

Architecture of Big Data Solutions

Guido Schmutz
Frankfurt, 13.12.2017



@gschmutz



guidoschmutz.wordpress.com

BASEL ▪ BERN ▪ BRUGG ▪ DÜSSELDORF ▪ FRANKFURT A.M. ▪ FREIBURG I.BR. ▪ GENF
HAMBURG ▪ KOPENHAGEN ▪ LAUSANNE ▪ MÜNCHEN ▪ STUTTGART ▪ WIEN ▪ ZÜRICH

trivadis
makes IT easier. 

■ Guido Schmutz

Working at Trivadis for more than 20 years

Oracle ACE Director for Fusion Middleware and SOA



Consultant, Trainer Software Architect for Java, Oracle, SOA and Big Data / Fast Data



Head of Trivadis Architecture Board

Technology Manager @ Trivadis

More than 30 years of software development experience

Contact: guido.schmutz@trivadis.com

Blog: <http://guidoschmutz.wordpress.com>

Slideshare: <http://www.slideshare.net/gschmutz>

Twitter: [@gschmutz](https://twitter.com/gschmutz)

Architektur of Big Data Solutions

gschmutz 21:08 on April 18, 2017

Tags: flink (1), kafka (59), kafka-connect (4), kafka-streams (17), spark-streaming (53), storm (39), streams (4)

Last week in Stream Processing & Analytics – 18.4.2017

This is the 62nd edition of my blog series blog series around Stream Processing and Analytics!

Every week I'm also updating the following two lists with the presentations/videos of the current week:

- [Presentations from SlideShare](#)
- [Videos from YouTube](#)

As usual, find below the new blog articles, presentations, videos and software releases from last week:

News and Blog Posts

General

- [Multi-Master Replication For Geo-Distributed Data: It's more than you think by Ellen Friedman](#)
- [Understanding Indicators of Attack \(IOAs\): The Power of Event Stream Processing in CrossStrike Falcon by Dan Brown](#)
- [Stream processing and messaging systems for the IoT age by Ben Lorica](#)

Apache Kafka / Kafka Streams / Confluent Platform

- [Creating a Data Pipeline with Kafka Connect API – from Architecture to Operations by Alexandra Wang](#)
- [Streaming Spring Boot Application Logs to ELK Stack – Part 1 by ksadayanamthu](#)
- [Streaming Spring Boot Application Logs to Apache Kafka – ELK/Kafka Stack – Part 2 by ksadayanamthu](#)



■ Agenda

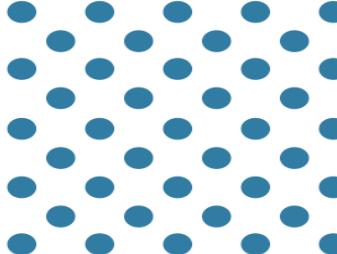
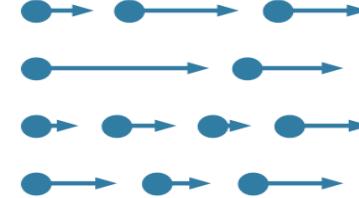
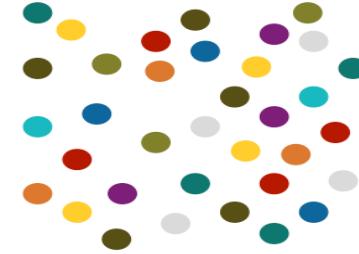
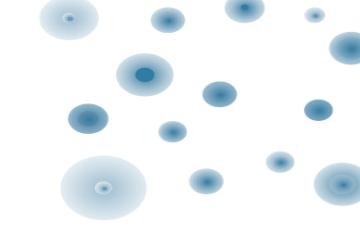
1. Introduction
2. Big Data & Fast Data Reference Architectures
3. Continuous Streaming Data Ingestion
4. Big Data & Cloud
5. Microservices Architecture
6. Big Data Ecosystem – many choices sorted!

Introduction

Architektur of Big Data Solutions

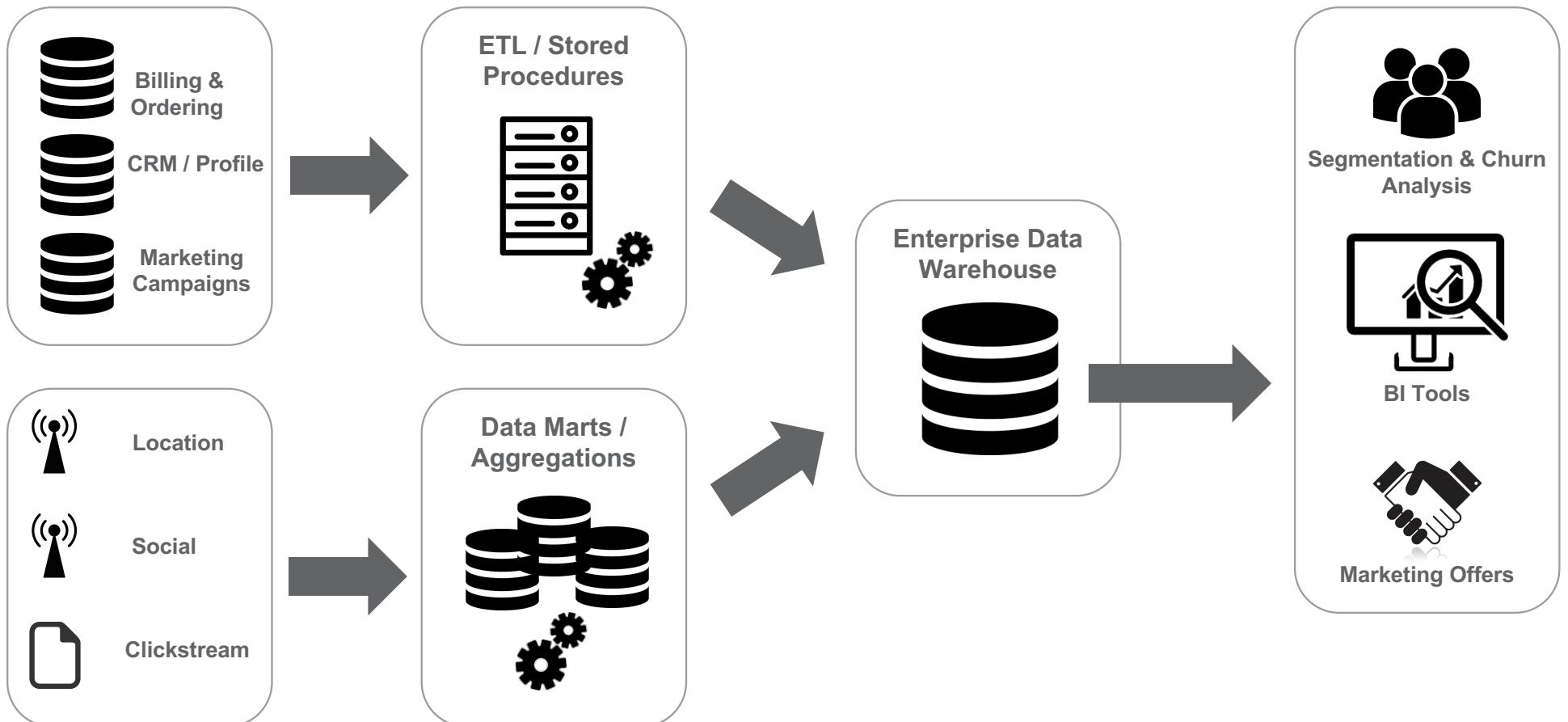


■ Big Data Definition (4 Vs) Characteristics of Big Data: Its Volume, Velocity and Variety in combination

Volume <i>Data at rest</i>	Velocity <i>Data in motion</i>	Variety <i>Data in many forms</i>	Veracity <i>Data in doubt</i>
 <ul style="list-style-type: none">▪ Tera- to Exabytes to process▪ Sensor and social data▪ New data stores	 <ul style="list-style-type: none">▪ Streaming data▪ (Milli)seconds to minutes to respond	 <ul style="list-style-type: none">▪ Structured and unstructured data▪ Text, numbers and multimedia	 <ul style="list-style-type: none">▪ Uncertainty due to data inconsistency and incompleteness, ambiguity, latency, deception and model approximations

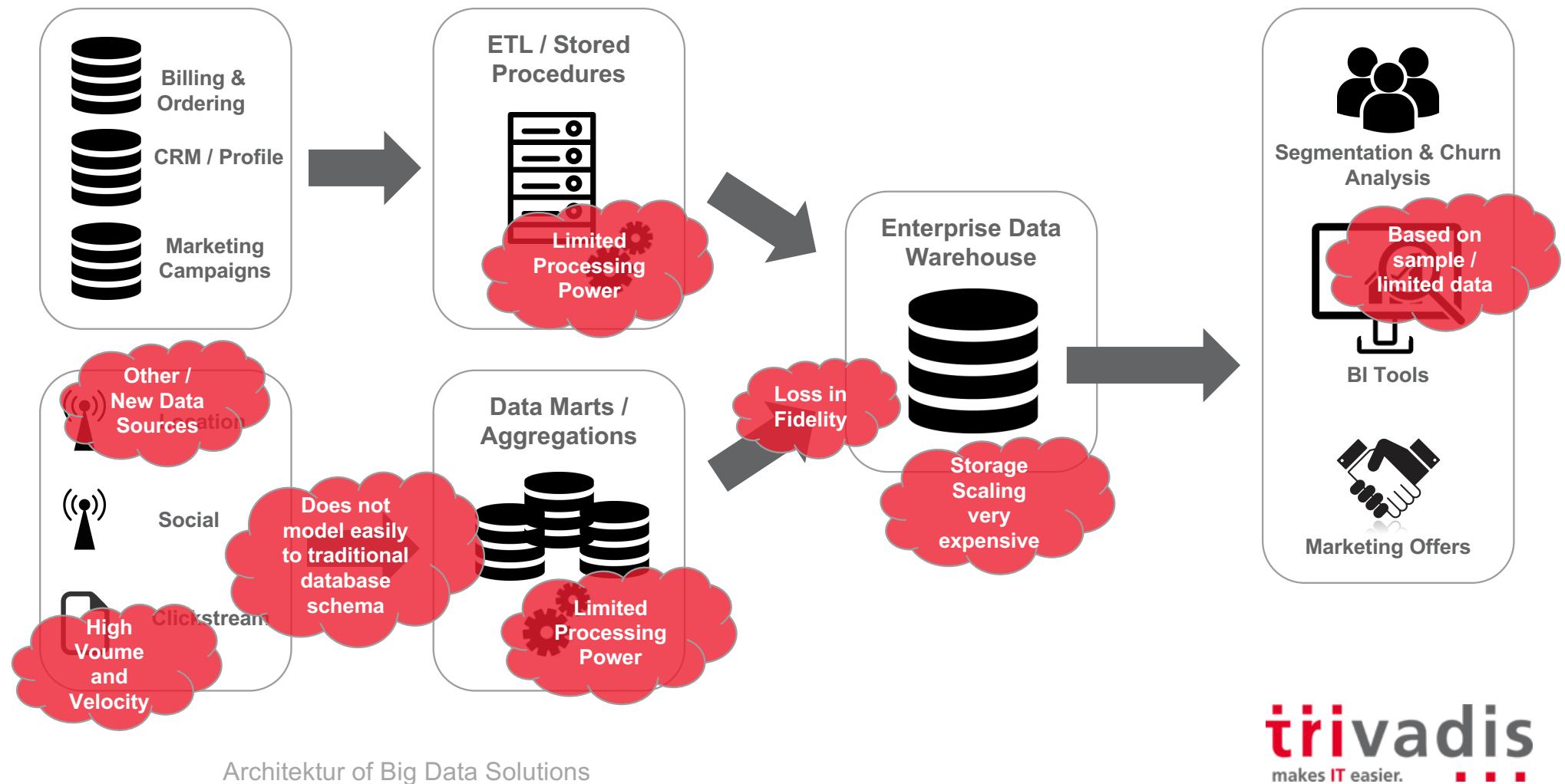
+ Time to action ? – Big Data + Real-Time = Stream Processing

■ Architektur von Big Data Lösungen



Architektur of Big Data Solutions

■ Traditional Flow Diagram - Challenges



■ Big Data to the rescue? Why is a structuring / architecture important?



■ Why talk about Big Data Architectures?

