BI in Azure one year live

Alexander Klein
Twitter: @SQL_Alex

SQLSaturday #865 - Rheinland

25.05.2019

SQLSaturday #880 - Munich
19.10.2019

SQLSATURDAY
✳ PASS

PASS
Microsoft
Data Platform
Community
DEUTSCHLAND e.V.

Who am I?

Independent BI Consultant

> 15 years experience of SQL Server

Focus on Microsoft BI Stack & Azure & AI

✉ a.klein@consulting-bi.de

📧 @SQL_Alex

🏠 consulting-bi.de

Next 60 Minutes

- BI ?
- Goal
- Toolbox
- How to
- Conclusion

Some theory first: "Modern" Data Warehouse

What is a Data Warehouse?

- Central component of Business Intelligence

- Stores data from different data sources

- Stores actual and historical data

- Offers an intuitively understandable data model

- Optimized for read performance

- Distinction between narrow definition of "Data Warehouse" as a data model/store and "Data Warehouse" as a system, containing also the ETL process and the data dictionary in a broader context

Different approaches

- Kimball (Dimensional approach)
  - Requirement-driven bottom-up approach
  - Start by identifying business needs and create data model as combination of data marts
  - Then try to identify data needed to answer those questions and transform into data model
  - Complexity lies in ETL process and effort caused by change requests to the data model
- Inmon (Normalized approach)
  - Data-driven top-down approach
  - Design data model from existing data
  - Business logic needed to find ways to answer business questions
  - Complexity lies in complexity of data extraction → Data Marts should be created
- Data Vault (Hub and spoke approach)
  - Stores every version of every piece of data
  - Designed to include every change in the data model
  - Complexity lies in massive growth and complexity of data extraction → Data Marts should be created
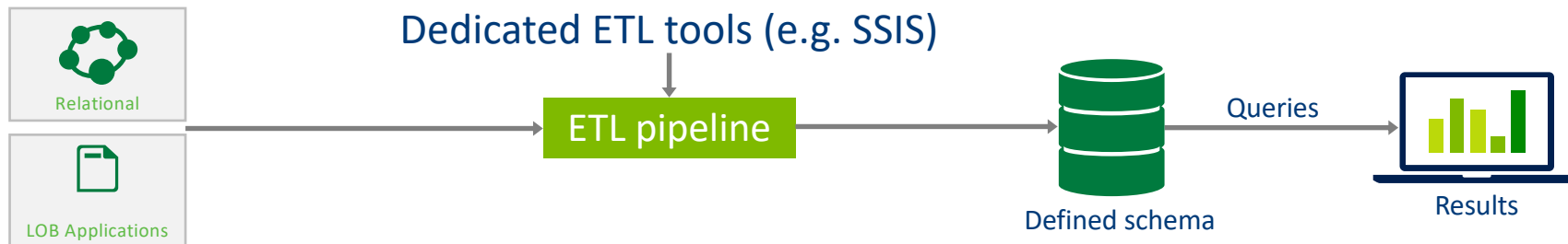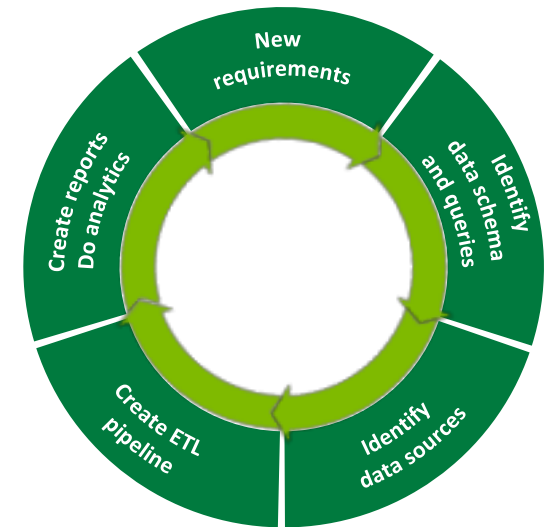
Important keywords

- Subject-oriented
  - to facilitate support of decision making processes
- Integrated
  - Removal of inconsistencies between data from different sources
- Time-variant
  - Long time horizon
- Nonvolatile
  - No creation, updates or deletion of data
- Summarized
  - Aggregations at different levels

Traditional business analytics process

1. Start with end-user requirements to identify desired reports and analysis

2. Define corresponding database schema and queries

3. Identify the required data sources

4. Create a Extract-Transform-Load (ETL) pipeline to extract required data (curation) and transform it to target schema ('*schema-on-write*')

5. Create reports. Analyze data

Relational

LOB Applications

Dedicated ETL tools (e.g. SSIS)

ETL pipeline

Defined schema

Queries

Results

All data not immediately required is discarded or archived

Why do we need "Modern"?

- Requirements for BI processes are changing
  - Integration of massively increasing amounts of data
  - Increasing velocity in addition
  - Variety of data is changing
    - Past: Structured data
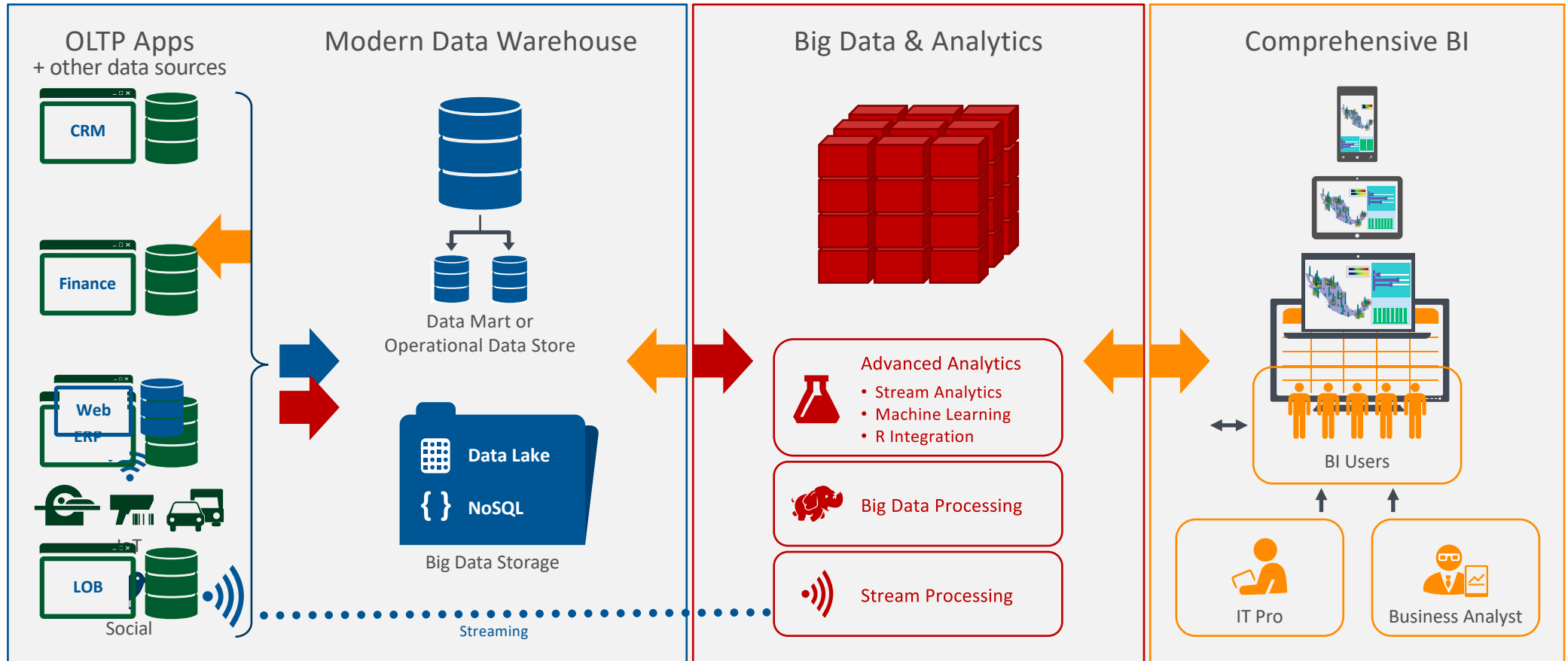   - Today: Structured, Unstructured, Semi-structured

