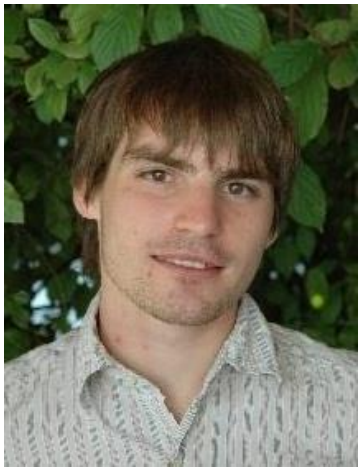
Deutsche
SQL Server Konferenz 2014
Community powered product knowledge by PASS Deutschland & Microsoft

PASS

# Big Data Scenario mit Power BI vs. SAP HANA

Gerhard Brückl

# About me



Gerhard Brückl

Working with Microsoft BI since 2006
Started working with SAP HANA in 2013

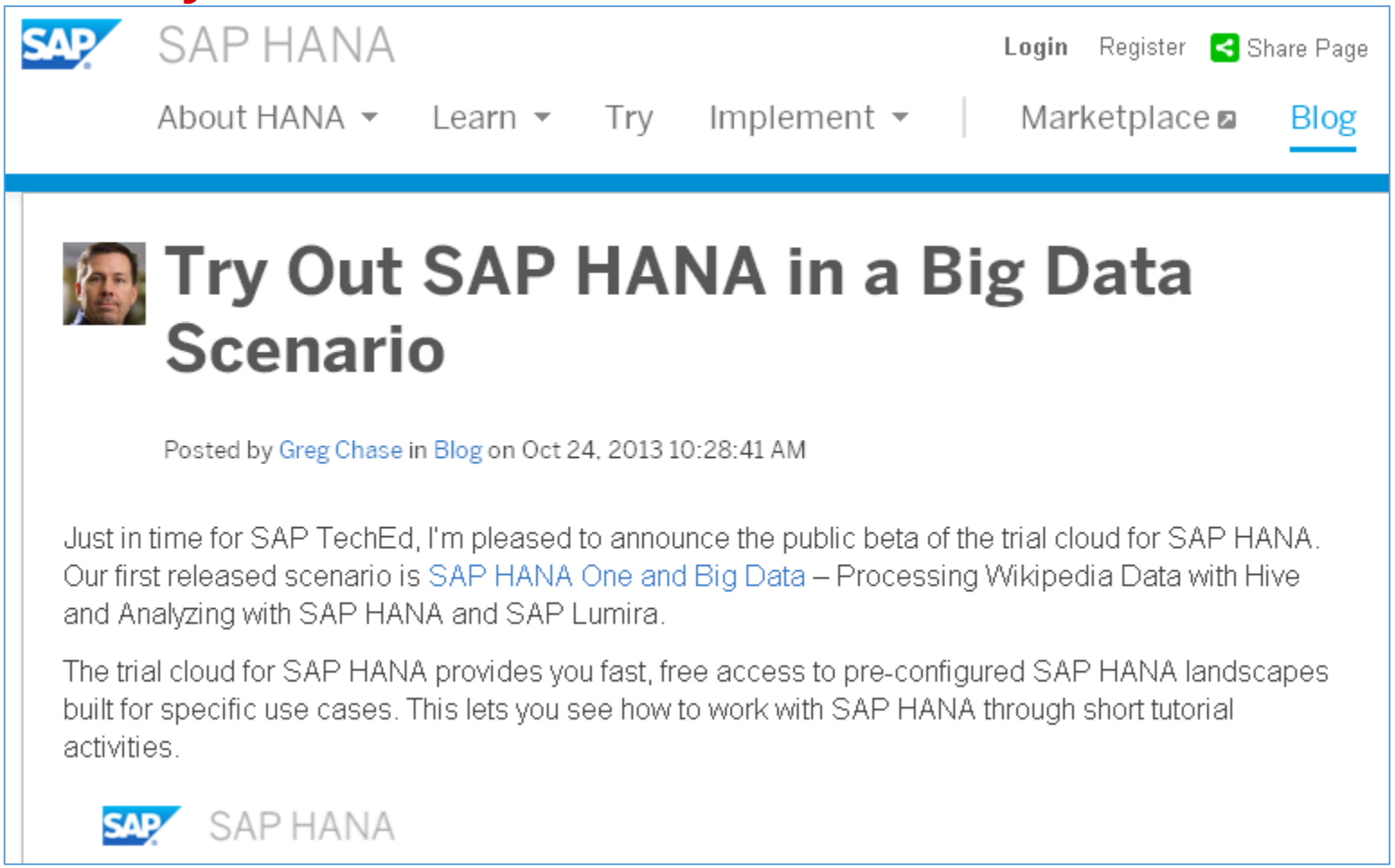
focused on Analytics and Reporting

Blog: blog.gbrueckl.at
