



# Big data-Projekte: Best practices aus der Praxis

Dr. Kathrin Spreyer  
*Big Data Engineer*

JAX / Big data days  
Mainz, 14.05.2014

- 1. Projekte**
- 2. Big**
- 3. Datenimport**
- 4. Datenschutz**
- 5. Testen**
- 6. Tooling**
- 7. Agilität**

**1. Projekte**



**2. Big**

**3. Datenimport**

**4. Datenschutz**

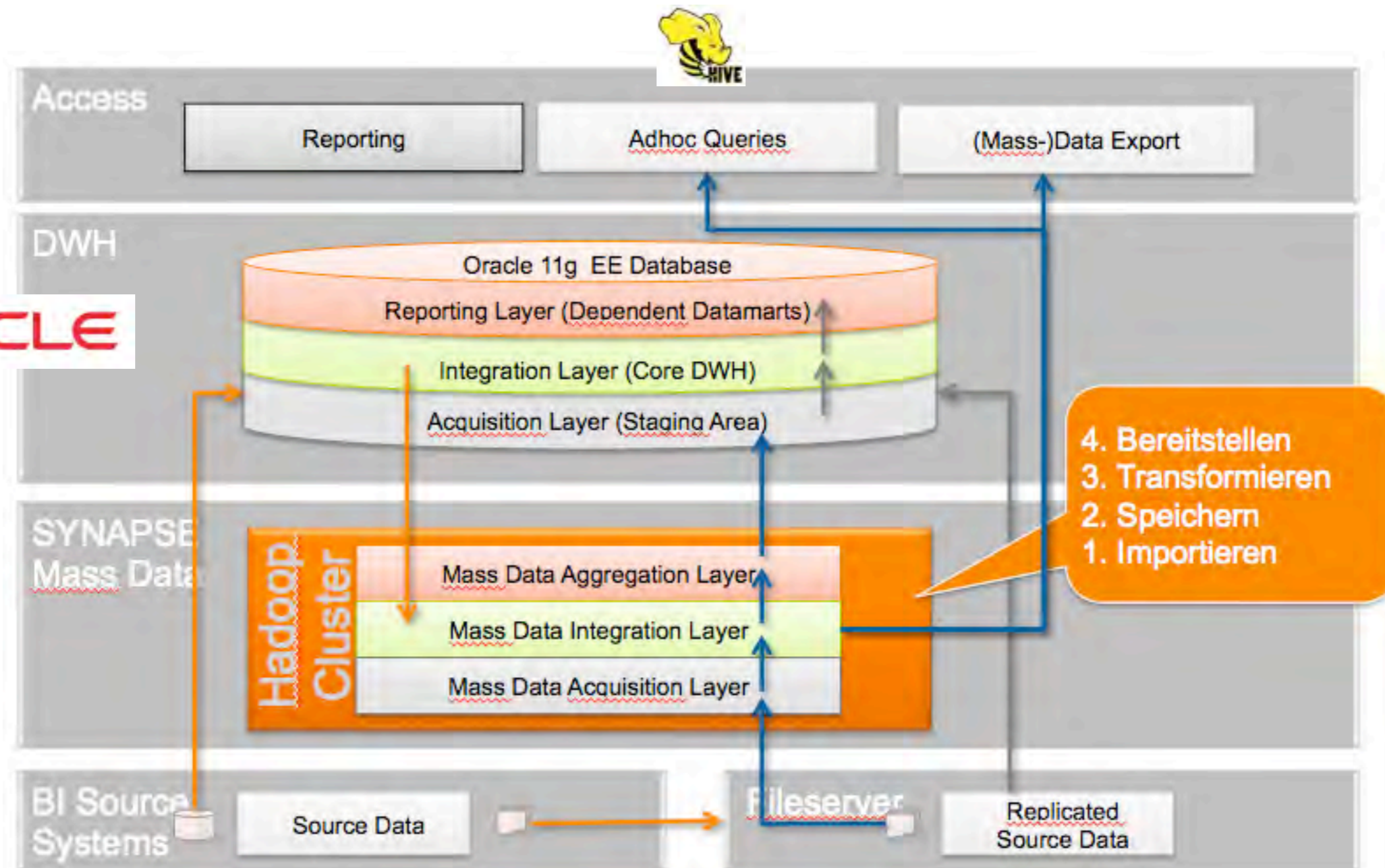
**5. Testen**

**6. Tooling**

**7. Agilität**

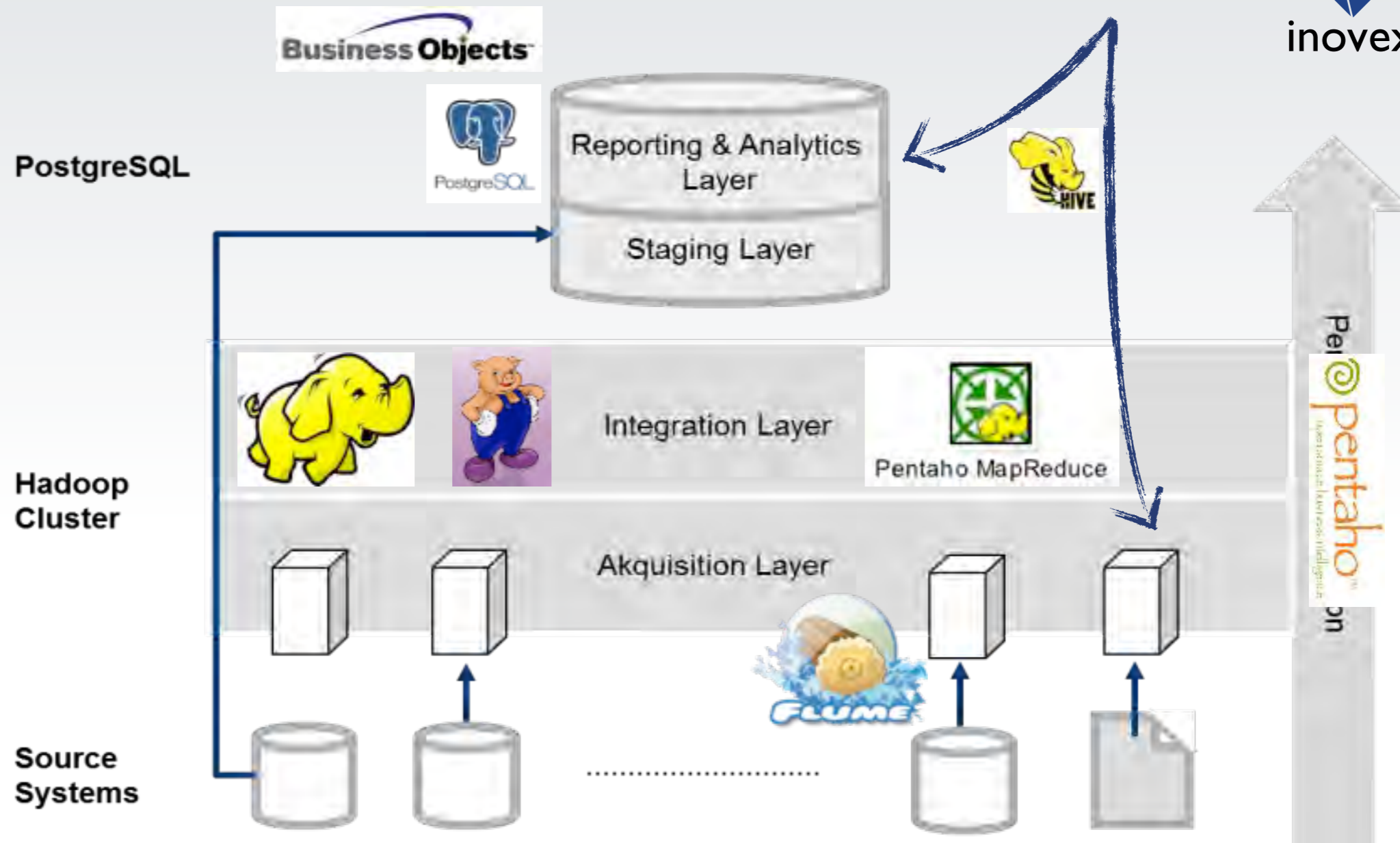
- **1&1**
  - 2009-2013
  - Web-Analytics
- **ProsiebenSat.1 Digital (PSD)**
  - 2013-heute
  - Integration von Reichweiten-,  
Vermarktungserlös- und  
Transaktionsdaten

