

ARGUS DATA LAKE

WIE GELINGEN KOMPLEXE ANALYSEN VON RIESIGEN DATENMENGEN?



INHALTSVERZEICHNIS

1 ARGUS DATA INSIGHTS SCHWEIZ AG

2 DEMO EXPLORATIVE DATENANALYSE

3 VISION

4 ARCHITEKTUR

5 ERFAHRUNGEN MIT DER CLOUD

6 ERFAHRUNGEN IM PROJEKT

7 IDEEN FÜR DIE ZUKUNFT

WAS IST DER ARGUS DATA LAKE

Daten zentral abgelegt

- Grundlage für Datenauswertung basierend auf Mediendaten.
- Befüllt aus Medienarten und -Kanälen die ARGUS DATA INSIGHTS Schweiz AG zur Verfügung stehen.

Daten angereichert (Enrichments)

- Angereichert um Informationen, welche sich aus den Daten mit Hilfe weiterer Quellen ergeben.
- Angereichert um Informationen, welche sich aus der Betrachtung der Daten insgesamt ergeben. Dazu kommen neben altbekannten statistischen Methoden auch Technologien wie NLP und ML zum Einsatz.

VISION

KUNDENNUTZEN & USE CASES

ARGUS Data Lake für Analytics Swiss Media unterstützt bei der Erkennung von Trends und Entwicklungsfähigkeit von Themen in Schweizer Medieninhalten.

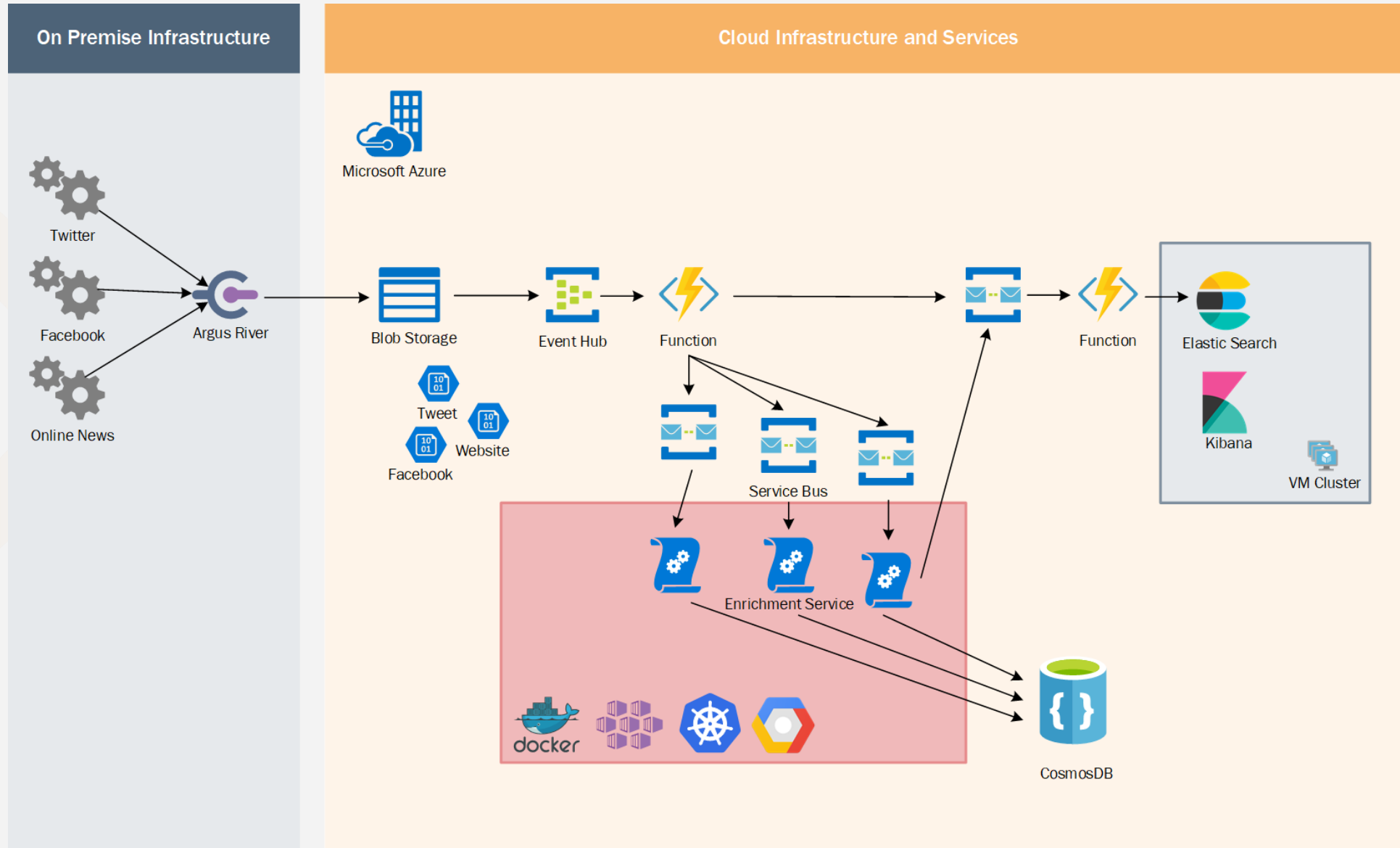
Der Fokus liegt anfänglich auf Inhalten aus Online News und Social Media.

