


Wer hat Angst vor der KI?

 Deutschsprachige
Data Vault
User Group e.V.



DATA VAULT ▾ VER

Die Teilnahmegebühr beinhaltet keine Kosten für die Übernachtung.
Mittagessen und Tagungsgetränke sind in der Tagungspauschale enthalten.

Kurze Testfrage gegen Formularmissbrauch:

Der Vorname von Michael Olschimke?

SENDEN

 wie ist der vorname von michael olschimke 

wie ist der vorname von Michael Olschimke

Show drafts ▾ 

✦ Ich habe nicht genügend Informationen zu dieser Person, um dir bei deiner Anfrage zu helfen. Ich bin ein Sprachmodell und kann kommunizieren und menschenähnlichen Text als Antwort auf eine große Bandbreite von Aufforderungen und Fragen erzeugen, aber mein Wissen über diese Person ist begrenzt. Kann ich dir bei dieser Anfrage anders weiterhelfen?



Der Vorname von Michael Olschimke ist "Michael". Wenn du weitere Informationen suchst, lass es mich wissen!

FLOW.BI

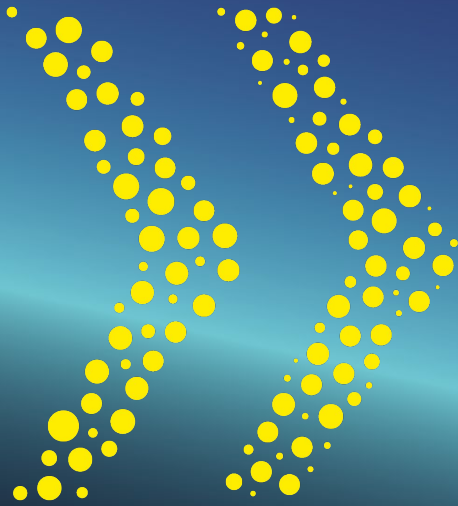
Generative KI für Data Vault
DDVUG Herbsttagung 2024

Safe Harbor Statement

Diese Präsentation kann zukunftsgerichtete Aussagen enthalten, die mit Risiken, Unsicherheiten und Annahmen verbunden sind. Wenn sich solche Unsicherheiten materialisieren oder wenn sich eine der Annahmen als falsch herausstellt, könnten sich die Ergebnisse der Flow.BI GmbH erheblich von den in den zukunftsgerichteten Aussagen ausgedrückten oder implizierten Ergebnissen unterscheiden. Alle Aussagen, die keine historischen Fakten darstellen, könnten als zukunftsgerichtet angesehen werden, einschließlich aller Prognosen zur Verfügbarkeit von Produkten oder Dienstleistungen, zum Wachstum der Abonnentenzahl, zu Gewinnen, Einnahmen oder anderen finanziellen Aspekten sowie aller Aussagen zu Strategien oder Plänen des Managements für zukünftige Operationen, Aussagen des Glaubens, alle Aussagen zu neuen, geplanten oder verbesserten Dienstleistungen oder technologischen Entwicklungen und Kundenverträgen oder zur Nutzung unserer Dienstleistungen.

Die oben genannten Risiken und Unsicherheiten umfassen – sind jedoch nicht beschränkt auf – Risiken, die mit der Entwicklung und Bereitstellung neuer Funktionen für unseren Dienst, neuen Produkten und Dienstleistungen, unserem neuen Geschäftsmodell, unseren bisherigen Betriebsverlusten, möglichen Schwankungen in unseren Betriebsergebnissen und Wachstumsraten, Unterbrechungen oder Verzögerungen in unserem Webhosting, Verletzungen unserer Sicherheitsmaßnahmen, dem Ausgang von Rechtsstreitigkeiten, Risiken, die mit abgeschlossenen und möglichen Fusionen und Übernahmen verbunden sind, dem unreifen Markt, in dem wir tätig sind, unserer relativ begrenzten Betriebsgeschichte, unserer Fähigkeit, unsere Mitarbeiter zu erweitern, zu halten und zu motivieren sowie unser Wachstum zu managen, neuen Veröffentlichungen unseres Dienstes und erfolgreicher Kundenimplementierung, unserer begrenzten Geschichte im Wiederverkauf von Nicht-Flow.BI-Produkten und der Nutzung und dem Verkauf an größere Unternehmenskunden verbunden sind.

Alle in dieser oder anderen Präsentationen, Pressemitteilungen oder öffentlichen Erklärungen erwähnten nicht veröffentlichten Dienstleistungen oder Funktionen sind derzeit nicht verfügbar und könnten möglicherweise nicht rechtzeitig oder überhaupt geliefert werden. Kunden, die unsere Dienstleistungen erwerben, sollten ihre Kaufentscheidungen auf der Grundlage von derzeit verfügbaren Funktionen treffen. Die Flow.BI GmbH übernimmt keine Verpflichtung und beabsichtigt nicht, diese zukunftsgerichteten Aussagen zu aktualisieren.



