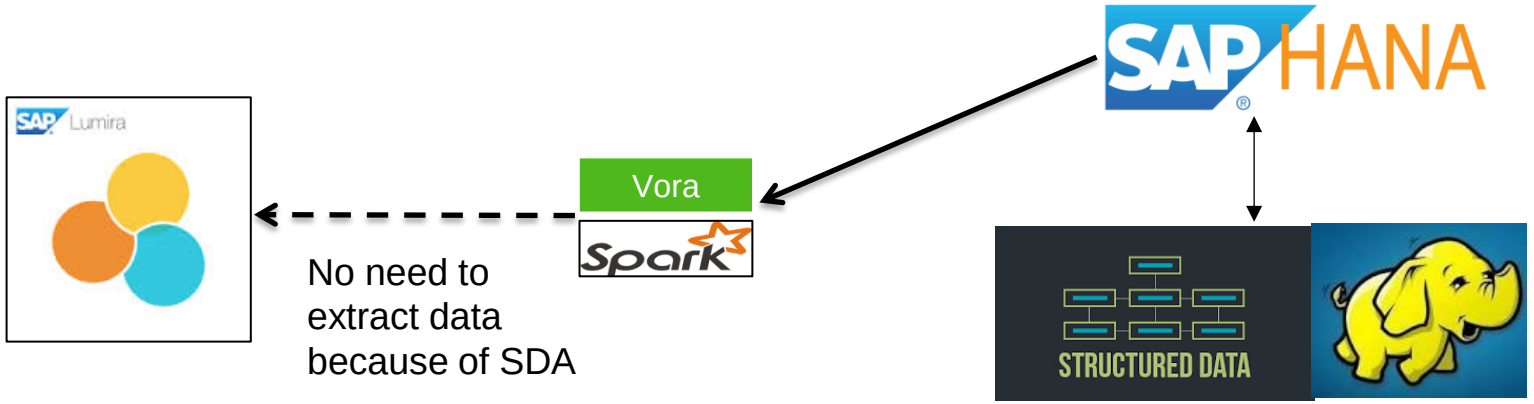
SAP HANA Vora is an in-memory query engine which leverages and extends the Apache Spark execution framework to provide enriched interactive analytics on Hadoop.

