Alexander Klein

PASS SQLSATURDAY

ETL in the Cloud

Save the date for exiting upcoming events

PASS Camp 2017

Main Camp **05.12. – 07.12.2017** (04.12. Kick-Off abends)
Lufthansa Training & Conference Center, Seeheim

SQL Konferenz 2018

PreCon: **26.02.2018**
MainCon: **27.02. – 28.02.2018**
Darmstadtium, Darmstadt

More information at PASS booth

ETL in the Cloud

Who am I?

Independent BI Consultant

> 15 years experience of SQL Server

Focus on Microsoft BI Stack

✉ a.klein@consulting-bi.de

@SQL_Alex

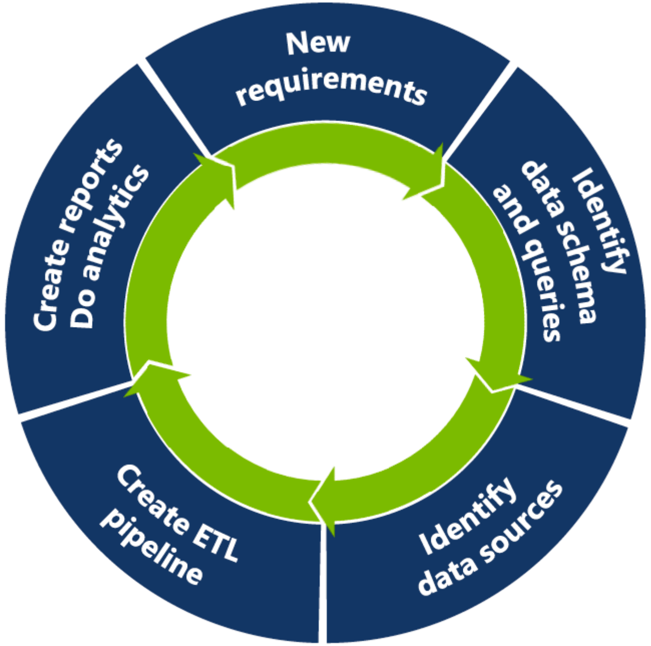
consulting-bi.de

Next 60 Minutes

- ETL
- Azure Logic App
- Azure Function
- Azure Data Factory
- Azure Data Lake
- Azure Stream Analytics
- Azure Automation / Runbook

Traditional business analytics process

1. Start with end-user requirements to identify desired reports and analysis
2. Define corresponding databases schema and queries
3. Identify the required data source
4. Create a Extrac-Transform-Load (ETL) pipeline to extract required data (curation) and transform it to target schema
5. Create reports and analyze data



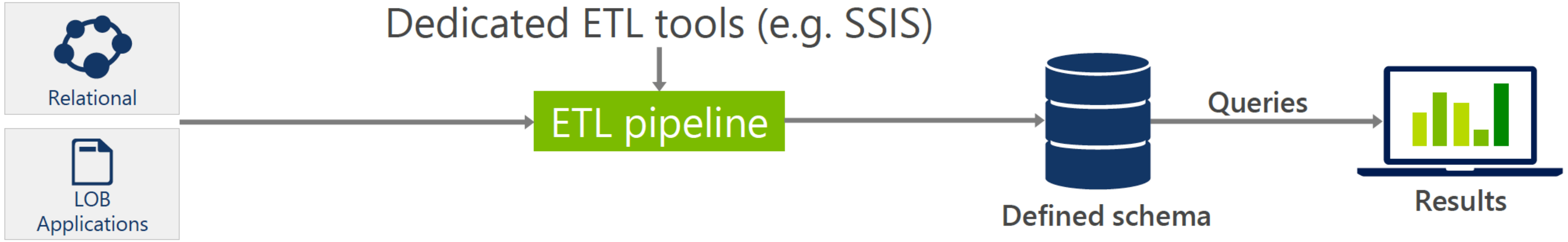
New requirements

Identify data schema and queries

Identify data sources

Create ETL pipeline

Create reports Do analytics



Dedicated ETL tools (e.g. SSIS)

