



Rapid Delivery | Predictable Outcomes



Velocity

Developed in conjunction with Microsoft UK & Corporate teams

2019

Features and Benefits

Features

- Direct data transfer without any Middleware (server-less)
- Uses only SAP (ABAP) and MS Azure technologies only
- Runs in SAP Application Layer using native SAP resource management capabilities
- Fast to deploy
- Enabler for multiple use cases beyond analytics
- Fully generic & parameterized solution to add tables to the lake - managed via configuration
- Support for Multi-SAP, Multi-Lake client landscapes
- Metadata of SAP tables (schemas) is also held in the Lake for enhanced data handling
- CDC Configurable at Table level
 - Batch Kill & Fill - (mass overwrite)
 - Batch Change Data Capture (delta processing)
 - Payload size
 - Compression techniques
- Highly secure data transfer mechanism
- Object Harmonisation - Auto rendering of logical Data Objects from raw SAP Tables - No need to understand SAP Tables

Benefits

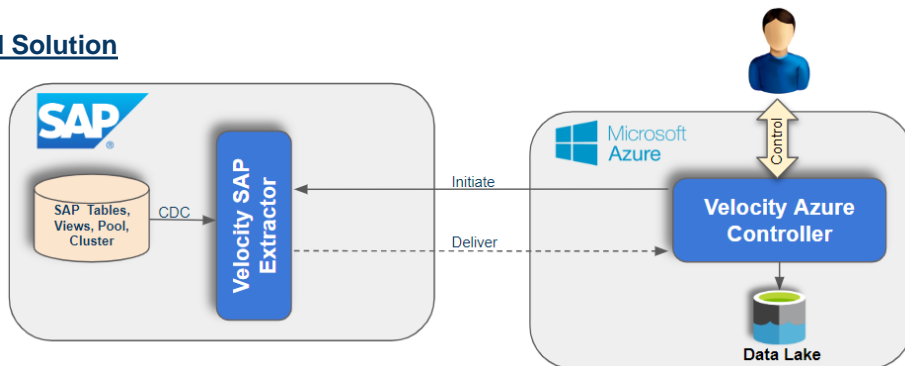
- Low Total Cost of Ownership(TCO)
- Faster Data Transfer Rates
- Increased reliability
- Highly Scalable and configurable
- Easy and fast to deploy & Maintain
- Little or no development required for new table extractions



Velocity - Direct Ingestion

- This is a server-less solution that has been developed using only native SAP and Azure technologies
- **Velocity** is comprised of two components that interact to optimise the data flow and provide parameterized control for the end users:
 - Azure Controller - which manages the entire CDC integration & solution
 - SAP Extractor - which takes commands on what to extract and when and pushes the data back
- The entire integration has a single point of user interaction which is via the Controller
- The direct connection utilises SAP OData RESTful API Service/Protocol recommended by SAP for downstream data consumption processes - this supports the bi-directional data flow
- With only two end points in the solution and data secured with HTTPS susceptibility to Man In The Middle Attacks is reduced. Our next release provides an additional encryption layer during the transfer.