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

SAP 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

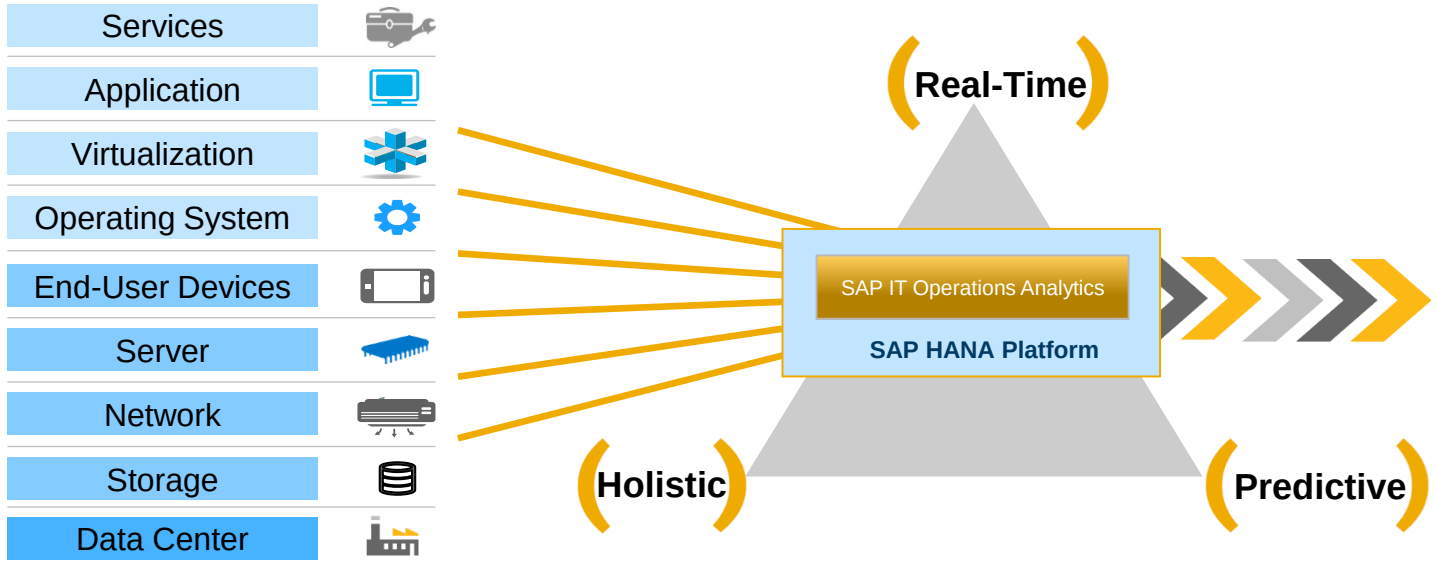
Zero Downtime

Security / Data Protection

SAP IT Operations Analytics

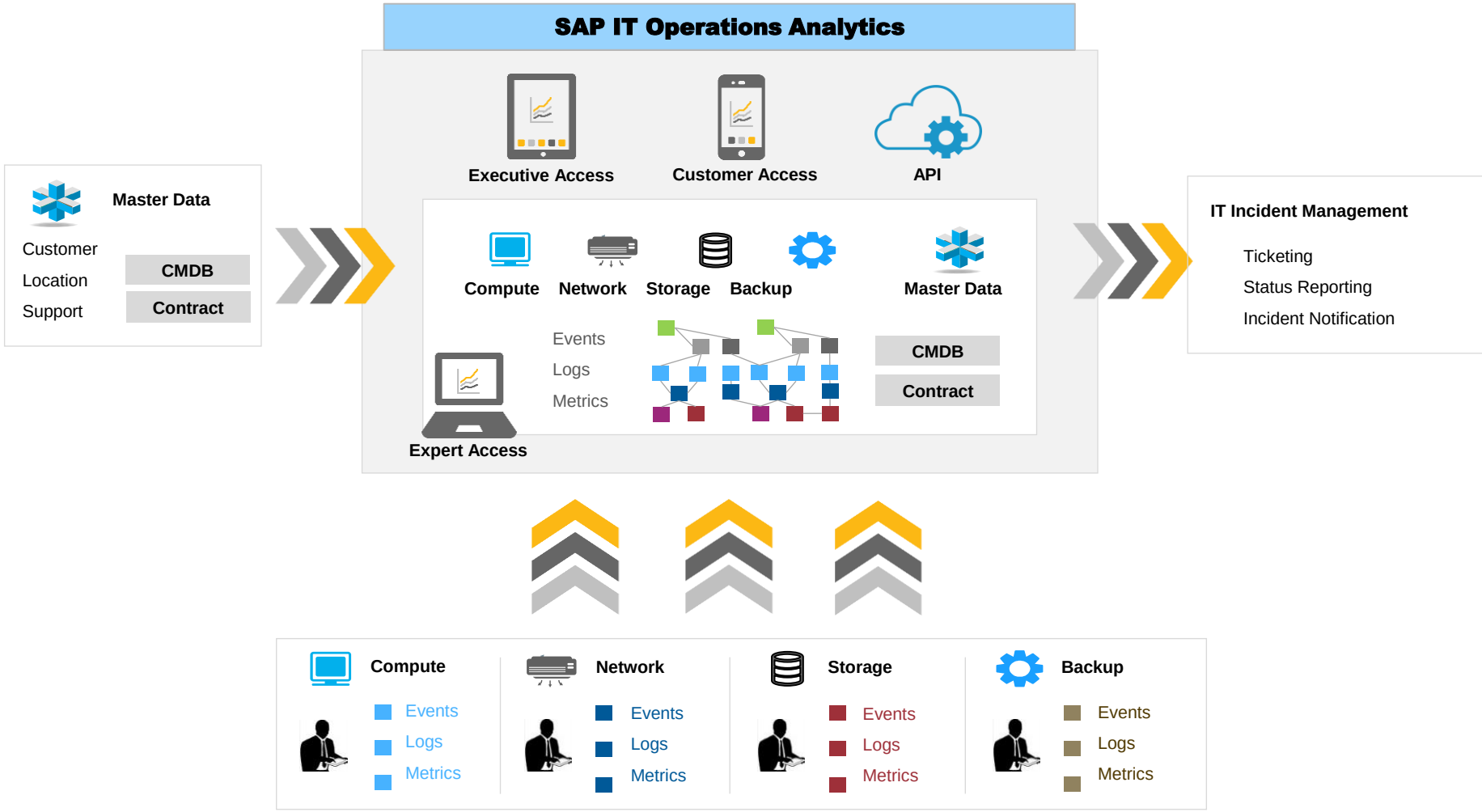
Now GA!

