

BIG DATA IM RETAIL-SEKTOR AM BEISPIEL KASSENBONDATEN

BUSINESS ANALYTICS DAY | 08.03.2017

REWE Systems GmbH | Jonas Freiknecht
inovex GmbH | Bernhard Schäfer



AGENDA

- 1 / Vorstellung REWE Systems GmbH und inovex GmbH
- 2 / Big Data @ REWE
- 3 / Bondatenverarbeitung auf Hadoop
- 4 / Beispiel Anwendung: Analytics Frontend
- 5 / Lessons Learned

KENNZAHLEN DER REWE SYSTEMS



VORSTELLUNG INOVEX GMBH




inovex

inovex ist ein IT-Projekthaus für die Digitale Transformation:

- Agile Development & Management
- Web · UI/UX · Replatforming · Microservices
- Mobile · Apps · Smart Devices · Robotics
- Big Data & Business Intelligence Platforms
- Data Science · Data Products · Search · Deep Learning
- Data Center Automation · DevOps · Cloud · Hosting
- Trainings & Coachings

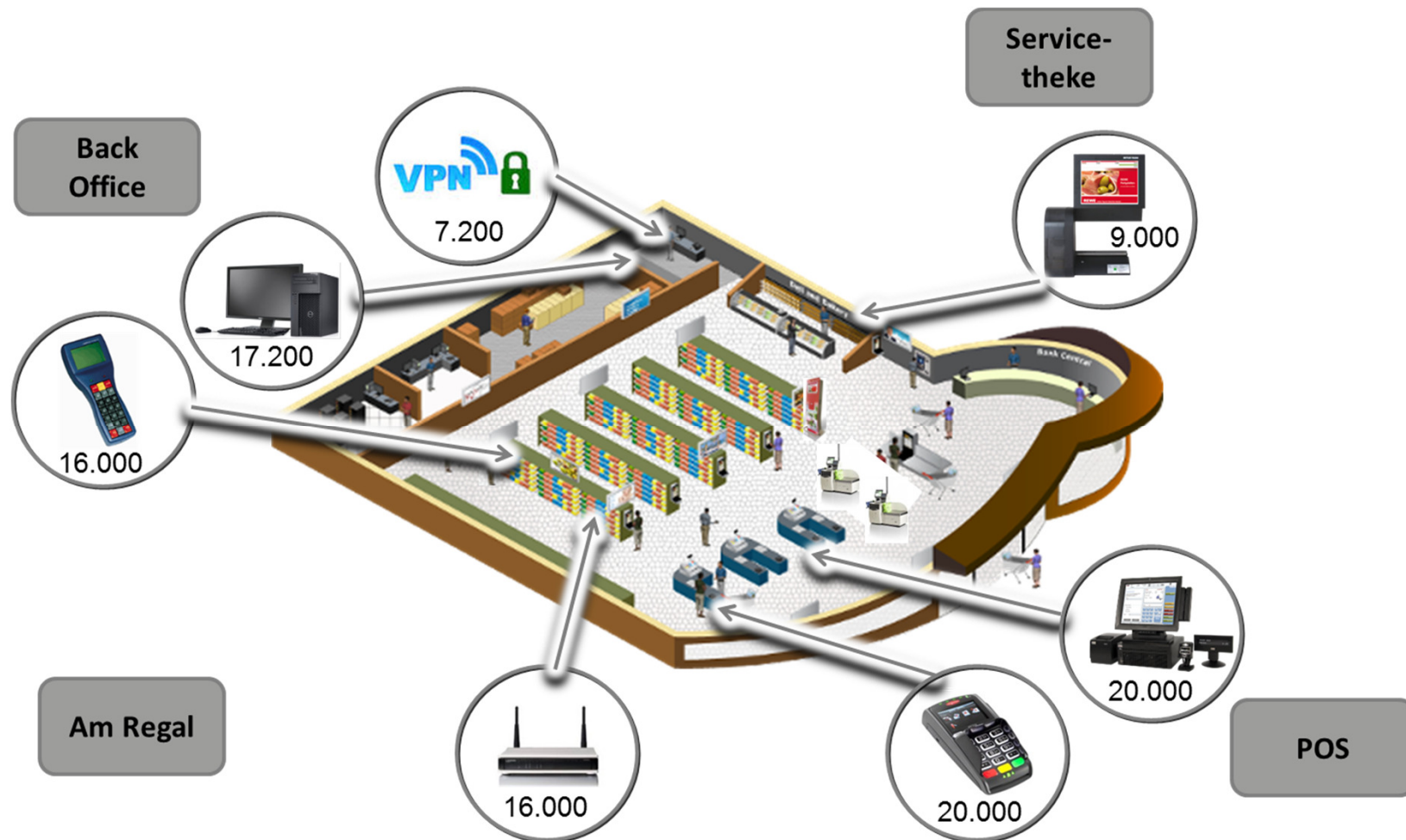
inovex gibt es in
Karlsruhe · Pforzheim ·
München · Köln · Hamburg.
Und natürlich unter
www.inovex.de