High Level Solution



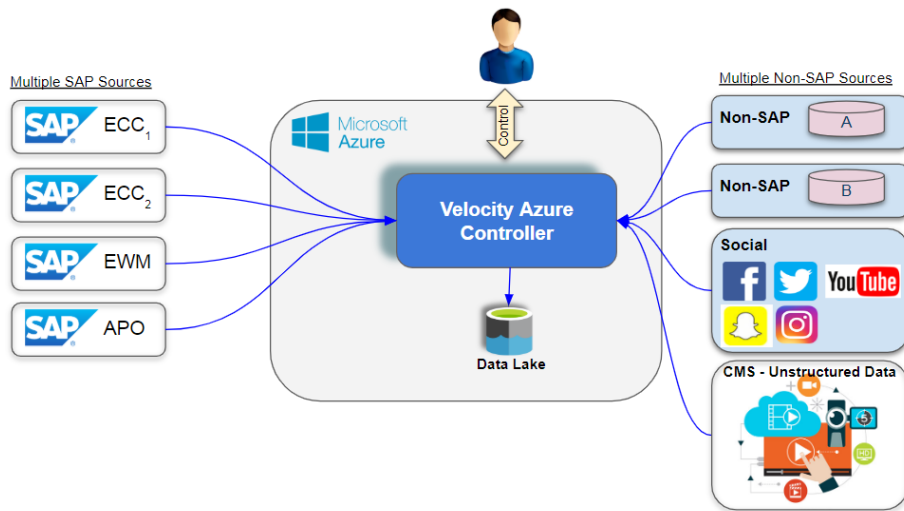
Multiple Connection Support

Multiple SAPs

- Velocity supports many SAP Systems to be connected to a single Azure instance
- Multiple clients from the same System can also be supported
- In this pattern each SAP system will have it's own Extractor installation, with a Single Azure Controller managing the landscape
- All SAP->Azure integration is managed via user configuration through the Controller portal within Azure
- There no technical limits on how far this can be scaled
- SAP integration is gully generic and managed exclusively via user interface

Non-SAP

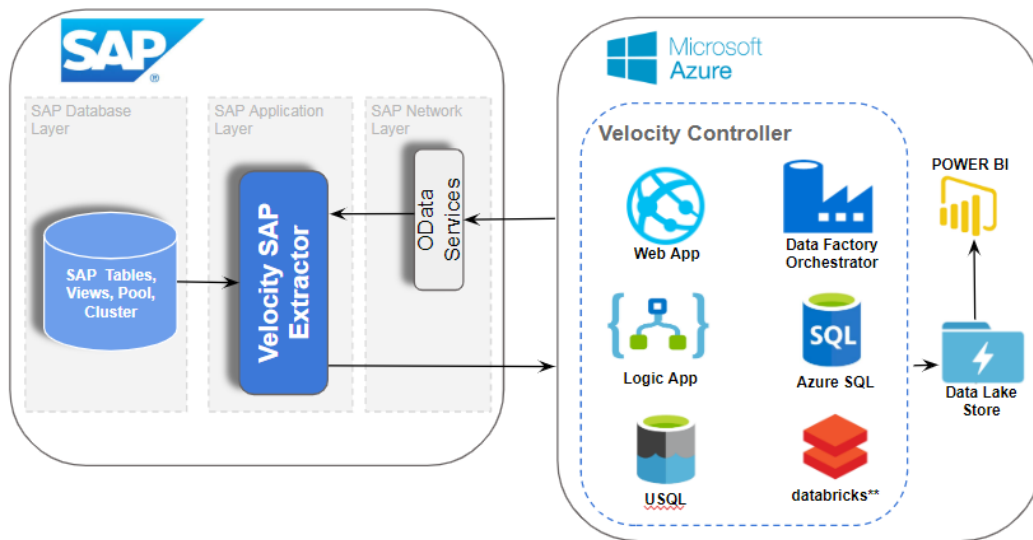
- Supported connections are via direct DB connection or via application APIs.
- Support for unstructured datasets from CMS systems eg Documents, Video, Audio streams etc for AI interrogation



Velocity Components

Velocity will transfer data directly from SAP Applications into Azure Cloud enabling business to do faster and near real-time data analytics and data visualizations.

The diagram below shows the high level architecture of Velocity, along with metrics against current industry standard SAP extraction method (BODS- Data Services).