eMail: gerhard@gbrueckl.at

pmOne

SSAS MAESTRO by Microsoft

# History




http://www.saphana.com/community/blogs/blog/2013/10/24/try-out-sap-hana-in-a-big-data-scenario

# Background

Wikimedia Foundation provides data for WikiPedia usage
- hourly page-view statistics
http://dumps.wikimedia.org/other/pagecounts-raw/

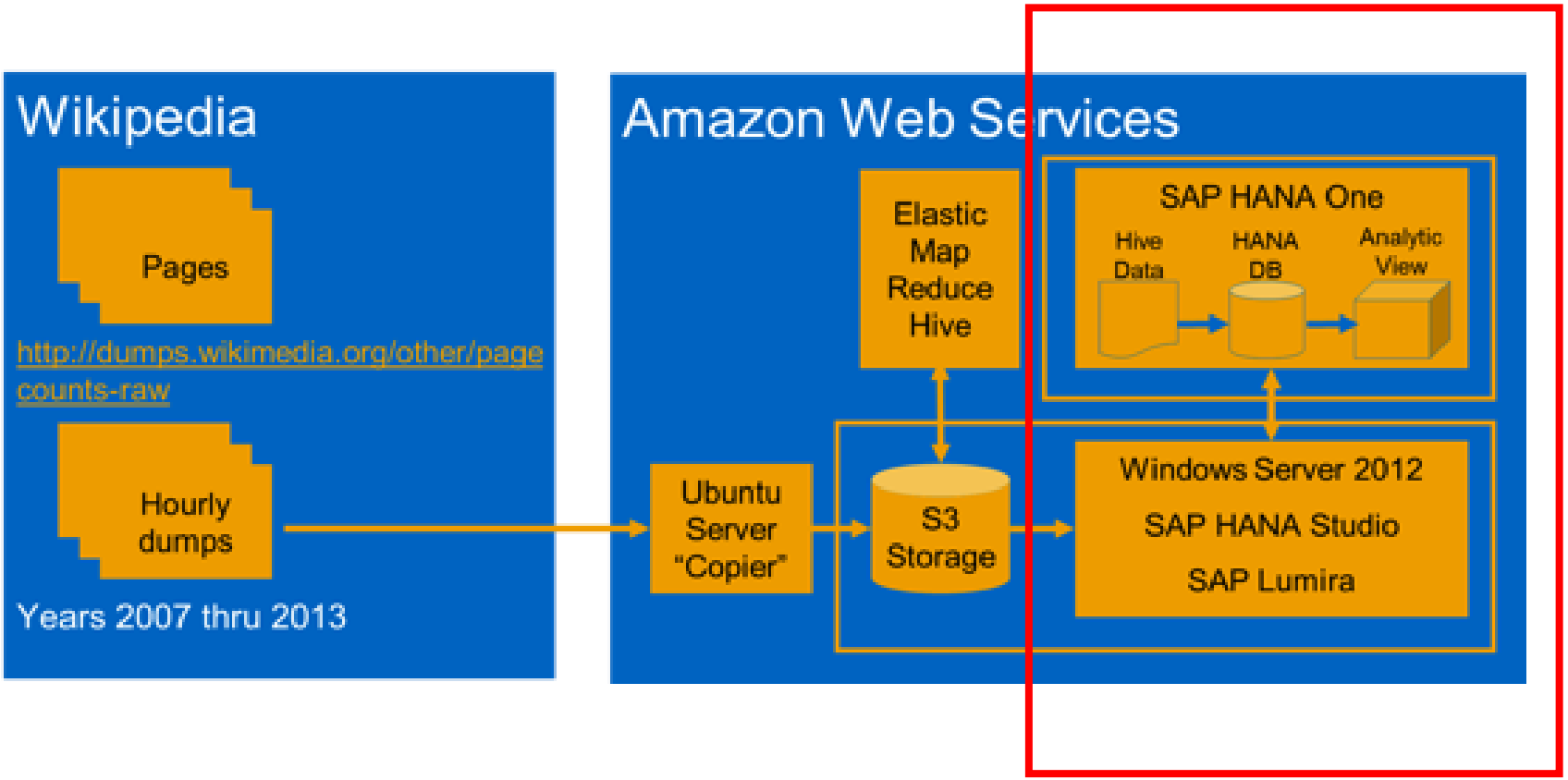
Loaded raw-files into Amazon S3 storage (~750 GB)

