

Startup Day with AWS and kreuzwerker

21.03.23, Berlin, Alte Münze

PROJECT
BLUEPRINT BERLIN

DRAWING NAME
CLIMB THE WALL

DESIGNED BY
O.U. STUDIO
Lin-Fee Chang 打明聯合
Chen-Hsiao-Tzu
Wu-Yi-Lin
Tsai-Pai-Huan

SCALE
1/1

DATE 2009/6/4

SHEET NO
1

DRAWING NO
A1-1



2010

Gründung der
kreuzwerker
GmbH



2013

Gründung der
kreuzwerker
Polska, Warschau



2017

Gründung der
kreuzwerker
Switzerland
AG, Zürich



2019

Eröffnung der
Niederlassung
in München



2020

Eröffnung der
Niederlassung
in Frankfurt
a.M.

A close-up photograph of a person's hands interacting with a silver laptop. One hand is pointing at the screen, while the other rests on the trackpad. A red semi-transparent banner is overlaid across the center of the image, containing white text. The background is blurred, showing a desk and a chair.

We advise, coach, plan, ...

... and implement

The background of the slide is a vibrant display of fireworks exploding against a black night sky. The fireworks are in various colors, including bright red, orange, yellow, and green. Some are large, spherical bursts, while others are more elongated and starburst-like. The trails of the fireworks are visible as thin lines of light.

**We like new challenges and
exciting problems**

Get your data in the Data Lake and derive valuable insights using NLP

Saba Sabrin

Team Manager - Data and Analytics,
kreuzwerker

Who am I?



Saba Sabrin

10 years of love with Software and Data engineering

Education

- B.Sc. in Computer Science & Engineering
- M.Sc. in Distributed Software System

Passion

- Building, Programming and Researching
- Music, history, traveling, exploring, fun-loving and being Mindful!



If you think of a data mart as a store of bottled water – cleansed and packaged and structured for easy consumption – the data lake is a large body of water in a more natural state. The contents of the data lake stream in from a source to fill the lake, and various users of the lake can come to examine, dive in, or take samples.

- Jamex Dixon (*CTO, Pentaho – 14 October 2010*)

Why AWS for your Data analytics?