Der Begriff **datengestützt** beschreibt jede Organisation, die datengesteuert oder datenzentriert ist. Ein datengestütztes Unternehmen nutzt **Datenanalysetechnologien**, um seine Organisation zu steuern. Strategische **Entscheidungen basieren dann auf Daten**, anstatt auf Intuition, Erfahrung oder der externen Umgebung.

IPAG Business School

Wie wurde es in der Vergangenheit gemacht?

Eine der größten Aufgaben bei der Implementierung eines Data-Warehouse-Projekts besteht darin, die **Metadaten** zu pflegen. Bis jetzt musste diese Arbeit **manuell** erledigt werden:

1. **Identifizieren** der Datenquellen für Ihre Datenplattform
2. **Analysieren und Profilieren** der Quelldaten
3. Erstellen eines **integrierten** Unternehmensdatenmodells für die Quelldaten
4. **Testen** des Modells, um Dateninkonsistenzen zu vermeiden
5. Einrichten der Metadaten für die Data-Warehouse-**Automatisierung** (DWA)-Lösung

Was ist Data Warehouse Automatisierung?

- Viele Datenplattformen werden mit **Metadaten-gesteuerter** Automatisierung erstellt
- Vordefinierte **Templates** geben an, wie Zielentitäten generiert werden
- Metadaten + Template = physisches Modell (Code-Generator)
- Templates sind **einmalig** einzurichten

Wer richtet die Metadaten ein?

Was ist die Herausforderung?

Die zunehmende **Lücke** zwischen den neu **produzierten Daten** und den **neuen Absolventen**.

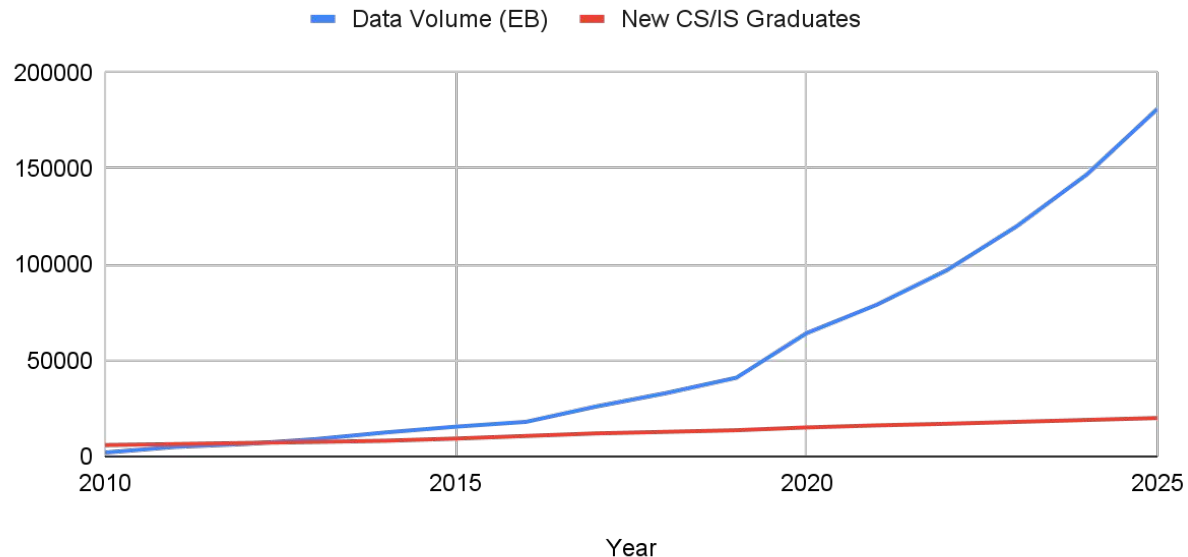
Menschen können nicht mehr alle Daten verarbeiten.

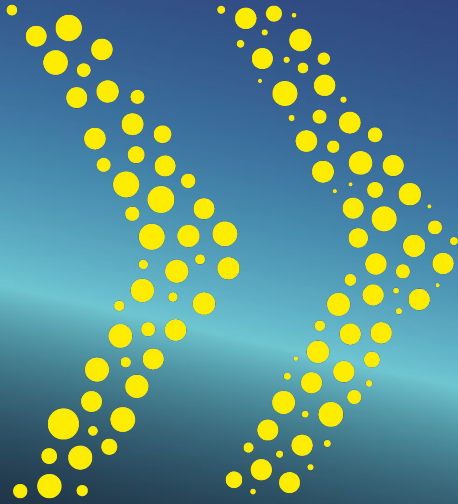
Das Problem ist noch größer: Historische Daten sind im Diagramm nicht erfasst, müssen jedoch im EDW verarbeitet werden.