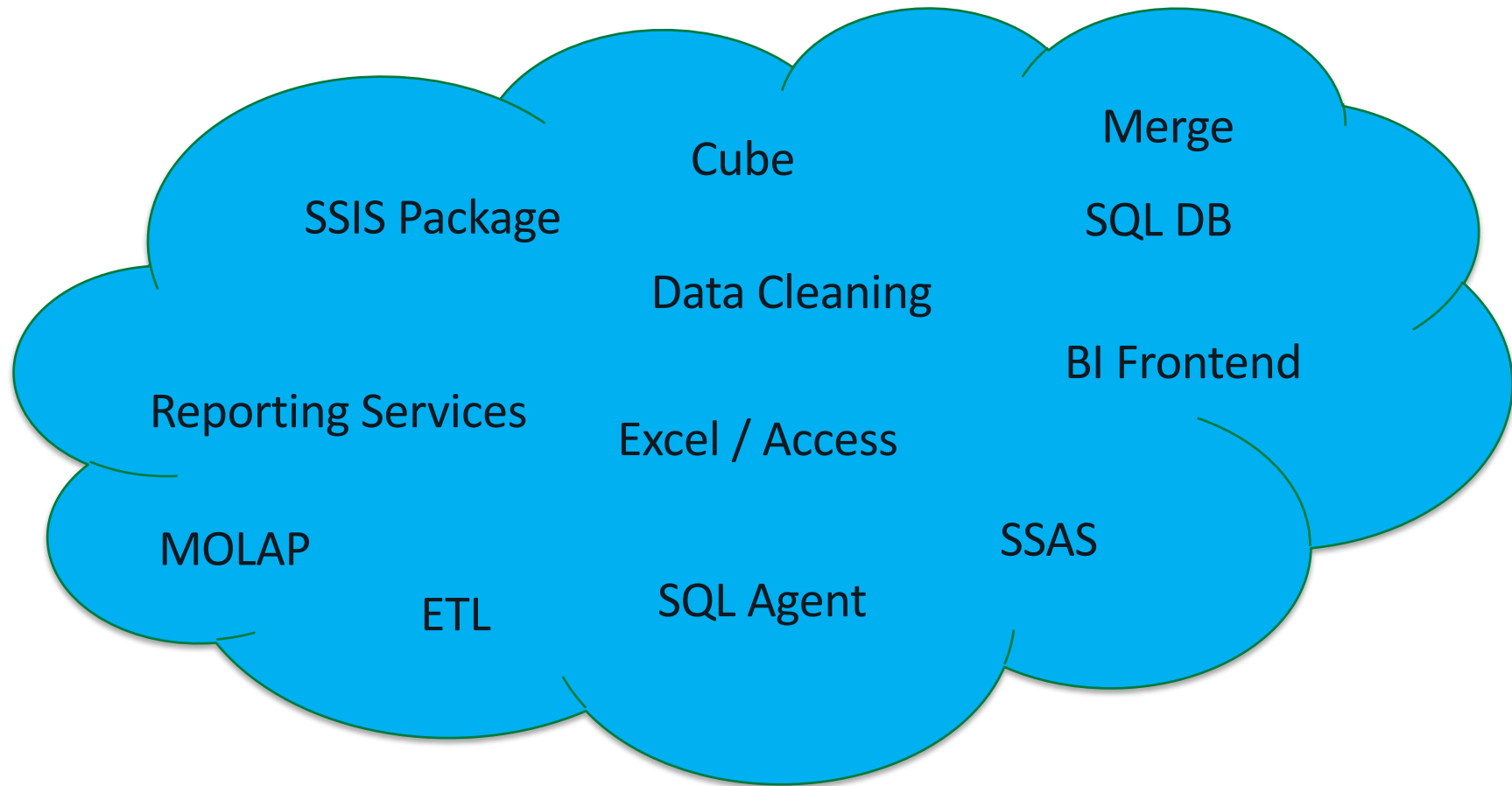# New big data thinking: All data has value

⚡ All data has potential value

⚡ Data hoarding

⚡ No defined schema—stored in native format

⚡ Schema is imposed and transformations are done at query time *(schema-on-read).*

⚡ Apps and users interpret the data as they see fit

```
Gather data          Store indefinitely          Analyze          See results
from all sources
```