Create HIVE-table on top of raw-files

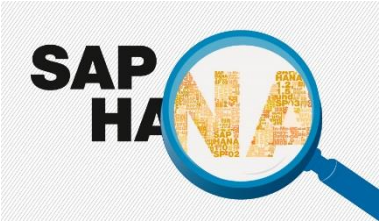
Extract important data from HIVE-table to CSV-files
WHERE [Page Count]/Hour >= 100

(Start analysis)

# Setup


Wikipedia
Pages
http://dumps.wikimedia.org/other/pagecounts-raw
Hourly dumps
Years 2007 thru 2013
Amazon Web Services
Elastic Map Reduce Hive
SAP HANA One
Hive Data
HANA DB
Analytic View
Ubuntu Server "Copier"
S3 Storage
Windows Server 2012
SAP HANA Studio
SAP Lumira

# Basic Comparison



| SAP HANA | | Microsoft Power BI |
|---|---|---|
| Enterprise DB Solution | **Use Case** | Self-Service BI Solution |
| SUSE Linux Enterprise (Server) | **Platform** | Microsoft Excel 365 / 2013 (Client) |
| In-memory Column Store | **Technology** | In-memory Column Store |
| Scalable | **Performance** | “Limited” |
| Scalable | **Storage** | Limited |
| Supported | **MPP** | No |
| | **…** | |
| Not Free | **Price** | Free |

# Is it a valid comparison?

No

# Why do we still compare it?
# And what does it have to do with Big Data?

Big Data scenarios usually start on a Green Field

→Vague imagination of what we want

→We can not know in advance what we get

→Maybe the goal changes/shifts because of new insights??

→Can we use the results at all??

**ROI – Was the investment worth it?**

The concept of Lab & Factory

# Lab & Factory – not a new Concept

New Idea

New Requirement

New Ingredient

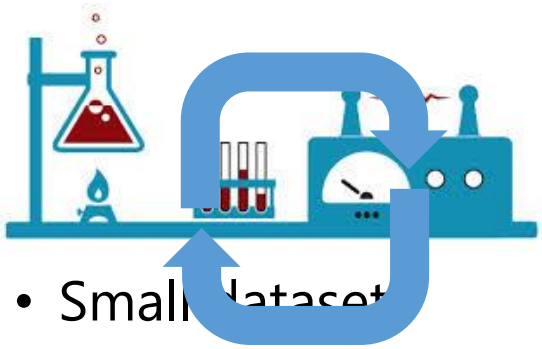
New Product

?

?

# Lab & Factory

## Laboratory



- Small dataset
- Full complexity
- Experimental approach
- No fixed results

## Factory

- Large dataset
- Necessary complexity
- Fixed outcome
- Scalability

# Identify Field – Lab or Factory?

Hourly page-view statistics over 3 months

Prefiltered using HADOOP and HIVE

What do we actually want to analyze?