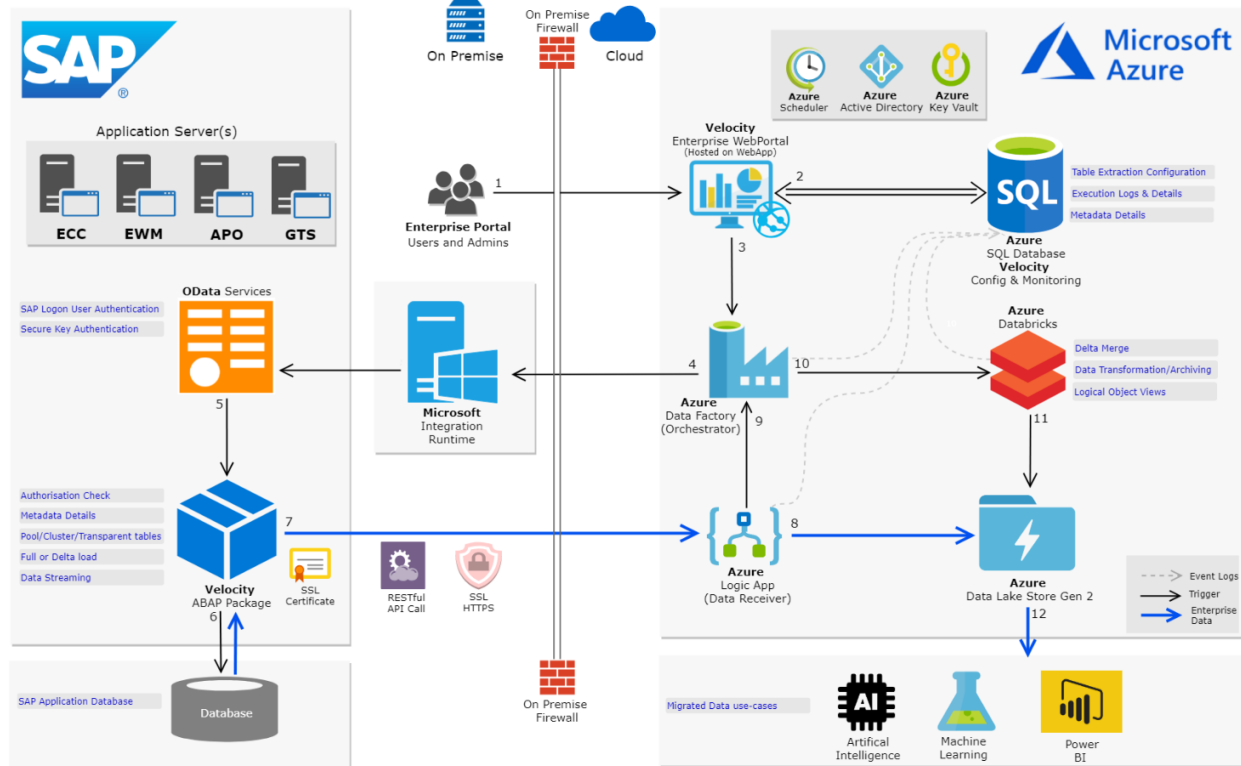


Load Performance per Single SAP Process (App Server)

	Velocity	SAP BODS
Transfer Rate	100 GB/hr	1 GB/hr

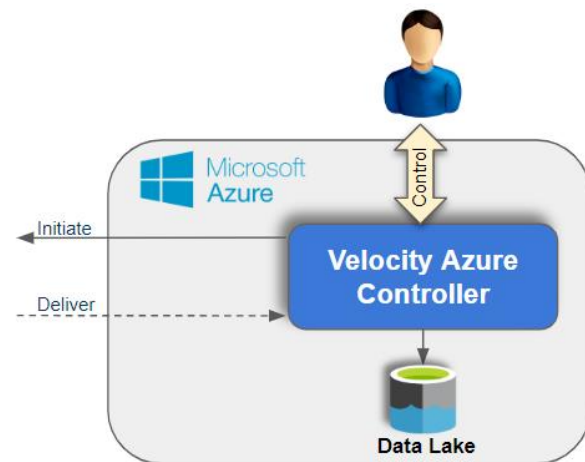
Enterprise Architecture

Kagool Velocity Enterprise Architecture



Velocity Controller

- End User Logon linked to Active Directory (for portal access)
- Controller has a User Interface allowing users to:
 - Maintain connections to SAP systems & Clients
 - Credential Definition
 - SAP Username/Passwords to be used for the SAP connection
 - Select SAP tables to syndicate to the Lake
 - Define of Selections & Projections
 - Define the characteristics of the CDC process
 - CDC method (Change Pointers, timestamp...)
 - Schedule (One-off, Weekly, Daily, Hourly, real-time)
 - Data Compression Activation
 - Batch Payload Size Adjustment (ie #Records or #KB)
 - Degree of Parallelism
 - Define the target details for the data to be stored in the Lake
 - Maintain table & field Blacklists
 - View the execution status of each interface (Active & Historic)
 - Success/Errors
 - Execution Timing
 - Performance charts