Iterate

Corporate BI: Transformative data architecture

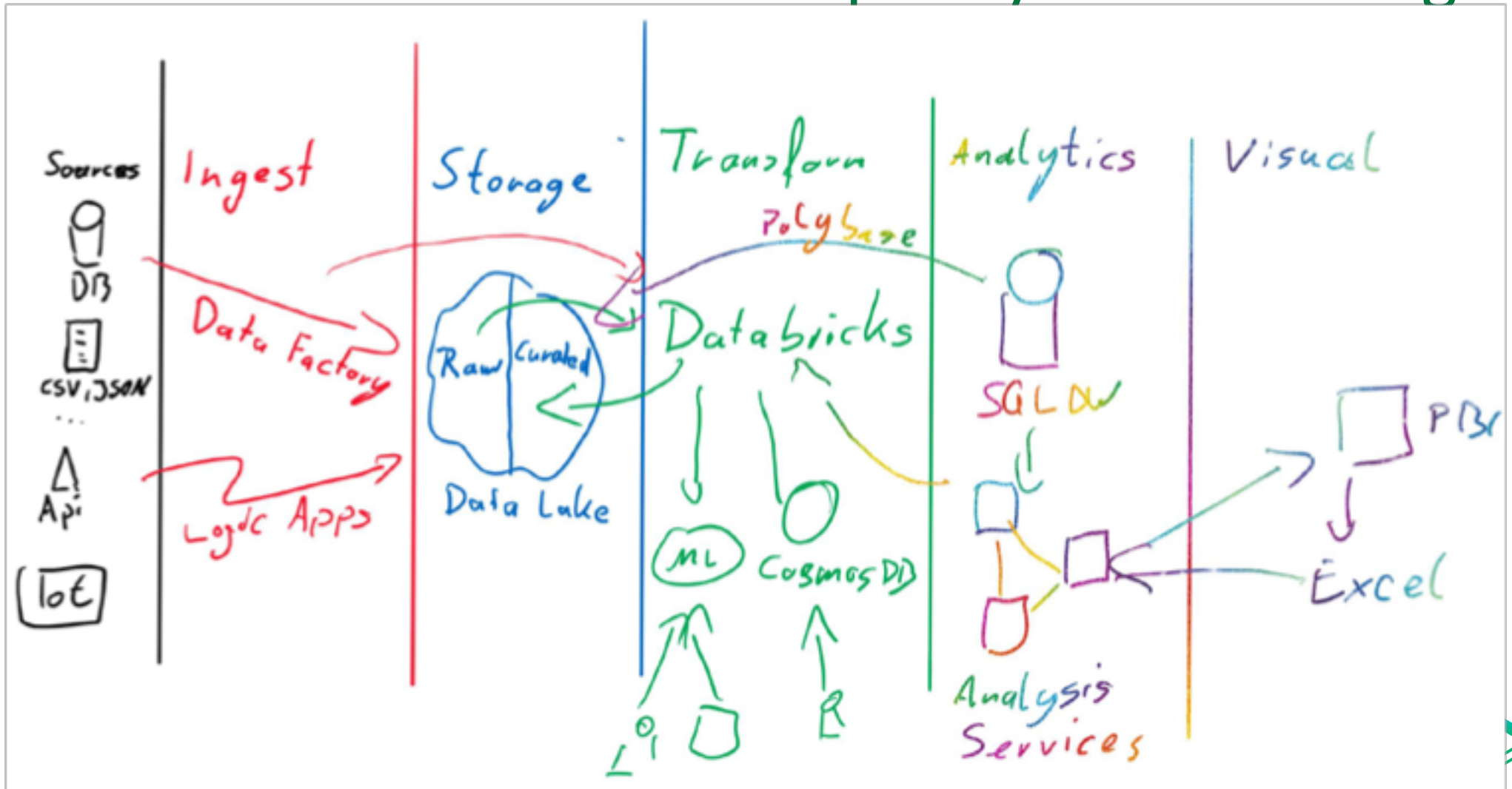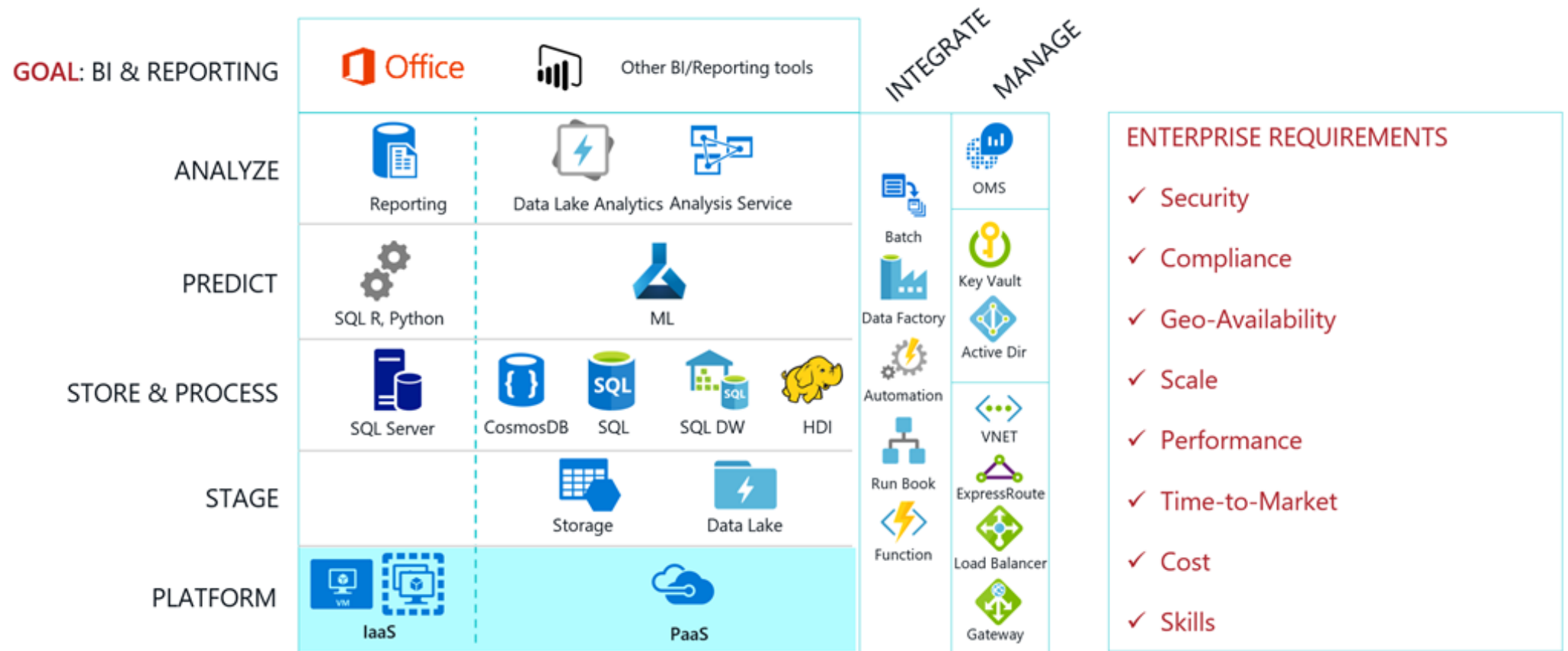# On Prime (classic)

Cloud

Goal

- Consolidation of Systems

- New IT System (Azure)

- Proses reengineering

- Paperless

- "Digitalisierung"

Toolbox BI in Azure – Example by Andre Essing

Toolbox BI in Azure

GOAL: BI & REPORTING — Office — Other BI/Reporting tools

INTEGRATE — MANAGE

ANALYZE	Reporting — Data Lake Analytics — Analysis Service	OMS
PREDICT	SQL R, Python — ML	Batch — Key Vault
STORE & PROCESS	SQL Server — CosmosDB — SQL — SQL DW — HDI	Data Factory — Active Dir
STAGE	Storage — Data Lake	Automation — VNET
PLATFORM	IaaS — PaaS	Run Book — ExpressRoute — Function — Load Balancer — Gateway

ENTERPRISE REQUIREMENTS

✓ Security

✓ Compliance

✓ Geo-Availability

✓ Scale

✓ Performance

✓ Time-to-Market

✓ Cost

✓ Skills

https://azure.microsoft.com/de-de/blog/technical-reference-implementation-for-enterprise-bi-and-reporting/

Toolbox

	Classic	Cloud
Store	SQL Server DB	Azure SQL DB Azure SQL DW Azure Data Lake
ETL	SSIS	Data Factory Logic App Stream Analytics Data Lake Analytics
Schedule	SQL Agent	Azure Automation
Cube	SSAS	Azure Analysis Services
Analyze & Reporting	SSRS / Excel	Power BI