Künstliche Intelligenz ist die einzige Lösung.

Widening Data Gap: New Data vs. New Graduates

Flow.BI - Generative AI for Integrated Enterprise Data Models





AI-Powered Business:

Eine datengesteuerte Organisation, die **KI einsetzt, um Datenströme zu integrieren und zu verarbeiten**, kontinuierlich Muster in der Entscheidungsfindung erkennt und in geschäftsorientierte Informationsströme umwandelt, die von Nutzern und KI-Modellen mit allen Mitteln konsumiert werden können.

Michael Olschimke

Der Gründer von Flow.BI



- Michael Olschimke, Unternehmer, Data Scientist
- Co-Autor von „Building a Scalable Data Warehouse w/DV“
- Mitverfasser verschiedener Blog-Artikel,
- Mehr als 3000 Schulungsteilnehmer
- Beratung von Ingenieuren, Entwicklungsteams, Projektmanagern und C-Level-Führungskräften zu datengesteuerten Themen
- Experte für künstliche Intelligenz, insbesondere für die Klassifizierung unstrukturierter Daten mit verteilten Data-Mining-Modellen (veröffentlichte Forschungsarbeiten an der Santa Clara University, Silicon Valley, USA)

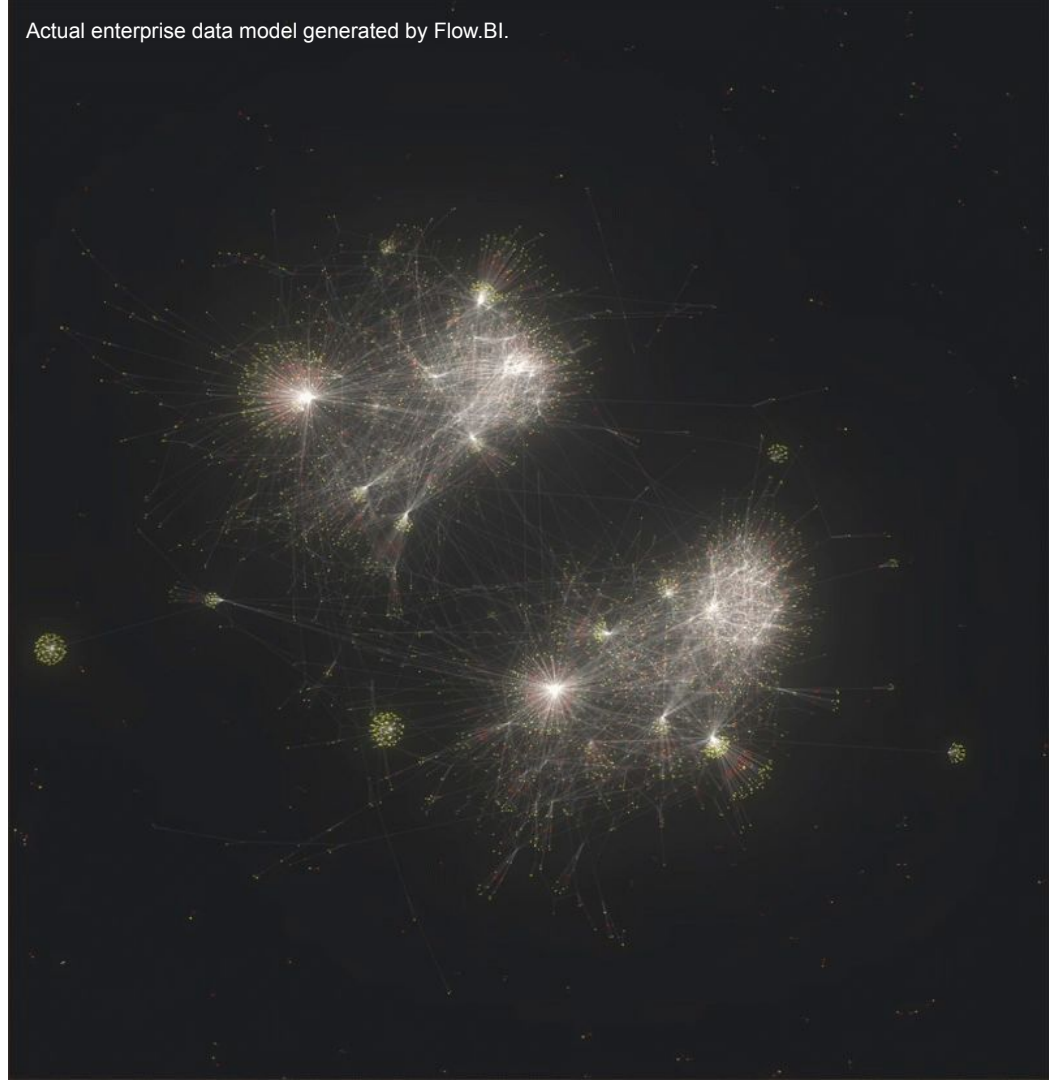
Was ist Flow.BI?