Relational

LOB Applications

ETL pipeline

Queries

Defined schema

Results

All data not immediately required is discarded or archived

ETL

Extraktion

der relevanten Daten aus verschiedenen Quellen

Transformation

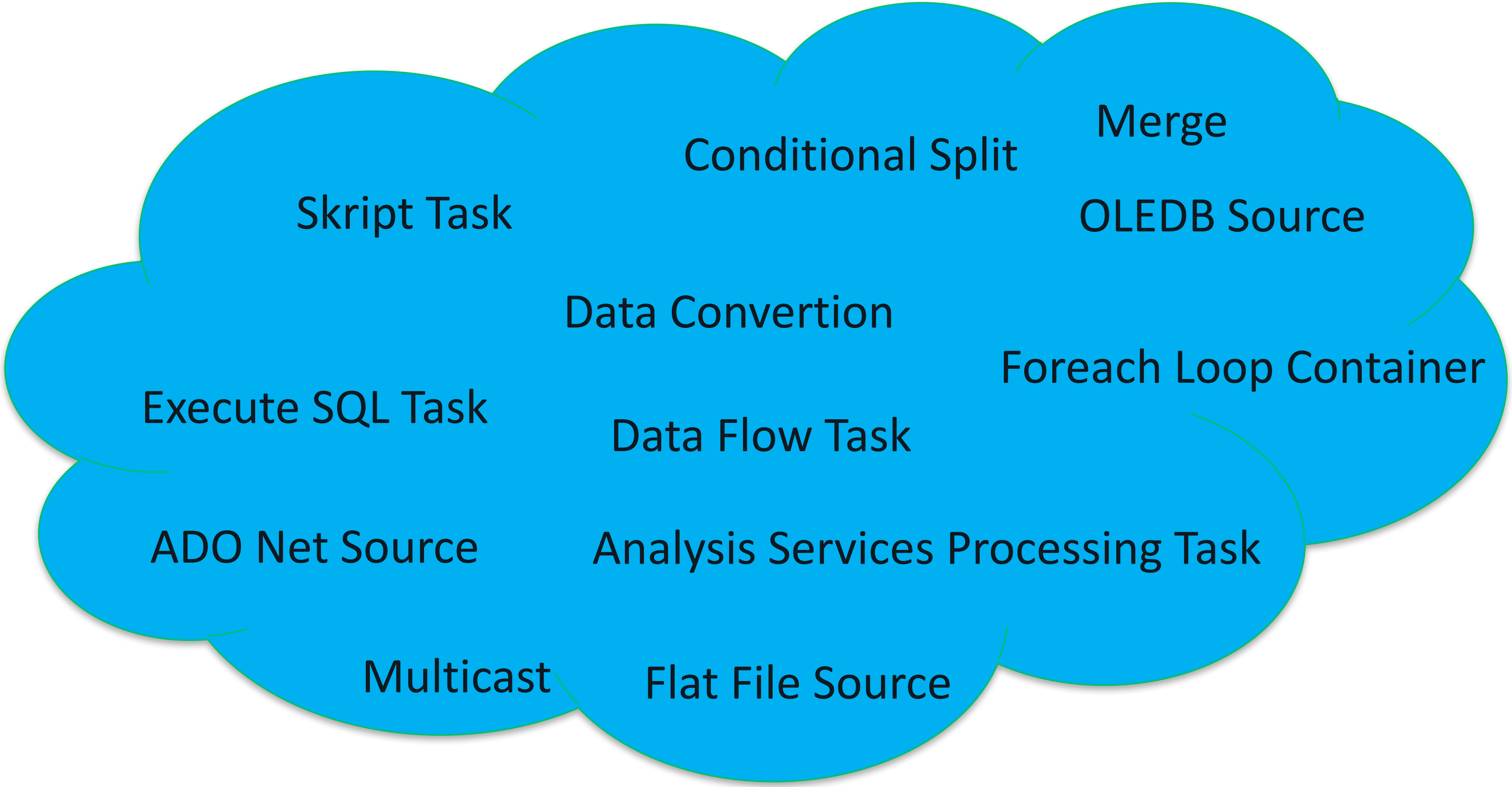
der Daten in das Schema und Format der Zieldatenbank

Laden

der Daten in die Zieldatenbank

Quelle: https://de.wikipedia.org/wiki/ETL-Prozess

On Prime (classic)