Self-Service



Scalability



Security



Cost



Automation

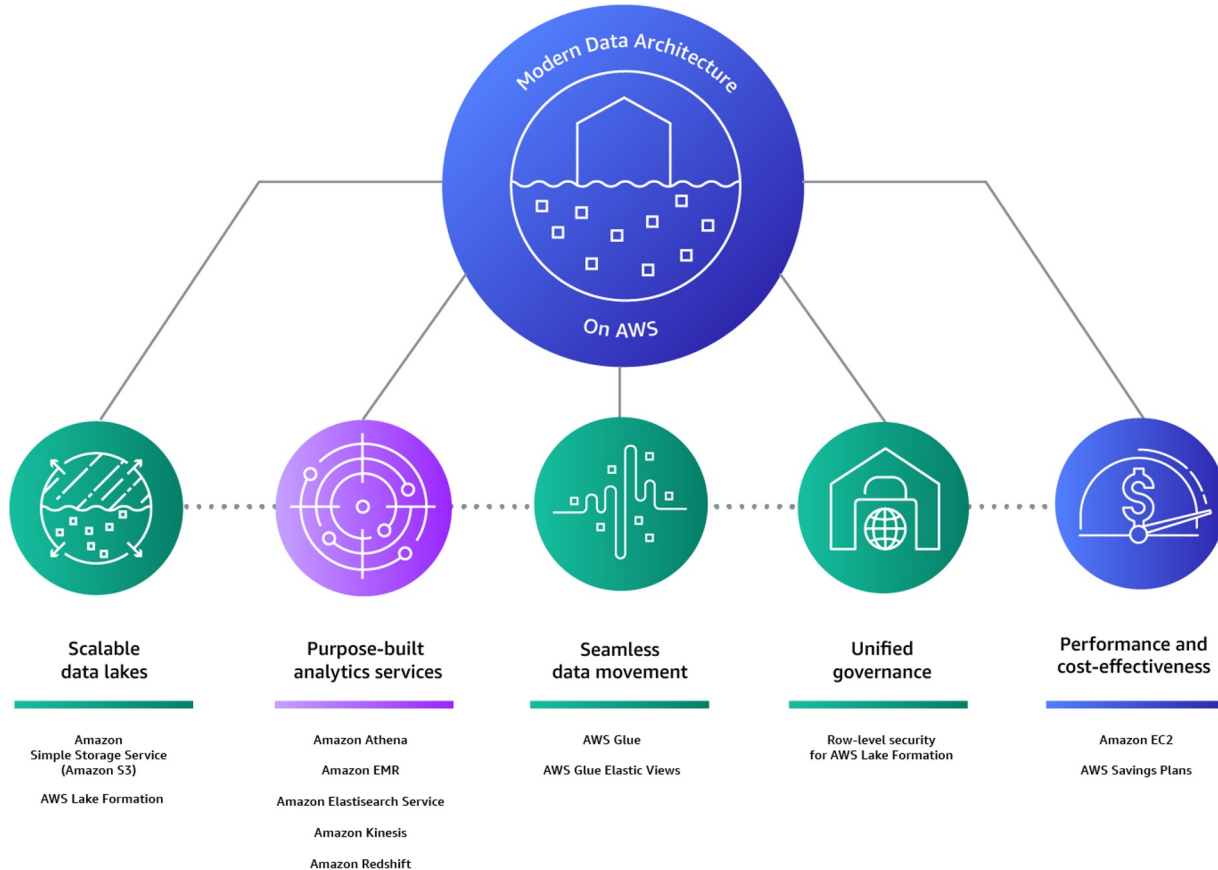


Innovation

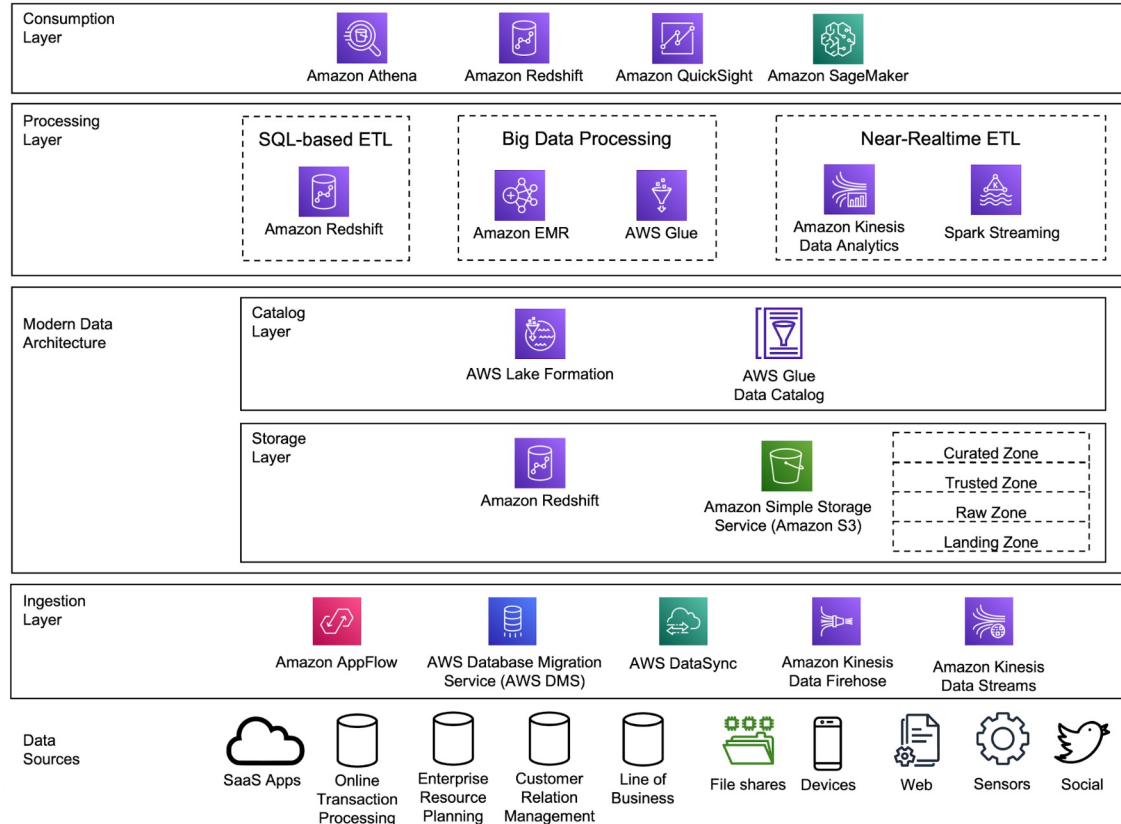


Sustainability

Time for Modernization



Modern Data Architecture



Where is your Data?