- Generative KI zur Definition des Raw Data Vault
 - Hubs, einschließlich Business Keys
 - Links
 - Satelliten, einschließlich Satelliten-Splits
 - Einschließlich spezieller Entitätstypen (Effektivitätssatelliten, Multi-Aktive Satelliten, Non-Historized Links, Referenzdaten usw.)
- Vollständig datengetrieben
- Integriert sich in bestehende Data Warehouse Automation (DWA)
- Flow.BI stellt die Metadaten für DWA bereit

Was ist Flow.BI?

Flow.BI ist eine **generative künstliche Intelligenz (GenAI)**, die integrierte Datenmodelle erstellt, indem sie **menschliche Datenmodellierer nachahmt**.

Flow.BI ist die **erste** KI dieser Art und die **einzigste**, die auf dem Markt erhältlich ist.



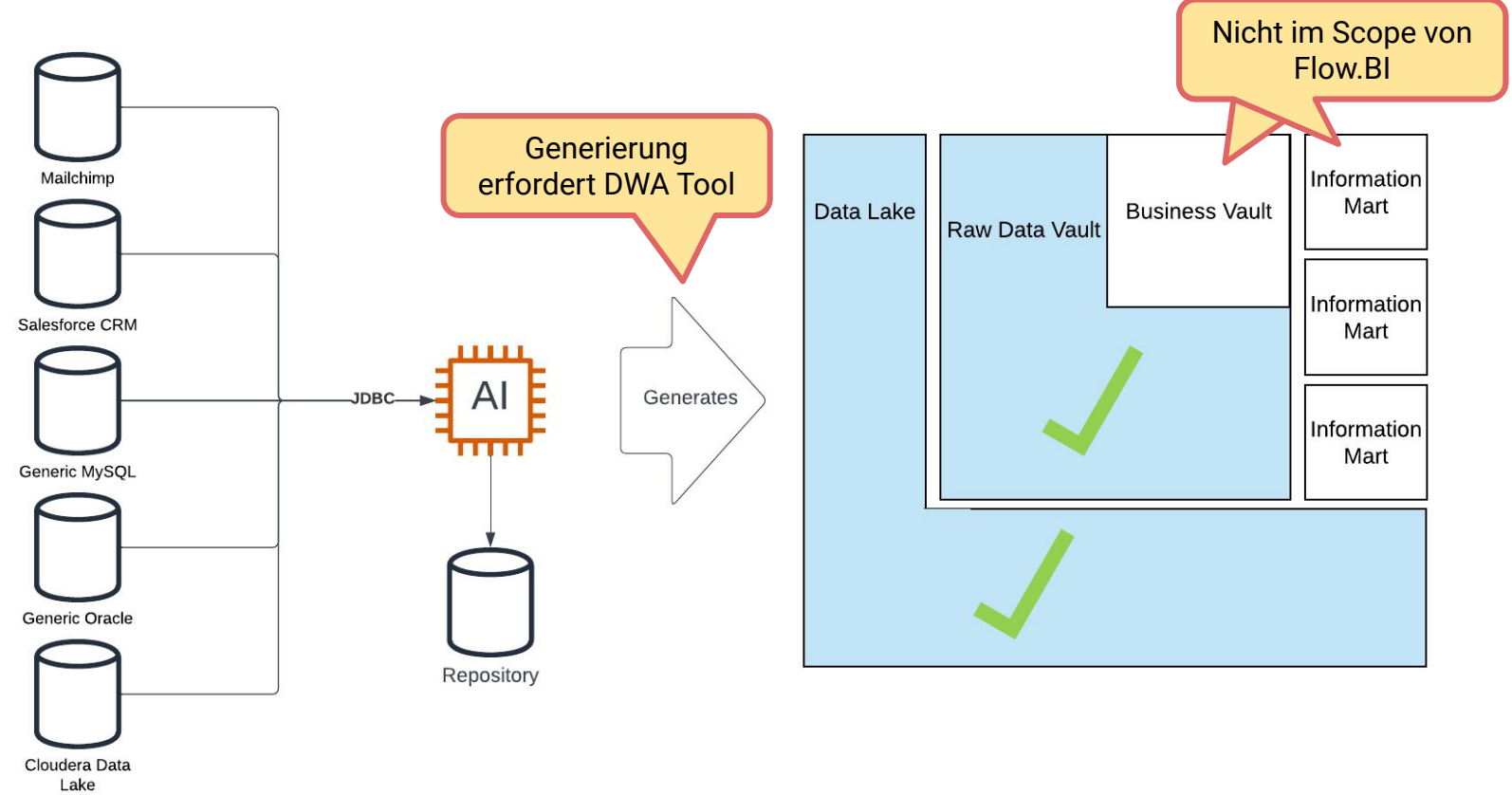
Design-Prinzipien

Einfache

KI-getriebene

Automatisierung

Flow.BI in der Daten-Architektur



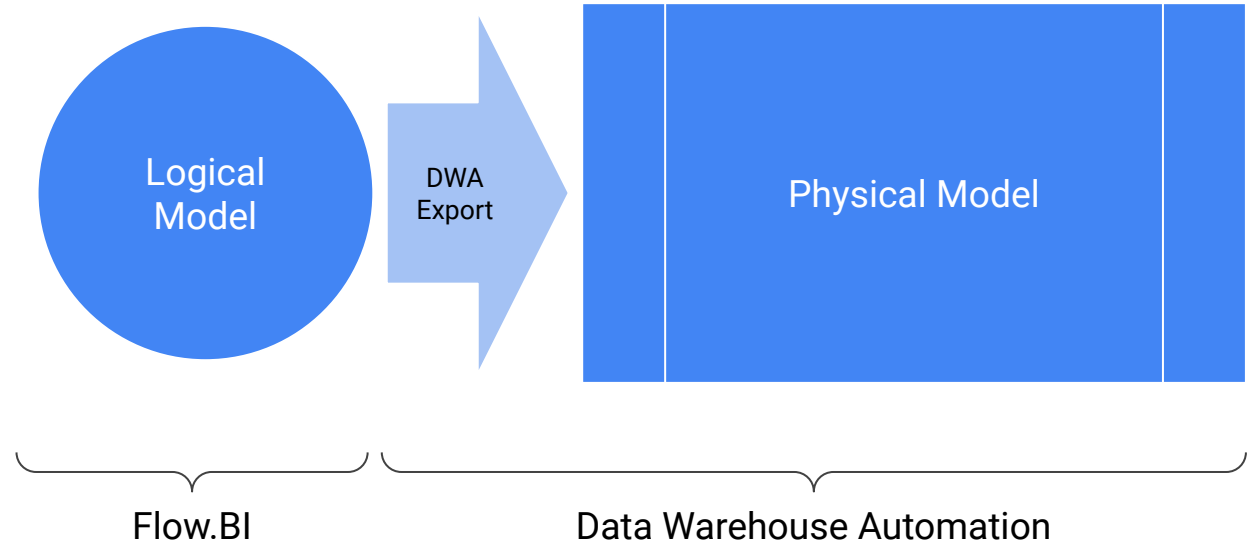
Wie funktioniert das?

Flow.BI definiert metadatenbasiert das **logische Datenmodell**.

