Daimler TSS

A company of Daimler AG

LECTURE @DHBW: DATA WAREHOUSE 02 STANDARD DWH ARCHITECTURE

ANDREAS BUCKENHOFER, DAIMLER TSS

DHBW
Duale Hochschule
Baden-Württemberg

ABOUT ME

Andreas Buckenhofer
Senior DB Professional
andreas.buckenhofer@daimler.com

Since 2009 at Daimler TSS
Department: Big Data
Business Unit: Analytics

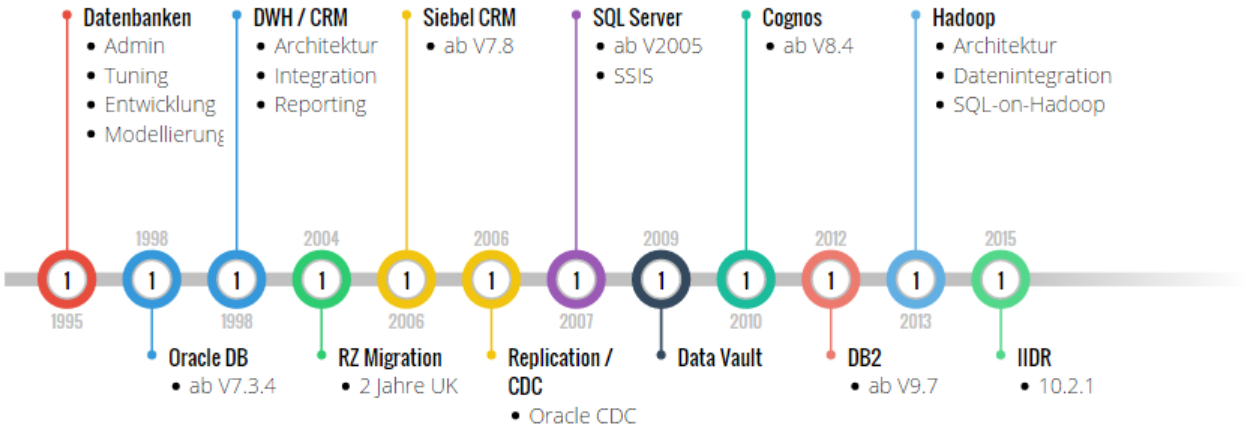
https://de.linkedin.com/in/buckenhofer

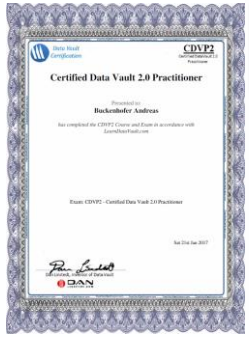
https://twitter.com/ABuckenhofer

https://www.doag.org/de/themen/datenbank/in-memory/

http://wwwlehre.dhbw-stuttgart.de/~buckenhofer/

https://www.xing.com/profile/Andreas_Buckenhofer2





ANDREAS BUCKENHOFER, DAIMLER TSS GMBH

“Forming good abstractions and avoiding complexity is an essential part of a successful data architecture”

Data has always been my main focus during my long-time occupation in the area of data integration. I work for **Daimler TSS** as Database Professional and Data Architect with over 20 years of experience in Data Warehouse projects. I am working with Hadoop and NoSQL since 2013. I keep my knowledge up-to-date - and I learn new things, experiment, and program every day.

I share my knowledge in internal presentations or as a speaker at international conferences. I'm regularly giving a full lecture on Data Warehousing and a seminar on modern data architectures at Baden-Wuerttemberg Cooperative State University DHBW. I also gained international experience through a two-year project in Greater London and several business trips to Asia.

I’m responsible for In-Memory DB Computing at the independent German Oracle User Group (DOAG) and was honored by Oracle as ACE Associate. I hold current certifications such as "Certified Data Vault 2.0 Practitioner (CDVP2)", "Big Data Architect“, „Oracle Database 12c Administrator Certified Professional“, “IBM InfoSphere Change Data Capture Technical Professional”, etc.

NOT JUST AVERAGE: OUTSTANDING.

As a 100% Daimler subsidiary, we give 100 percent, always and never less. We love IT and pull out all the stops to aid Daimler's development with our expertise on its journey into the future.

Our objective: We make Daimler the most innovative and digital mobility company.

INTERNAL IT PARTNER FOR DAIMLER

+ Holistic solutions according to the Daimler guidelines
  + IT strategy
  + Security
  + Architecture
+ Developing and securing know-how
+ TSS is a partner who can be trusted with sensitive data

As subsidiary: **maximum added value** for Daimler

+ Market closeness
+ Independence
+ Flexibility (short decision making process, ability to react quickly)

