

# Die Datenfabrik - oder wie Daten Schritt für Schritt veredelt werden.

Raphael Branger Jan Riedo

# ?

### Stellen Sie Fragen in der virtuellen Plattform

- Loggen Sie sich auf der virtuellen Plattform mit Ihrem persönlichen Link ein (aus Bestätigungs-Mail MAKE BI)
- Kicken Sie in der Lobby auf der Session K2 auf den Button "Teilnehmen"



Kicken Sie in der **Session K2** auf der rechten Seite in der Navigation auf "Q&A" MOBILE: Klicken Sie der Navigation unten zuerst auf "Interaktion", danach auf das **Sprechblasen-Symbol** 



Geben Sie im entsprechenden Feld Ihre Frage ein und schicken Sie diese ab, indem Sie auf das **Pfeil-Symbol** klicken

Die Referenten werden am Ende Ihrer Präsentation vereinzelt auf Fragen eingehen. Nicht beantwortete Fragen können gerne bei den diversen Networking-Möglichkeiten mit den entsprechenden Referenten aufgegriffen werden.





#### Wer sind wir?



Jan Riedo Head of Innovation Center Polymechaniker, BSc Maschinentechnik MSc Biomedical Engineering

jriedo@it-logix.ch



Raphael Branger
Principal Consultant Data & Analytics
MA Information Management

rbranger@it-logix.ch





#### Warum kostet mein erstes Dashboard 80'000.-?

Das Dashboard ist schnell gemacht. Kostspielig ist in der Regel die Datengewinnung und –verarbeitung. Heute schauen wir für einmal hinter die Kulissen. Kommen Sie mit!







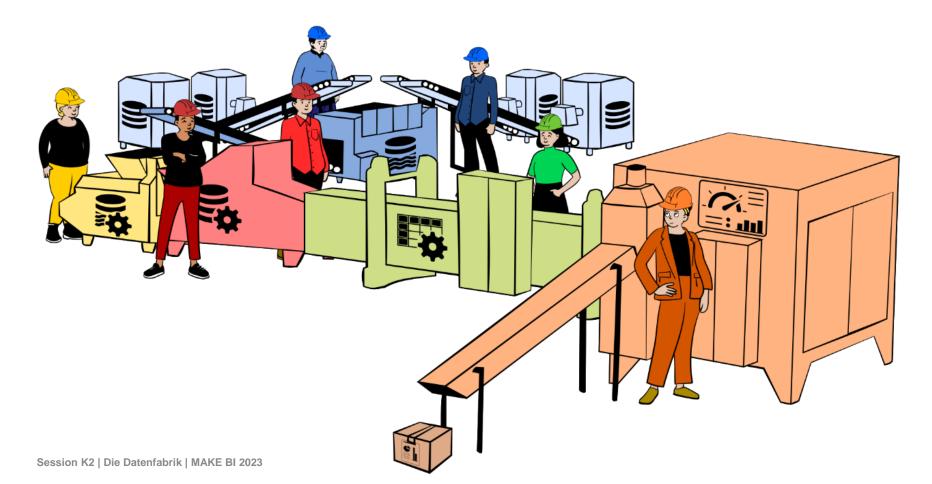
Willkommen auf der Baustelle "Datenfabrik" der Muster AG!





#### Was soll mit dem Bau der Datenfabrik erreicht werden?

- > Standardisierte, wiederholbare, qualitätsgesicherte Datenversorgung
- > Integration und Harmonisierung von Daten verschiedener Quellen
- Effiziente Entwicklung neuer Datenprodukte







### **Rapid Prototyping**

Bevor wir mit dem Bau einer Produktionslinie beginnen, müssen wir einen

Prototypen entwickeln.

Rapid Prototyping umfasst folgende Aufgaben:

- **Data Profiling**
- **Data Modeling**
- **Data Product Mocking**
- Writing Work Items