ETL - Azure Logic App

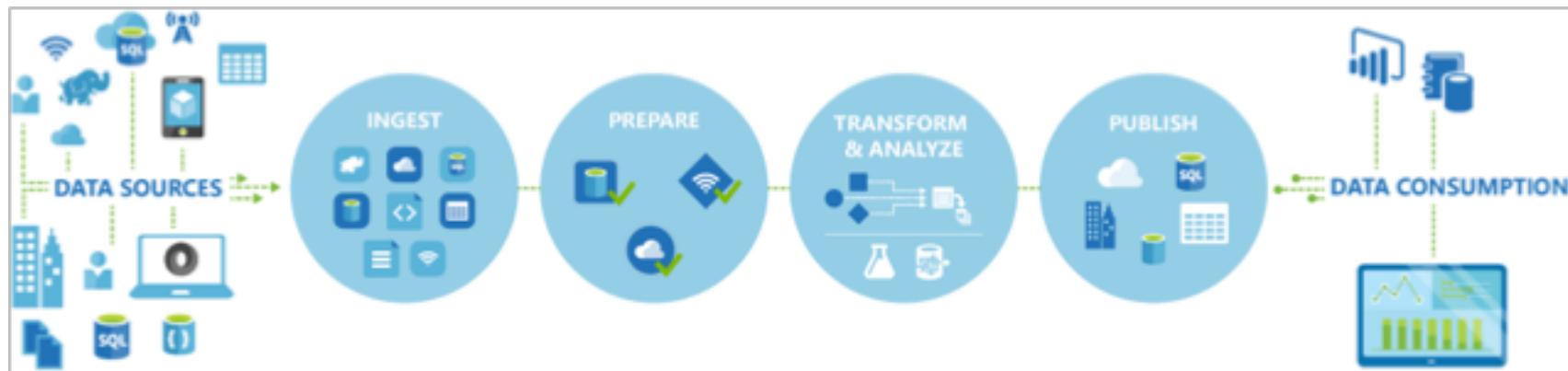iPaaS (integration Platform as a Service)

Workflow

Connectors

Trigger

Actions

Enterprise Integration Pack

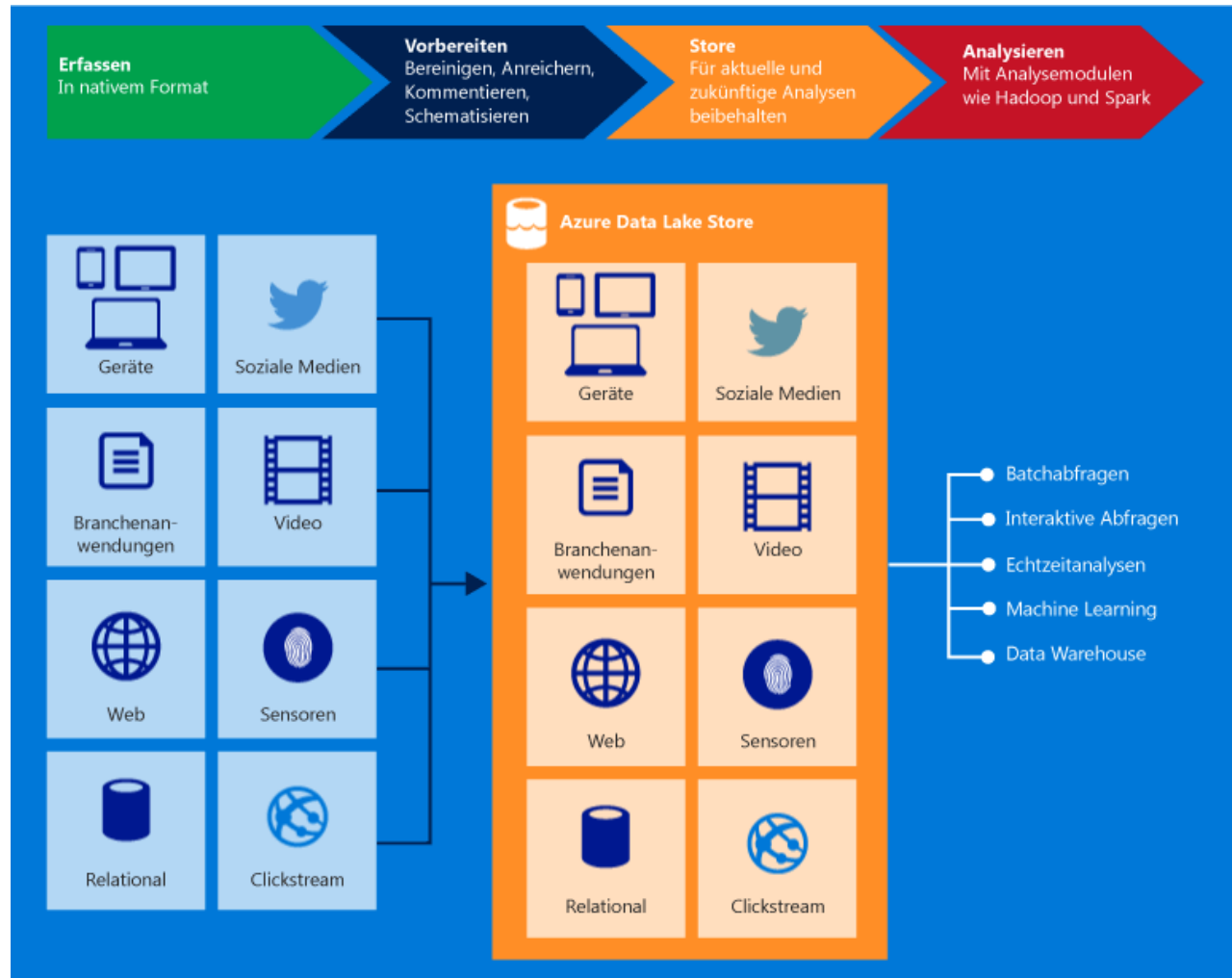# ETL - Azure Data Factory (ADF)