Choosing the right architecture is key for any (big data) project

Big Data is still quite a rather young field and therefore a “moving target”

no standard architectures available which have been used for years

In the past years, some **architectures and best practices have evolved**

Know your use cases before choosing your architecture / technologies

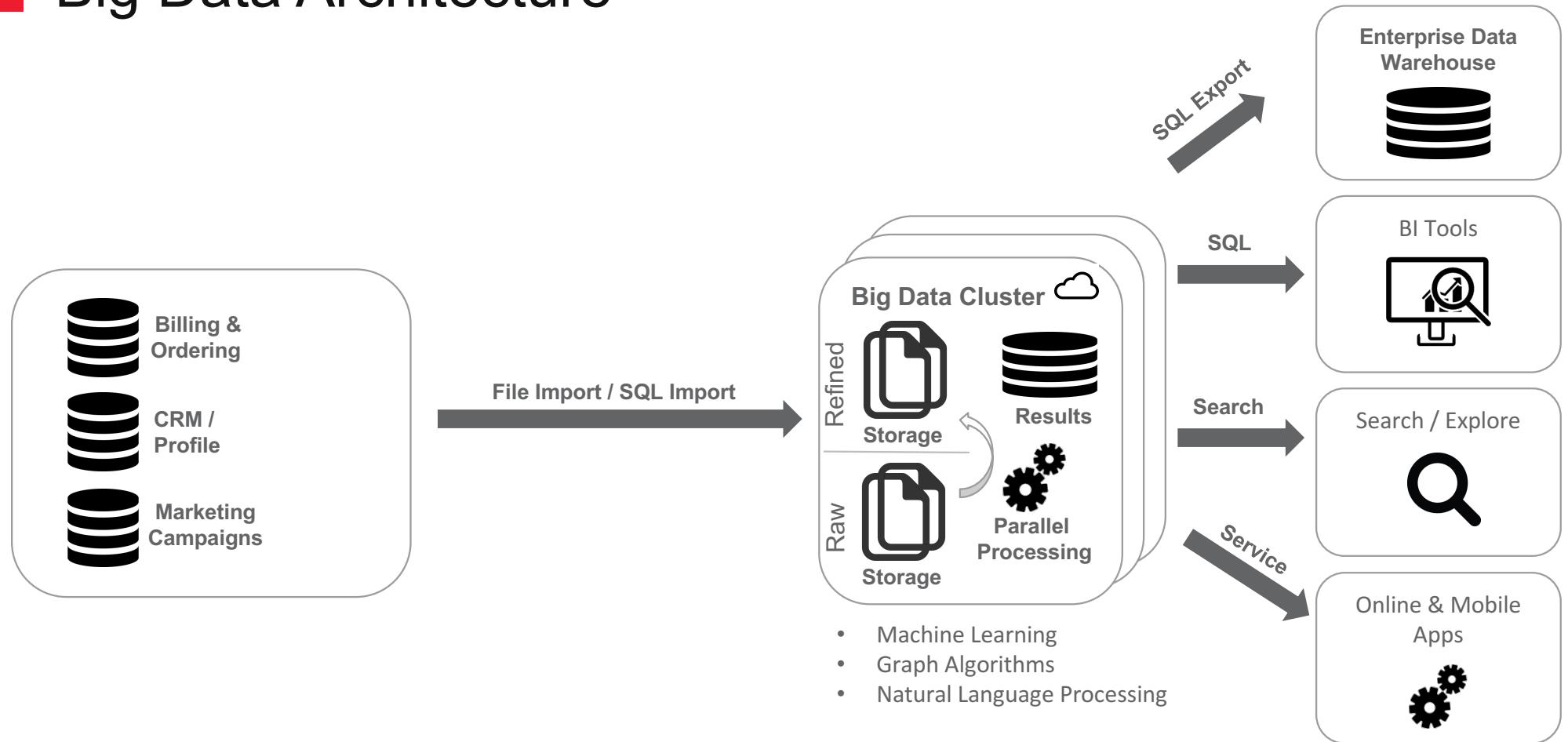
To have a **reference architecture** in place **helps in choosing the right/matching technologies**

Big Data & Fast Data Reference Architectures

Architektur of Big Data Solutions

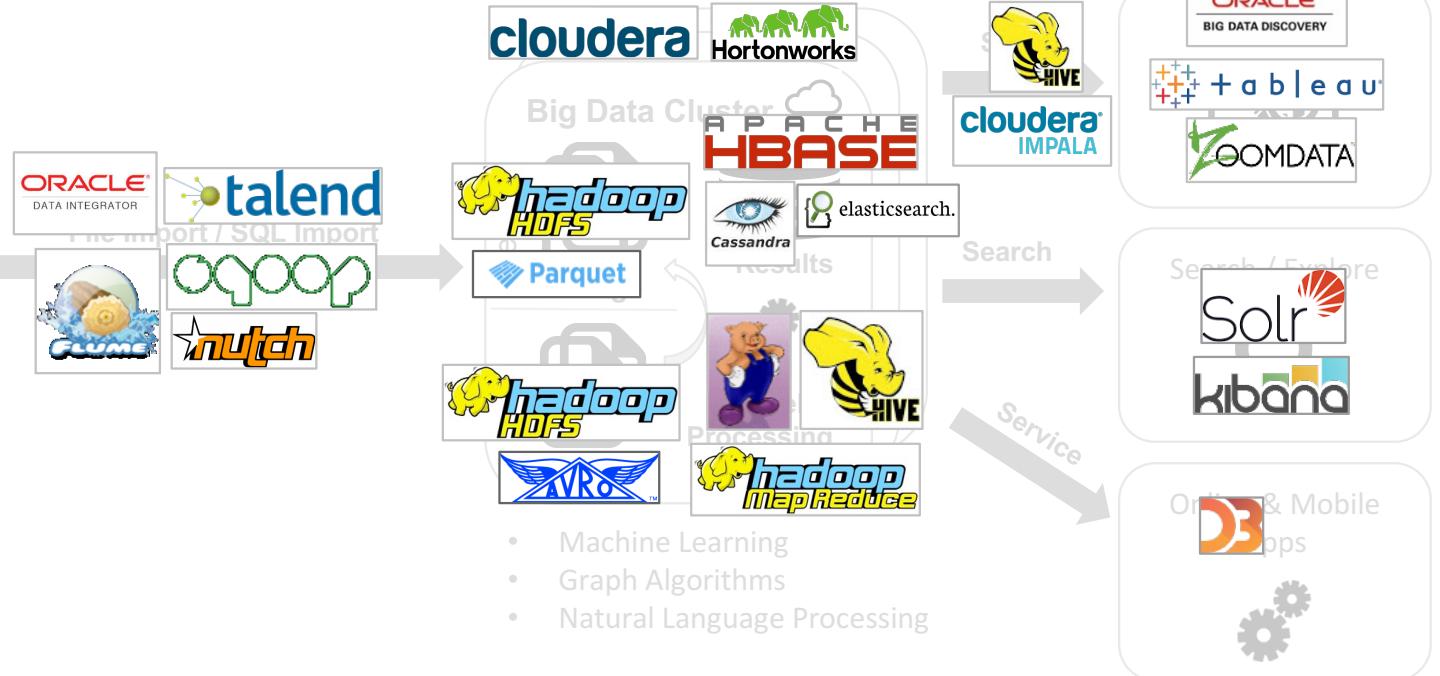
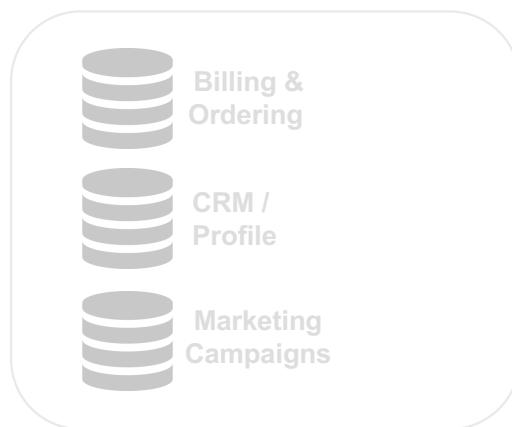
trivadis
makes **IT** easier. 

■ Big Data Architecture



Architektur of Big Data Solutions

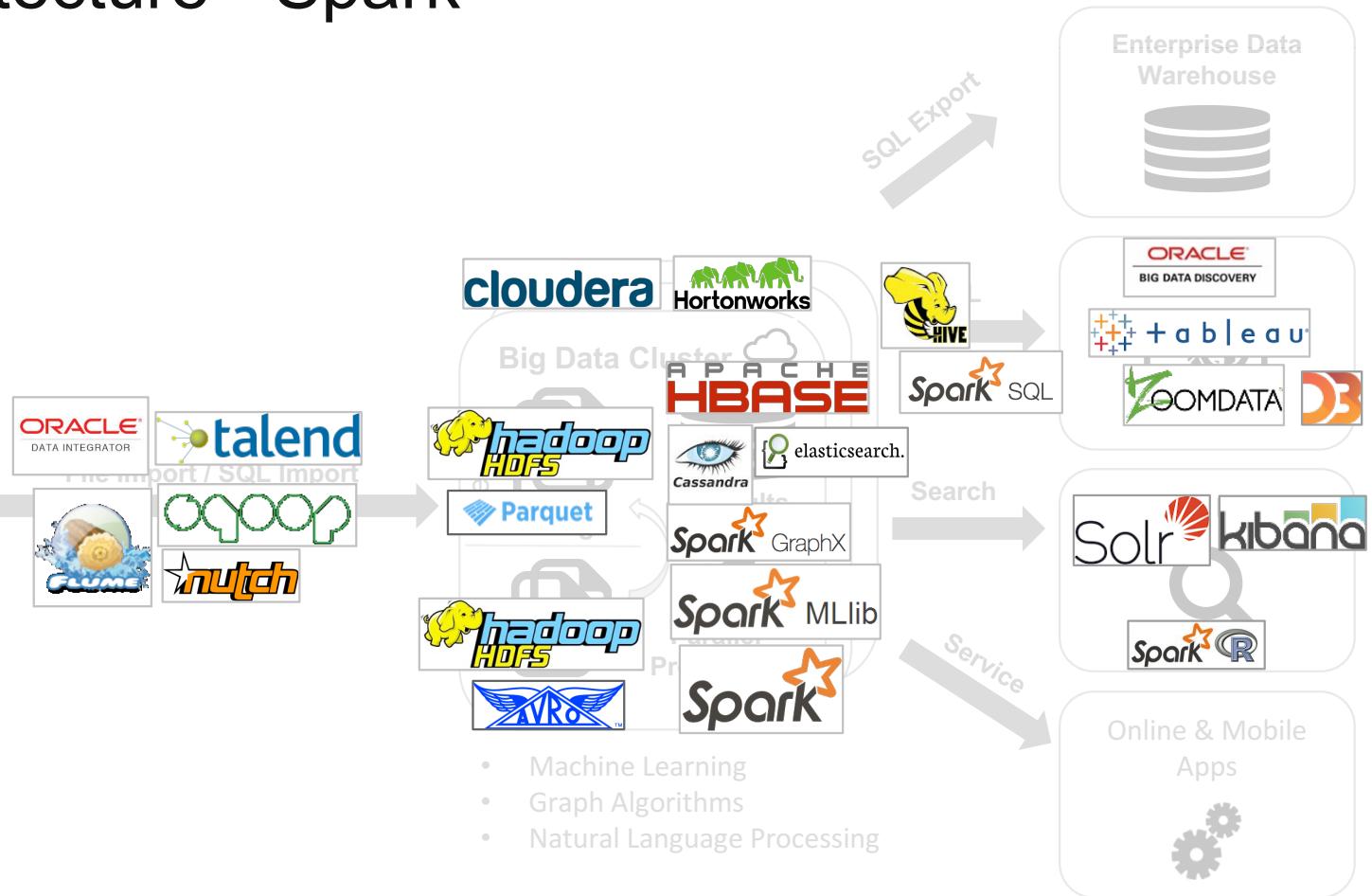
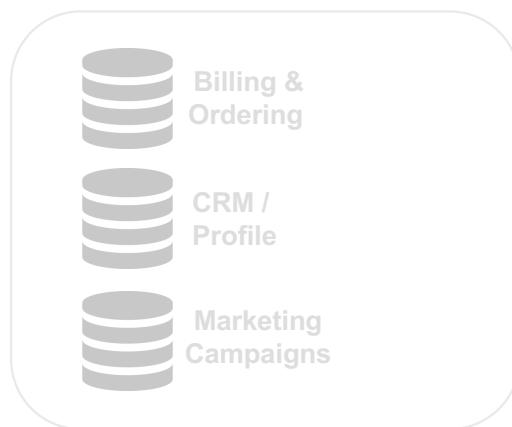
■ Big Data Architecture - Hadoop



Architektur of Big Data Solutions

trivadis
makes IT easier.