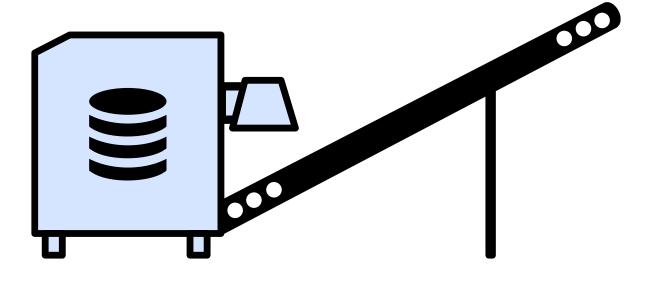
12

Vorstellung der Maschinentypen für eine Datenproduktionslinie







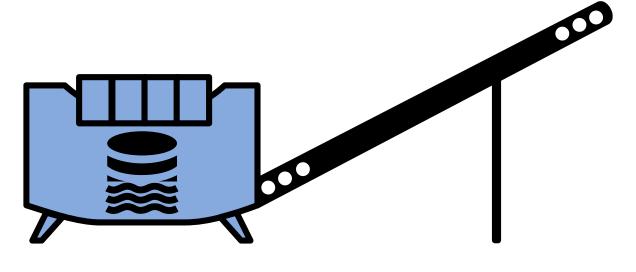


- Maschinentyp: Extraktor
- Aufgaben:
  - Extrahieren von Daten aus Quellsystemen
  - Delta-Detection
- Output: Rohdaten
- Beliebte Hersteller:
  - Microsoft
  - Theobald
  - Cdata
- Häufig eingesetzte Modelle:
  - Theo Xtract 5000 (speziell f
    ür SAP-Quellen)
  - > ADF AZ 35 (ideal für Cloud-Umgebungen)





#### **Multitemperierter Datensee**

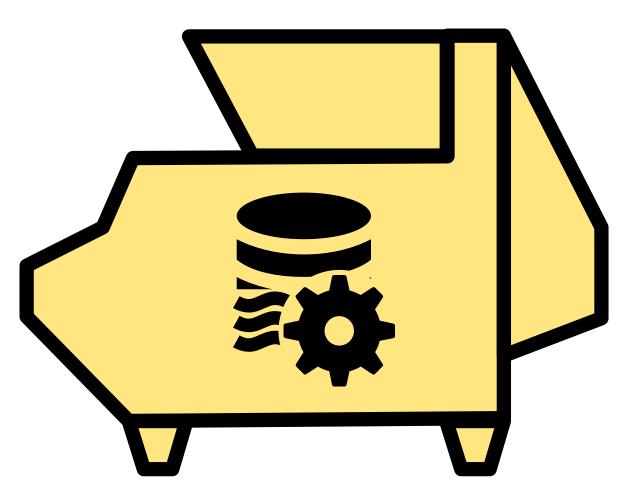


- Maschinentyp: Multitemperierter Datensee
- Aufgaben:
  - Langfristiges Speichern der extrahierten Daten zu günstigen Preisen
  - Unterschiedliche Lagerzonen (cold, warm, hot)
- Output: Archivierte Rohdaten
- Beliebte Hersteller:
  - Microsoft
  - Google
  - Amazon
- > Häufig eingesetzte Modelle:
  - DTL Gen2
  - > 1Lake 300





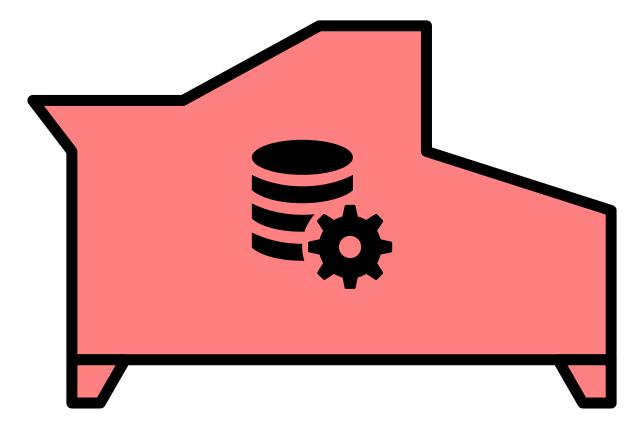
#### **Datenpersistierer**



- Maschinentyp: Datenpersistierer
- Aufgaben:
  - Business Key Validierung
  - Datentyp-Mapping auf Zieldatenbank
  - Historisierung der quell-nah modellierten Daten
  - Erkennung gelöschter Datensätze im Quellsystem
- Output: Historisierte Daten (Persistent Stage)
- Beliebte Hersteller:
  - WhereScape
  - > IT-Logix
- Häufig eingesetzte Modelle:
  - Basismodell:
    - > WS 3D 8
    - WS RED 8 DST
  - Erweiterungen
    - > ITX PSA Auto 2000