Cloud-based data integration service that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation.
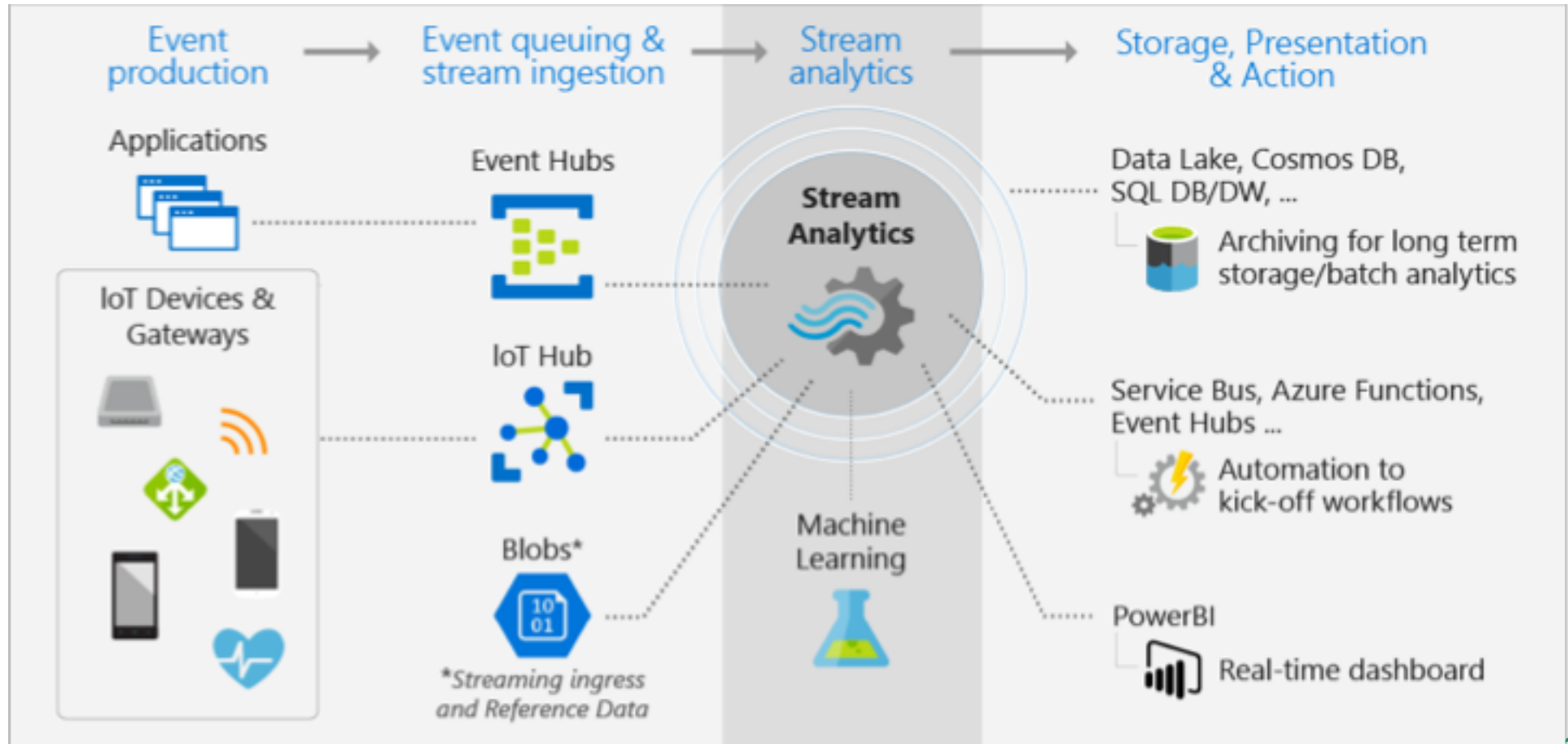
ETL - Azure Data Factory (ADF)
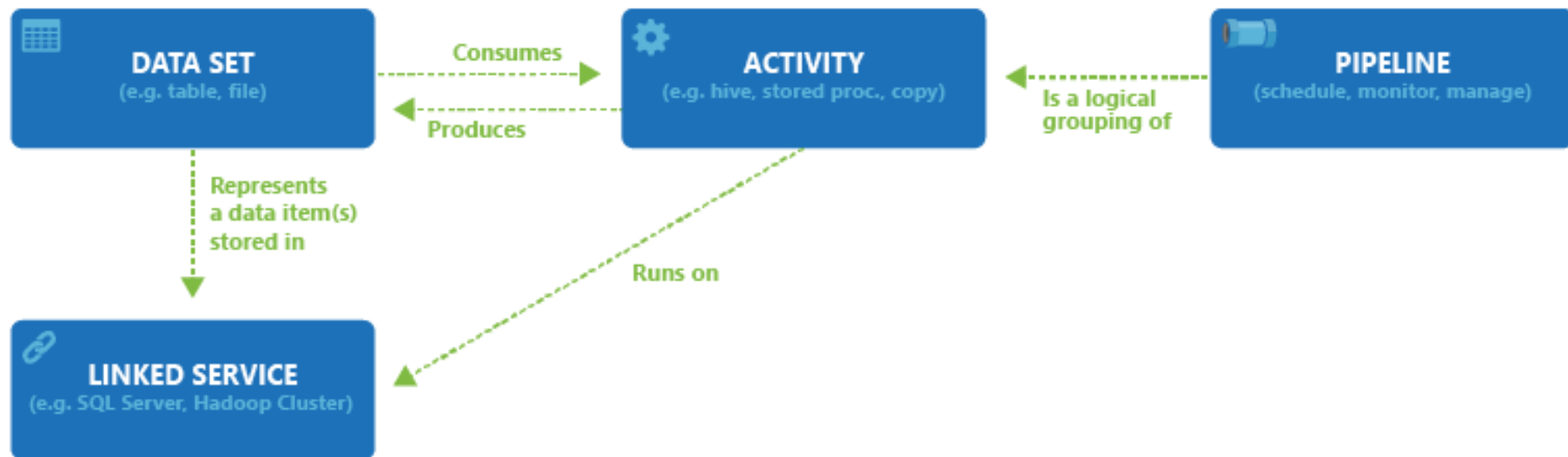
ETL - Azure Data Lake (ADL)

ETL - Azure Stream Analytics (ASA)

Event production → Event queuing & stream ingestion → Stream analytics → Storage, Presentation & Action

Applications · Event Hubs

IoT Devices & Gateways · IoT Hub · **Stream Analytics** · Machine Learning

Blobs*

*Streaming ingress and Reference Data

Data Lake, Cosmos DB, SQL DB/DW, ... — Archiving for long term storage/batch analytics

Service Bus, Azure Functions, Event Hubs ... — Automation to kick-off workflows

PowerBI — Real-time dashboard

https://docs.microsoft.com/en-in/azure/stream-analytics/stream-analytics-introduction

ETL - Azure Function

Azure Functions is a solution for easily running small pieces of code, or "functions," in the cloud. You can write just the code you need for the problem at hand, without worrying about a whole application or the infrastructure to run it.

Azure Automation

Azure Automation is a software as a service (SaaS) application that provides a scalable and reliable, multi-tenant environment to automate processes with runbooks and manage configuration changes to Windows and Linux systems using Desired State Configuration (DSC) in Azure, other cloud services, or on-premises.

Azure Runbook

Types:

graphical runbook

PowerShell runbook

Deep learing, machine learing and AI

Deep
learing

Machine
learing

AI

AI .. ML ?

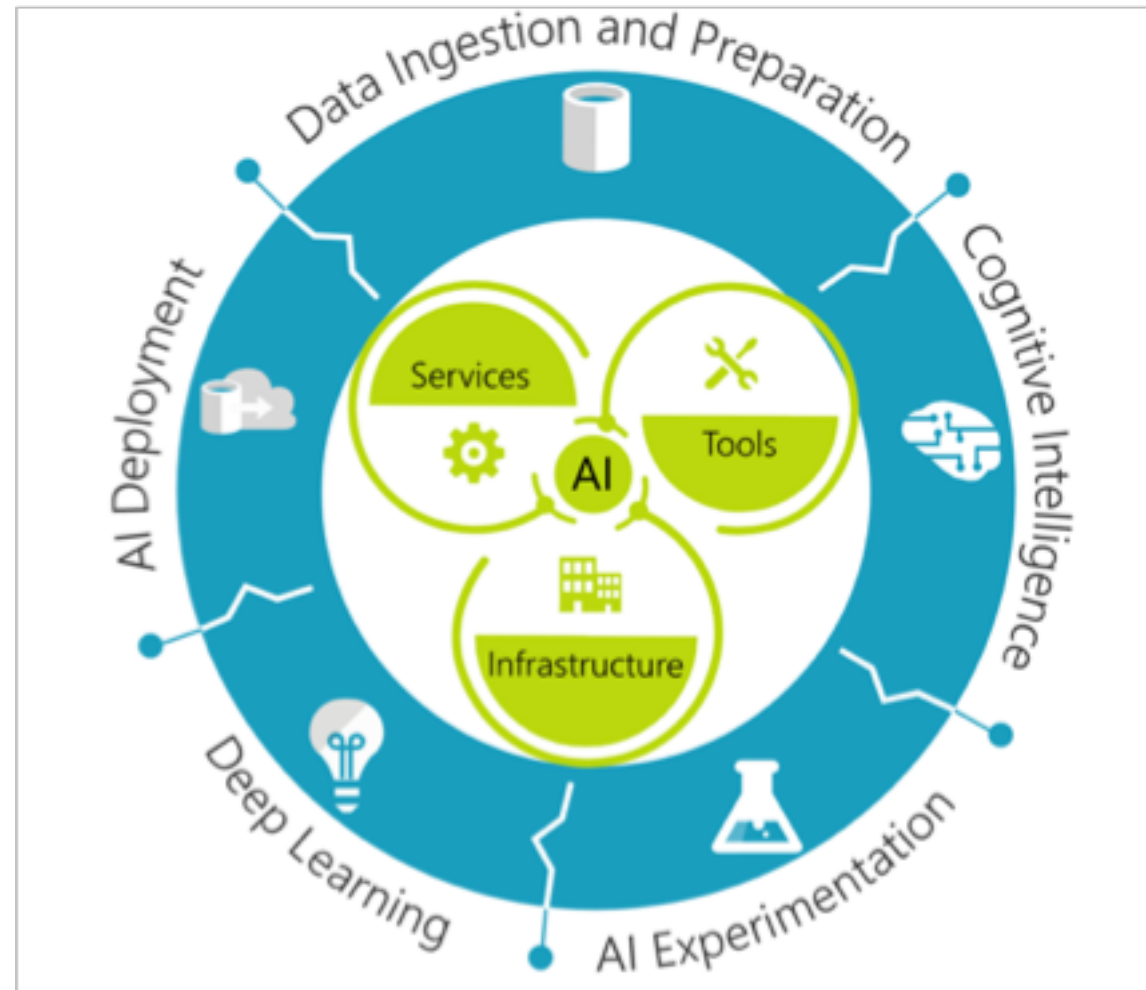|	Raising money	Hiriring developers
AI	✓	
ML		✓