LOCATIONS

Daimler TSS Germany

7 locations

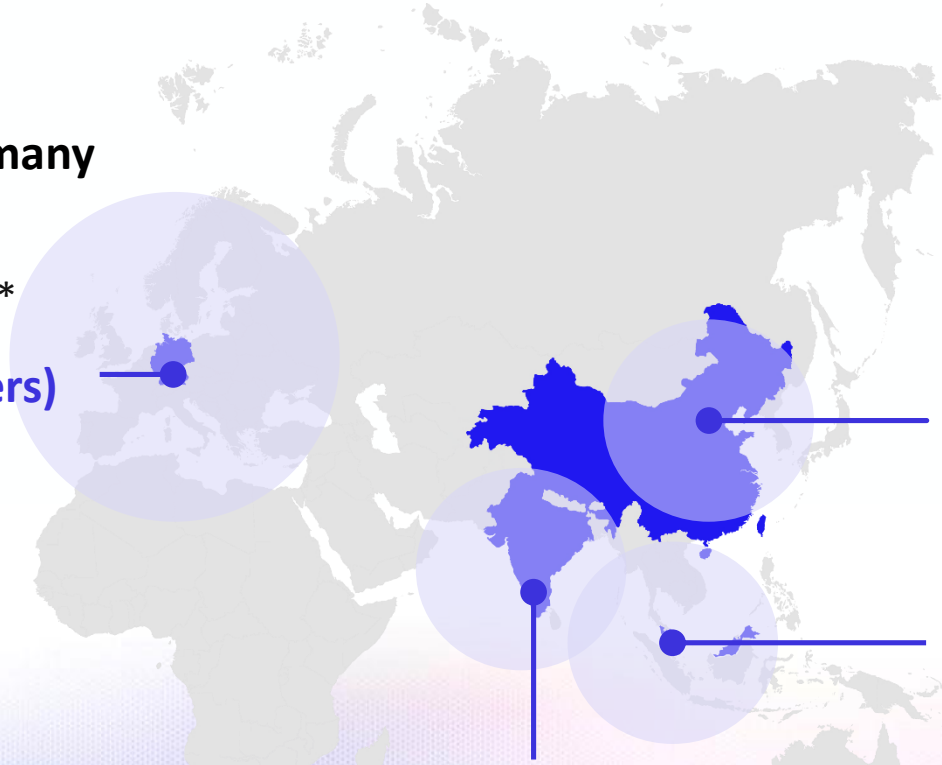
1000 employees*

Ulm (Headquarters)

Stuttgart

Berlin

Karlsruhe



Daimler TSS China
Hub Beijing
10 employees

Daimler TSS Malaysia
Hub Kuala Lumpur
42 employees

Daimler TSS India
Hub Bangalore
22 employees

* as of August 2017

DWH LECTURE - LEARNING TARGETS

- Describe different DWH architectures
- Explain DWH data modeling methods and design logical models
- Name DB techniques that are well-suited for DWHs
- Explain ETL processes
- Specify reporting & project management & meta data requirements
- Name current DWH trends

PURPOSE: WHY ARE DWH ARCHITECTURES USEFUL?

- Specific implementation can follow an architecture
  - Architecture describes an ideal type. Therefore an implementation may not use all components or can combine components
- Better understanding, overview and complexity reduction by decomposing a DWH into its components
  - Can be used in many projects: repeatable, standardizable
- Map DWH tools into the different components and compare functionality
- Functional oriented as it describes data and control flow

EXAMPLES OF DATA WAREHOUSES IN THE INDUSTRY

Apple: multiple Petabytes

- Customer insights: who's who and what are the customers up to

Walmart: 300TB (2003), several PB today

- It tells suppliers, "You have three feet of shelf space. Optimize it."

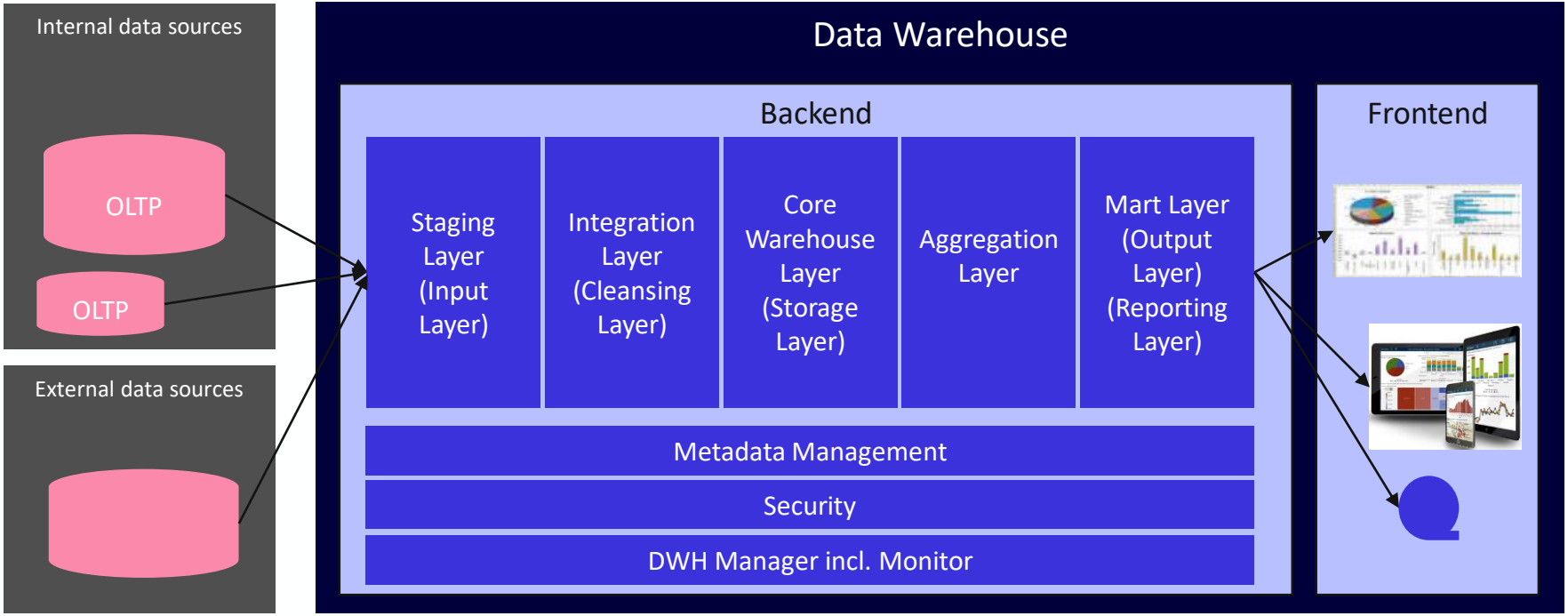
eBay: >10PB, 100s of production DBs fed in

