Unleash the Power of SAP Data with High Business Agility and Low Cost

Leveraging Azure Data Platform and AecorSoft

Abstract

This whitepaper describes how SAP customers can bring the huge volume complex data out of SAP for advanced analytics in the most effective, efficient and secure manner, by using capabilities from Azure data platform and AecorSoft solution. This paper covers from core components to reference architecture, and advises the roadmap towards the modern cloud based end-to-end solution.

Published: April 2020

AecorSoft Inc.

info@aecorsoft.com

Executive Summary

Being able to effectively drive business values from the massive enterprise data is more important than ever to the IT enabled business agility. In the era of digital transformations, businesses are shifting IT enterprise data solutions from on-premise to Cloud, to take advantage of the elasticity, scalability, security, and supportability of Cloud Storage and Compute power.

When it comes to SAP data, most of the data sources are still reside in on-premise systems or IaaS systems. Bringing the massive and complex SAP data to cloud for scale-out analytics has been a challenge on the critical path to many SAP customers.

By leveraging Azure Data Platform and AecorSoft Integration Service solution, it is now possible to integrate with SAP data sources securely and productively, and help SAP customers make the most out of their SAP data.

SAP Integration Challenges

Integrating with SAP is a constant challenge to many SAP customers.

Data Model Complexity: SAP's data models are powerful for transactional postings, yet complex for analysis. On SAP ECC, there are more than 120K system tables with sophisticated relationship to keep data integrity throughout the business workflows. There are both normalized and de-normalized data models, and there are tabular and multi-dimensional data structures. Some tables have binary fields which cannot be interpreted directly but have to be translated through multiple function calls. Deep understanding of SAP system metadata and data models is extremely important.

Security and Compliance: SAP offers strong security control and governance on its data. In most cases, the SAP data integration needs to go through the application layer, which is not as straightforward as working with database directly. This requires rich knowledge of SAP application layer APIs.

Extraction Efficiency: Doing incremental loads (or "delta loads") is the only viable option when working with large tables, but the delta logic on some SAP tables is not easy. Some large tables do not even offer "Changed On" timestamp fields. In consideration of data modeling, logic of delta calculation, and logic of table joining, in many cases, the SAP standard extractors would be the better option to extract data, but then, working with SAP extractors is another big challenge.

Overview of Traditional Solutions

In the past, the majority of SAP customers had to use SAP BW to build BI solution. Although BW itself is a well-engineered data warehouse product, the adoption of BW has never been a smooth journey to most BW users, due to the sophistication, complexity, latency of BW, and the lengthy project life cycles. When a BW project is finally completed, a lot of subsequent stabilization and optimization efforts would follow afterwards. From the Total Cost of Ownership and Data Timeliness standpoints, BW customers seek better solution to replace BW.

BOBJ Data Services (BODS) was a traditional choice for SAP ETL solution, for the customers who wanted to extract SAP ERP data or migrate BW data to other data platforms. However, BODS is also a heavy-weight platform to work with, and requires proprietary technical skills to develop and maintain. Because BODS is not an SAP NetWeaver product, for customers who want to extract SAP data onto Microsoft data platform, they will have to spin up multiple engineering and support teams with SAP skill, BODS skill, and Microsoft skill, to be able to start the train. Then, the train requires a lot of energy to run, and still stay on the ground without reaching the cloud. Internally, BODS has limited SAP ERP and BW metadata discovery, and limited delta extraction management. The high cost, high latency, and low agility make BODS less than ideal either.

There are other tools on the market which offer limited capability to extract from SAP tables only, and that is far from enough. A number of such tools also require custom configurations and custom ABAP code on SAP in order to work, which is counter-productive when it comes engineering and business agility.

Why Azure

Microsoft Azure is one of the best cloud data platforms on the market, with rich structured and unstructured data storage support, enterprise data security, advanced data modeling, versatile predictive and prescriptive analytics, machine learning and artificial intelligence capabilities, and powerful dashboard and information visualization, all lead to effective and efficient enterprise data management experiences in the cloud.