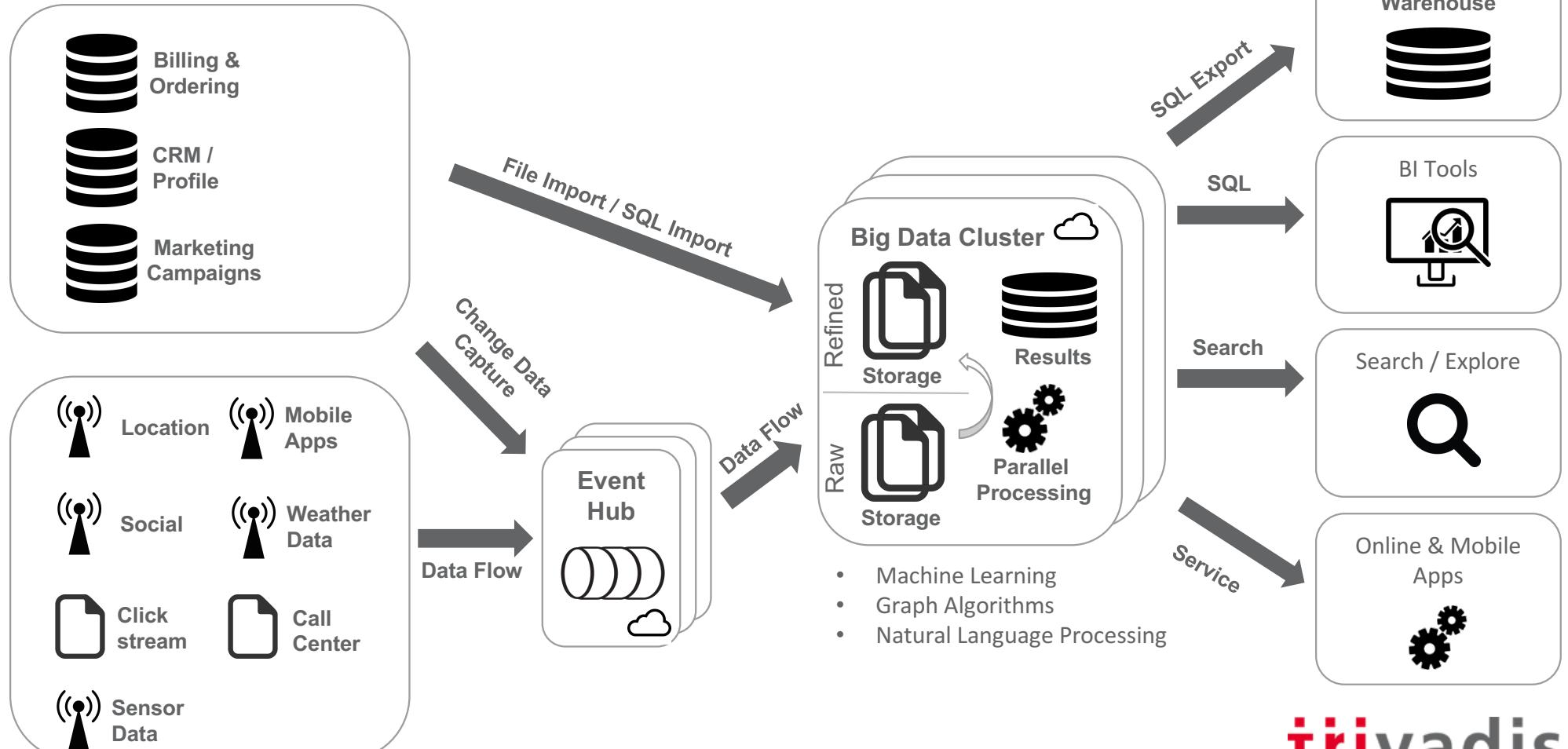
■ Big Data Architecture - Spark



Architektur of Big Data Solutions

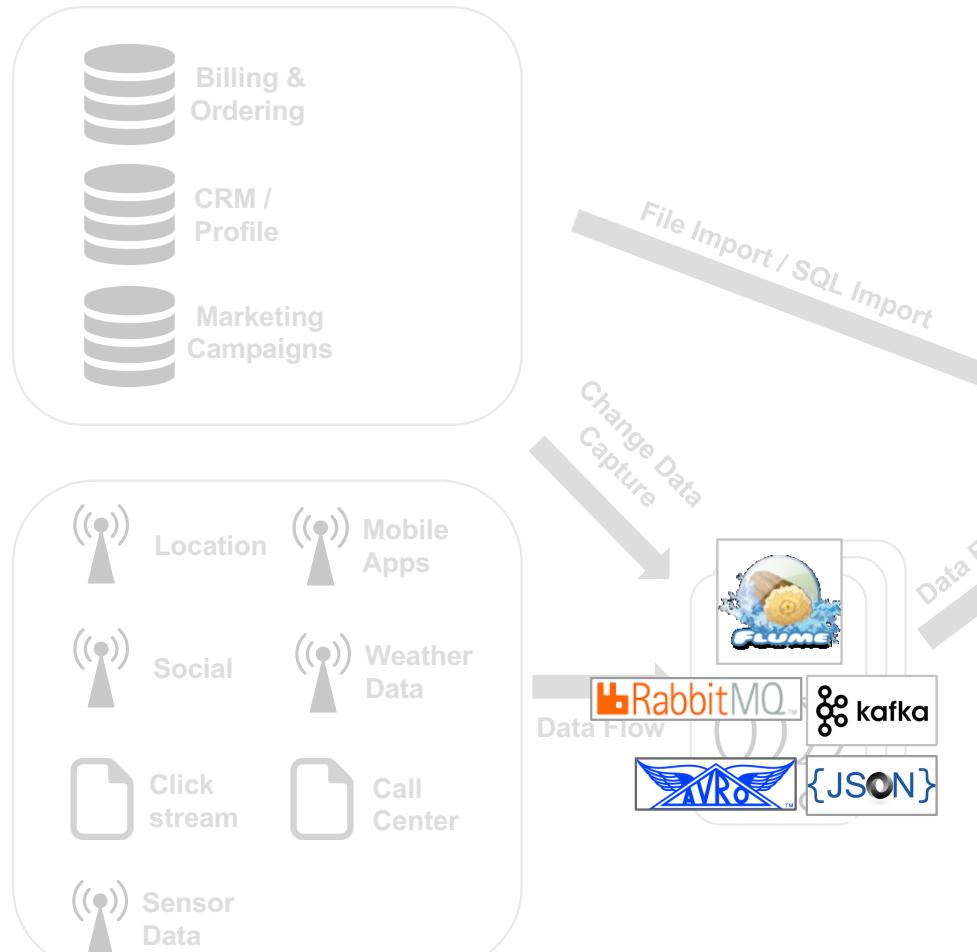
trivadis
makes IT easier. ■■■

■ Event Hub for handling streaming data

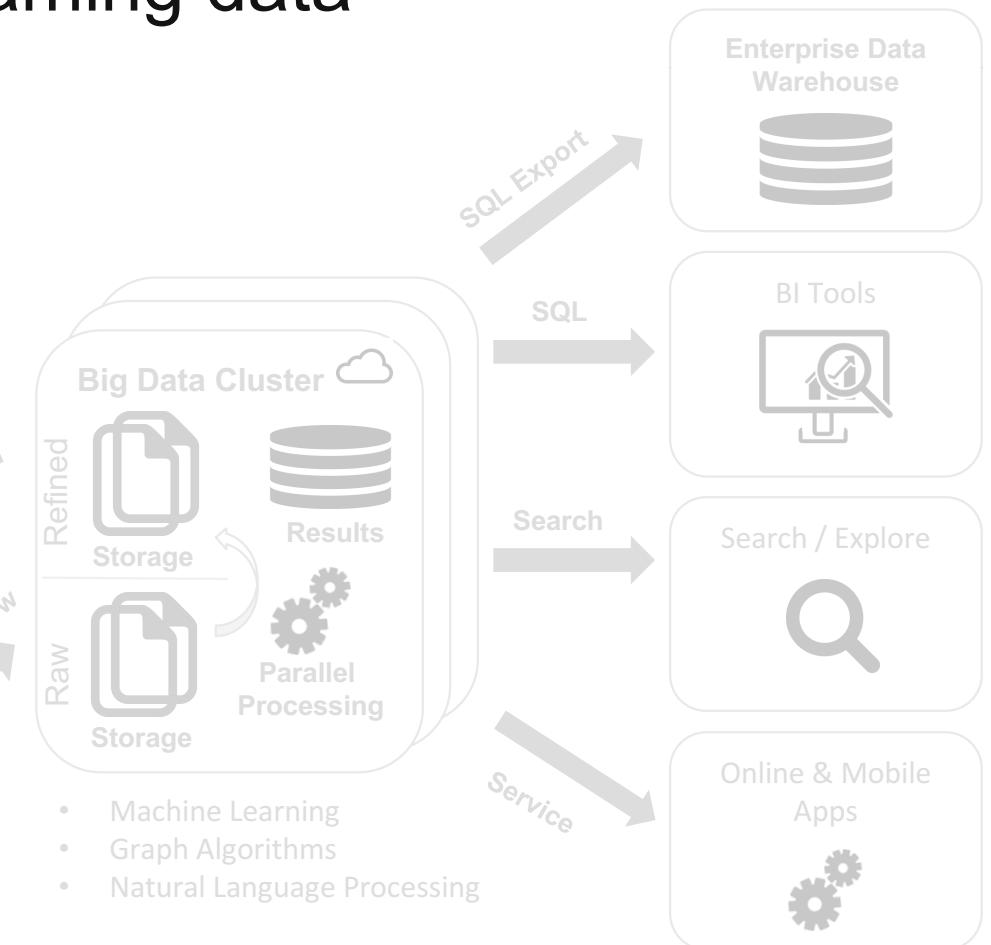


Architektur of Big Data Solutions

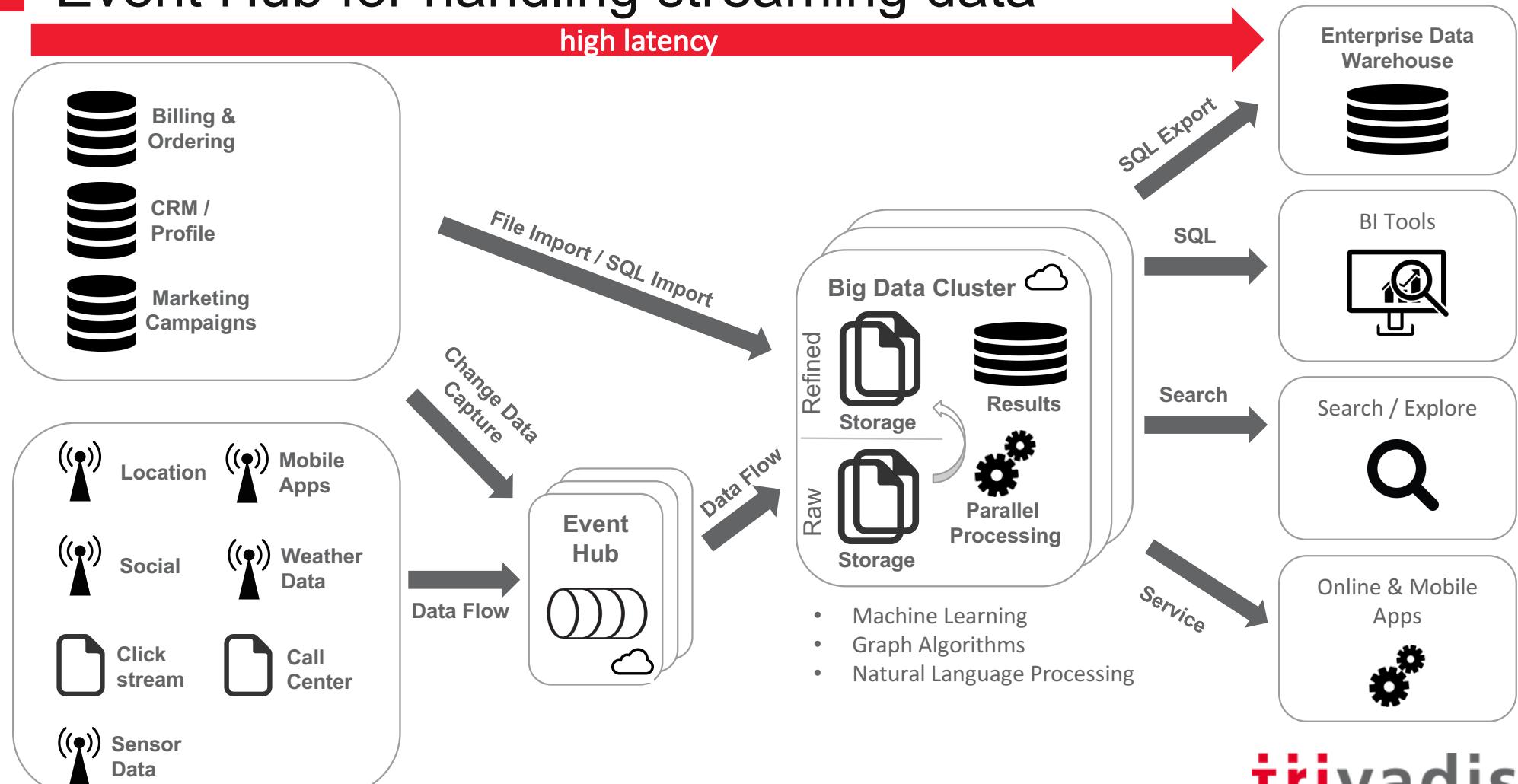
■ Event Hub for handling streaming data



Architektur of Big Data Solutions



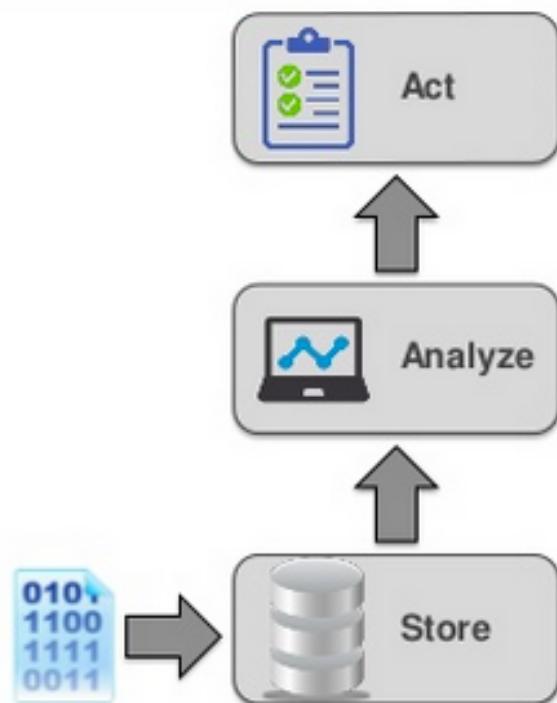
Event Hub for handling streaming data



Architektur of Big Data Solutions

■ “Data at Rest” vs. “Data in Motion”