Wir nutzen Technologien,
um unsere Kunden glücklich
zu machen. *Und uns selbst.*

AGENDA

- 1 / Vorstellung REWE Systems GmbH und inovex GmbH
- 2 / **Big Data @ REWE**
- 3 / Bondatenverarbeitung auf Hadoop
- 4 / Beispiel Anwendung: Analytics Frontend
- 5 / Lessons Learned

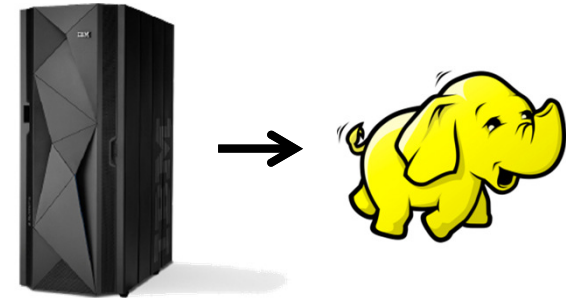
IT-EINSATZ IM MARKT 2010



BIG DATA @ REWE SYSTEMS – DAS „WARUM“

Entlastung des Mainframes

- Auslagerung rechenintensiver Jobs nach Hadoop (*ETL Offloading*)
- Vorverarbeitung der Daten in Hadoop
- Vermeidung teurer explorativer Datenanalysen auf dem Mainframe



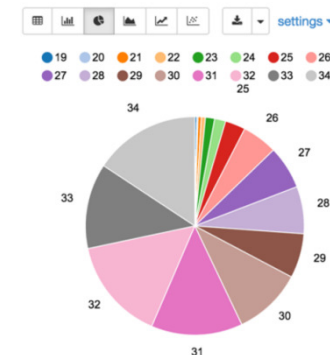
Konsolidierung der Daten und Werkzeuge

- Zusammenführen der Vielzahl an Datenquellen (*Data Lake*)
- Längere Vorhaltezeit der Rohdaten und Aggregate
- Abteilungsübergreifende Standard Tools