> **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.**



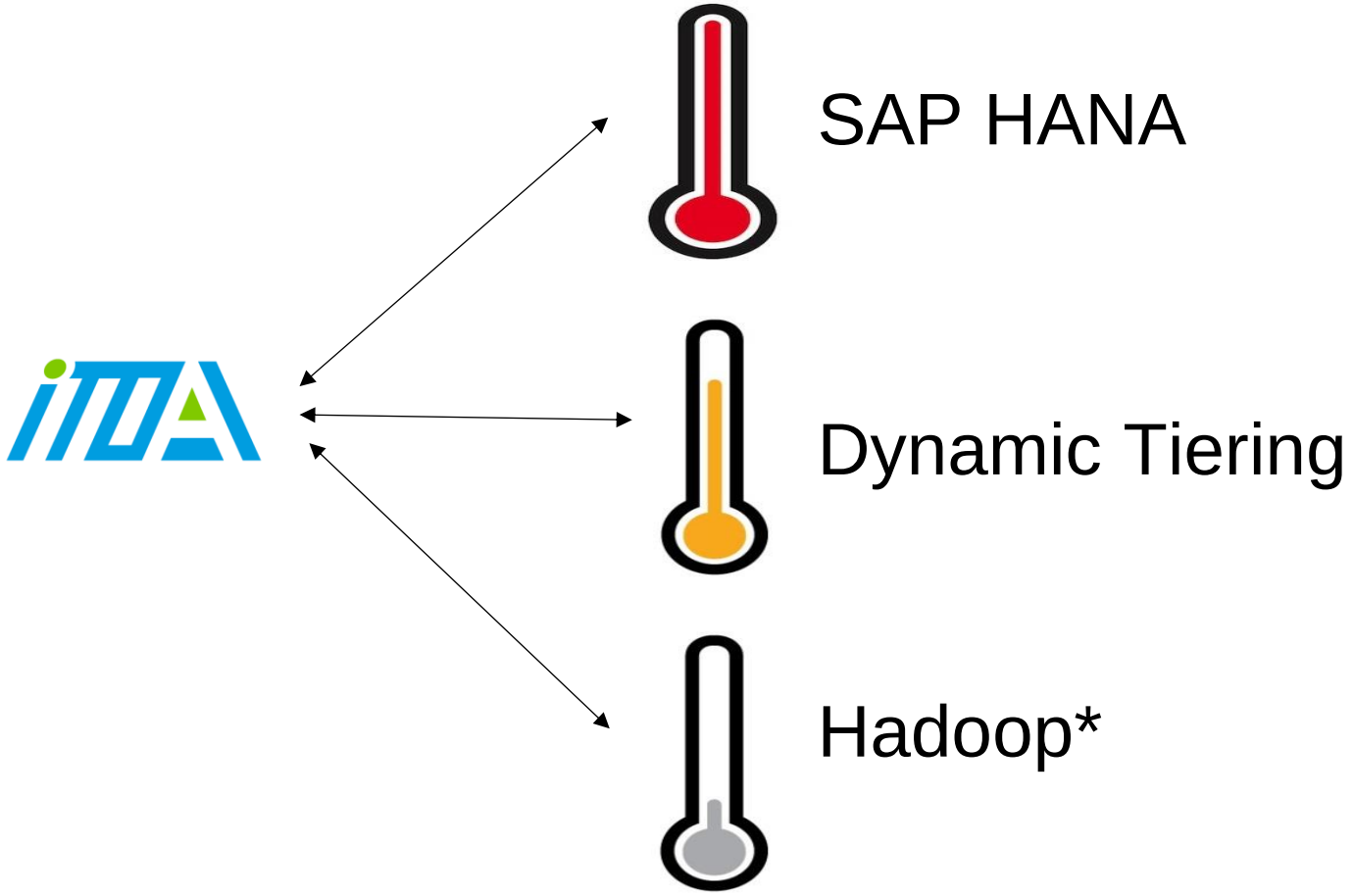
Root Cause Analytics

"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

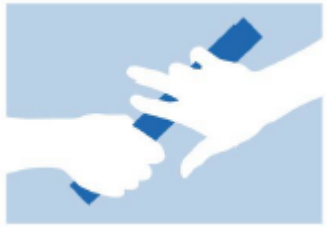
Realizing the impact on your data center

Responsive IT Operations

Maximizing IT Responsiveness

by having a holistic view within and allover 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
@AngelaHarveySAP

© 2016 SAP SE or an SAP affiliate company. All rights reserved.