Data at Rest

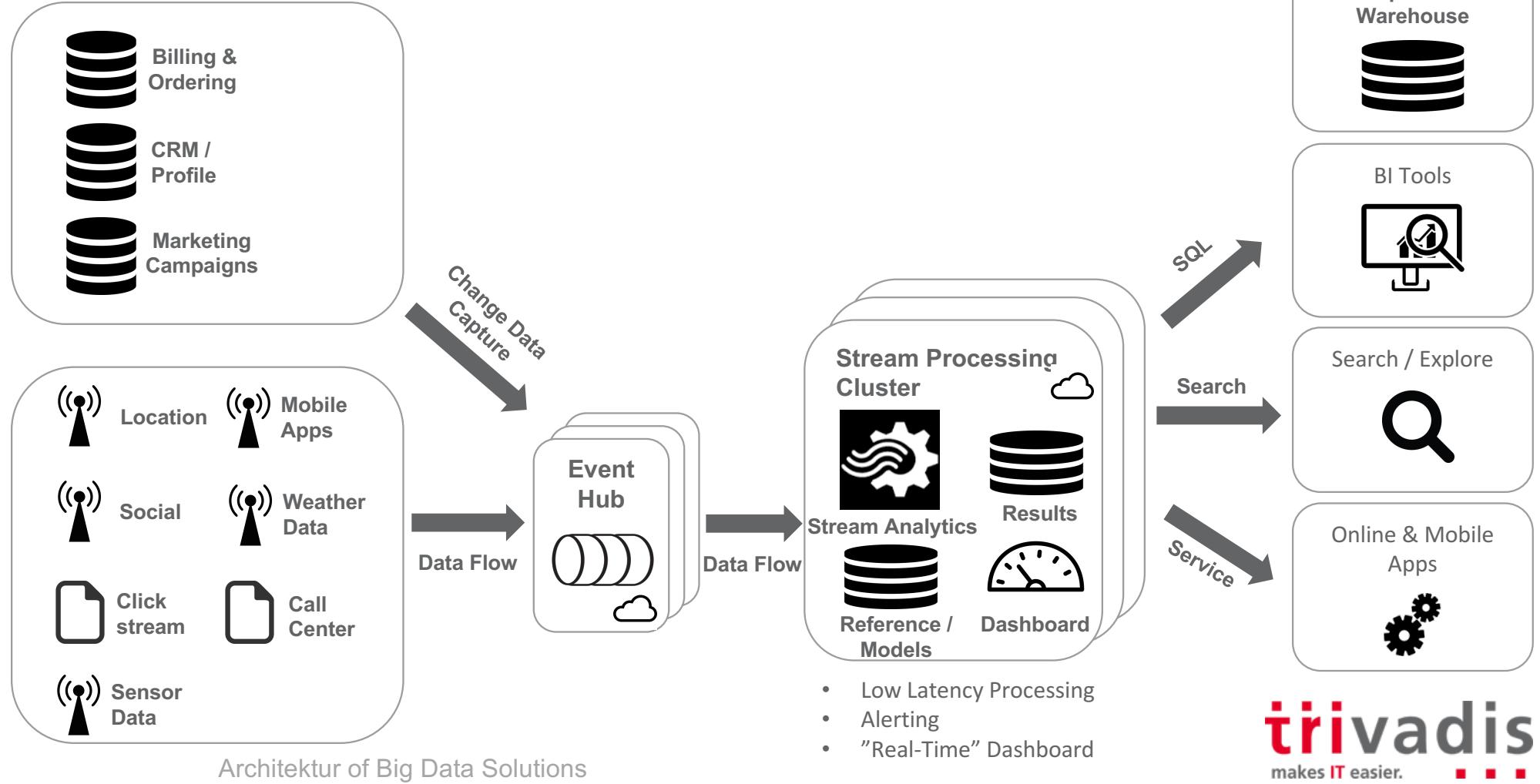


Data in Motion

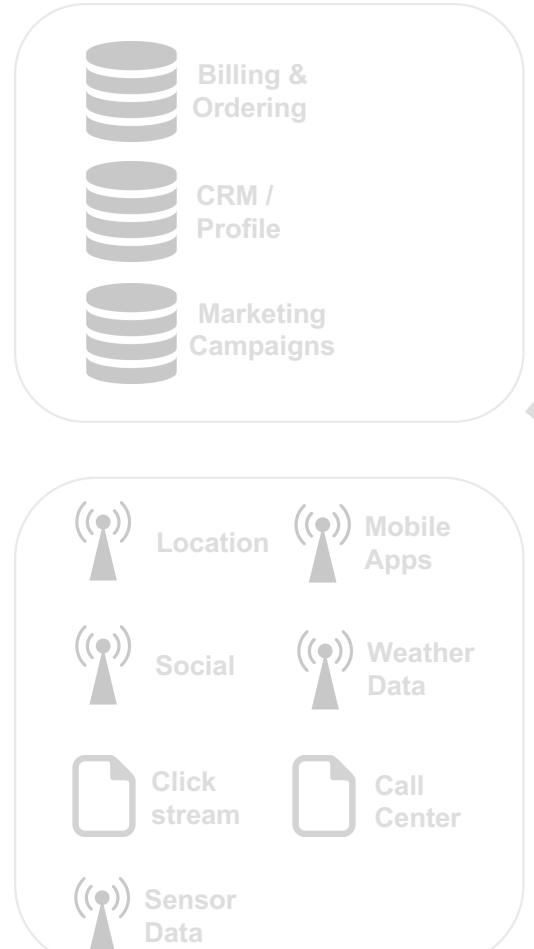


Architektur of Big Data Solutions

■ Streaming Analytics Architecture



■ Streaming Analytics Architecture – Open Source

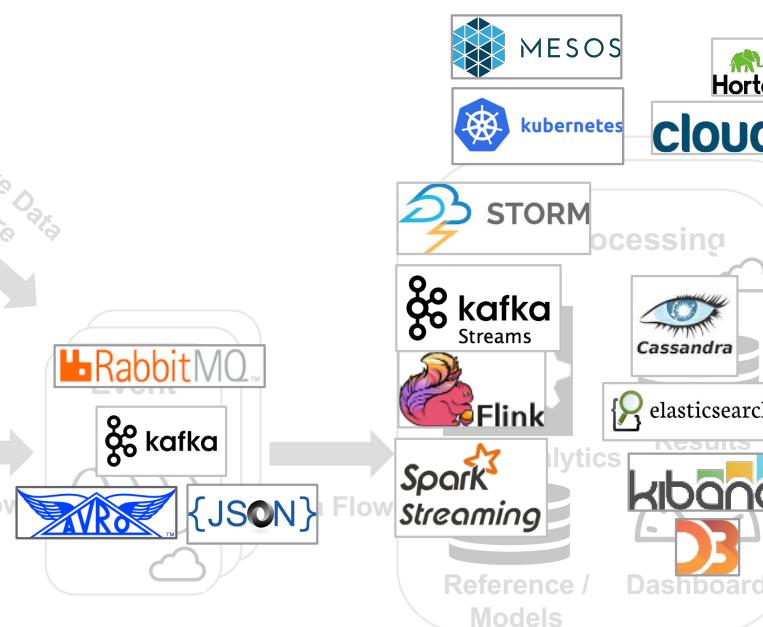


Change Data
Capture

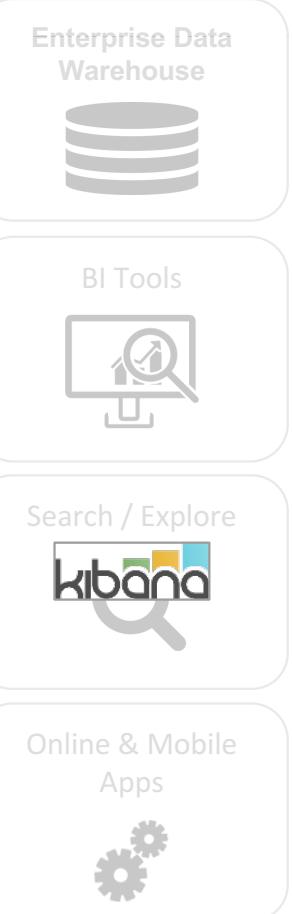
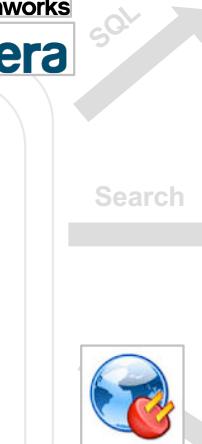
Data Flow