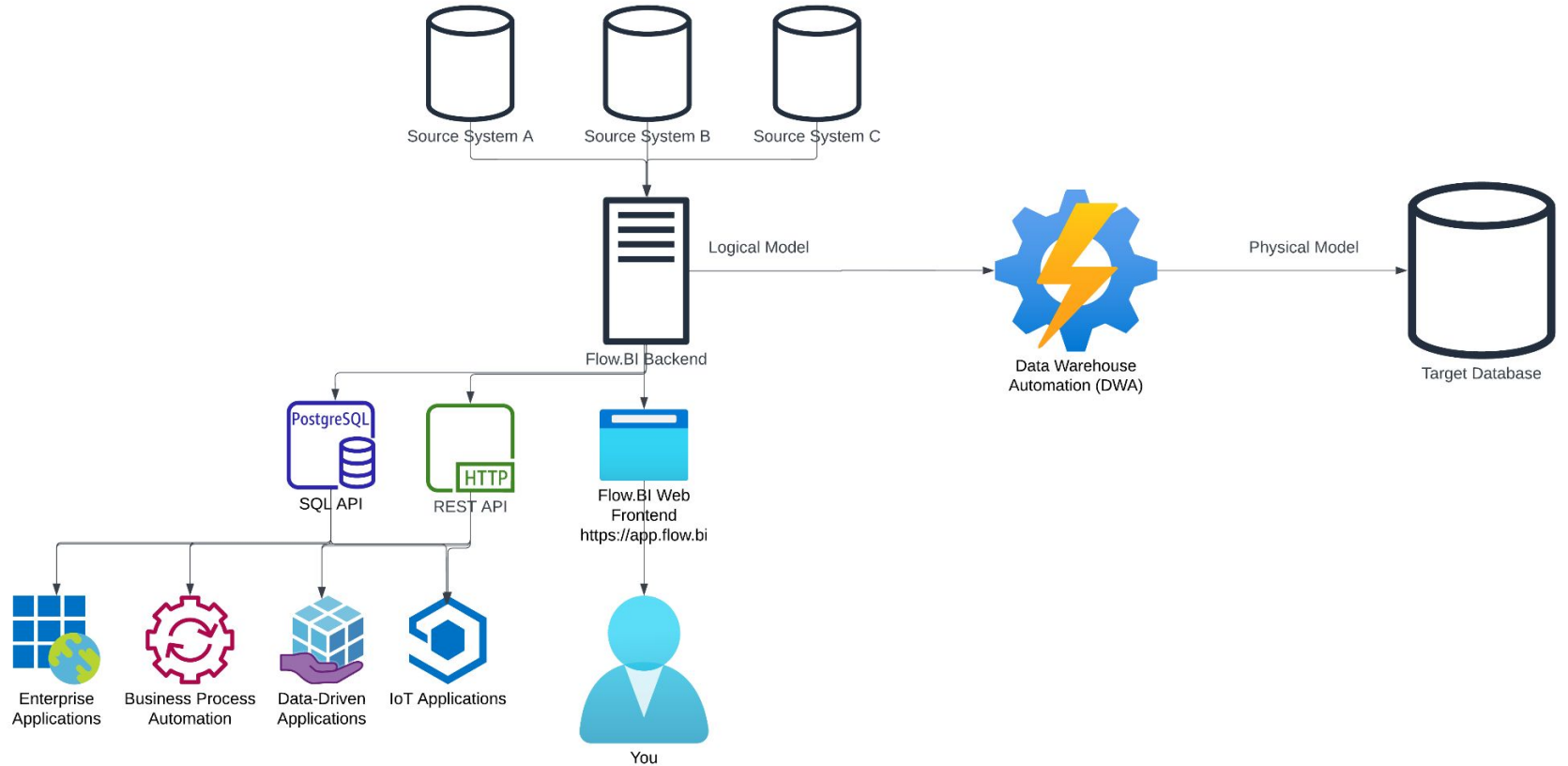
Das **physische Modell** wird durch ein Data Warehouse automationstool wie beispielsweise

- Datavault Builder
- dbt
- Iris
- ...

generiert.



Architecture Overview



APIs

REST API

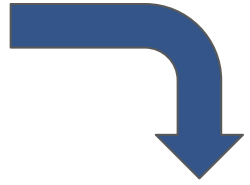
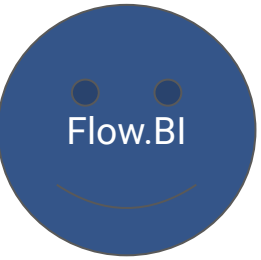
- API-first-Ansatz
- REST API bietet die Funktionalität der GUI
- "Script everything"
- Jede Aktion in der Benutzeroberfläche kann geskriptet werden
- REST-API ermöglicht Aktualisierungen, Löschen usw.

SQL API

- Indirekter Zugriff auf das Backend-Repository
- Data Vault 2.0-Modell mit vollständiger Historie
- Semistrukturiert: erwartet JSON, XML, Arrays
- PostgreSQL-basiert
- Nur Lesezugriff

dbt Integration

- dbt ist eine Open Source Lösung (Freemium) welches DV generieren kann
- Basiert auf Python



Funktionen

- Der Benutzer verbindet die **JDBC-Datenquelle** (im Wesentlichen ohne Einschränkung, aber die Treiber müssen vom Flow.BI-Team implementiert werden)
- Flow.BI identifiziert **Konzepte** wie z.B. Kunde, Produkt, Filiale usw. in den angeschlossenen Datenquellen
- Flow.BI generiert die **Metadaten des integrierten logischen Modells** (über alle angeschlossenen Datenquellen)
 - Flow.BI entscheidet über spezifische Data Vault Entitäten
- **Integration mit DWA** zur Erstellung des physischen Modells

Was sind die Vorteile?

- Reduzierung des technischen Entwicklungsaufwands
- Konzentration der Experten auf das Schaffen von Geschäftswerten
- Prozess der Datenlieferung:
 - effektiver
 - schneller
 - kostengünstiger
- Kürzere Zeit bis zur Auslieferung
- Weniger Risiko

Flow.BI macht Fehler!

- Flow.BI **imitiert den menschlichen Datenmodellierer einschließlich menschlicher Fehler**
- Anstatt dies zu verhindern, hat das Modell Vorkehrungen für den Umgang mit solchen Fehlern
- Flow.BI erzeugt **immer ein gültiges Data-Vault-2.0-Modell**
 - Solche Modelle können refactored werden
 - Ableitung eines besseren Modells im Business Vault möglich

Flow.BI erlaubt keine eigene Modellierung!