Merge

Conditional Split

OLEDB Source

Skript Task

Data Convertion

Foreach Loop Container

Execute SQL Task

Data Flow Task

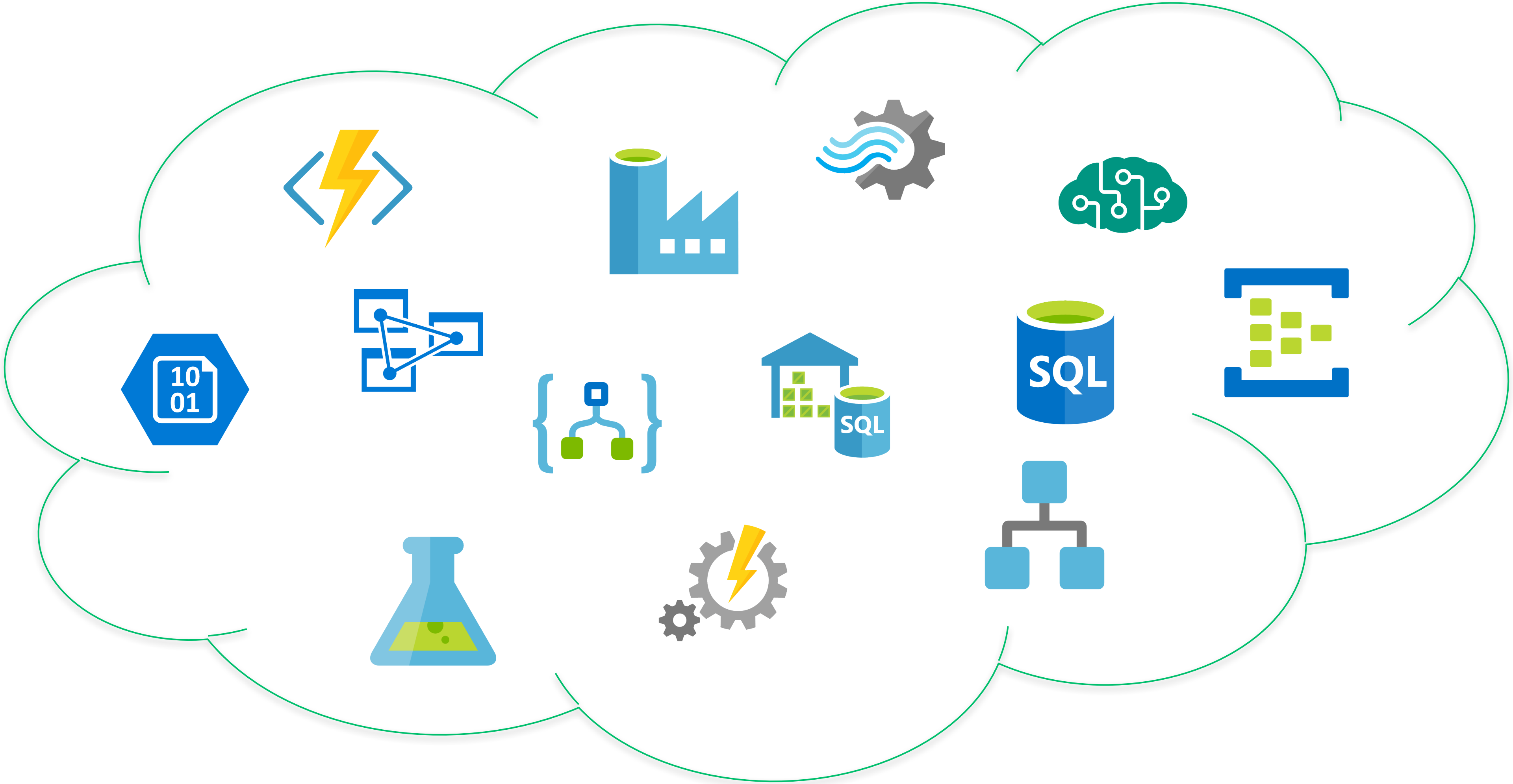
ADO Net Source

Analysis Services Processing Task

Multicast

Flat File Source

Cloud



Azure Logic App

iPaaS (integration Platform as a Service)

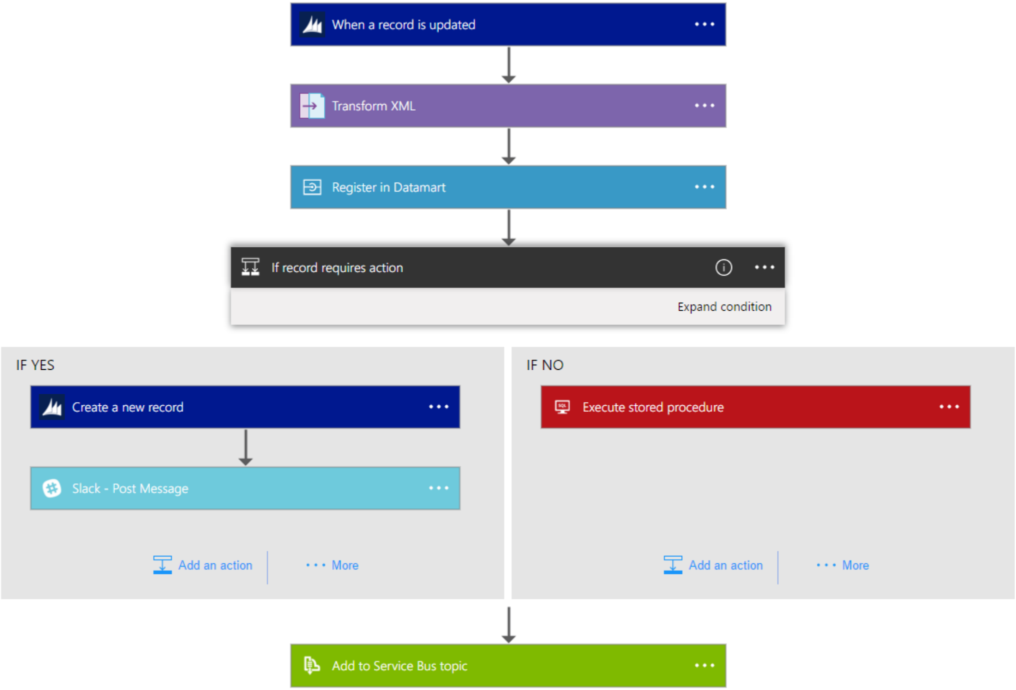
Workflow

Connectors

Trigger

Actions

Enterprise Integration Pack



Azure Logic App

Connectors:

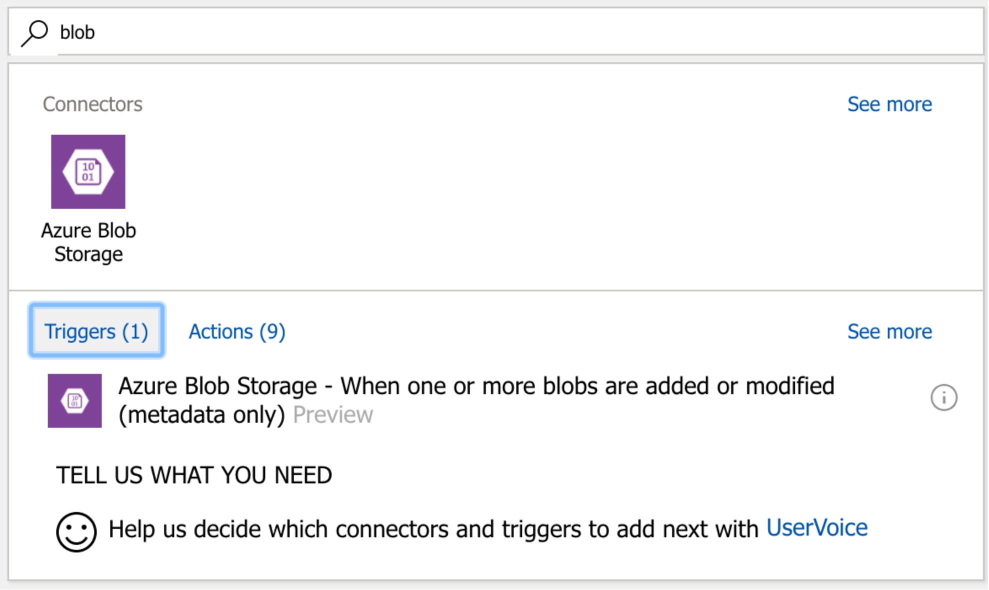
- FTP
- HTTP
- SQL Server
- Azure Blob Storage
- Office 365 Outlook
- Event Hub
- Service Bus
- Twitter
- Power BI
- Salesfroce
- Cognitive Services
- Dynamics 365 CRM / NAV
- Google Drive
- Youtube
- Informix
- DB2
- …