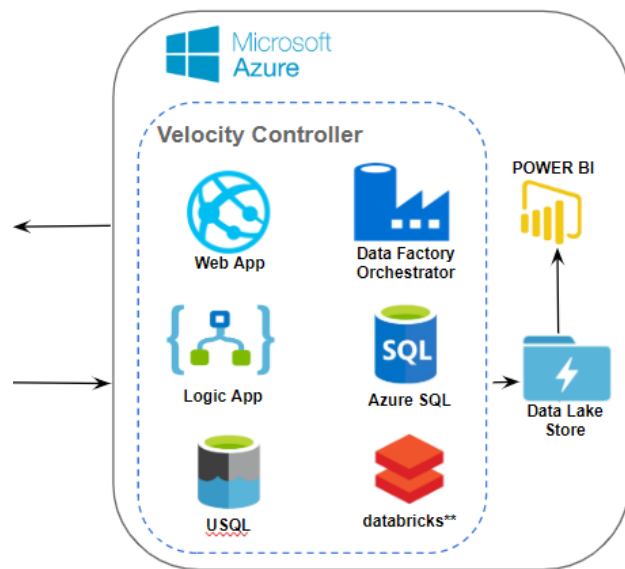


Velocity Controller....cont

The Controller manages the entire syndication process. It allows users to enter their extraction requirements and then controls the execution.

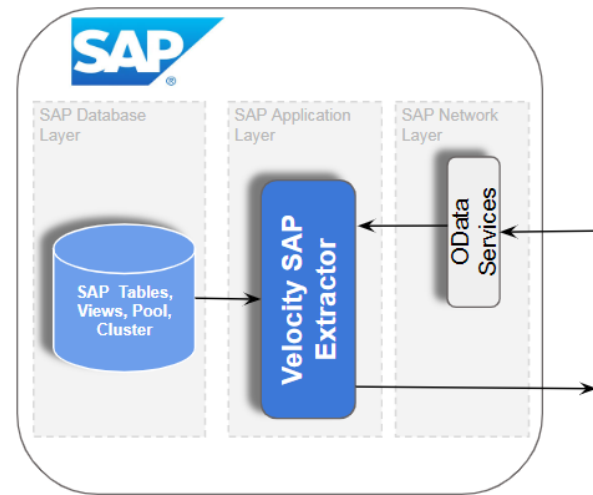
The Controller is a combination of multiple native components developed across the Azure Application Stack:

- Web App
 - This is an interactive front end to the solution for admins to maintain jobs and users to view status on the CDC jobs
- Data Factory
 - Performs the end to end Data Orchestration
- Logic App
 - End point for receiving data from SAP (OData Service)
- Azure SQL
 - Maintains metadata and execution logs/stats
- USQL
 - CDC Merge operations
- Databricks
 - This integration is available from Release 1.1 (Nov-18)



Velocity Extractor

- The Extractor is a slave process that takes extraction commands from the Controller
- It is an ABAP object that sits within the application layer of the SAP system.
- It will extract the requested table object and package the data into payload sizes that align to what is defined by the user in the Controller portal
- CDC Method
 - All known SAP CDC methods have been designed into the process
 - Where no identifiable method is available then Kill & Fill is used - generally small custom tables
- Extraction
 - Data is extracted via application layer and not database layer
 - Data from non-Standard tables is auto translated
- Data Delivery
 - Payload size management e.g. packets of 50k records or 50KB
 - Data is compressed where required
- SAP native resource management capabilities limit/control the extraction



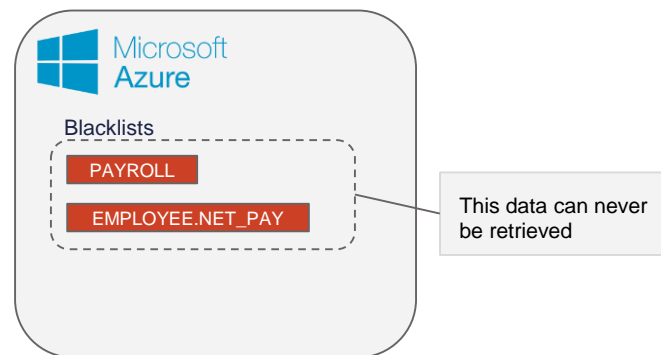
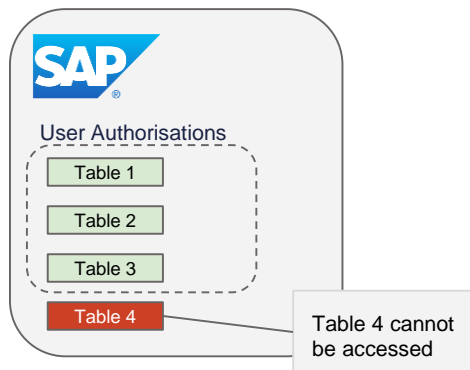
Security Features