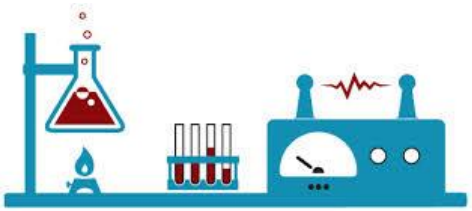
Dataset: Small / pre-filtered

Complexity: not going to change

Output: not defined

| Laboratory | Factory |
|---|---|
| X | |
| X | |
| X | |

# The laboratory

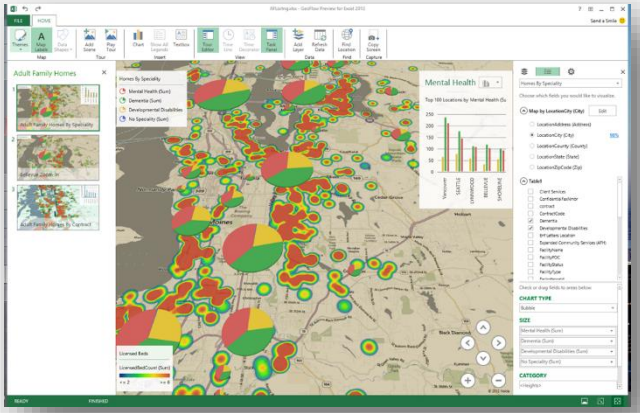


*Try & Error*

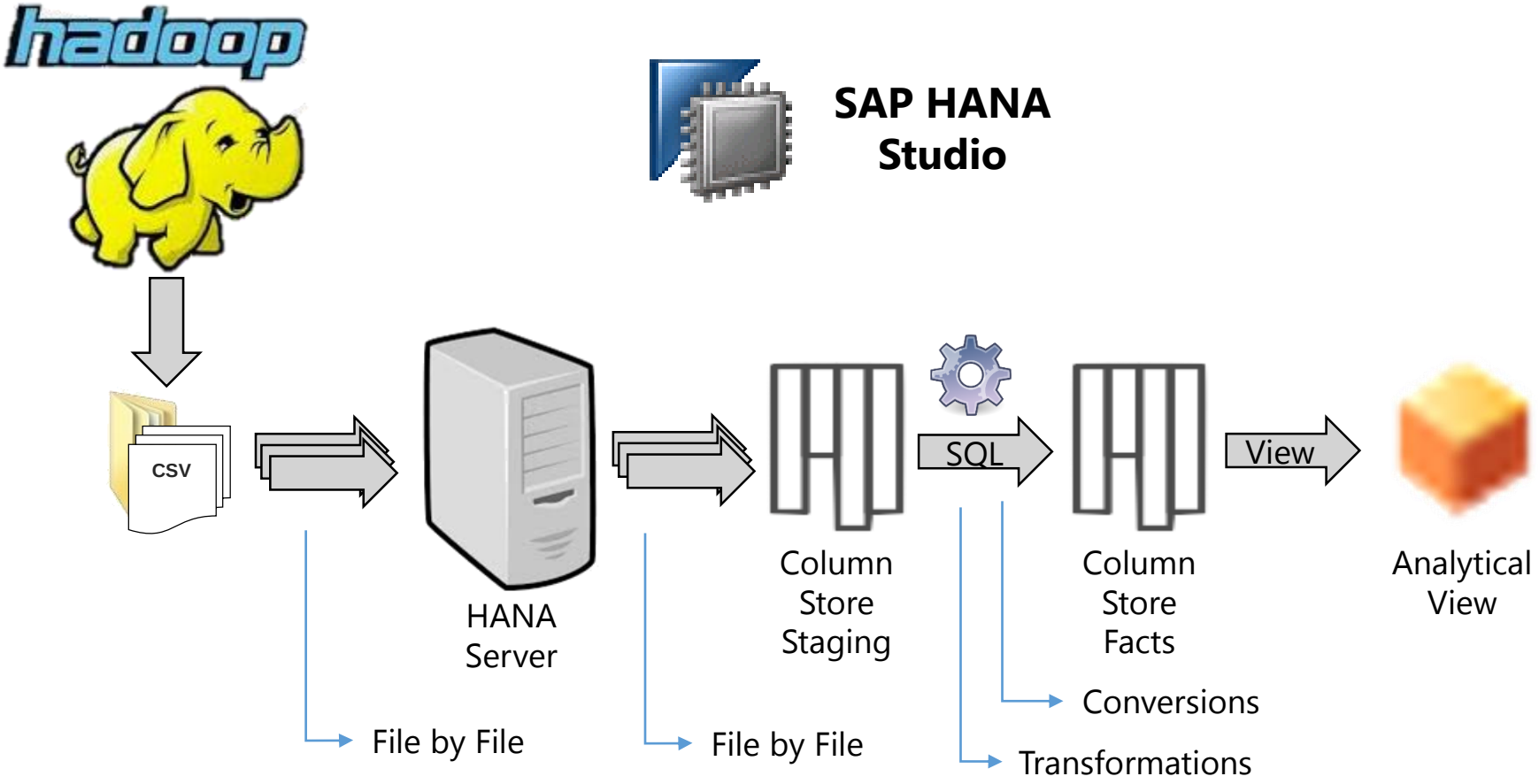
*Fast Iterations*

*Complex Analysis*

*Visualization*

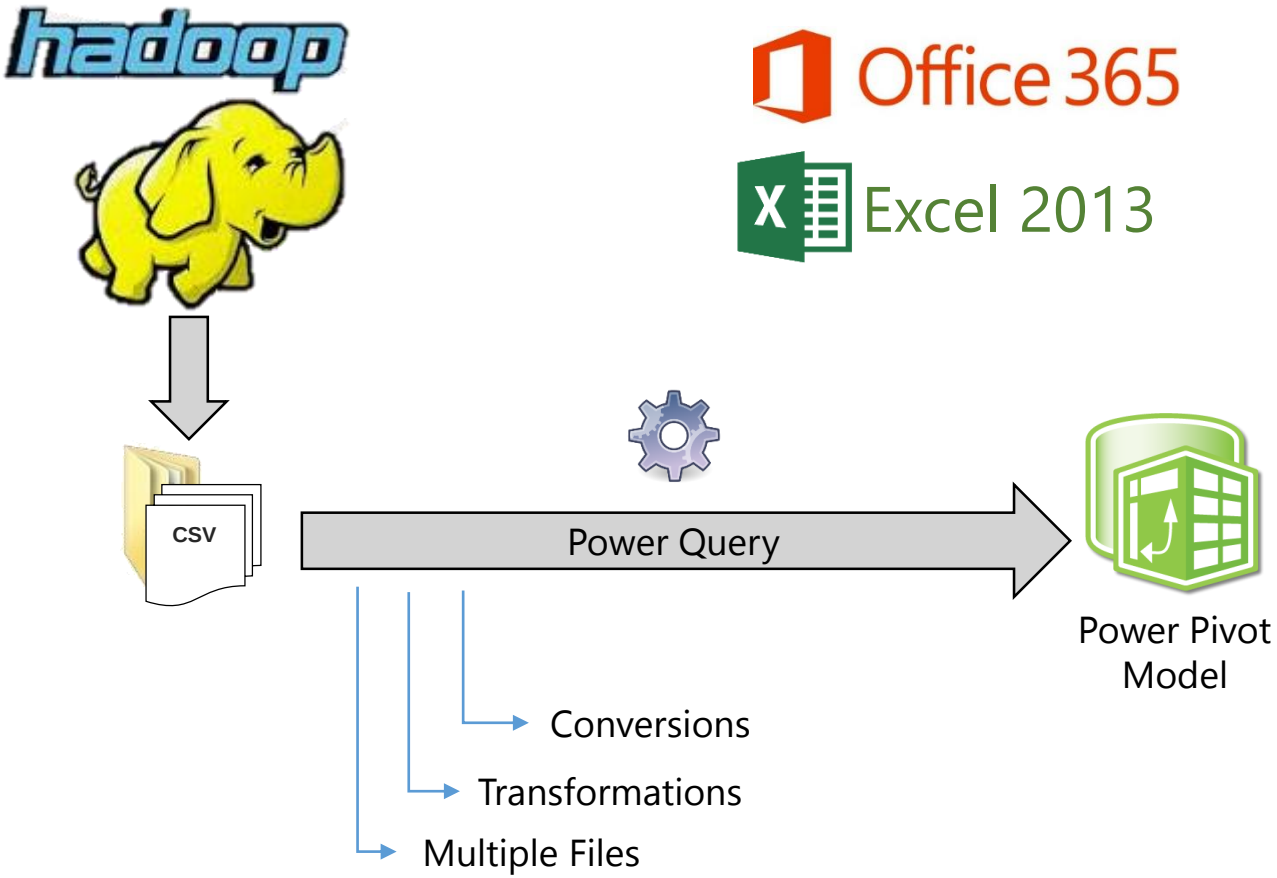


# Workflow – SAP HANA



# DEMO
# SAP HANA

# Workflow – Microsoft Power BI



# DEMO
# Microsoft Power BI

# Try & Error

- SAP HANA
  - Finding errors in data can be tricky
  - Data Transformations and Error Handling requires specific skills (SQL, ...)

- Microsoft Power BI
  - Errors are highlighted during import already
  - Data Transformations and Error Handling is quite easy

# Fast Iterations

- SAP HANA
  - Reloading data is quite complex
  - Upload files one by one
  - Import files one by one
  - Manually trigger transformation steps

- Microsoft Power BI
  - Simple reload of data
  - Multiple files at once
  - One clean process

# Complex Analysis & Visualization

- SAP HANA
  - (Native) integration of **R** in SAP HANA
  - Integrated visual analysis tool is quite basic
  - Use separate front end (SAP LUMIRA) for Visualization
  - Other user experience

- Microsoft Power BI
  - Medium complex analysis possible
  - Fully integrated into Microsoft Excel
    - Charts, Pivot Tables, conditional formatting, ...
    - integrate able with all other Excel Add-In
  - Power View & Power Map for interactive visual analysis