Azure Logic App

Trigger:

Listener waiting for event A ….

e.g. new file created on a blob storage



blob

Connectors

See more

Azure Blob Storage

Triggers (1)

Actions (9)

See more

Azure Blob Storage - When one or more blobs are added or modified (metadata only) Preview

TELL US WHAT YOU NEED

Help us decide which connectors and triggers to add next with UserVoice

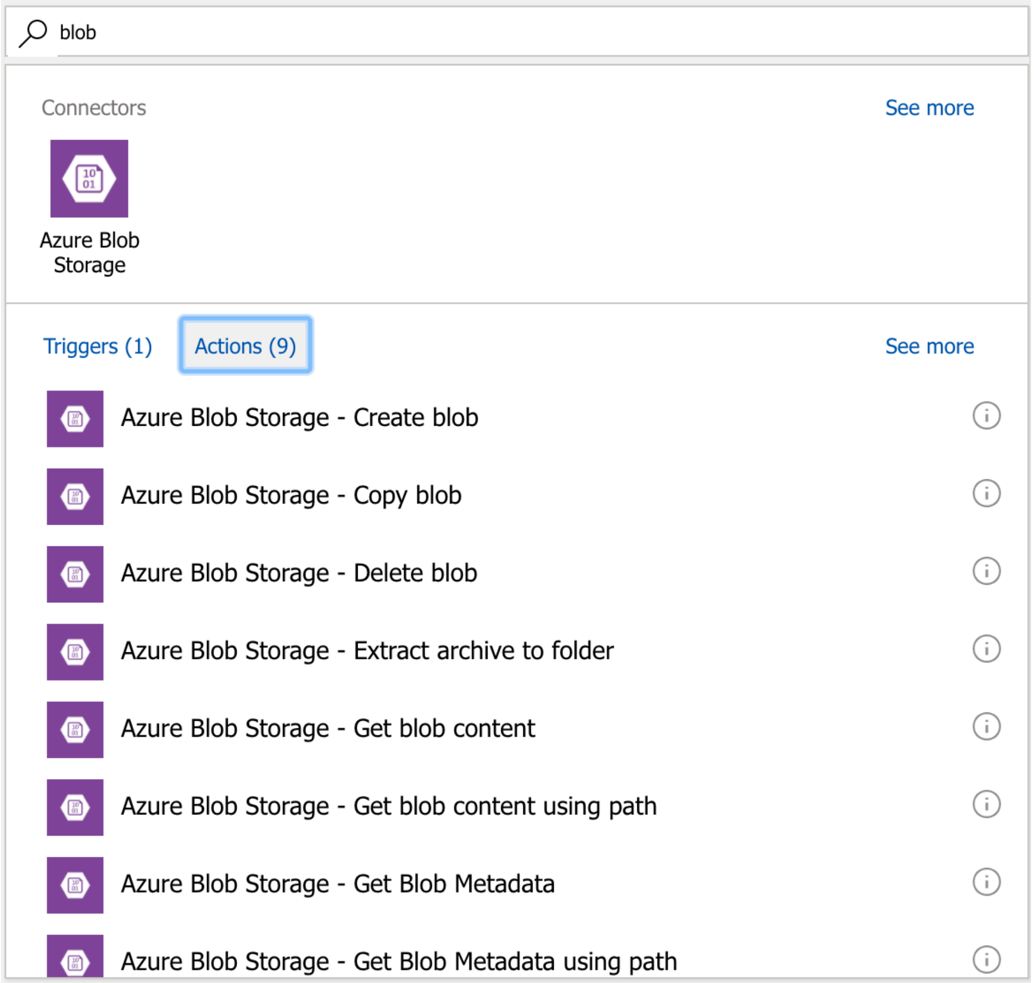
Azure Logic App

Action:

Follows after each trigger.

What to do when a trigger act.

e.g. copy blob



blob

Connectors

See more

Azure Blob Storage

Triggers (1)

Actions (9)

See more

Azure Blob Storage - Create blob

Azure Blob Storage - Copy blob

Azure Blob Storage - Delete blob

Azure Blob Storage - Extract archive to folder

Azure Blob Storage - Get blob content

Azure Blob Storage - Get blob content using path

Azure Blob Storage - Get Blob Metadata

Azure Blob Storage - Get Blob Metadata using path

Azure Logic App

Integrationskonto-Connectors:

- A52
- EDIFACT
- XML
- X12

Azure Logic App

Demo ...

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 Function

Language:

- C#
- F#
- Node.js
- Python
- PHP
- Batch
- Bash
- any executable

Azure Function

Integrations:

- Azure Cosmos DB
- Azure Event Hubs
- Azure Mobile Apps (tables)
- Azure Notification Hubs
- Azure Service Bus (queues and topics)
- Azure Storage (blob, queues, and tables)
- GitHub (webhooks)
- On-premises (using Service Bus)
- Twilio (SMS messages)