# Architektur 1&1



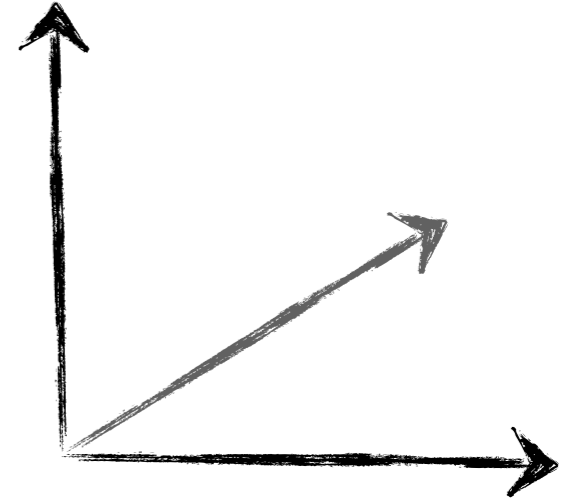
## Lösungsansatz

Hybrides System aus relationaler Datenbank und Hadoop Cluster



# Was heißt "Big?"

**3V**

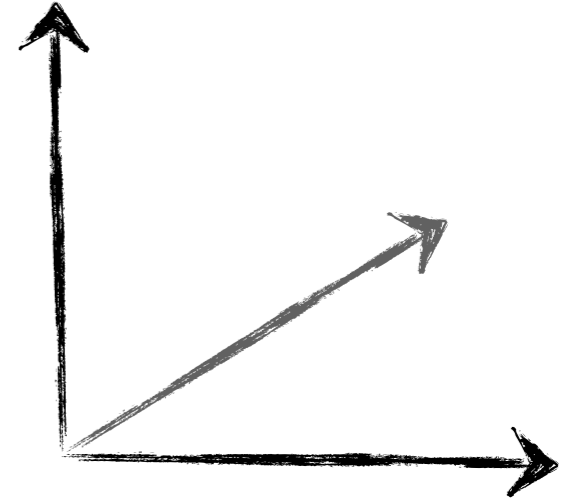


# Was heißt "Big?"

**3V**

**PARALLELISIERUNG**

**SKALIERBARKEIT**





**ZWISCHENSPEICHERUNG**

**(N)RT**



**SKALIERBARER  
TRANSPORT**

**KOORDINATION M. QUELLSYSTEMEN**

**COMPUTERSICHERHEIT**

**MULTI-TENANCY**

**COMPUTERSICHERHEIT**

**MULTI-TENANCY**

**NUTZEN VS. ANONYMITÄT**

**PII**

**ANONYMISIERUNG**

The Netflix logo, consisting of the word "NETFLIX" in a white, bold, sans-serif font with a black outline, set against a solid red rectangular background.

**NETFLIX**

**KORREKTHEIT**  
**ROBUSTHEIT**  
**PERFORMANZ**

## **REGRESSIONS-/AKZEPTANZTESTS**

**KORREKTHEIT**

**ROBUSTHEIT**

**PERFORMANZ**

## REGRESSIONS-/AKZEPTANZTESTS

KORREKTHEIT

ROBUSTHEIT

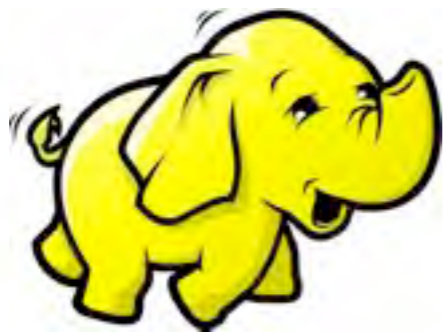
PERFORMANZ

DATENMENGE

TESTDAUER

WARTBARKEIT (FACHSEITE)

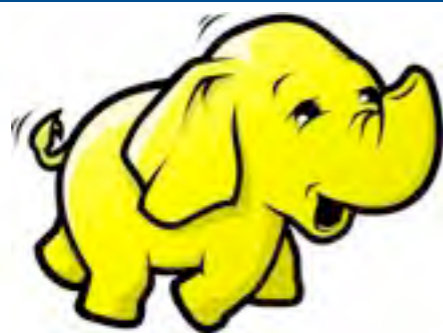
# Tooling



cascading







cascading



**DOWNSTREAM-ANFORDERUNGEN**

**DATENMODELLIERUNG**

**RT?**

**NUTZERAKZEPTANZ**

**REIFEGRAD**





## **REFACTORING MIT NEUEN TECHNOLOGIEN**

**ITERATIV**

**INKREMENTELL**

## REFACTORING MIT NEUEN TECHNOLOGIEN

ITERATIV

INKREMENTELL

DATEN-EXPLORATION

ANFORDERUNGEN ERWACHSEN AUS DATEN

**LUECKEN**

**JUNGE TECHNOLOGIEN**

1. U.a. **Hadoop-Entwickler-Training**

2. **1-3 Tage**

3. **Inhouse** oder **offen**

4. Offene Hadoop-Trainings 2014:

~~18.-20. März (Köln)~~

24.-26. Juni (München)

18.-20. November (Karlsruhe)

[www.inovex.de/trainings/offene-trainings/](http://www.inovex.de/trainings/offene-trainings/)

**Fragen?**

**Meinungen?**