- Get better understanding of customers

Most DWHs are much smaller though. For huge and small DWHs: High challenges to **architect** + develop + maintain + run such complex systems

https://gigaom.com/2013/03/27/why-apple-ebay-and-walmart-have-some-of-the-biggest-data-warehouses-youve-ever-seen/ and http://www.dbms2.com/2009/04/30/ebays-two-enormous-data-warehouses/

LOGICAL STANDARD DATA WAREHOUSE ARCHITECTURE



DATA SOURCES

- Providing internal and external data out of the source systems
- Enabling data through Push (source is generating extracts) or Pull (BI Data Backend is requesting or directly accessing data)
  - Example for Push practice (deliver csv or text data through file interface; Change Data Capture (CDC))
  - Example for Pull practice (direct access to the source system via ODBC, JDBC, API and so on)

STAGING LAYER

- “Landing Zone” for data coming into a DWH
- Purpose is to increase speed into DWH and decouple source and target system (repeating extraction run, additional delivery)
- Granular data (no pre-aggregation or filtering in the Data Source Layer, i.e. the source system)
- Usually not persistent, therefore regular housekeeping is necessary (for instance delete data in this layer that is few days/weeks old or – more common - if a correct upload to Core Warehouse Layer is ensured)
- Tables have no referential integrity constraints, columns often varchar

INTEGRATION LAYER

- Business Rules, harmonization and standardization of data
- Classical Layer for transformations: ETL = Extract – TRANSFORM – Load
- Fixing data quality issues
- Usually not persistent, therefore regular housekeeping is necessary (for instance after a few days or weeks or at the latest once a correct upload to Core Warehouse Layer is ensured)
- The component is often not required or often not a physical part of a DB

CORE WAREHOUSE LAYER

- Data storage in an integrated, consolidated, consistent and non-redundant (normalized) data model
- Contains enterprise-wide data organized around multiple subject-areas
- Application / Reporting neutral data storage on the most detailed level of granularity (incl. historic data)
- Size of database can be several TB and can grow rapidly due to data historization
- Write-optimised layer

AGGREGATION LAYER

- Preparing data for the Data Mart Layer to the required granularity
  - E.g. Aggregating daily data to monthly summaries
  - E.g. Filtering data (just last 2 years or just data for a specific region)
- Harmonize computation of key performance indicators (measures) and additional Business Rules
- The component is often not required or often not a physical part of a DB

DATA MART LAYER

- Read-optimised layer: Data is stored in a denormalized data model for performance reasons and better end user usability/understanding
- The Data Mart Layer is providing typically aggregated data or data with less history (e.g. latest years only) in a denormalized data model
- Created through filtering or aggregating the Core Warehouse Layer
- One Mart ideally represents one subject area
- Technically the Data Mart Layer can also be a part of an Analytical Frontend product (such as Qlik, Tableau, or IBM Cognos TM1) and need not to be stored in a relational database

METADATA MANAGEMENT, SECURITY, MONITOR

- Metadata Management
  - “Data about Data”, separate lecture
- Security
  - Not all users are allowed to see all data
  - Data security classification (e.g. restricted, confidential, secret)
- DWH Manager incl. Monitor
  - DWH Manager initiates, controls, and checks job execution
  - Monitor identifies changes/new data from source systems, separate lecture

THANK YOU

Daimler TSS

Daimler TSS GmbH
Wilhelm-Runge-Straße 11, 89081 Ulm / Telefon +49 731 505-06 / Fax +49 731 505-65 99
tss@daimler.com / Internet: www.daimler-tss.com/ Intranet-Portal-Code: @TSS
Domicile and Court of Registry: Ulm / HRB-Nr.: 3844 / Management: Christoph Röger (CEO), Steffen Bäuerle