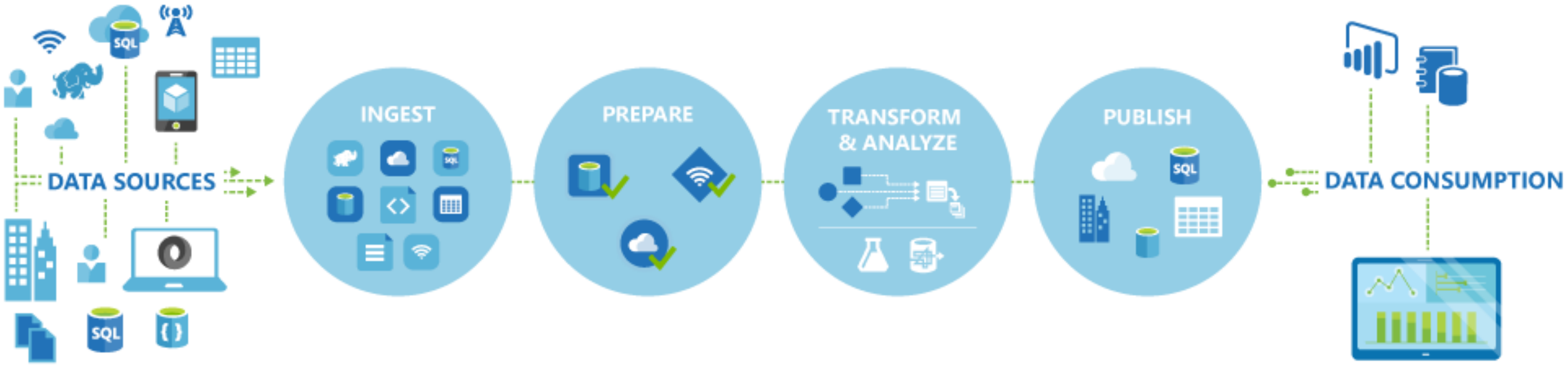
Azure Function

What can I do:

- BlobTrigger
- EventHubTrigger
- Generic webhook
- GitHub webhook
- HTTPTrigger
- QueueTrigger
- ServiceBusQueueTrigger
- ServiceBusTopicTrigger
- TimerTrigger

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.



Azure Data Factory (ADF)

Pipeline:

A data factory may have one or more pipelines. A pipeline is a group of activities. Together, the activities in a pipeline perform a task.

Azure Data Factory (ADF)

Activity:

Activities define the actions to perform on your data. For example, you may use a Copy activity to copy data from one data store to another data store.

Azure Data Factory (ADF)

Data transformation activity	Compute environment
Hive	HDInsight [Hadoop]
Pig	HDInsight [Hadoop]
MapReduce	HDInsight [Hadoop]
Hadoop Streaming	HDInsight [Hadoop]
Spark	HDInsight [Hadoop]
Machine Learning activities: Batch Execution and Update Resource	Azure VM
Stored Procedure	Azure SQL, Azure SQL Data Warehouse, or SQL Server
Data Lake Analytics U-SQL	Azure Data Lake Analytics
DotNet	HDInsight [Hadoop] or Azure Batch

Azure Data Factory (ADF)

Datasets:

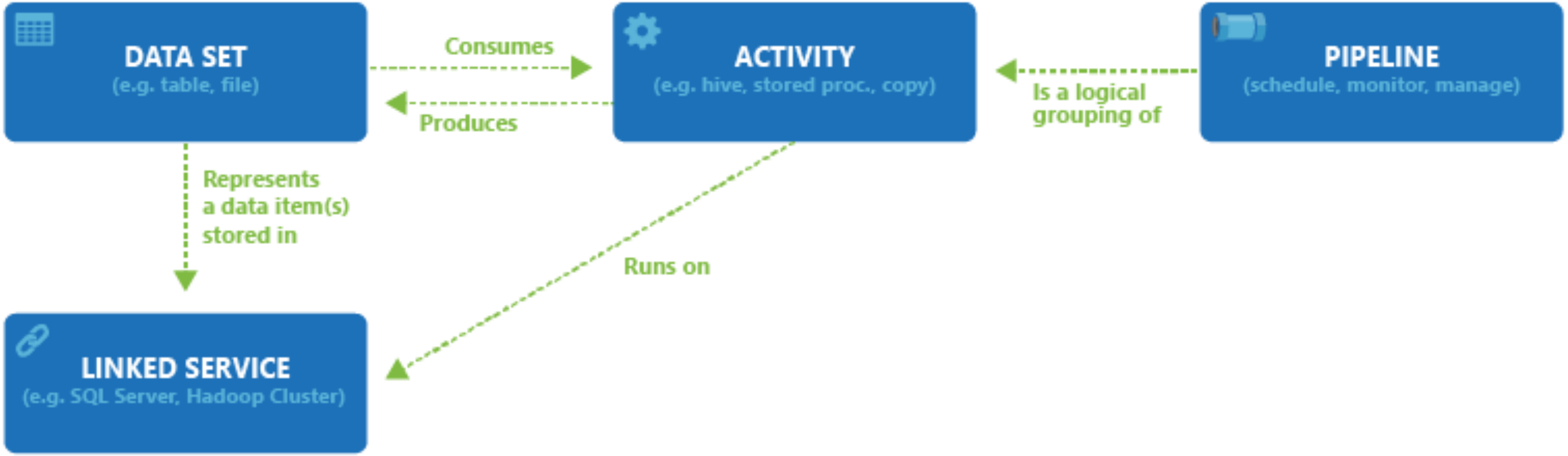
An activity takes zero or more datasets as inputs and one or more datasets as outputs. Datasets represent data structures within the data stores, which simply point or reference the data you want to use in your activities as inputs or outputs.

Azure Data Factory (ADF)

Linked services:

Linked services are much like connection strings, which define the connection information needed for Data Factory to connect to external resources. Think of it this way - a linked service defines the connection to the data source and a dataset represents the structure of the data.