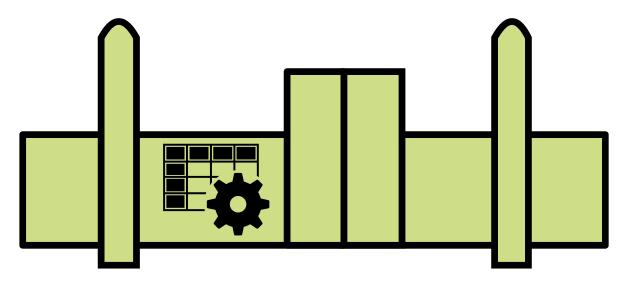
#### **Datenhomogenisierer**



- Maschinentyp: Datenhomogenisierer
- Aufgaben:
  - Integration Daten verschiedener Quellen in gemeinsame Zielentität
  - Supernatural Key Management
  - Historisierung der dimensional modellierten Daten
  - Umsetzung Geschäftsregeln für die Berechnung von Basiskennzahlen
- Output: Integrierte, homogenisierte, dimensional modellierte Daten
- Beliebte Hersteller:
  - WhereScape
  - > IT-Logix
- Häufig eingesetzte Modelle:
  - Basismodell:
    - > WS 3D 8
    - WS RED 8 D&F
  - Erweiterungen
    - ITX Core Auto 2000



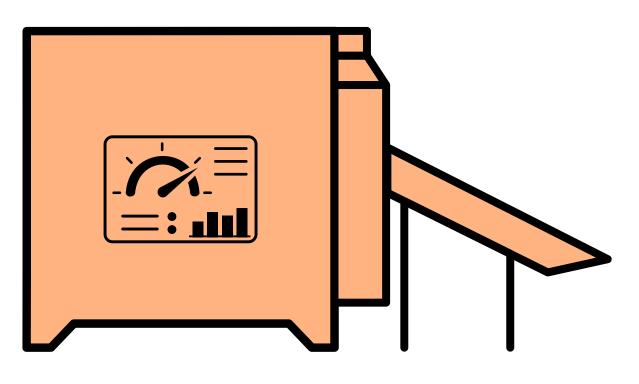
#### **Datenaufbereiter**



- Maschinentyp: Datenaufbereiter
- Aufgaben:
  - Themenspezifische Zusammenstellung von Datenobjekten
  - Umsetzung von themenspezifischen Kennzahlen und Filterlogiken
  - Umsetzung temporaler Kennzahlen
  - Übersetzung der Bezeichnungen von Datenobjekte in Fachsprache inkl. Mehrsprachigkeit
- Output: Themenspezifisch aufbereitete Daten
- Beliebte Hersteller:
  - Microsoft
- Häufig eingesetzte Modelle:
  - > PBX Premium P2
  - > SSAS 2022 Pro



#### **Datenvisualisierer**



- Maschinentyp: Datenvisualisierer
- Aufgaben:
  - Visualisierung von Daten in Form von Tabellen und Diagramme
  - Ermöglichen eines interaktiven Benutzererlebnisses
  - Bereitstellen von Daten-Schnittstellen
- Output: Datenprodukt
- Beliebte Hersteller:
  - Microsoft
  - Inforiver
  - Zebra Bl
- Häufig eingesetzte Modelle:
  - PBX Desk Pro
  - SSRS Page Reporter 3000
  - > XLS 365
  - IBCS Standardizer Flex



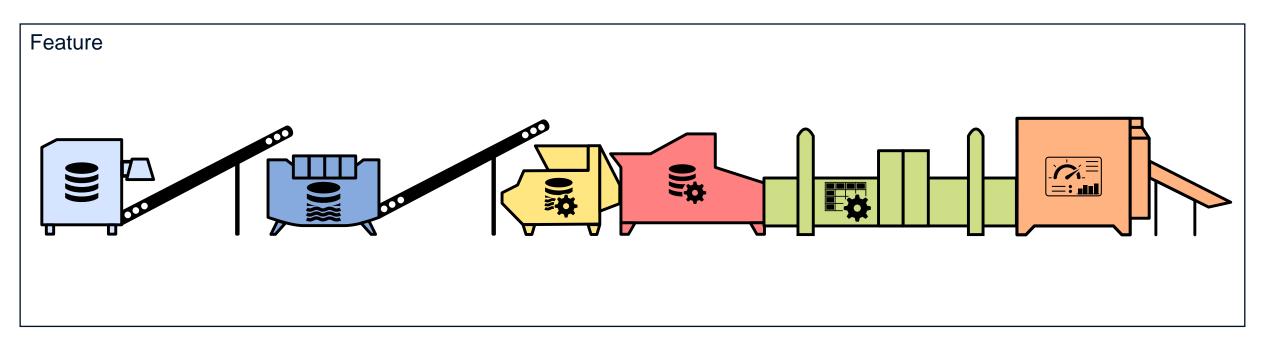
## Inkrementeller Bau der Datenfabrik





#### **Feature = End-to-End Inkrement**

Das Endergebnis erstreckt sich von der Quelle bis zum fertigen Datenprodukt.