Microsoft AI Platform

Use a comprehensive set of flexible AI services for any scenario, and enterprise-grade AI infrastructure that runs AI workloads anywhere at scale. Modern AI tools designed for developers and data scientists help you create AI solutions easily, and with maximum productivity.

AI Service (#Annainabox)

Accelerate the development of AI solutions with high-level services. Use your preferred approach adapted to your targeted scenario, and achieve maximum productivity and reliability.

Cognitive Services

Use AI to solve business problems. Infuse your apps, websites, and bots with intelligent algorithms to see, hear, speak, and understand natural methods of communication.

Azure Bot Service

Accelerate development for conversational AI. Integrate seamlessly with Cortana, Office 365, Slack, Facebook Messenger, and more.

Machine Learning Services

Model AI algorithms and experiment with ease. Customize based on your requirements.

AI Infrastructure

AI compute

Flexible compute services from virtually limitless scale to the edge.

Apache Spark for Azure HDInsight

Take advantage of Apache Spark in the cloud for mission critical deployments.

Data Science Virtual Machines

Use a friction-free data science environment that contains popular tools for data exploration, modeling, and development activities.

Batch AI training

Experience unlimited, elastic scale-out deep learning. Run large-scale, massively parallel GPU-enabled AI development.

Azure Container Service (AKS)

Scale and orchestrate containers using Kubernetes, DC/OS, or Docker Swarm.

AI on data

AI-enable your data platform

Data Lake Store

Run data transformations and AI on petabyte-scale.

SQL Database

Use R, Python, and native machine learning in an industry-leading SQL DB.

Azure Cosmos DB

Integrate AI with a globally distributed, multi-model database service.

AI Tools

Machine Learning Studio

Easily build, deploy, and manage predictive analytics solutions

Azure Machine Learning Workbench

Visual AI-powered data wrangling, experimentation, and lifecycle management.

Visual Studio Code Tools for AI

Build, debug, test, and deploy AI with Visual Studio Code on Windows and Mac.

Azure Notebooks

Organize your datasets and Jupyter Notebooks in a centralized library for data science and analysis.
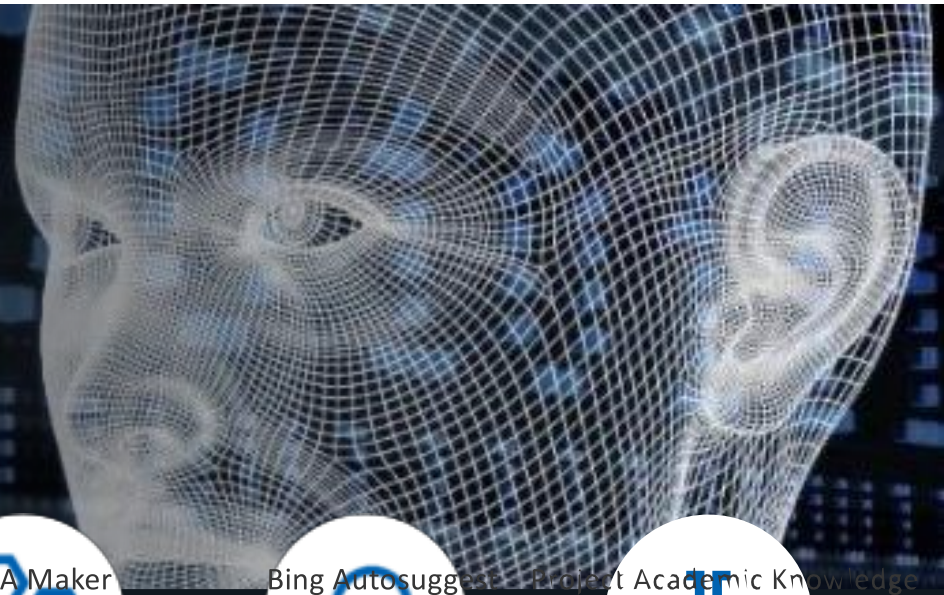
Other popular open source tools

Support for Jupyter Notebooks, PyCharm, and more.
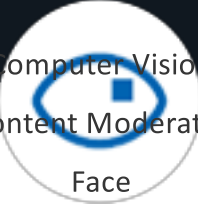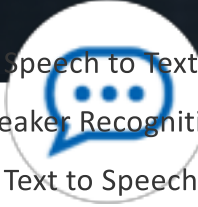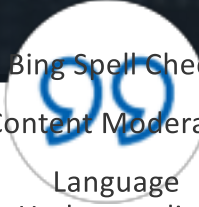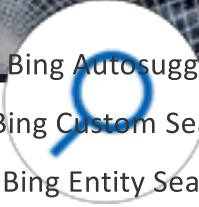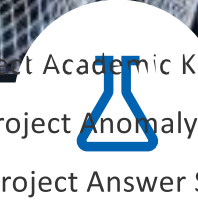
AI Toolkit for Azure IoT Edge

Deploy deep learning models and AI to run locally on IoT devices through pre-built models.