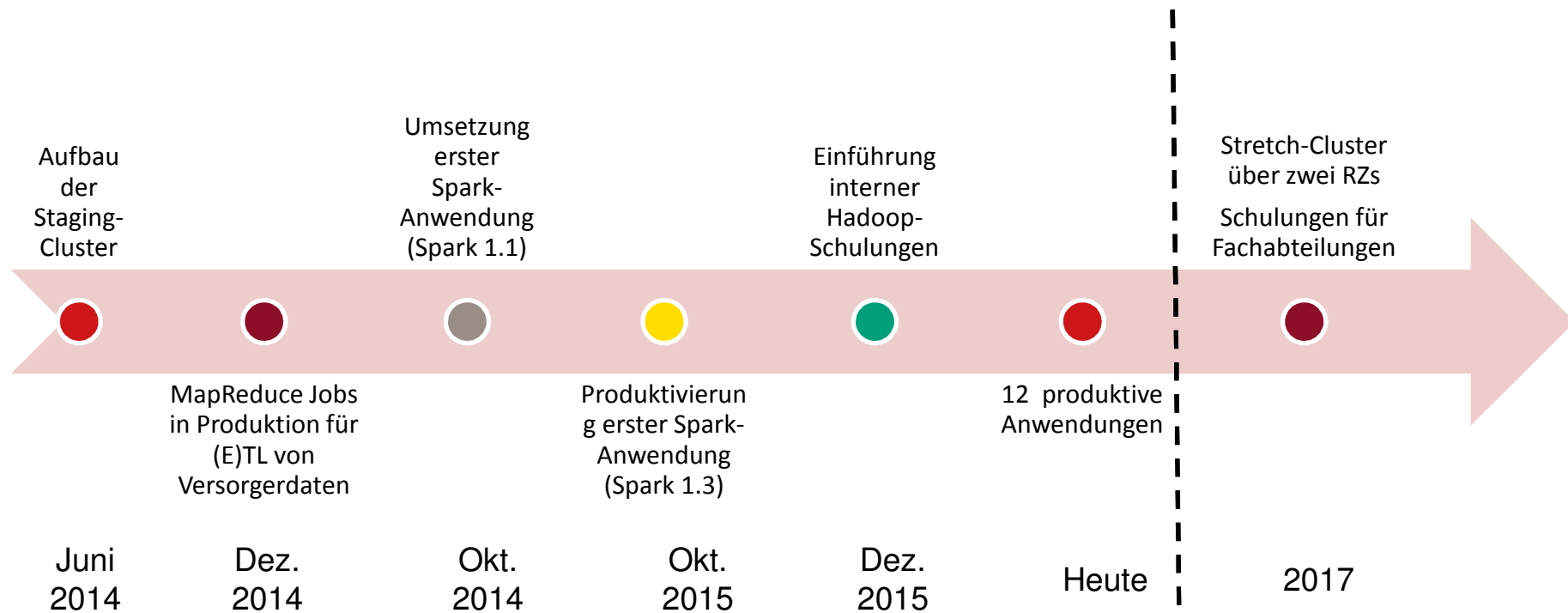


Erschließung neuer Use Cases

- Analysen über eine Vielzahl von Datenquellen (*Apache Drill*)
- Skalierbares Machine Learning (*Apache Spark*)
- Vorgefertigte Analyse-Templates für Fachanwender (*Apache Zeppelin*)