- User Logon linked to Active Directory
- Multiple roles within the Azure Controller control access to the configuration. Roles are as follows:
 - Super Admin
 - Creates and maintains connections to systems
 - Maintains Blacklists
 - Create and maintain users
 - General Admin
 - Creates and maintains table extract jobs
 - Guest User
 - View Only
- SAP->Azure Data transmission
 - Direct Connection between end-points without any intermediate systems
 - Improved protection against Man In The Middle Attacks
 - Uses HTTPS Protocol
 - Next release will include additional AES Encryption capability as the data is transferred over the network



Security Features...cont

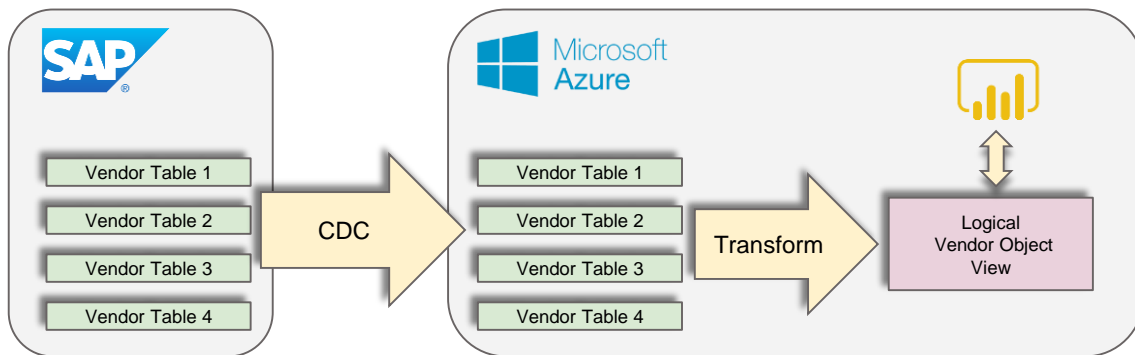
- Two methods to control what data can be transferred to the Lake:
 - **SAP User Authorisations** restricts access to tables within the SAP System
 - **Azure Blacklists** provide additional control to the process, maintained within Azure by Super Admins
 - **Table Blacklists**
 - Restricts tables from being extracted from SAP eg PAYROLL table
 - Even if the SAP Authorisation allows it
 - **Field Blacklists**
 - Suppresses individual fields during syndication process to prevent the field value from getting into the Lake
 - Example - Allows users to Syndicate the EMPLOYEE Table to the lake, but the NET_PAY field would be automatically always suppressed during the data transfer



Harmonize: Logical Object Maintenance

- The syndication process works at table level continuously maintaining the SAP tables in the Lake
- In parallel a Logical Object (Entity) view for all SAP Common Objects is maintained eg Customer, Vendor, Sales Order, Invoice....etc. This is performed by consolidating the raw SAP tables into single or fewer tables
- This means the Analytics users do not need to understand the complex SAP Table Structures when working with the data, they just use the Logical Entity Object Views
- Velocity will be deployed with a suite of Logical Object Transformations for SAP Standard objects. These will need some adjustment depending on the level of customisation on each client site however this should be a small activity.