Microsoft Cognitive Services

Give your apps a human side

Vision

Computer Vision

Content Moderator

Face

Video

Video Indexer

Custom Vision Service

Speech

Speech to Text

Speaker Recognition

Text to Speech

Speech Translation

Language

Bing Spell Check

Content Moderator

Language Understanding

Text Analytics

Translator Text

Knowledge

QnA Maker

Custom Decision

Search

Bing Autosuggest

Bing Custom Search

Bing Entity Search

Bing Image Search

Bing Web Search

Bing Video Search

Bing Visual Search

Labs

Project Academic Knowledge

Project Anomaly Finder

Project Answer Search

Project Conversation Learner

Project Entity Linking

Project Event Tracking

Project Ink Analysis

Project Local Insights

Project Gesture

Project Knowledge Exploration

Project Personality Chat

Project URL Preview

Microsoft Cognitive Services
Give your apps a human side

Vision
Computer Vision

Content Moderator

Face

Video

Video Indexer

Speech
Speech to Text

Speaker Recognition

Text to Speech

Speech Translation

Language
Bing Spell Check

Content Moderator

Language Understanding

Text Analytics

Translator Text

Knowledge
QnA Maker

Custom Decision

Search
Bing Autosuggest

Bing Custom Search

Bing Entity Search

Bing Image Search

Bing Web Search

Bing Video Search

Bing Visual Search

Labs
Project Academic Knowledge

Project Anomaly Finder

Project Answer Search

Project Conversation Learner

Project Entity Linking

Project Event Tracking

Project Ink Analysis

Project Local Insights

Project Gesture

CUSTOMIZATION

| Custom Vision Service | Custom Speech Service | Language Understanding | Custom Decision Service | Bing Custom Search |

Text Analytics

- Language detection
  Identify the language, 120 supported laguages

- Key phrase extraction
  Extract key pharses from a piece of text and retieve topics

- Sentiment analytics
 Understand if a record has positive or negative sentiment

Text Analytics

Sentinent

The API returns a numeric score between 0 and 1. Scores close to 1 indicate positive sentiment and scores close to 0 indicate negative sentiment. Sentiment score is generated using classification techniques. The input features of the classifier include n-grams, features generated from part-of-speech tags, and word embeddings. English, French, Spanish and Portuguese text are supported.

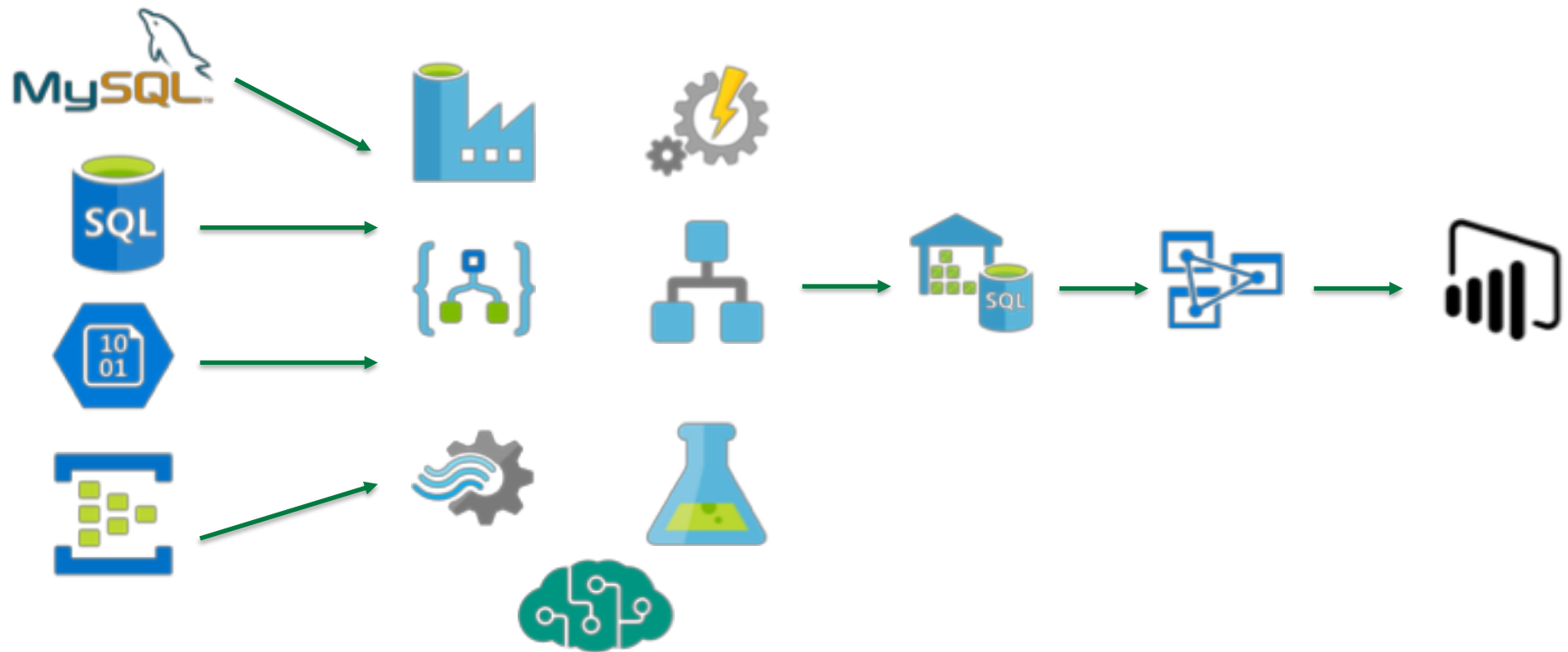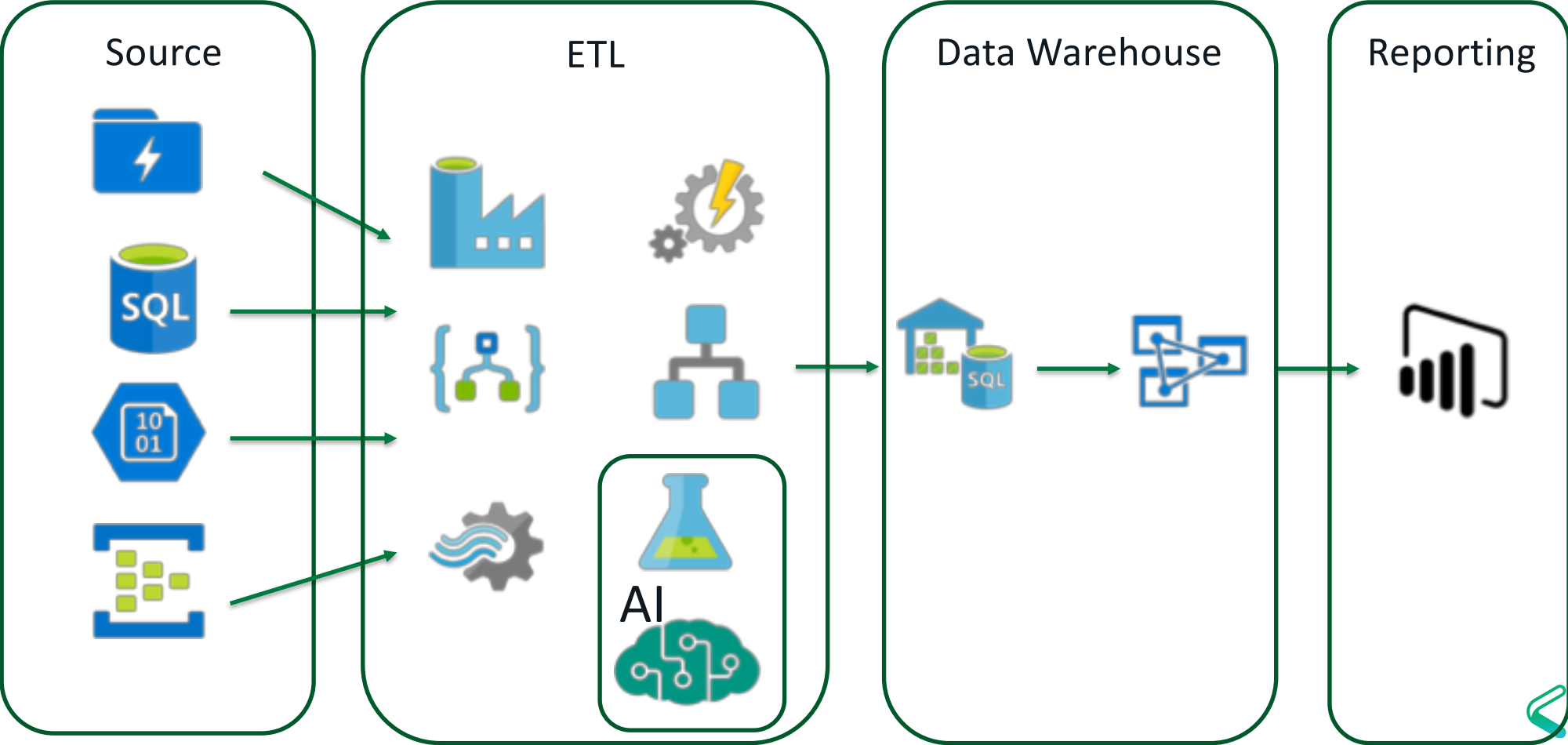# Azure Machine Learning (#Annastyle)
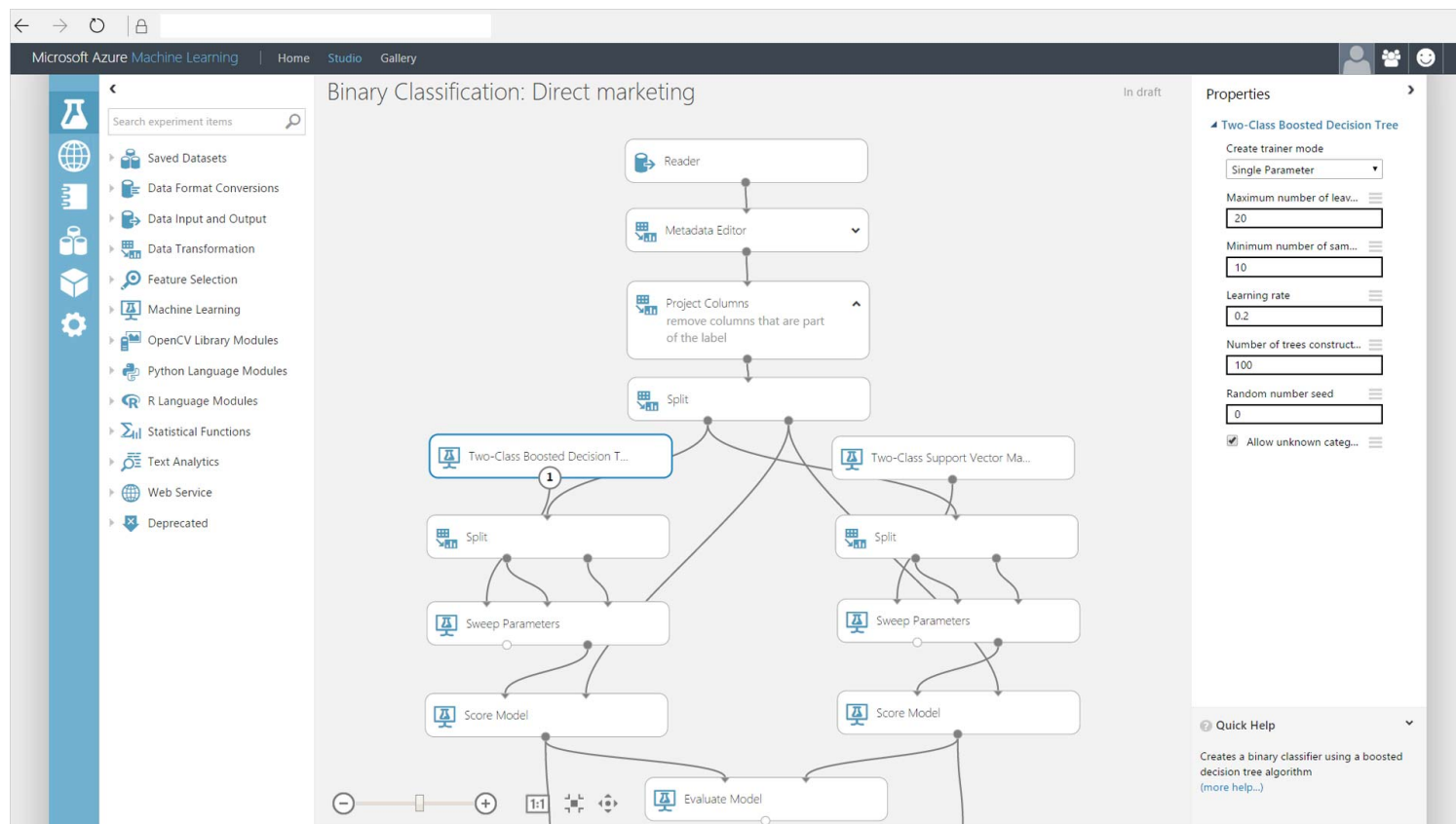
5 questions