Architektur of Big Data Solutions

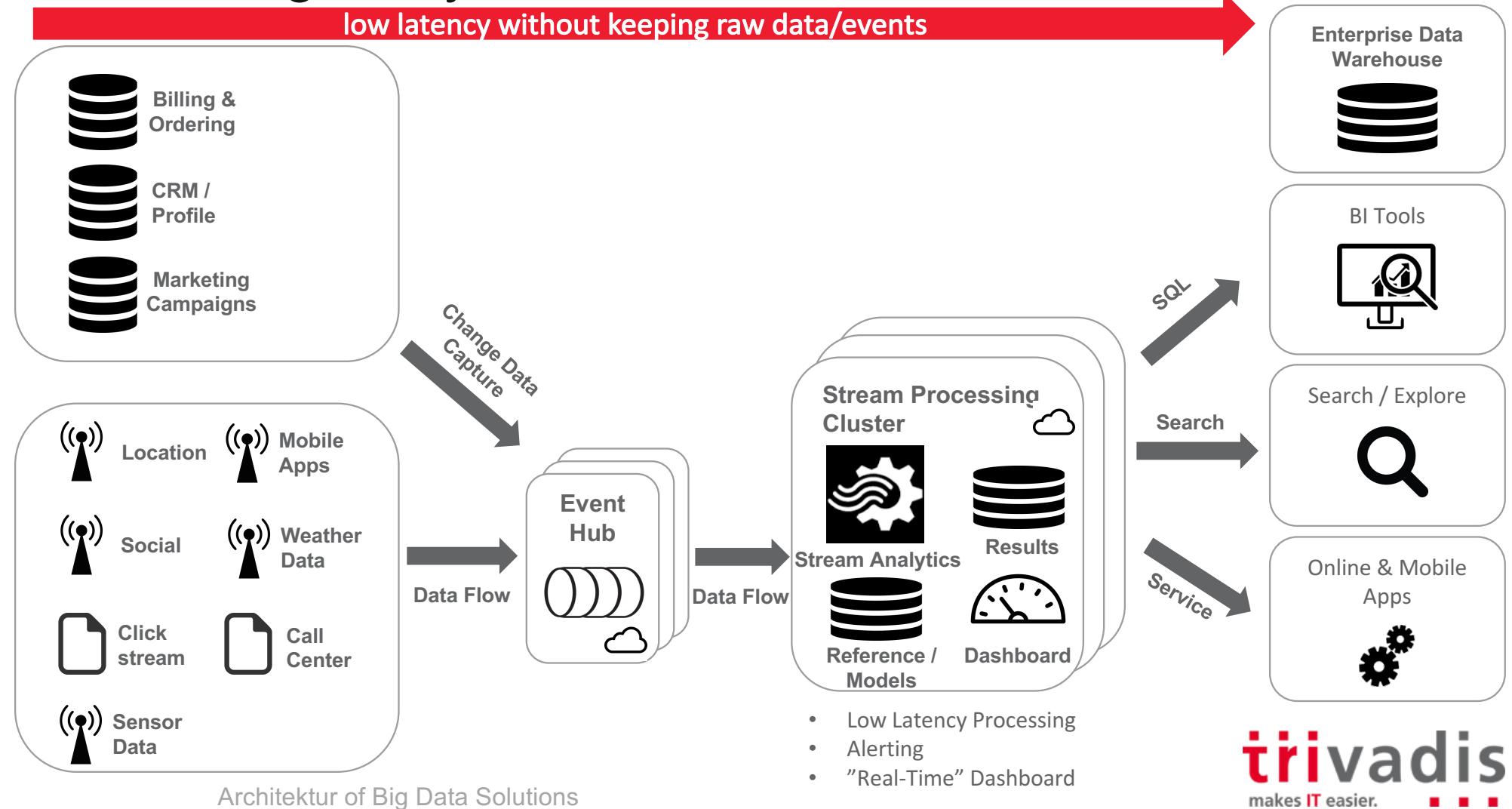
Flow



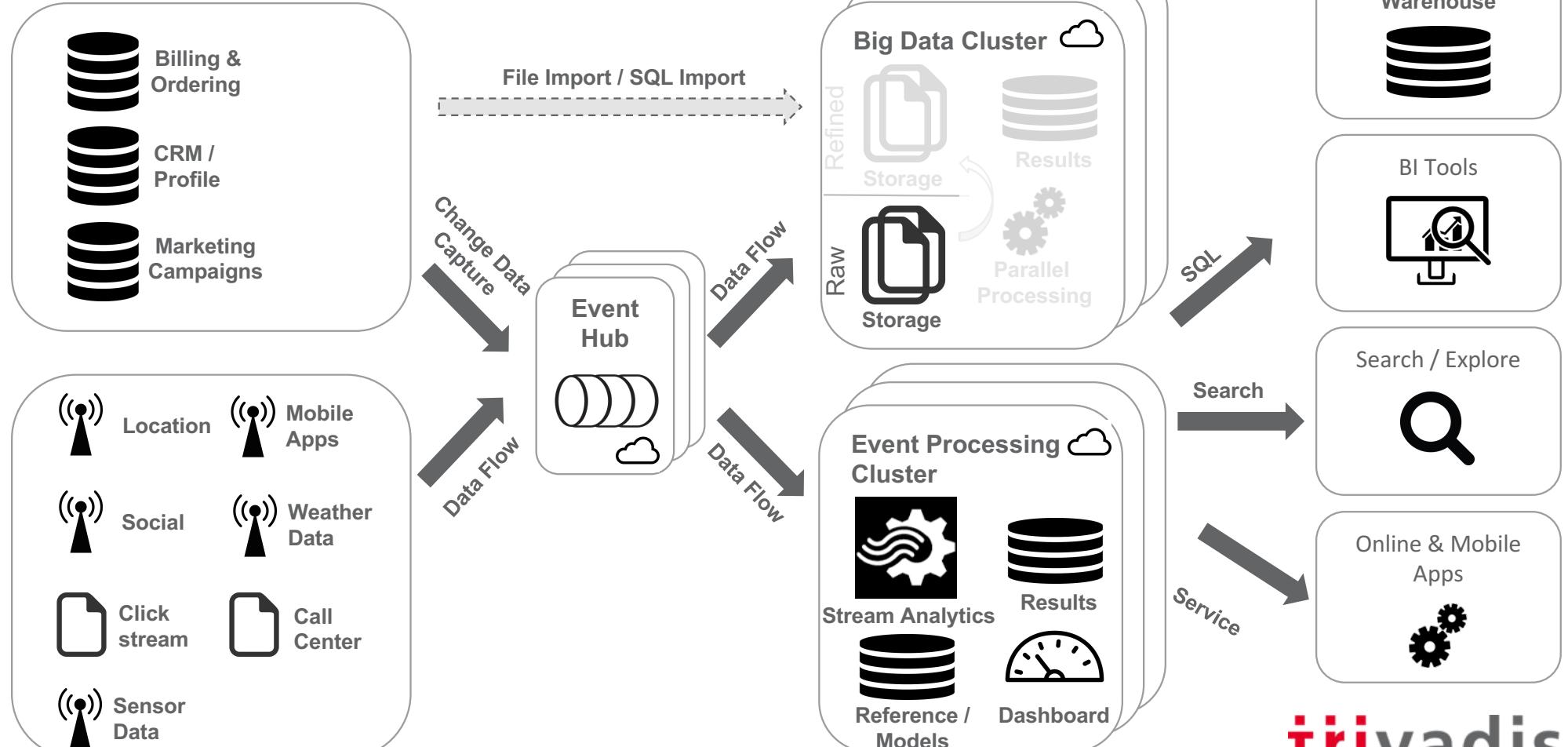
- Low Latency Processing
- Alerting
- "Real-Time" Dashboard



Streaming Analytics Architecture

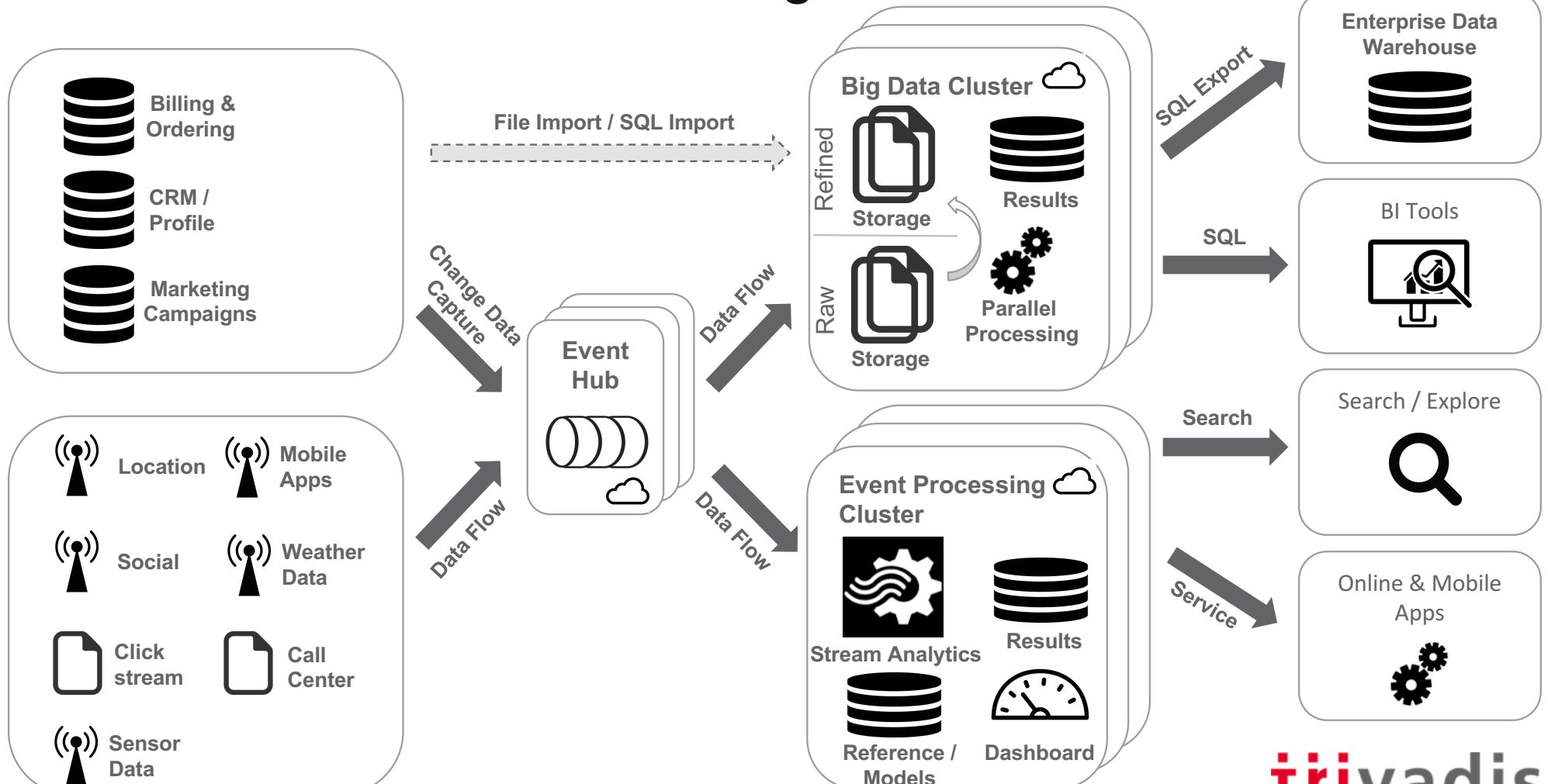


Keep raw event data



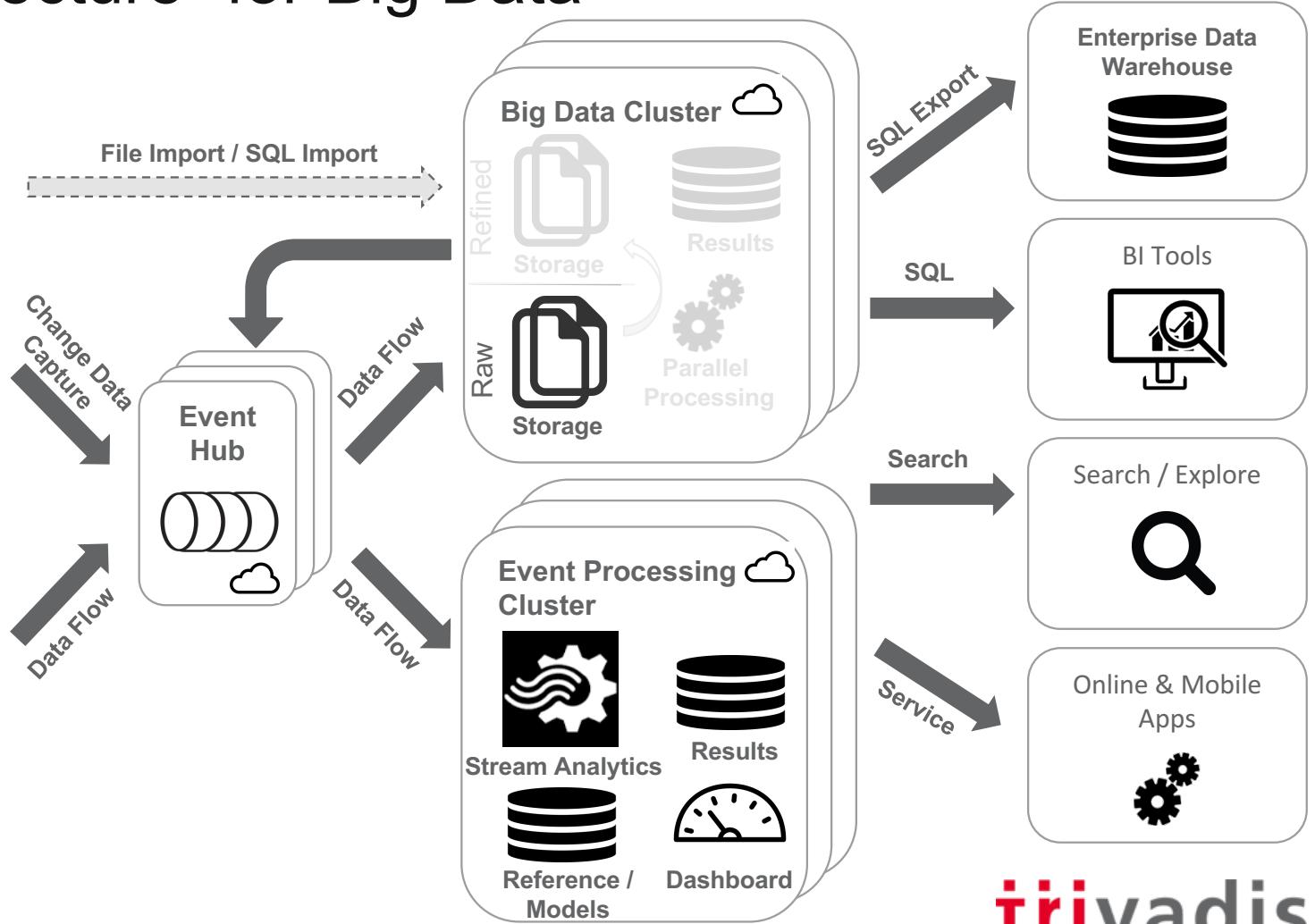
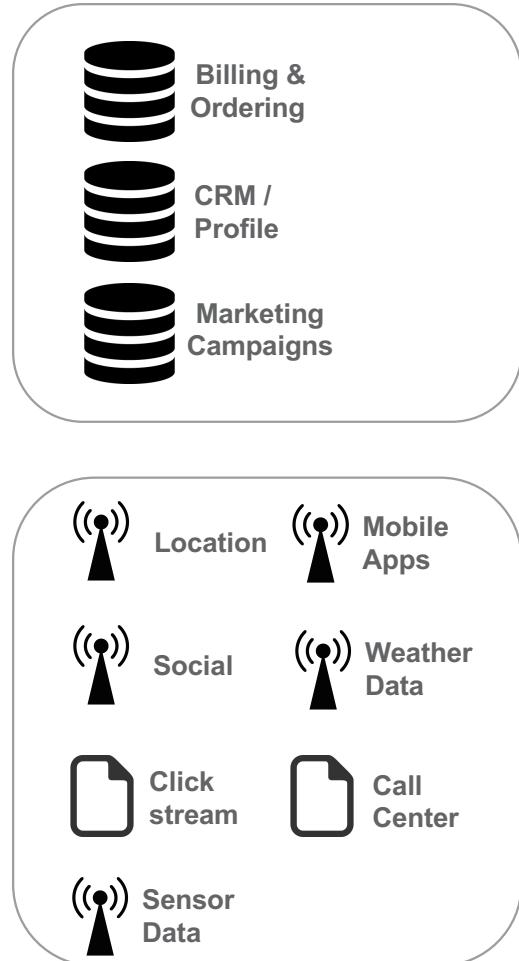
Architektur of Big Data Solutions