Azure Data Factory (ADF)



DATA SET
(e.g. table, file)
Consumes
Produces
ACTIVITY
(e.g. hive, stored proc., copy)
Is a logical grouping of
PIPELINE
(schedule, monitor, manage)
Represents a data item(s) stored in
Runs on
LINKED SERVICE
(e.g. SQL Server, Hadoop Cluster)

Azure Data Factory (ADF)

Source:

- Azure Storage
- FTP
- HTTP
- Amazon S3
- HDFS
- Oracle
- SAP BW
- SAP HANA

Sink:

- Azure Blob Storage
- Azure Data Lake
- Azure SQL DB
- Azure SQL DW
- Azure Cosmos DB
- Oracle
- Filesystem

Azure Data Factory (ADF)

Demo …

Azure Data Lake (ADL)

Azure Data Lake Store

Azure Data Lake Analytics

HDFS for the Cloud

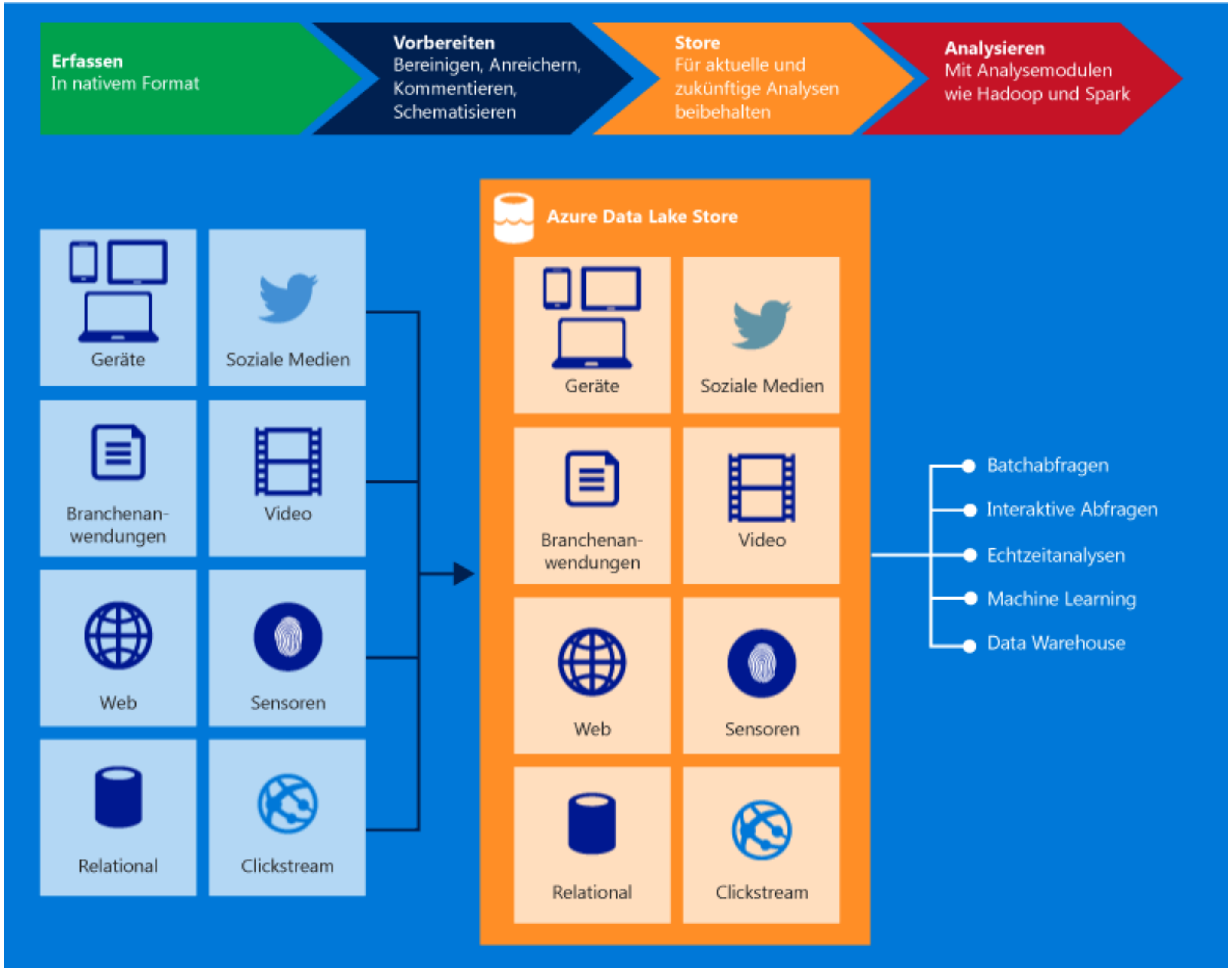
Hadoop

Spark

Always encrypted

Big Data

Azure Data Lake Store (ADLS)



Azure Data Lake Analytics (ADLA)