BIG DATA @ REWE SYSTEMS – HISTORIE

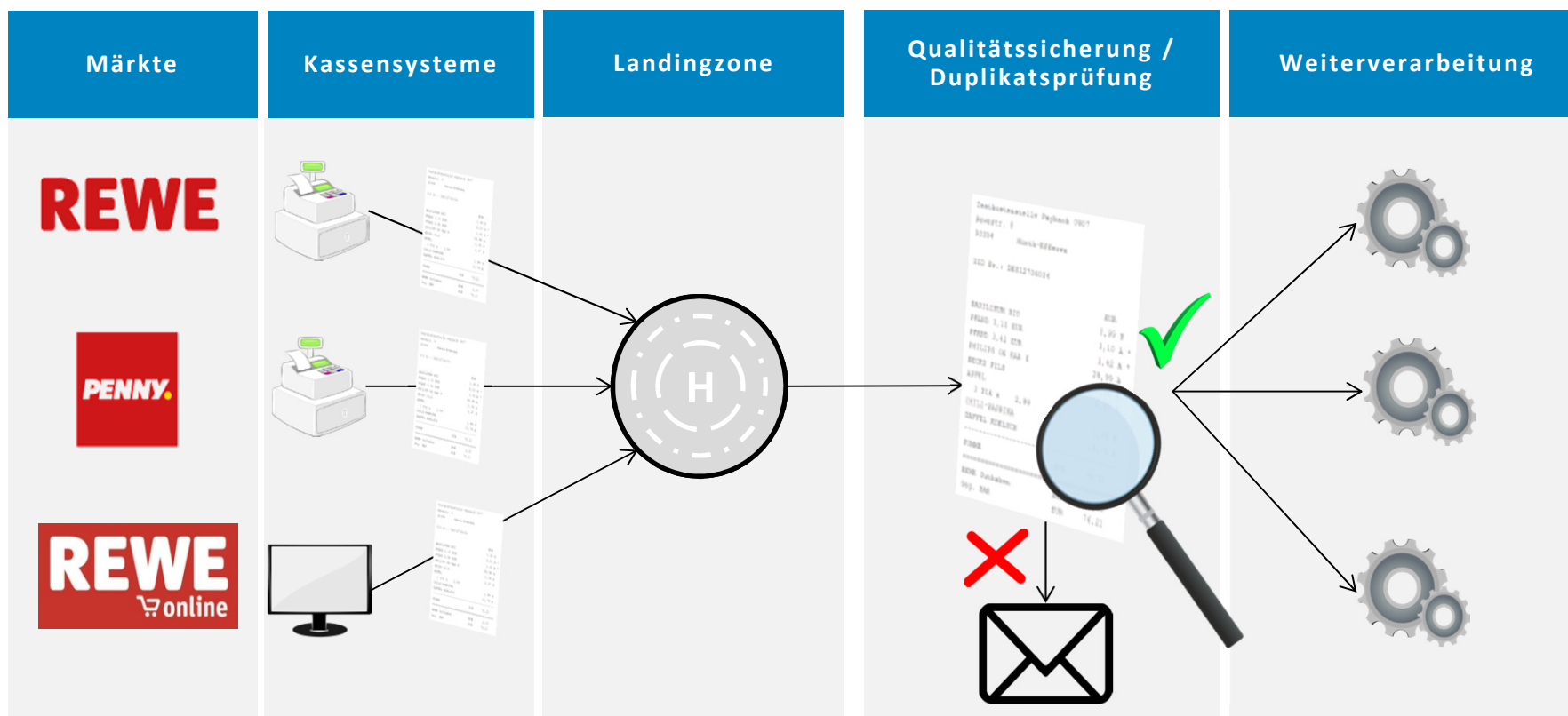


AGENDA

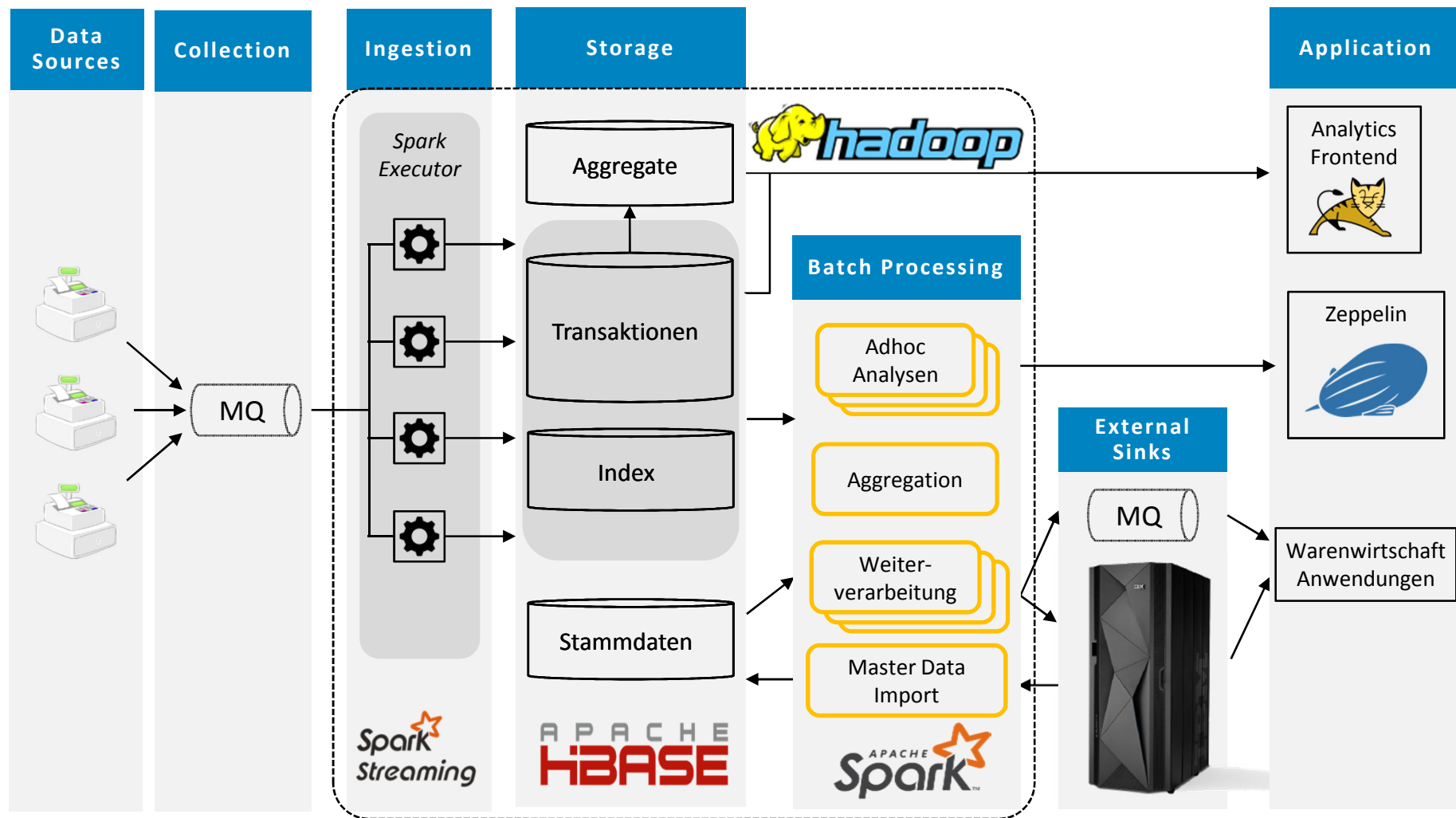
- 1 / Vorstellung REWE Systems GmbH und inovex GmbH
- 2 / Big Data @ REWE
- 3 / Bondatenverarbeitung auf Hadoop**
- 4 / Beispiel Anwendung: Analytics Frontend
- 5 / Lessons Learned