Weitere Kanäle werden nach und nach integriert.

- Neu können **Analysen auf historischen Daten** erstellt werden.
- Die Qualität der **gewonnenen Aussagen** ist **besser als bei globalen Anbietern**, da die Methoden stärker auf die Schweiz ausgerichtet sind.

TREIBER FÜR ARCHITEKTURENTSCHEIDE

- **Erweiterbarkeit** für neue
 - Enrichments
 - Kanäle oder Quellen
 - Interessante Daten
 - Produkte
- **Skalierbarkeit**
 - Alle Dienste sind manuell oder automatisch skalierbar.



MENGENGERÜST

Initialer Ausbau (IST)

- ~2 TB / Jahr
- 7 Mio Items / Monat -> ~3 Items/s
- 10 Enrichments

Finaler Ausbau (Idee)

- 92 TB / Jahr
- 75 Mio Items / Monat -> ~30 Items/s
- 10++ Enrichments

ERFAHRUNGEN MIT DER CLOUD

- **Ressourcenknappheit**
- **Pricing**
 - Pricing je nach Service schwierig → total unbekannt
 - Preise der einzelnen Services teilweise überraschend (+/-)
 - Verträge führen teilweise zu erheblichen Rabatten
 - Nicht zwingend günstiger als on-premises
- **Flexibilität, Prototyping**
 - Cloud-Umzug in 30 Minuten

ERFAHRUNGEN IM PROJEKT

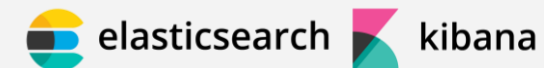
Plan 1: Data Warehouse & Power BI

- Ziel: Minimierung den Entwicklungskosten im Frontend Bereich
- Erkenntnis: Nicht für solch grosse Datenmengen geeignet



Aktuell:

- Elasticsearch & Kibana: Datenbank optimiert für Suche und Aggregationen
- Benutzerschnittstelle für Analysten (Experten)
- Kibana ist KEINE Benutzerschnittstelle für Endkunden



AUSBLICK

IDEEN FÜR DIE ZUKUNFT

- Weitere Datenquellen einbinden
- Data Lake als Content Provider
- Data Lake als Plattform für bestehende Prozesse
- Erweiterung durch zusätzliche Anreicherungsschritte
- Managed NLP / AI weiterhin beobachten



 **ARGUS DATA INSIGHTS®**
WISSEN ZUM ERFOLG