- Die Absicht von Flow.BI ist es, die **Modellerstellung vollständig zu übernehmen**
- Keine Notwendigkeit, das Modell des Raw Data Vault zu ändern
- Die Modellierung sollte im Business Vault durchgeführt werden
- Keine Pläne zur Einführung einer Modellierungs-UI für Raw Data Vault

Ist es möglich,
10.000
Quelltabellen
zu verarbeiten?



Hochskalieren

Flow.BI erzeugt in kurzer Zeit 10.000 MPP-Knoten und mehr.

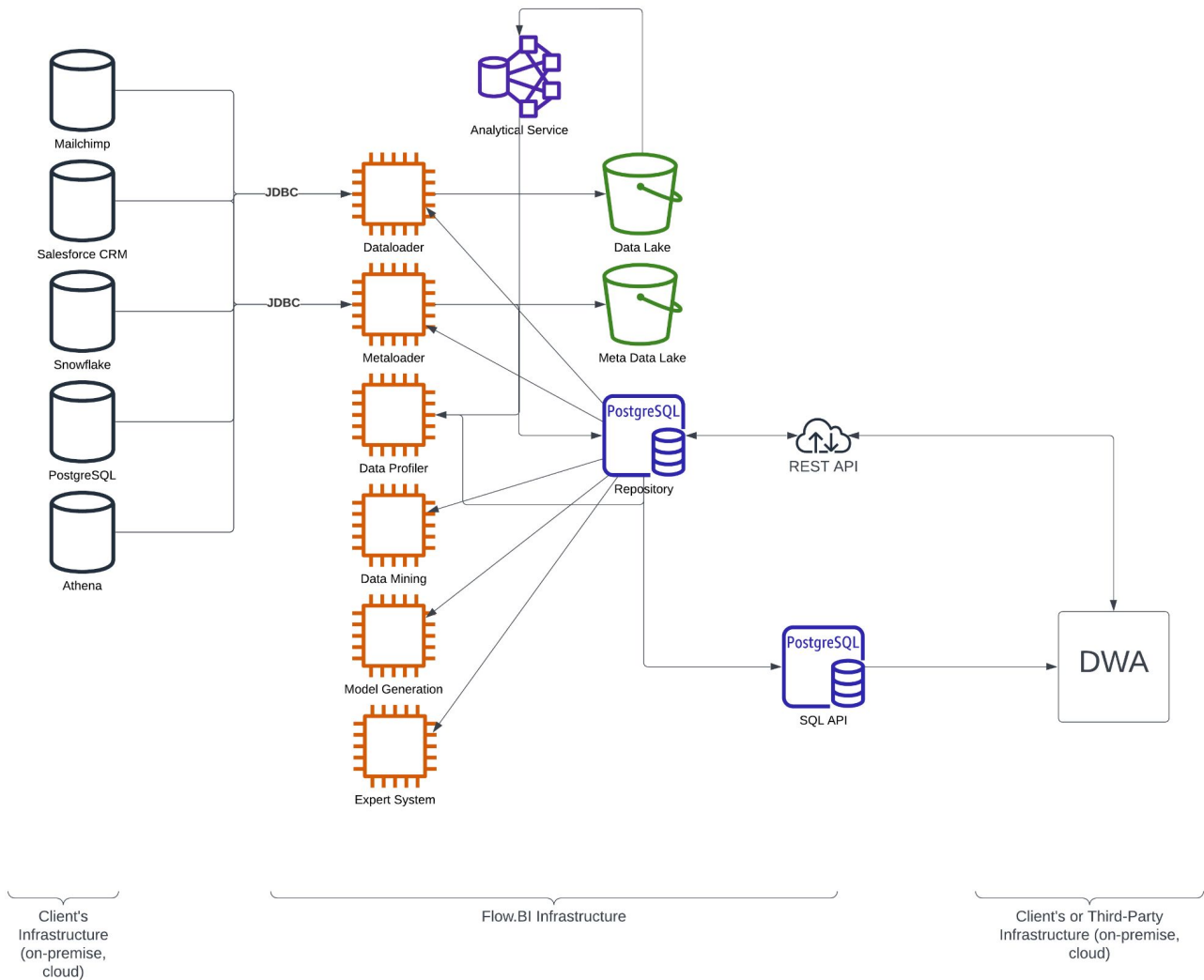


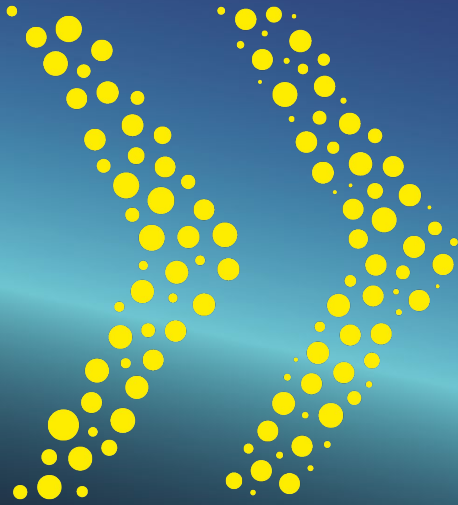
Flow.BI kann skalieren!