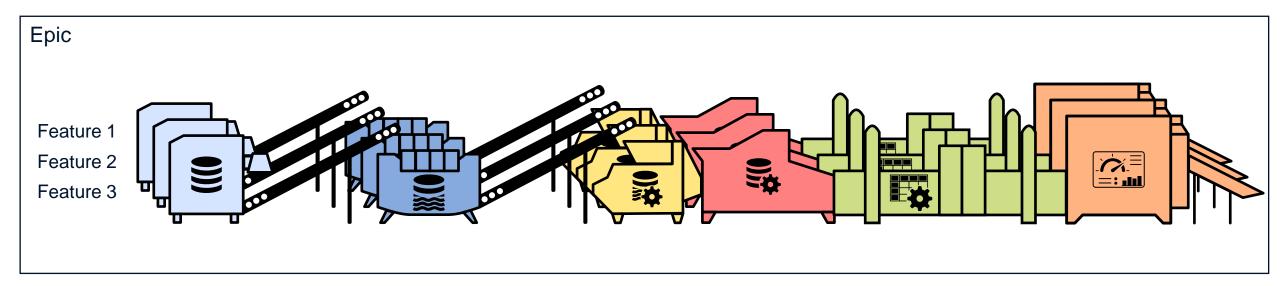




#### **Single Piece Entwicklung**

Pro Release werden oft mehrere Features umgesetzt.

Beim Bau dieser Features empfehlen wir, eine "Single" bzw. "One" Piece Flow:
Ein Feature (= One Piece) nach dem anderen wird entwickelt.

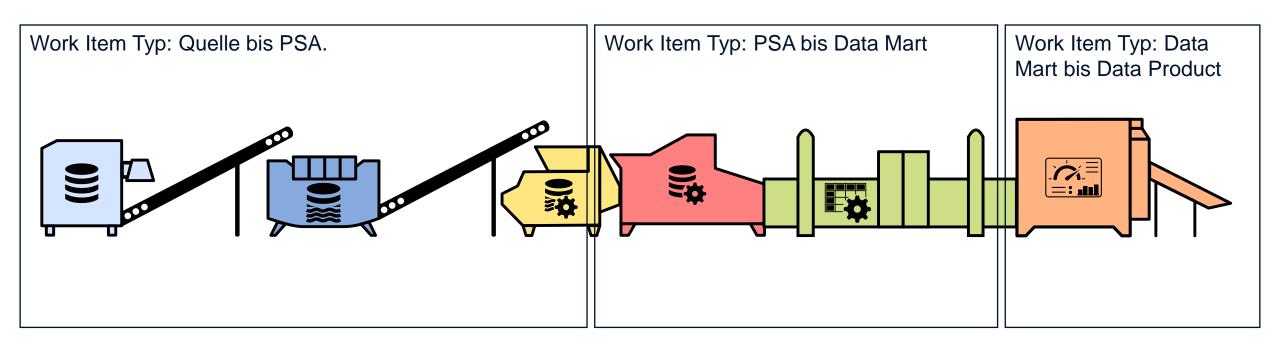






#### Wie ist der Bau der Datenproduktionslinie organisiert

Generalizing Specialists können alle "Maschinen" bedienen. Alternativ können pro Work Item Typ entsprechende Spezialisten eingesetzt werden.