BONDATENVERARBEITUNG AUF HADOOP – HINTERGRUND

Anforderung: REWE Group-weite, qualitätsgesicherte Landing-Zone für Bondaten



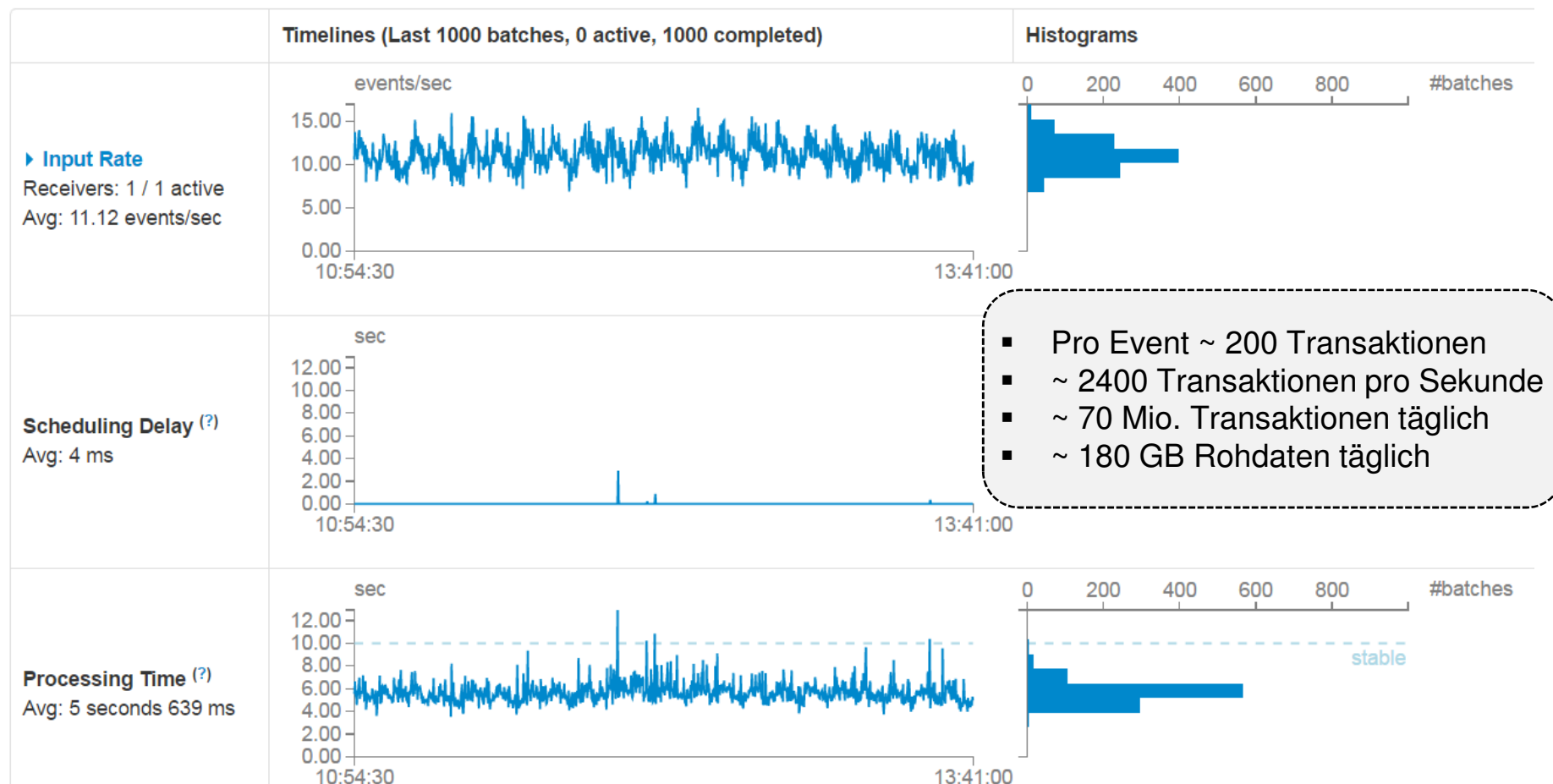
BONDATENVERARBEITUNG AUF HADOOP – ARCHITEKTUR DATA HUB