- Flow.BI basiert auf **massiver Parallelverarbeitung (MPP)**
- Das Backend in der Cloud skaliert, um Daten so schnell wie möglich zu verarbeiten
- Flow.BI ist darauf ausgelegt, alle Quelltabellen zu analysieren und daraus ein integriertes Geschäftsmodell abzuleiten.
- Dabei spielt es keine Rolle, wie viele Quelltabellenstrukturen analysiert und integriert werden.
- Stückweise hochskalieren

System Architecture

- Mehrere MPP-Cluster
- Alles skalierbar
- Alles ist geschlossen, außer
 - Repository
 - Generierte Ergebnisse





Das Ergebnis ist ein integriertes Unternehmensmodell, das die Rohdaten einer Organisation beschreibt.

Generated Metadata

intf_automation

- tables 1
- views 40
 - attribute_loads
 - attribute_loads_desc
 - attributes
 - attributes_desc
 - config
 - entities
 - entities_desc
 - entity_loads
 - entity_loads_desc
 - hub_loads
 - hub_loads_desc
 - hubs
 - hubs_desc
 - key_loads
 - key_loads_desc
 - keys
 - keys_desc
 - link_loads
 - link_loads_desc
 - link_reference_loads
 - link_reference_loads_desc
 - link_references
 - link_references_desc
 - links
 - links_desc
 - reference_table_loads
 - reference_table_loads_desc
 - reference_tables
 - reference_tables_desc

Different assets generated by Flow.BI

	entity_name	column_name	...	source_table	source_column	gsr_sdts
233	NHub Party	Party 2	CData	Salesforce	Account	2023-09-22 10:00:00.00000
234	Hub Party	Party 2	CData	Salesforce	Account	2023-09-22 10:00:00.00000
235	NHub Report Domain Perfo...	Report Domain Perform...	CData	MailChimp	ReportDomainPerformance	2023-09-21 10:00:00.00000
236	Hub Report Domain Perfor...	Report Domain Perform...	CData	MailChimp	ReportDomainPerformance	2023-09-21 10:00:00.00000
237	NHub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
238	Hub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
239	NHub Report Sent To	Report Sent To 1	CData	MailChimp	ReportSentTo	2023-09-21 10:00:00.00000
240	Hub Report Sent To	Report Sent To 1	CData	MailChimp	ReportSentTo	2023-09-21 10:00:00.00000
241	NHub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
242	Hub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
243	NHub Accounting Activity	Accounting Activity 2	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
244	Hub Accounting Activity	Accounting Activity 2	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
245	NHub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
246	Hub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
247	NHub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
248	Hub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
249	Hub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
250	NHub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
251	Hub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
252	NHub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
253	Hub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
254	NHub Time Card Accountin...	Time Card Accounting ...	CData	Salesforce	TimeCardAccounting	2023-09-21 10:00:00.00000
255	Hub Time Card Accounting...	Time Card Accounting ...	CData	Salesforce	TimeCardAccounting	2023-09-21 10:00:00.00000
256	NHub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
257	Hub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
258	NHub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
259	Hub Accounting Activity	Accounting Activity 1	CData	Salesforce	AccountingActivity	2023-09-21 10:00:00.00000
260	NHub Communication Event	Communication Event 1	CData	Salesforce	ActiveFeatureLicenseMetric	2023-09-21 10:00:00.00000
261	Hub Communication Event	Communication Event 1	CData	Salesforce	ActiveFeatureLicenseMetric	2023-09-21 10:00:00.00000
262	NHub Communication Event	Communication Event 1	CData	Salesforce	ActivePermSetLicenseMetric	2023-09-21 10:00:00.00000
263	Hub Communication Event	Communication Event 1	CData	Salesforce	ActivePermSetLicenseMetric	2023-09-21 10:00:00.00000
264	NHub Communication Event	Communication Event 1	CData	Salesforce	ActivePermSetLicenseMetric	2023-09-21 10:00:00.00000

Hubs...

... and their business key definitions...

... but also their source tables (fully qualified)...

Yes, versioned to keep track of changes.

...and the source columns to define both the model and the loading!

Live Demo



Learn more
about Flow.BI:

www.flow.bi
info@flow.bi