■ “Lambda Architecture” for Big Data



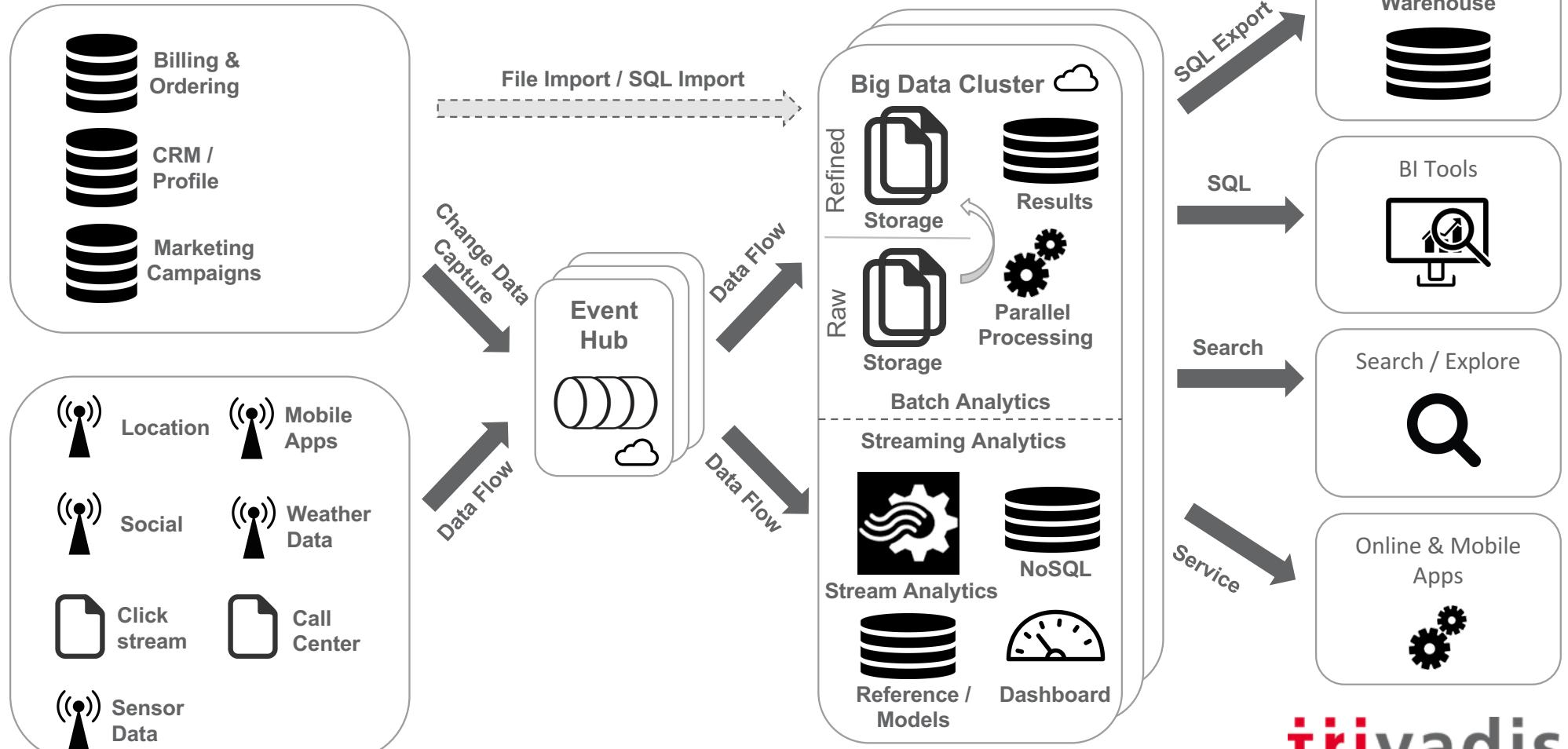
Architektur of Big Data Solutions

■ “Kappa Architecture” for Big Data



Architektur of Big Data Solutions

■ “Unified Architecture” for Big Data



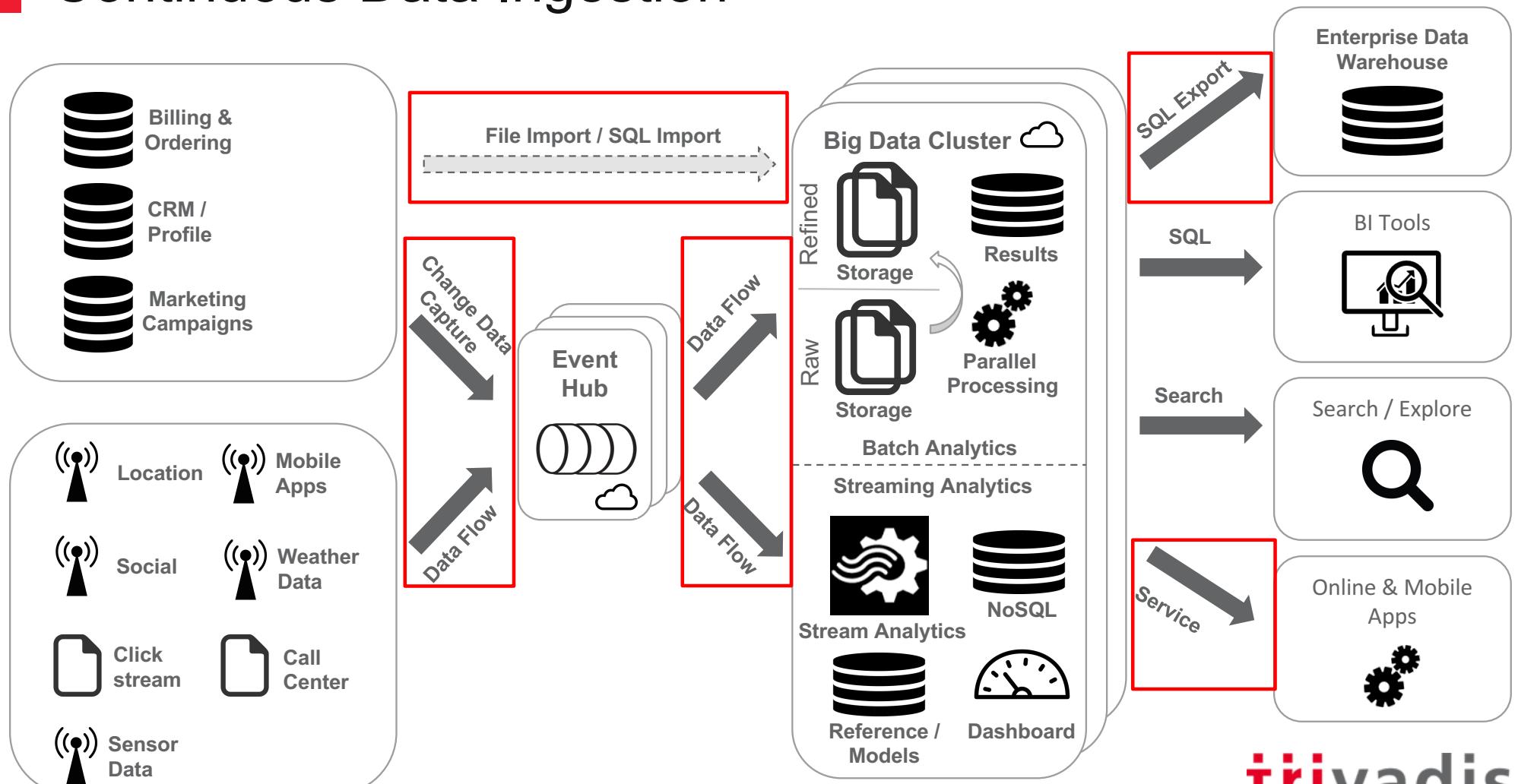
Architektur of Big Data Solutions

Continuous Streaming Data Ingestion

Architektur of Big Data Solutions

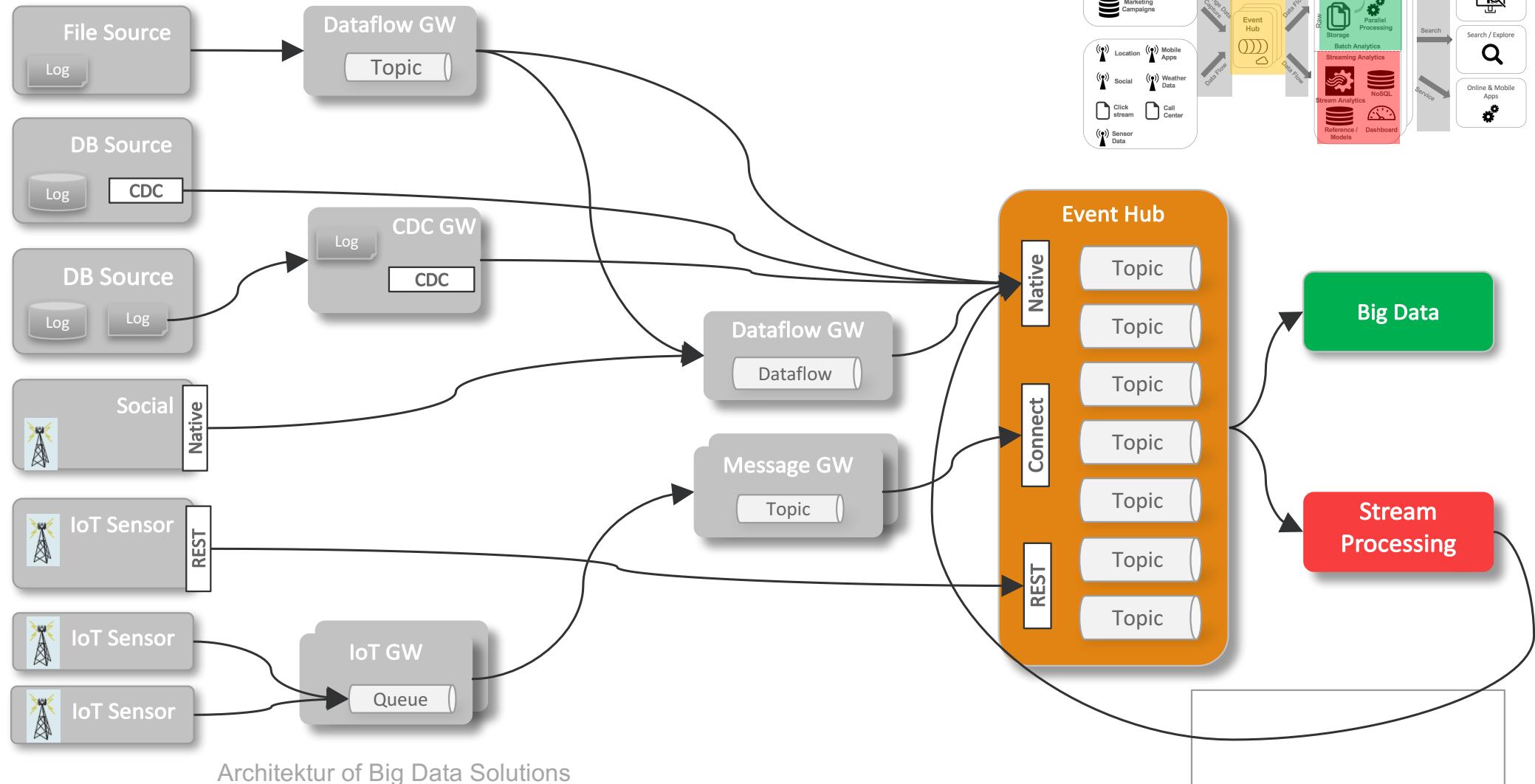
trivadis
makes **IT** easier. 

Continuous Data Ingestion



Architektur of Big Data Solutions

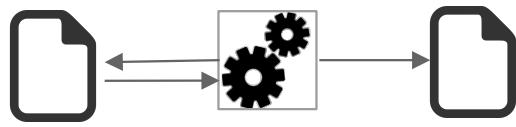
Continuous Streaming Data Ingestion



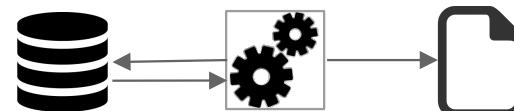
Architektur of Big Data Solutions

Continuous Streaming Data Ingestion

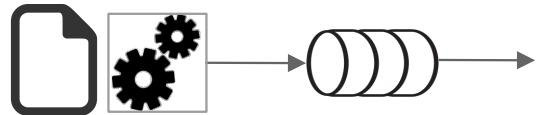
File Polling



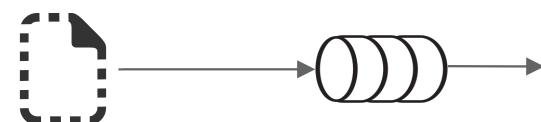
SQL Polling



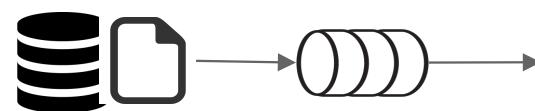
File Stream (File Tailing)



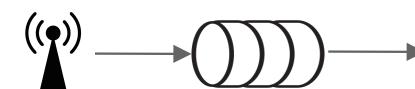
File Stream (Appender)



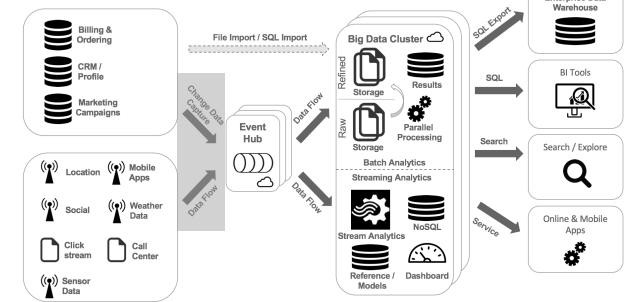
Change Data Capture (CDC)