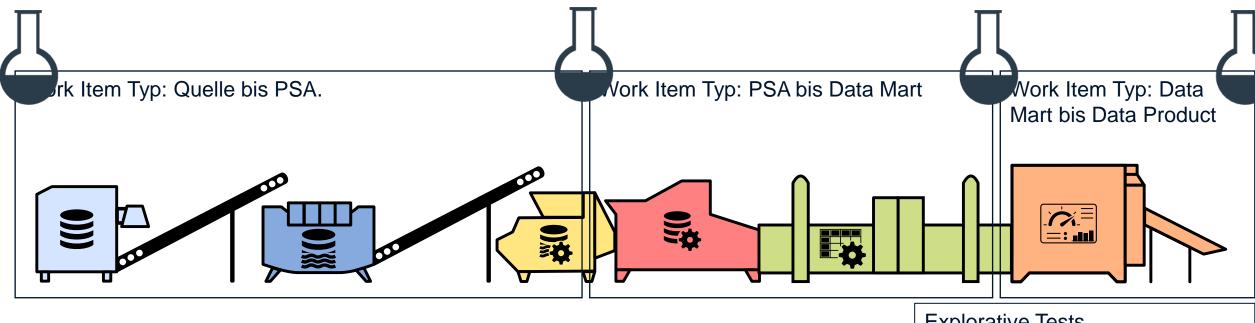
24

#### Qualitätssicherung

**BiG** EVAL High Data Quality. Mastered. swiss made software



Pro Work Item Typ machen auch unterschiedliche Arten von Qualitätssicherungsmassnahmen Sinn!



**Explorative Tests** 

Business Rules, Kennzahlen-Vergleiche

Row Counts, BK-Checks, Konsistenzprüfungen, Performance-Tests



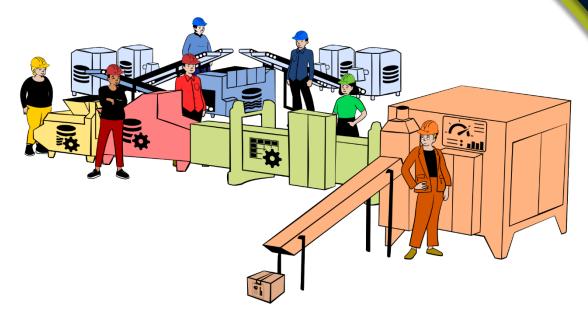
## **Zusammenfassung und Fazit**

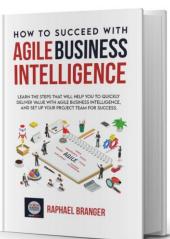




#### **Zum Mitnehmen...**

- Alle sprechen von "Daten als Assets" das bedeutet auch, in die Dateninfrastruktur zu investieren.
- Daten müssen aufbereitet und veredelt werden, um Entscheide mit grosser Auswirkung datengestützt treffen zu können.
- Die Grundkosten für den Aufbau der Datenfabrik darf man nicht einfach dem ersten sichtbaren Datenprodukt anlasten.
- Heute haben Sie das Beispiel einer möglichen Produktionslinie gesehen – je nach Bedarf sind auch kürzere oder längere Produktionslinien möglich.
- Die Analogie der Datenfabrik hilft, ein abstraktes Thema anschaulich zu illustrieren und Mitstreiter zu gewinnen.
- Diese und weitere Analogien sind im Buch "How to Succeed with Agile Business Intelligence" ausgeführt.









# Wann dürfen wir Sie in Ihrer Datenfabrik besuchen?



Raphael Branger
Principal Consultant Data & Analytics
VR / Partner

rbranger@it-logix.ch



#### Jan Riedo

Head of Innovation Center Senior Consultant Data & Analytics Associate Partner

jriedo@it-logix.ch



Twitter



LinkedIn



Xing



YouTube



Blog



#### Bewerten Sie diese Session jetzt in der virtuellen Plattform

- Loggen Sie sich auf der virtuellen Plattform mit Ihrem persönlichen Link ein (aus Bestätigungs-Mail MAKE BI)
- Kicken Sie in der Lobby auf der Session K2 auf den Button "Teilnehmen"



Kicken Sie in der **Session K2** auf der rechten Seite in der Navigation auf "**Umfragen" MOBILE**: Klicken Sie der <u>Navigation unten</u> zuerst auf "**Interaktion"**, danach auf das **Umfragen-Symbol** 



Beantworten Sie die vier Fragen, indem Sie <u>pro Frage</u> auf das zutreffende Feld klicken und Ihre Auswahl anschliessend mit einem Klick auf den Button "**Abstimmen**" bestätigen

Die von Ihnen ausgewählten Bewertungen werden weder für die anderen Teilnehmenden noch für die Referenten ersichtlich sein. Das Ergebnis der Auswertung wird nach dem Event in aggregierter und anonymisierter Form an die jeweiligen Referenten kommuniziert werden.