# An efficient laboratory

Power View

Power Pivot

Power Map

Microsoft Excel

Power Query

Unstructured Data

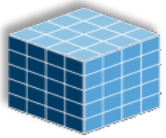
Structured Data

WebServer-Logs
Sensor-Data

Data Sources

Databases

SAP



Cubes

# The factory

*Automated*

*Host it and run it*

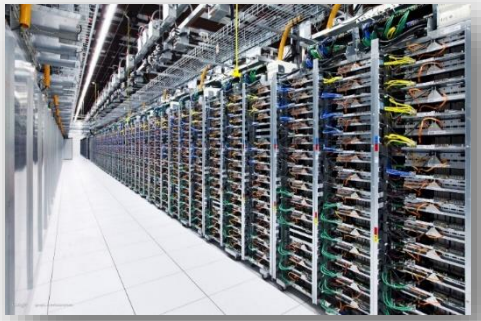
*At Enterprise Scale*

*For Real-time Enterprise*

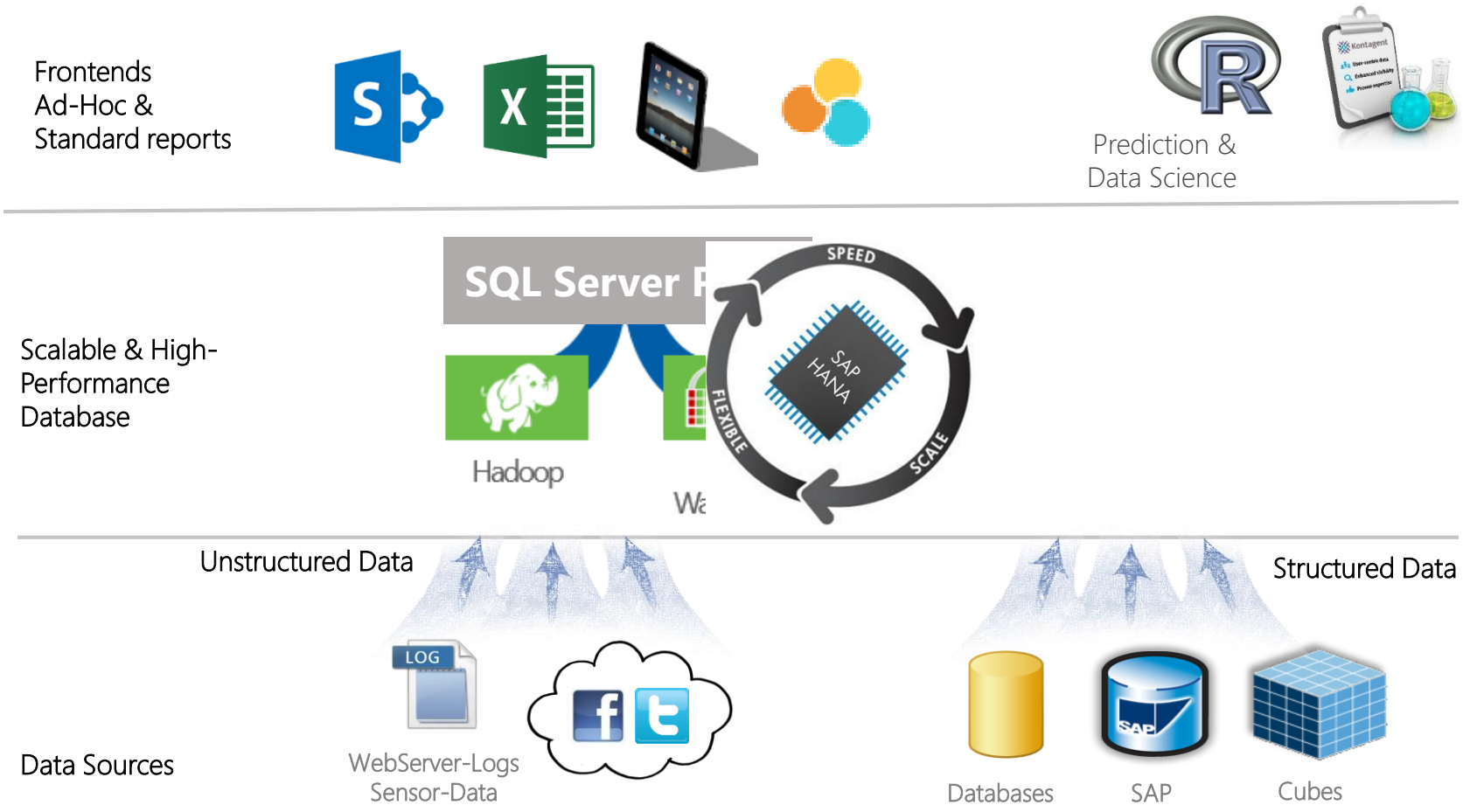
Easy to consume

Integrated in the business process

Analyze on mass data

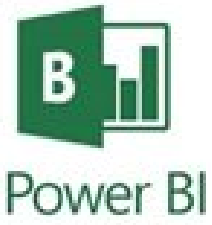


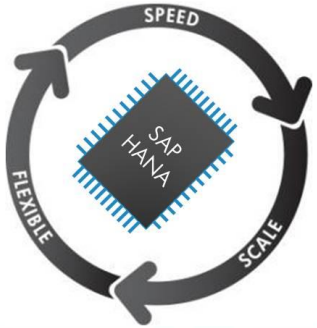
# A possible Factory Architecture



# Conclusion

1. Identify field!
2. Identify goals!
3. Find solution!
4. Provision solution:



Deutsche
SQL Server Konferenz 2014
Community powered product knowledge by PASS Deutschland & Microsoft

PASS

# Vielen Dank!

Gerhard Brückl
gerhard@gbrueckl.at
email