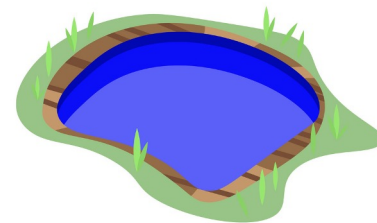
- Sea or Lake or Pond?
- Get your data from Sea and move into the Lake and create rich "Ponds"



Distribute
d Data (Big
Data)



Data Lake
(Amazon S3)



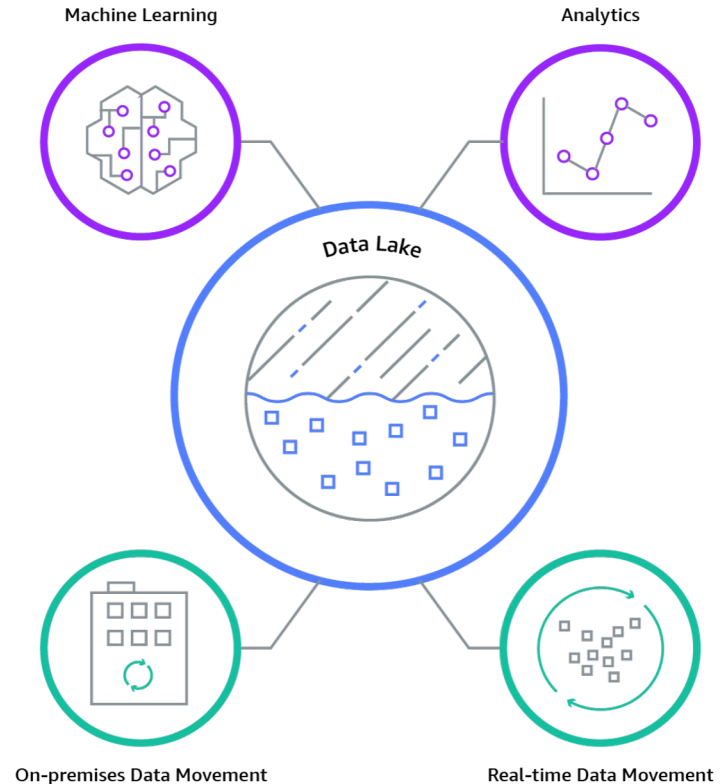
Data Analytics
(AWS Redshift,
Athena)

Why Data Lake?

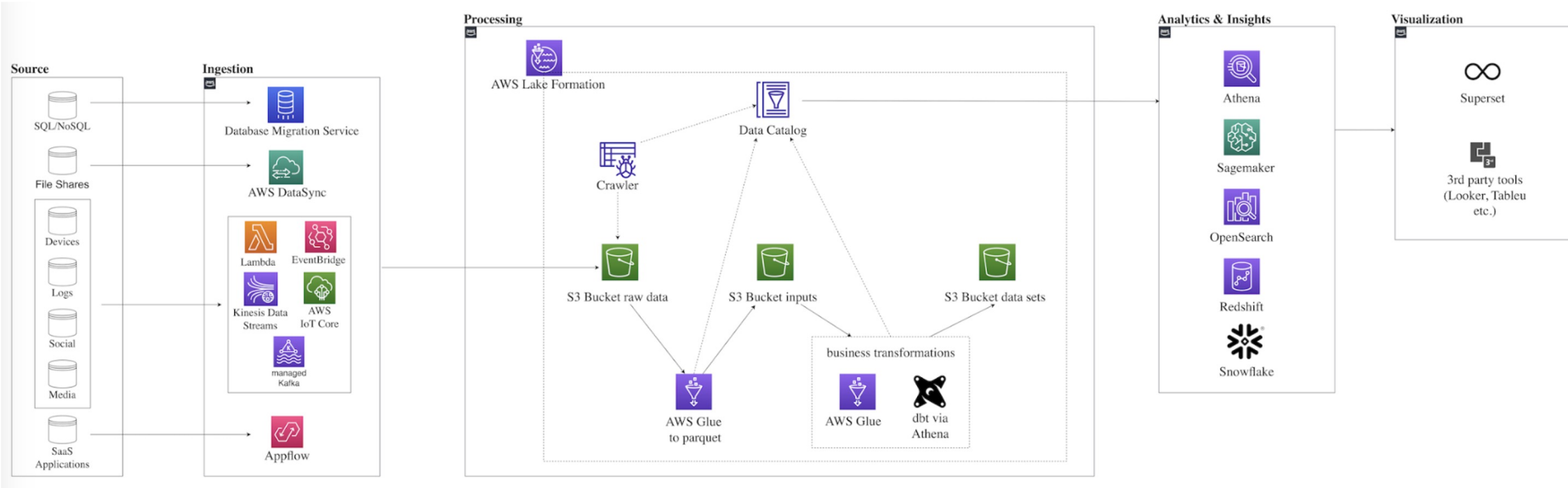
- Business reason
 - Generate Value from data to increase your Revenue
 - Predict future and act wisely!...
- Technology reason
 - Unified place for all the “data needs”
 - Get rid of “Data Silo” and “Data Swamp”
 - Build the Data Platform