U-SQL

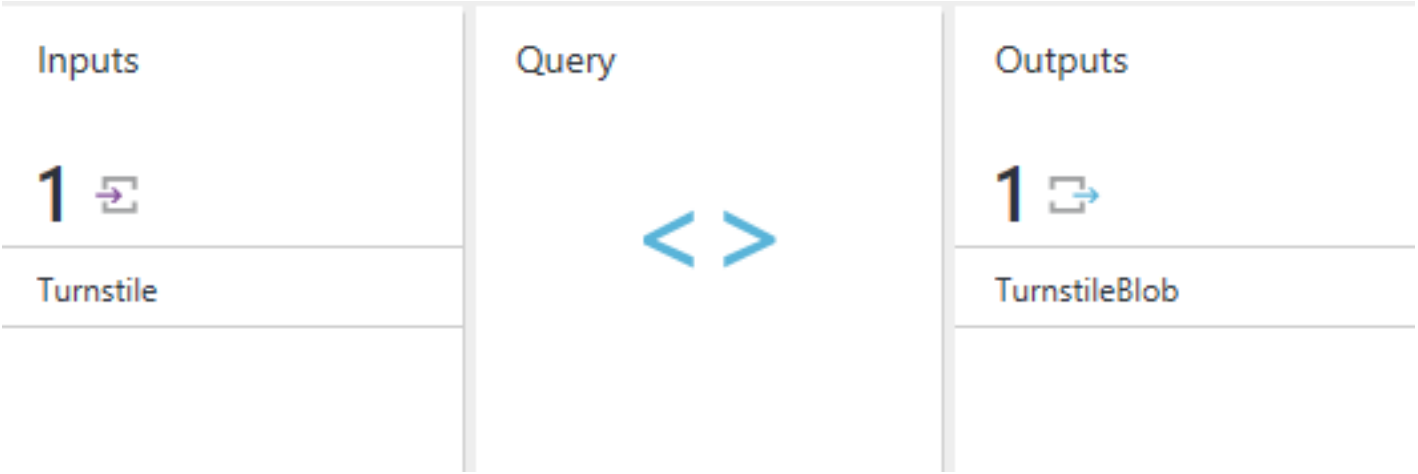
Dynamic scaling

HDFS for the Cloud

Azure Stream Analytics (ASA)

Real-time event processing engine

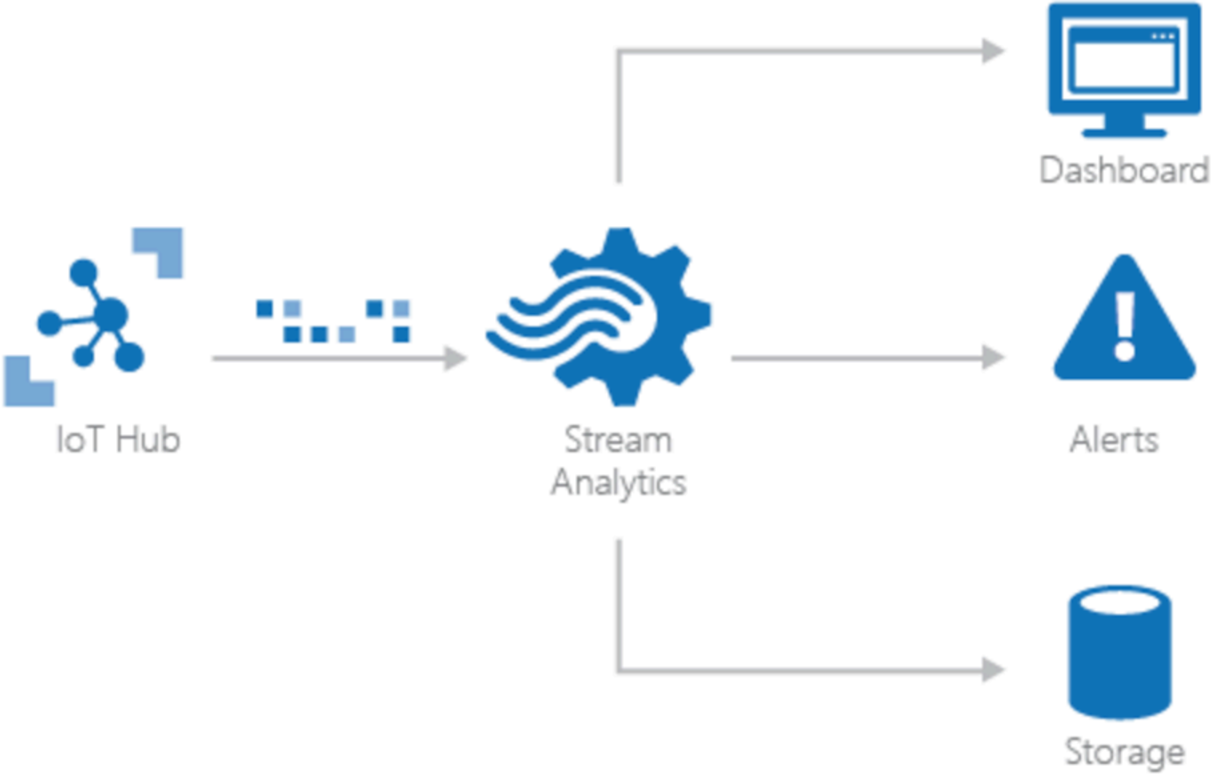
SQL syntax



Azure Stream Analytics (ASA)

Real-time event processing engine

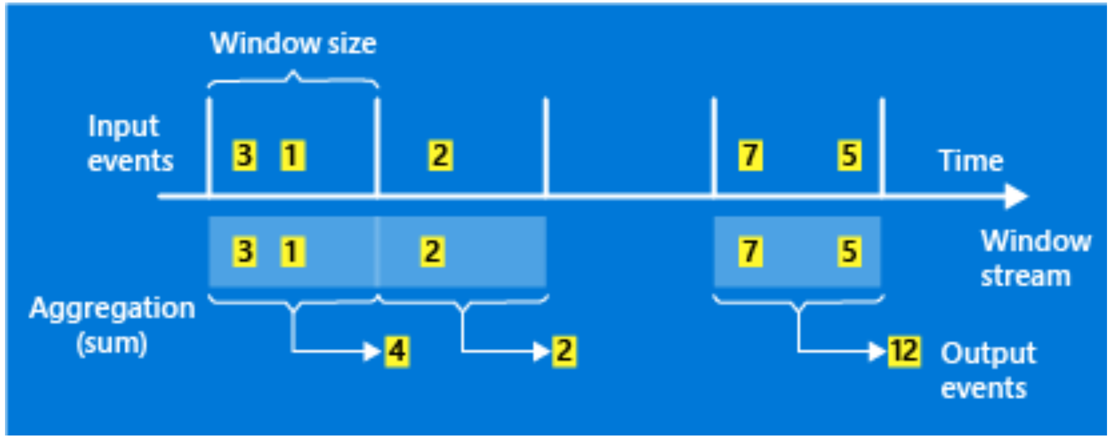
SQL syntax



Azure Stream Analytics (ASA)