BONDATENVERARBEITUNG AUF HADOOP – SPARK STREAMING UI

Streaming Statistics

Running batches of 10 seconds for 11 weeks 5 hours since 2016/12/06 08:26:24 (667168 completed batches, 44140538 records)



AGENDA

- 1 / Vorstellung REWE Systems GmbH und inovex GmbH
- 2 / Big Data @ REWE
- 3 / Bondatenverarbeitung auf Hadoop
- 4 / Beispiel Anwendung: Analytics Frontend**
- 5 / Lessons Learned

[Folie entfernt]



[Folie entfernt]



[Folie entfernt]

[Folie entfernt]

AGENDA

- 1 / Vorstellung REWE Systems GmbH und inovex GmbH
- 2 / Big Data @ REWE
- 3 / Bondatenverarbeitung auf Hadoop
- 4 / Beispiel Anwendung: Analytics Frontend
- 5 / Lessons Learned

LESSONS LEARNED – HADOOP TECHNOLOGIEN

Spark

- Big Data *Swiss Army Knife*
- Datenquellen:
 - JDBC: DB2, ...
 - Strukturierte Dateien in HDFS: CSV, JSON, Parquet, ...
 - NoSQL Datenbanken: HBase



Spark Streaming

- Flexible, robuste und skalierbare Streaming Engine
- Für kritische 24/7 Applikationen sehr viel Detailwissen erforderlich¹

¹ <https://www.inovex.de/blog/247-spark-streaming-on-yarn-in-production/>

LESSONS LEARNED – HADOOP TECHNOLOGIEN (2)

HBase

- Skaliert für sehr große Datenmengen
- Höherer Entwicklungsaufwand im Vergleich relationalen DBs
- SQL-Zugriff über Apache Drill möglich

The logo for Apache HBase, with "APACHE" in a small, spaced-out font above "HBASE" in a larger, bold, red font.The logo for Apache Drill, featuring a stylized grey arrow pointing upwards next to the word "APACHE" in a small font above "DRILL" in a larger, blue font.The logo for Apache Maven, with the word "Maven" in a bold, black font and a stylized red and yellow feather icon to its right.

Continuous Integration

- Integrationstests mit Spark Local Mode & HBase Mini Cluster
- Einfache Integration in bestehendes Buildsystem (*Quickbuild*)

LESSONS LEARNED – BIG DATA SERVICE @ REWE

Start Simple

- Fokus auf wenige Kernkomponenten aus Hadoop Ökosystem: *Spark, HBase, Oozie, (Hive)*
- Agile Entwicklung
- *Java* anstatt *Scala/Python* → Vereinfacht die Ausbildung von Big Data Entwicklern

Standardisierung

- Bereitstellung von Java Libraries für etablierte Patterns
- Namenskonventionen für *HDFS* und *HBase*

Training

- Schulungsprogramm Java Entwickler → Big Data Entwickler
 - Vermittlung Konzepte über verteilte Systeme
 - *Java 8 Lambda Streams* ähnlich zu *Spark RDD API*



**Vielen Dank für Ihre
Aufmerksamkeit!**



**REWE
SYSTEMS**

**WE ARE
HIRING!**

<https://www.rewe-systems.com/ihre-karrierechancen/jobangebote/>