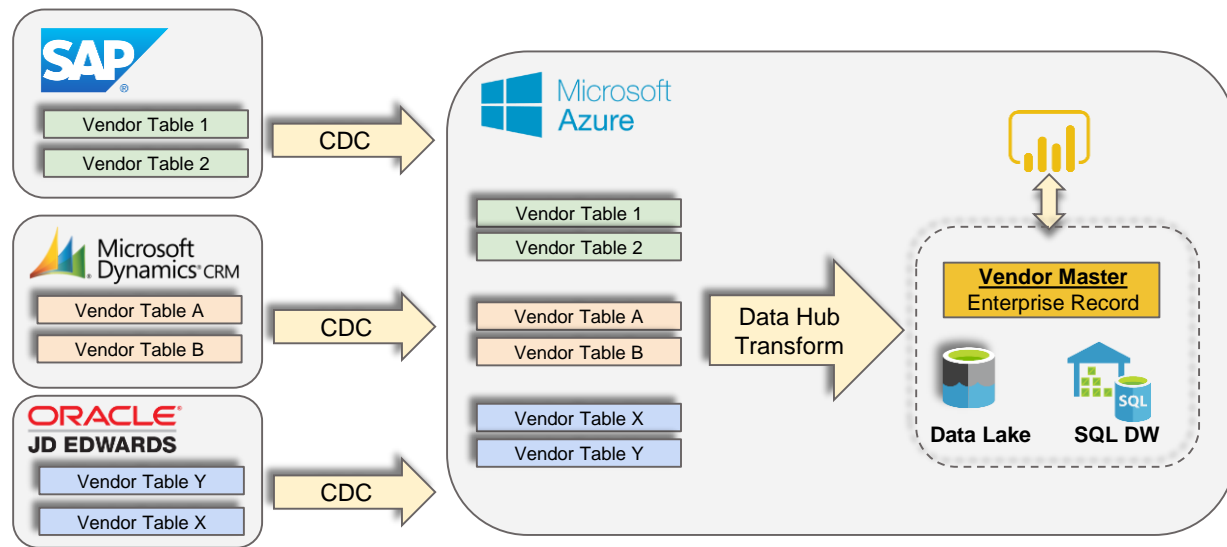
Example Master CDC



Enterprise Data Hub (Available in Jan-19 Release)

- Data from landscape systems syndicate data in Siloed (system unique) fashion to the Lake
- Data Hub Transformation performs a Match & Merge process to generate Enterprise Level Master Datasets (Enterprise Unique). The Data Hub Process can push Enterprise level data in Lake format (csv) or into a SQL database of choice
- Transactional Datasets can then be merged and folded under the single Enterprise Level dataset

Example Master CDC



Rapid Deployment

- Velocity uses standard native SAP and Azure technology - nothing else.
 - The Controller uses Azure Technology
 - The Extractor uses SAP technology.
- Since this is a serverless solution deployment is straightforward, quick and can be performed by client support teams with Kagool supervision:
 - **SAP Side**
 - Create new SAP user
 - Transports
 - OData Service Activation
 - Velocity Extractor
 - User Authorisations
 - Licence Key Install
 - Configuration
 - **Azure Side**
 - Data Factory Pipelines
 - Logic App Scripts
 - Azure SQL Scripts
 - Data Lake Configuration
 - Web Application (User Interface)



Additional Use Cases

The following additional Use Cases have been identified beyond syndication for Analytics purposes

- **SAP Archiving**
 - Make use of the cheaper storage medium that the Lake offers
 - Process can syndicate aged data and delete the data out of the source SAP system

- **iDoc Error Management**
 - Can be notoriously slow and tedious to navigate this data in SAP.
 - Data can be better reviewed and managed in the Lake



Release 1.0

Performance

- Delta gross payload sizes of >100GBs tested
- Configurable Parallel Payload processing control
- Transfer Compression

Change Data Capture (CDC)

- Support for all known delta management patterns within SAP
- CDC control method maintained at Table level

Connectivity

- Connection from 1 Azure source to multiple SAP and non-SAP systems

Configurability

- The following parameters are configurable by end users within Azure:
 - SAP tables to extract,
 - Interface parameters (CDC method, #Parallel streams, Payload packet size, Compression, Scheduling)
 - Blacklist maintenance

Security

- Sensitive field suppression
- User roles for Table addition, connection maintenance, execution log/status viewing
- Blacklisting control of tables & fields
- Active Directory for controller user login

Harmonisation

- Raw SAP tables are harmonised into logical objects in the Lake for many SAP standard objects.



Release Roadmap

End-Nov-2018

- Incorporation of Databricks
- Maintenance of non-SAP Connection credentials managed via Azure Controller User Interface
- Language Pack for Controller
- Redo-log source extraction (Oracle, DB2, MSSQL)

January-2019

- Harmonization process extended to provide Data Hub capabilities (Object consolidation across multiple sources) using databricks and USQL database
- HTTPS additional encryption feature
- Data Profiling/Quality & Score Carding