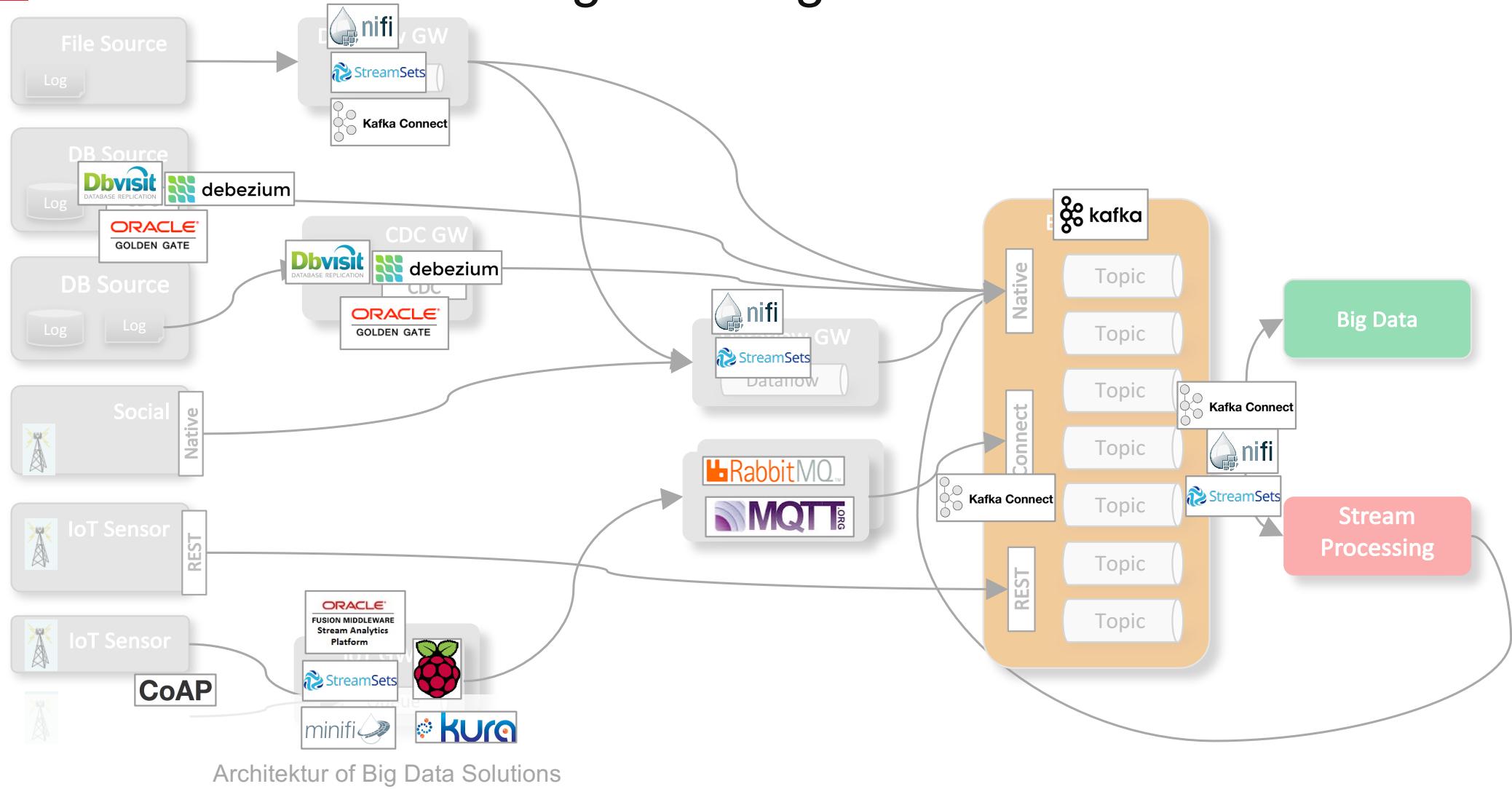
Sensor Stream



Architektur of Big Data Solutions



Continuous Streaming Data Ingestion



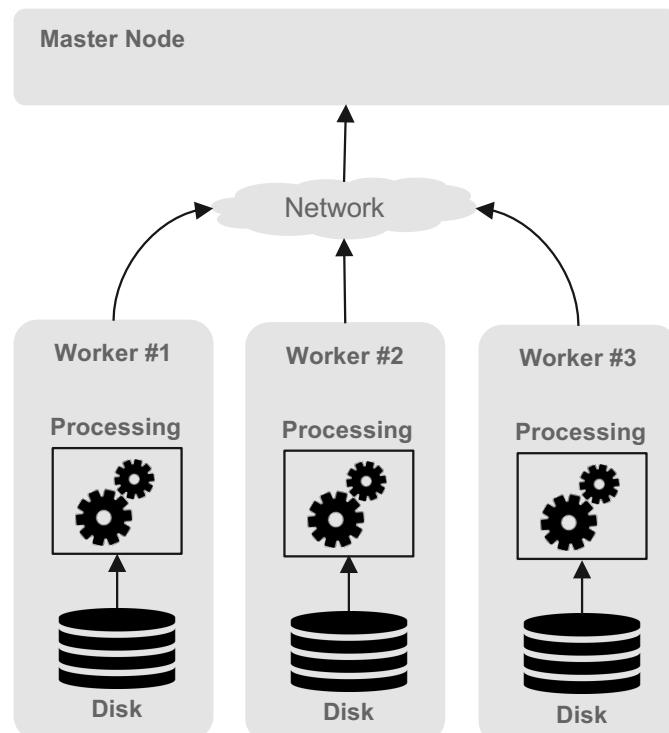
Big Data & Cloud

Architektur of Big Data Solutions

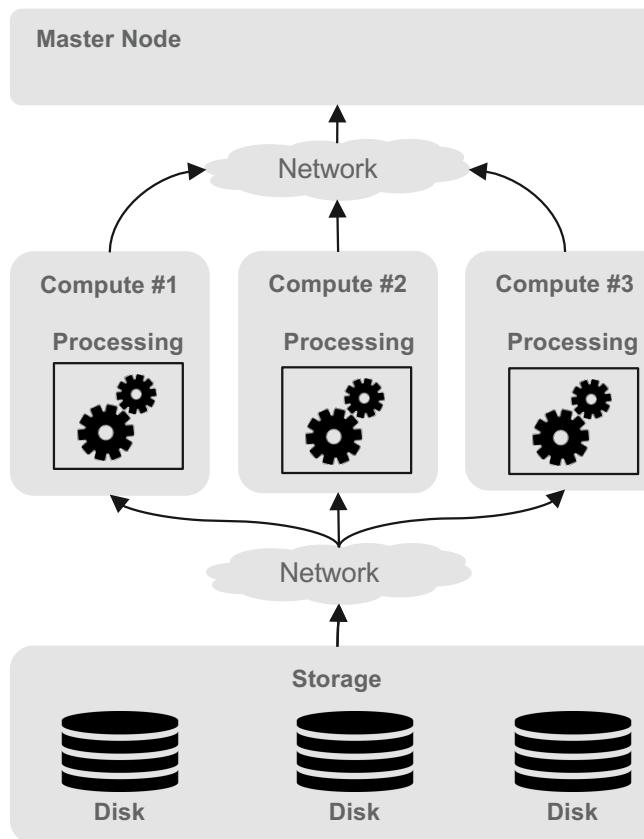
trivadis
makes **IT** easier. 

■ Data Locality vs. Compute/Storage Separation

Data Local Compute



Separate Compute and Storage

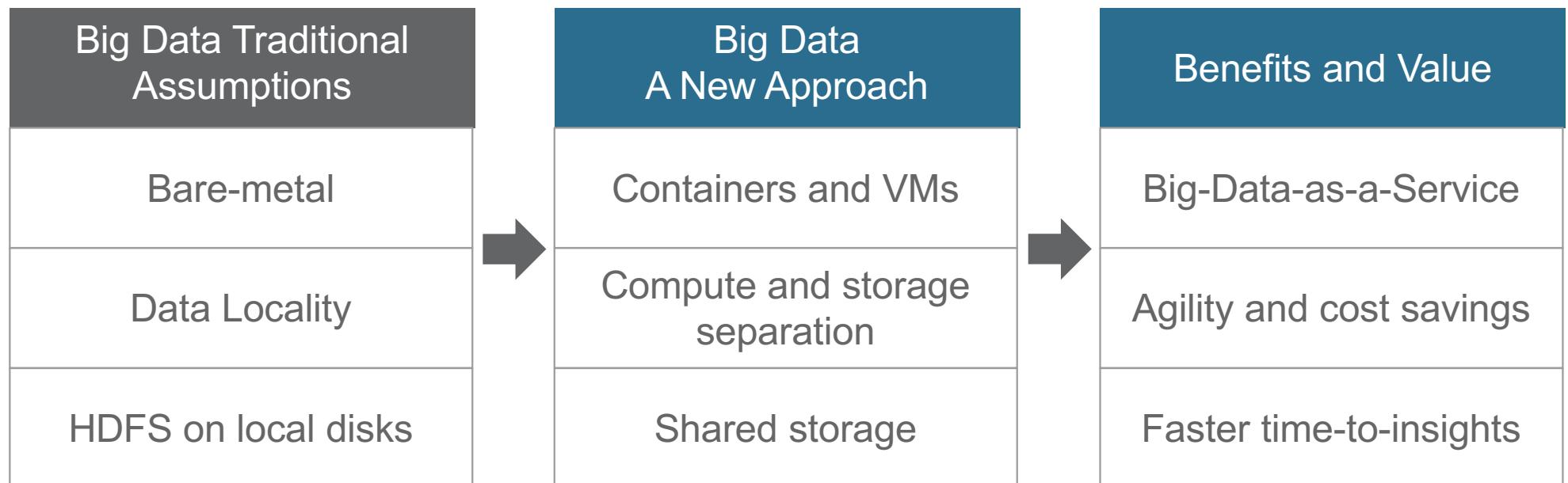


Architektur of Big Data Solutions

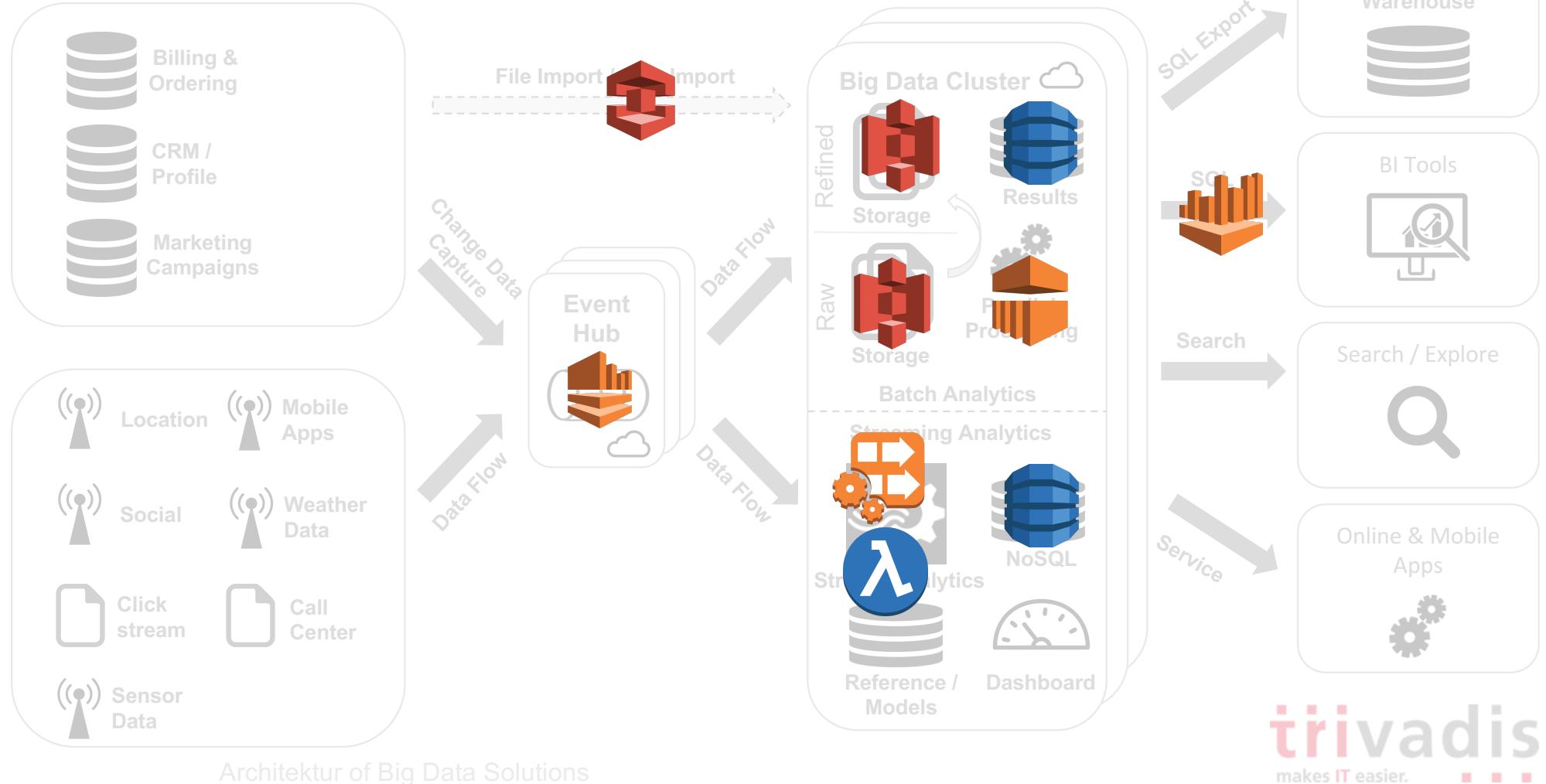
Separation of compute and storage – the fundamental difference

- store data in Object Storage instead of DFS
- bring up Compute nodes only for data processing
- multiple workloads on separate clusters can access same data

■ A new way to Manage Big Data



■ Big Data & Cloud - Amazon WebServices (AWS)