Amazon Data Lake

- Identify data sources
- Move the Data workload
- Define management and Life-Cycle Policy
- “Analyse” with the power of Analytics and AI (ML, NLP)



- A Data Analytics Platform
 - Distributed system architecture
 - Scalable **data lake**
 - Supports built in data transformation with **dbt** and **AWS Athena**
 - Developed on top of **AWS Glue** and
 - Batch Data Processing (Available in alpha version)
 - Deployment Stack using AWS CDK



Data Source:
Any/All

Raw Data Type:
Object

Intermediate Data Type: Parquet

Stay tuned for Upcoming Features



- Real-time (Streaming) Data Processing
- Large Scale Parallel data processing
- Data Warehouse and AI integrations
- Visualization integrations with Tableau, Quicksight

Possibilities with Data Lake

- Build
 - Data Product for your business use case
 - Data Service for cross platform use case
- Analyse data
 - Data Warehousing
 - Natural Language Processing
 - Machine Learning

The art of interaction between computer and human language with the help of Linguistics and Machine Learning.

- Amazon Textract
- Amazon Translate
- Amazon Comprehend
- Integrate with Open source NLP toolkits (NLTK, Stanford Core NLP)

Natural Language Processing

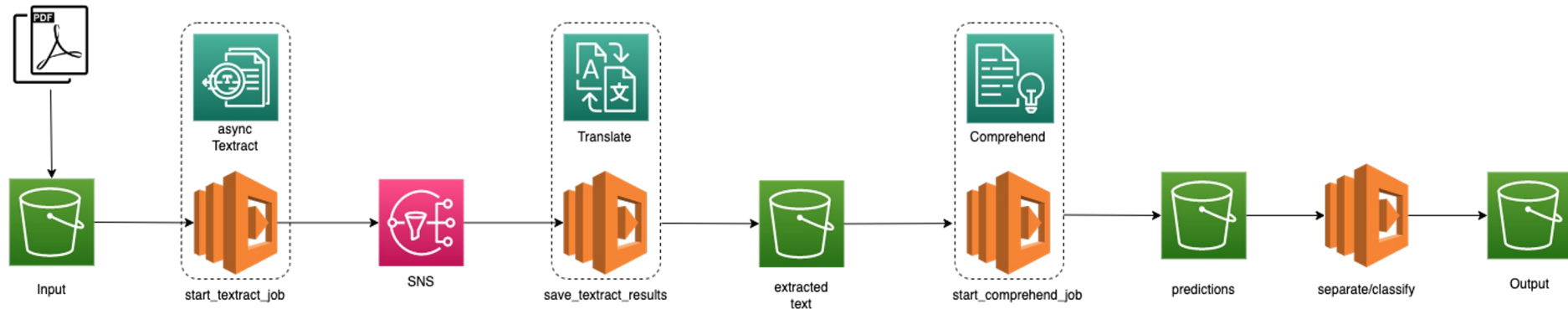
The art of interaction between computer and human language with the help of Linguistics and machine learning.

Amazon Textract can extract text, handwriting and form data from scanned/digital Documents.

Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text.

NLP in Action

- A Logistics case study for Document analytics



Summary

- Bring your Use-Case
 - **OR** build your Use-Case!
- Start small and **Scale-Up** gradually
- Derive meaningful **Value** and keep improving!

Thank You!

Questions? saba.sabrin@kreuzwerker.de

Looking forward to unlock the Future!

Please take a
moment and rate
your experience -
>



Thank you for attending our Startup Day!



kreuzwerker

kreuzwerker GmbH
Ritterstr. 12
10969 Berlin
aws@kreuzwerker.de

www.kreuzwerker.de
Fon +49 30 609 838 80
Fax +49 30 609 838 899