- Is this A or B? uses classification algorithms
- Is this weird? uses anomaly detection algorithms
- How much? or How many? uses regression algorithms
- How is this organized? uses clustering algorithms
- What should I do now? uses reinforcement learning algorithms

How to

BI landscape in Azure

Month 1 - 4

- Which tools are the right ones?
- POC
- First ETL with Data Facotry
- Create first reports in Power BI

Month 5 - 6

- Go live first company

- Build more reports

- Stabilization ETL process

- Present current status to management

Month 7 - 10

- Connect new data sources

- Azure ML

- Bulid more reports

Month 11

- Service time of DW

- Connect new data sources

- Bulid more reports

Month 12

- Go live second company

- Stabilization ETL process

- Connect new data sources

Conclusion

- Have plan B
- If plan B fails have plan C
- If plan C fails have flight ticket somewhere nice or at least place to hide

- just kidding

Conclusion - pros

- Rethink of all your process and workflows
- Create development environments in minutes
- Short development time
- Develop in small size & scale up in production
- Test new technologies without of any special need of hardware
- Safe money

Conclusion - pros

- Central logging of services

- Powershell as master language

- Releasetime of services (new features voting)

- Responsetime (questions / tickets)

Conclusion - cons

- Learn different languages

- Max 32 concurrent query Azure SQL DW

- Error messages

- SQL send mail

- TSQL in Azure SQL DW

- Database projects in Visual Studio for Azure SQL DW

Always keep in mind

- Use Key Vault for store credentials

- Error handling

- Notification

- Data delivery

Question! Question?

Thank you for your attention

✉ a.klein@consulting-bi.de

𝕋 @SQL_Alex

🏠 consulting-bi.de

Appendix

Azure BI & Reporting
https://azure.microsoft.com/de-de/blog/technical-reference-implementation-for-enterprise-bi-and-reporting/

Azure Logic App
https://docs.microsoft.com/en-us/azure/logic-apps/logic-apps-overview

Azure Data Factory
https://azure.microsoft.com/en-us/services/data-factory/

Azure SQL DW
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-overview-what-is

Azure Analysis Services
https://docs.microsoft.com/en-us/azure/analysis-services/analysis-services-overview