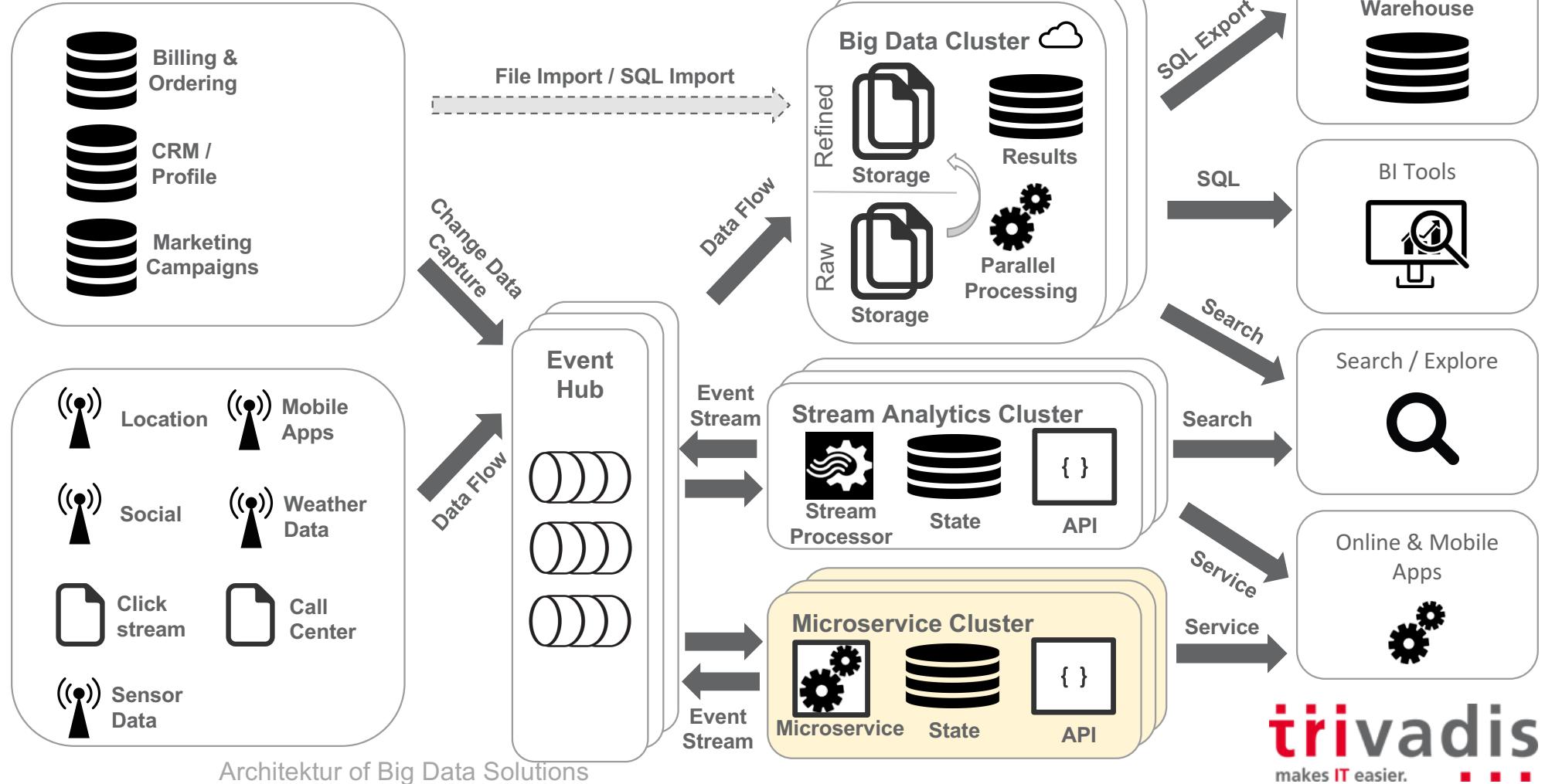


Microservices Architecture

Architektur of Big Data Solutions

trivadis
makes **IT** easier. 

Asynchronous Microservice Architecture

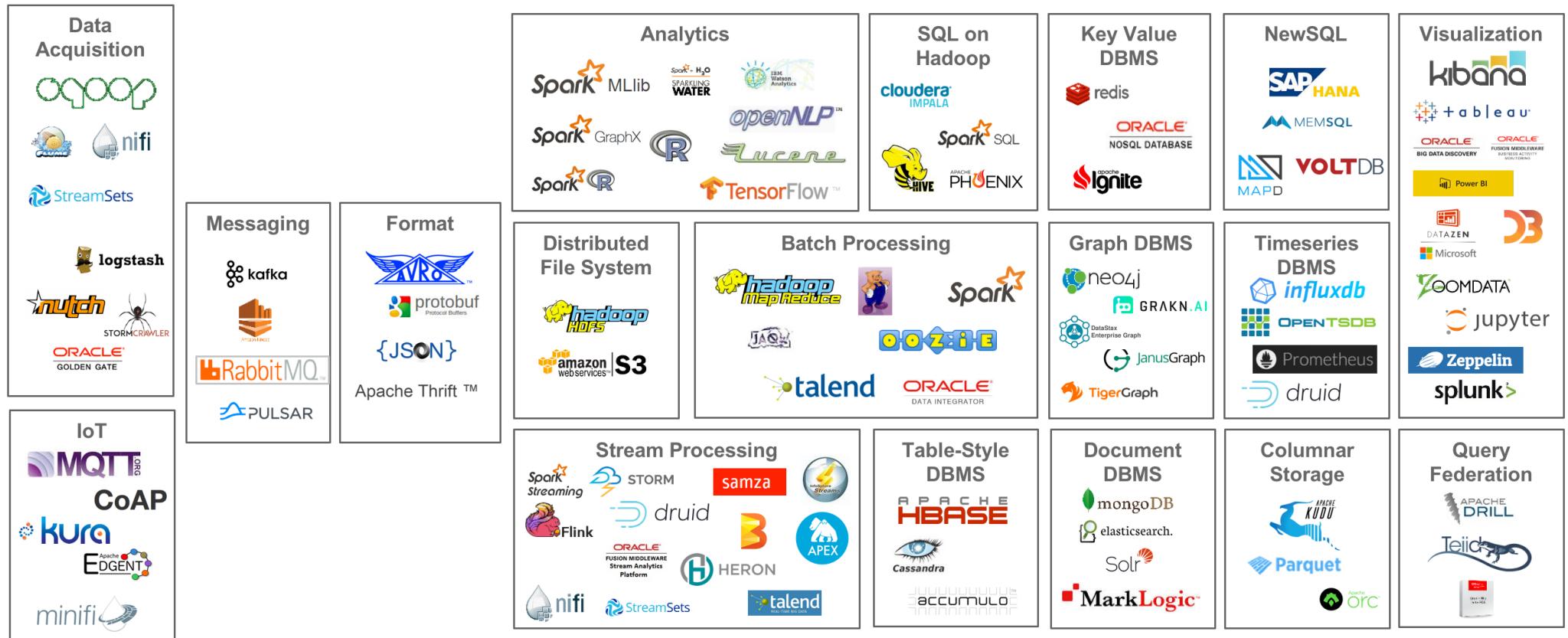


Big Data Ecosystem – many choices sorted!

Architektur of Big Data Solutions

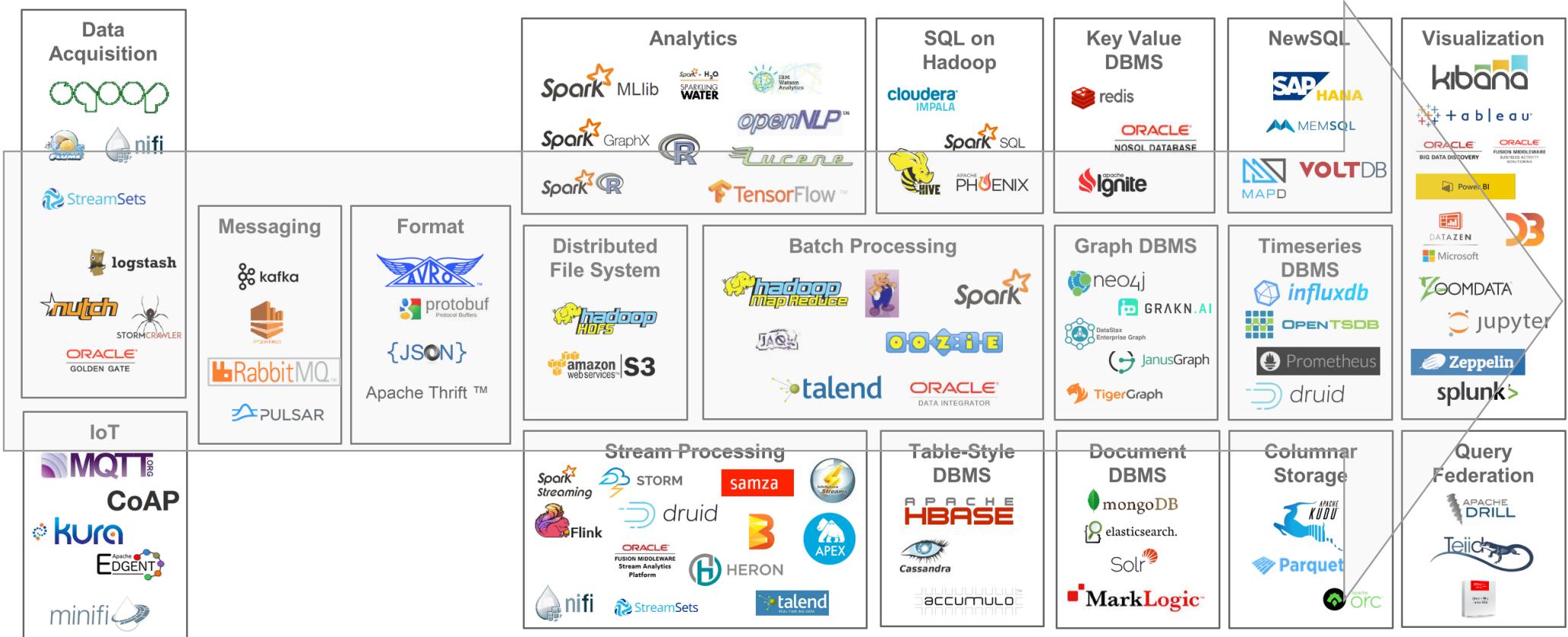
trivadis
makes **IT** easier. 

■ Big Data Ecosystem – many choices sorted!

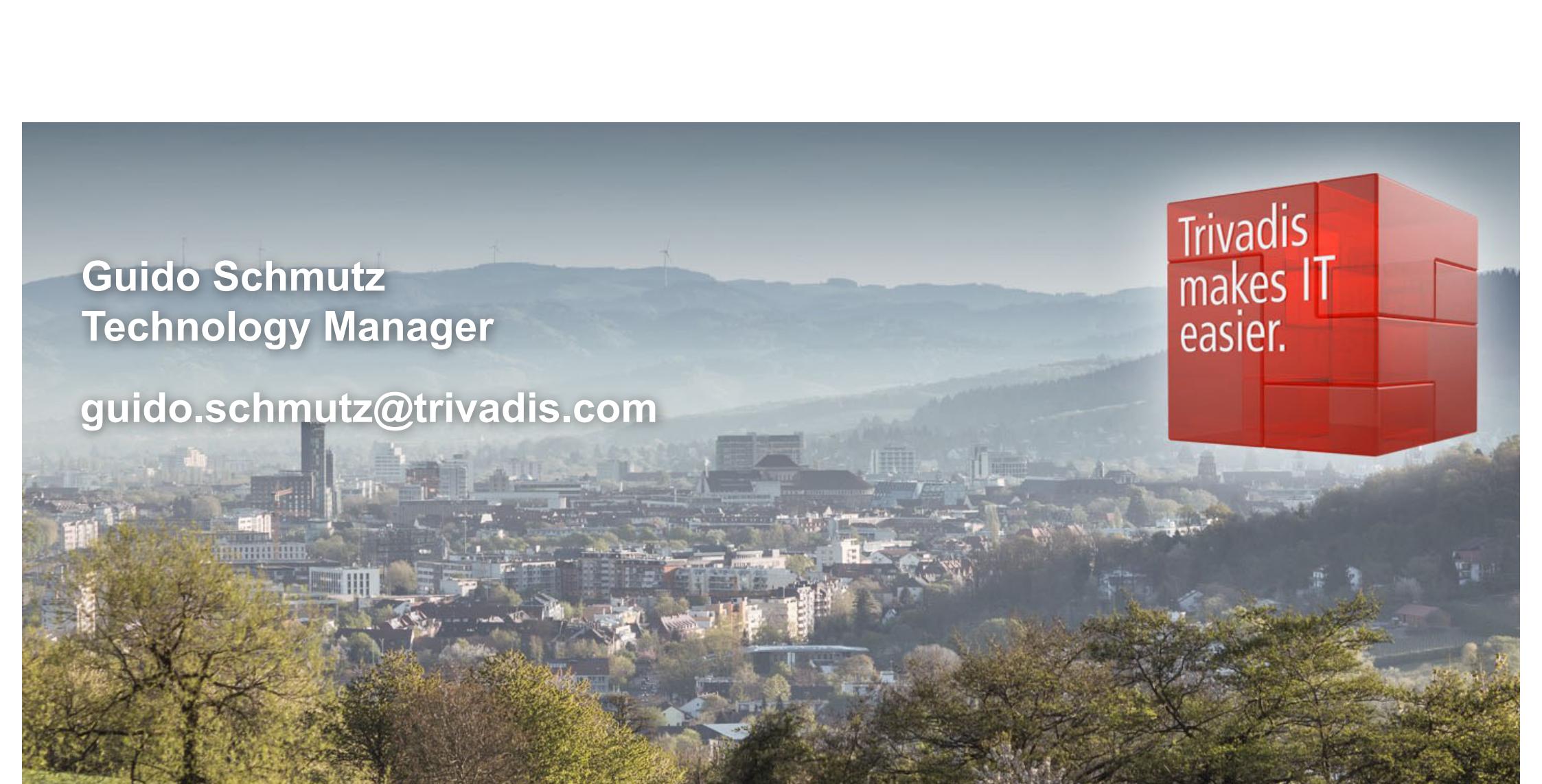


Architektur of Big Data Solutions

■ Big Data Ecosystem – many choices sorted!



Architektur of Big Data Solutions



Guido Schmutz
Technology Manager

guido.schmutz@trivadis.com