Grouping:

- Tumbling Window
- Hopping Window
- Sliding Window



Azure Stream Analytics (ASA)

Source:

- Azure Event Hub
- Azure IoT Hub
- Azure Blob Storage

Supported formats:

- Avro
- JSON
- CSV

Sink:

- Azure Blob Storage
- Azure Data Lake Store
- Azure Document DB
- Azure Event Hub
- Azure Table Storage
- Azure SQL DB
- Azure Service Bus Queue
- Power BI

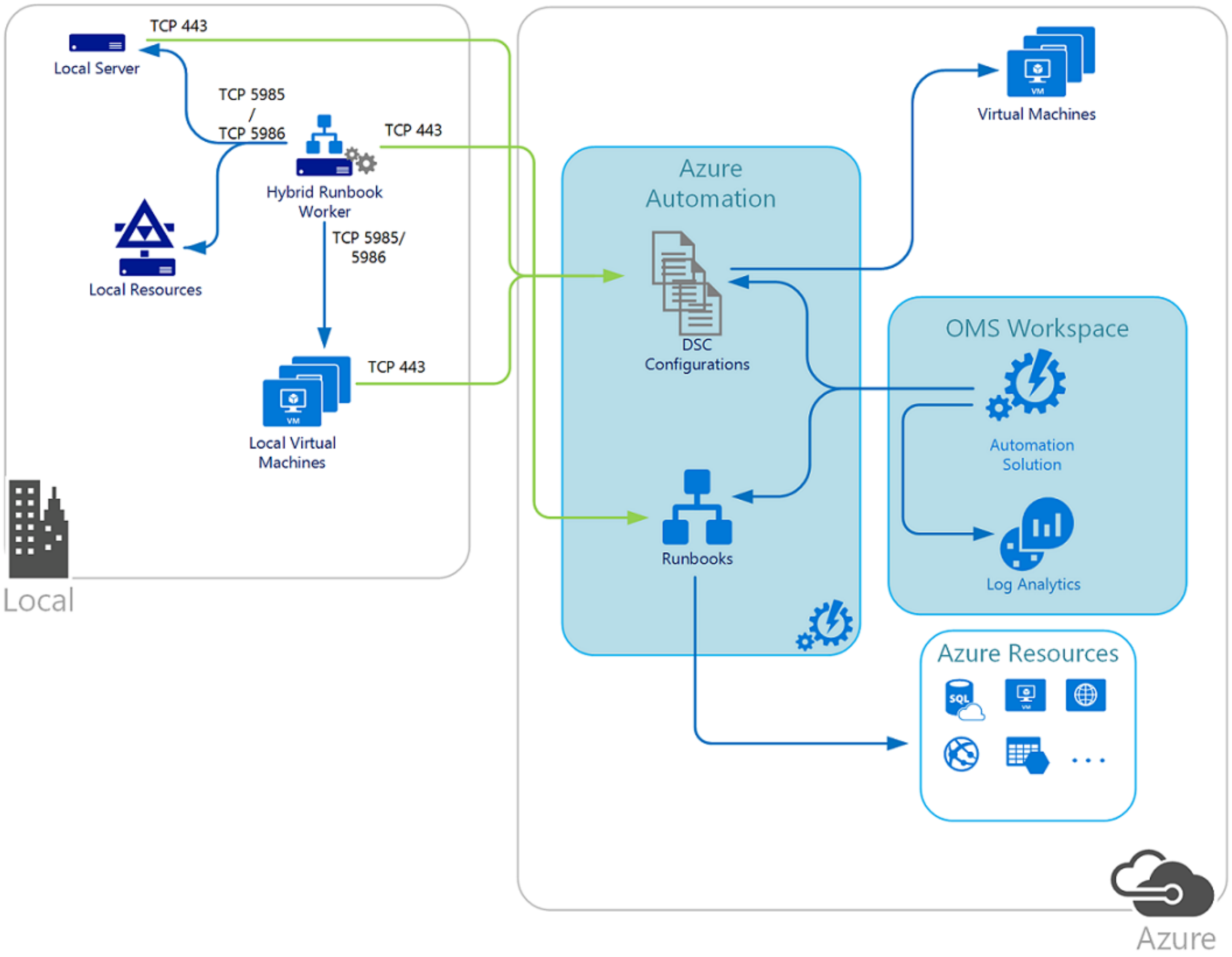
Azure Stream Analytics (ASA)

Demo …

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 Automation



Azure Runbook

Types:

graphical runbook

PowerShell runbook

Visual Studio & TFS

- Azure Data Factory
- Azure Data Lake
- Azure Logic App *
- Azure Stream Analytics (not all sources and destinations supported !)

Deployment options

	Azure Portal	Visual Studio	PowerShell
Azure Data Factory	X	X	X
Azure Data Lake	X	X	X
Azure Function	X		X
Azure Logic App	X	(X)	X
Azure Stream Analytics	X	(X)	X

Always keep in mind

- Error handling
- Notification
- Reporting
- Data delivery

Question! Question?

ETL in the cloud

Thank you for your attention
Tak for din opmærksomhed
Tack för din uppmärksamhet
Takk for oppmerksomheten
Takk fyrir athyglina
Vielen Dank für Ihre Aufmerksamkeit