



## CONSIST PROJEKT REFERENZ

# Freie Bahn für Daten – in nur 90 Tagen zum Big Data Lake

### Ein Big Player in der Logistik – Die Story in Kürze

Das Logistik-Unternehmen: eine Holding mit drei Teilkonzernen, 15 angeschlossenen Teilgesellschaften und 29 Milliarden bewegter Gütertonnen pro Jahr in mehr als 23.000 Transporteinheiten – kurzum, mit einer gigantischen Menge an Infrastrukturdaten. Ein Konzern, der nun verstärkt die Big-Data-Ambitionen seiner Geschäftsbereiche aufgreift und diese darin unterstützt, Data Science im gesamten Unternehmen zu etablieren. Denn gigantische Datenmengen bergen ein gigantisches Potenzial, einzelne Aufgaben, Geschäftsprozesse, Unternehmensteile oder ganze Unternehmen auf deren Basis weiterzuentwickeln.

Innerhalb der verschiedenen Teilkonzerne existieren daher bereits unterschiedliche Data Tools und -Technologien. Drei verschiedene Bu-

siness-Intelligence-Plattformen waren die Folge. Dies brachte eine hohe Komplexität, hohe Integrationskosten und lange Umsetzungsintervalle mit sich. Vor diesem Hintergrund entstand der Wunsch nach einer Konsolidierung hin auf eine zentrale Big-Data-Plattform. Konzernweit wurde daher das Thema Big Data und, in nächster Konsequenz, Data Science, zu einem zentralen Thema des zugehörigen Competence Centers gemacht.

Um alle Konzerneinheiten einvernehmlich in diesem Prozess mitzunehmen, sollte ein Pilot gestartet werden, der die Vorteile einer einheitlichen Daten-Plattform eindrucksvoll vor Augen führt. In nur 90 Tagen sollte ein konzeptionell grob beschriebenes Toolset in eine funktionierende Pilotapplikation mit realen

Daten umgesetzt werden. Consist begleitete diesen Prozess architektonisch und „hands on“ in der praktischen Umsetzung. Gemeinsam mit dem agilen Konzern-Team entstand so der Pilot „Güterwagenmonitoring\_ETA (Estimated Time of Arrival)“.

### Die Aufgabe

- ◆ Etablierung eines konzernweiten Konzeptes für den Umgang mit Big-Data-Herausforderungen – Festlegung der Methodik, Technologie und Infrastruktur
- ◆ Aufsetzen eines On-Premise Big Data Lakes und Pilotierung eines Use Cases auf dessen Basis
- ◆ Gestalten des Wegs zu einer Unternehmensplattform (Enterprise Data Lake)



## Die Herausforderung

- ◆ Die sehr unterschiedliche Struktur von Data Lakes bedingt wichtige Festlegungen für die Auswahl der technischen Plattform, Art der Distributionen und Anzahl der Knoten für die Skalierung der Plattform.
- ◆ Die Klärung dieser Fragen im Kontext mit wechselseitigen Abhängigkeiten, deren schnelle Abstimmung im Konzern und das sofortige Aufsetzen eines Piloten bedurfte eines erfahrenen Teams.
- ◆ Hoher Zeitdruck durch anstehende erste echte Trials und Teststellungen in Konzernteilen sowie gestiegenem Bedarf für derartige Systeme

## Die Lösung mit Consist

- ◆ Entwicklung eines Konzeptes, orientiert am Branchenstandard der Lambda-Architektur, als Grundlage eines konzernspezifi-

schen Enterprise Data Lakes für Batch- und Realtime-Verfahren:

- ◆ Definition eines geeigneten Frameworks, das im Rahmen einer Teststellung konfiguriert und evaluiert wurde
- ◆ Implementierung einer Ingestion-Pipeline zur Verarbeitung der Daten

## Eingesetzte Technologien

- ◆ Ingestion-Pipeline via Logstash, Kafka und Flume (unter Einsatz von Consist-Interceptoren): Aufbereitung der relevanten Daten und finale, wohlpartitionierte Übertragung ins HDFS
- ◆ Aggregation und Analyse der Daten mittels Hive
- ◆ Überführung der Analyseergebnisse via Sqoop in eine relationale Datenbank
- ◆ Zeitliche Orchestrierung der Einzelprozesse durch Oozie

## Besondere Stärken von Consist

- ◆ Ganzheitliches Know-how in Big-Data-Technologien, gepaart mit fachlichem Bahn-Know-how
- ◆ Effiziente, technologisch fundierte Projektdurchführung in time und in budget
- ◆ Gemeinsam mit dem agilen Team des Competence Centers gelang es sehr schnell eine technische und fachliche Umgebung für den Data Lake zu schaffen.

## Kundennutzen

- ◆ Hohe Akzeptanz für eine zentrale konzernweite Big-Data-Plattform
- ◆ Data Lake mit Pilot-Use-Case, der als Nukleus für die Produktplattform verwendet und gleichzeitig produktionsähnlich weiterbetrieben werden kann
- ◆ Vermeidung höheren Zeit- und Kostenaufwandes durch weitere Teillösungen dank Konsolidierung auf eine zentrale Big-Data-Plattform

**CONSIST**  
Business Information Technology

**We do IT  
for you.**

### Ihr Kontakt:

Asmus Hammer  
hammer@consist.de  
0431 3993-637  
Consist Software Solutions GmbH  
Christianspries 4, 24159 Kiel  
www.consist.de