Advantage of Azure Data Factory V2

Azure Data Factory (ADF) is the cloud based, serverless, scalable, highly available data integration service in Azure. Although ADF comes with many connectors out of the box, it is not enough to meet the velocity and variety requirement of integration with complex SAP data. The support for SAP data is very limited, and requires large amount of custom configurations and development on SAP.

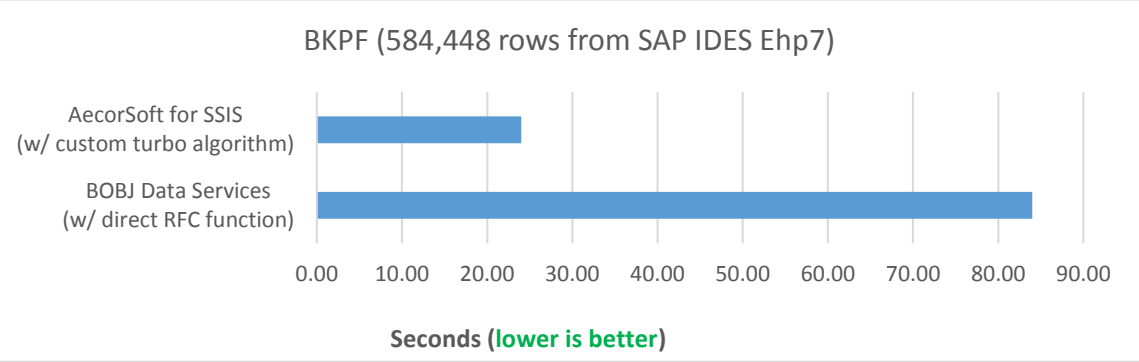
Comparing to ADF V1, the ADF V2 technology is a leap forward, with SSIS support through the Integration Runtime (IR). Customers are now able to lift and shift SSIS packages from on-premise to Azure, to fully take advantage of both the Azure scalability and the powerful SSIS extraction, transformation, and loading capabilities. This is especially valuable to customers who have already invested in SSIS. However, the SAP connectivity limitation remains in ADF V2.

Why AecorSoft

AecorSoft is world's leading software vendor on integration product solution surrounding SAP. AecorSoft Integration Service™ for SSIS and ADF v2, which is based upon the same data engine from AecorSoft product family officially certified by SAP, enables seamless integration and in-memory streaming experience to bring massive data out of SAP onto Azure.

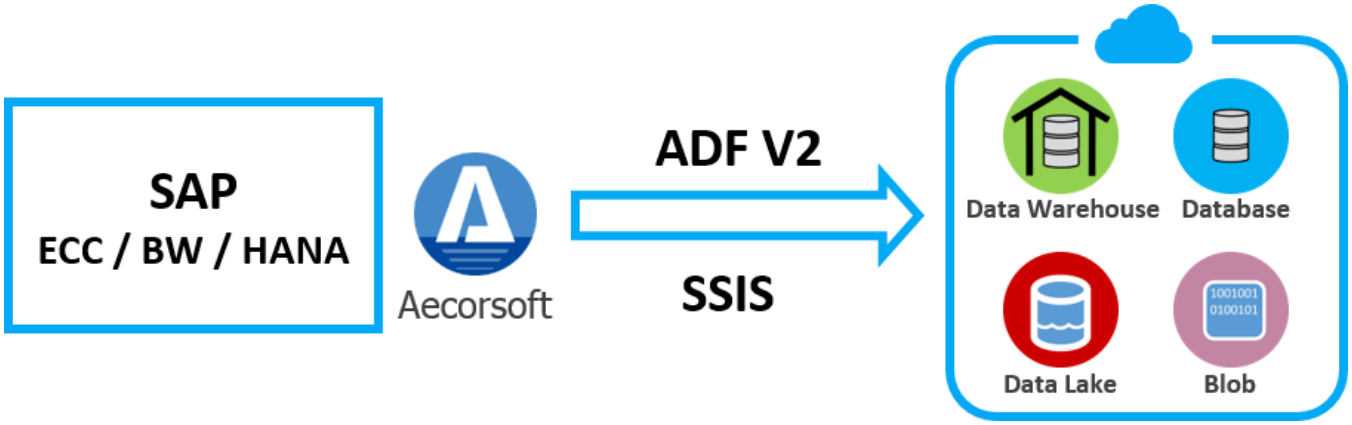
The value differentiators of AecorSoft are:

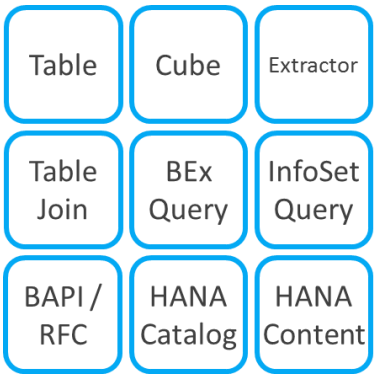
- ***Business Agility***: super easy-to-use, plug-n-play experience without custom configuration or custom ABAP code on SAP, ideal for agile releases and continuous IT project delivery;
- ***Security***: full compliance to SAP Security model at application layer, without any backend DB 'hacking';
- ***Variety***: full support of many SAP object types, including Table, InfoCube, BEx Query, Delta Queue Extractor, InfoSet Query, BAPI/RFC, CDS View;
- ***Performance***: complete in-memory streaming without internal staging on disk; much faster than BODS;
- ***Efficiency***: powerful delta extraction;
- ***Future-ready***: Complete PaaS experience via Azure Data Factory v2 SSIS Integration Runtime.



AecorSoft Integration Service makes "instant datawarehousing" possible, by significantly reducing the integration efforts with SAP data sources. It is the best companion to SSIS and ADF when it comes to integration with SAP. BI professionals are now able to navigate and discover SAP metadata at ease, and treat SAP just like ordinary data source. Results can be achieved within minutes!

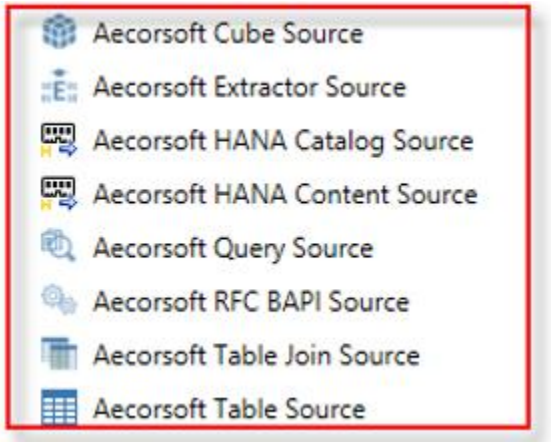
The Journey of Bringing SAP data to Azure: High Level Architecture





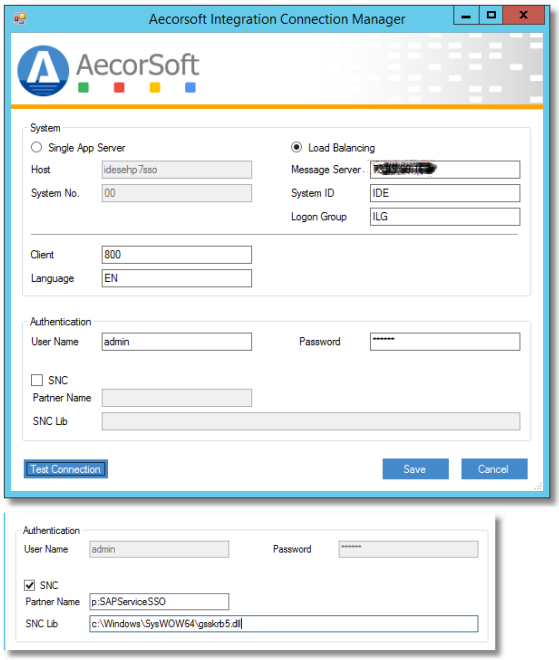
The Journey of Bringing SAP data to Azure: Core Component Features

AecorSoft Integration Service consists of a set of SAP Data Source Components:



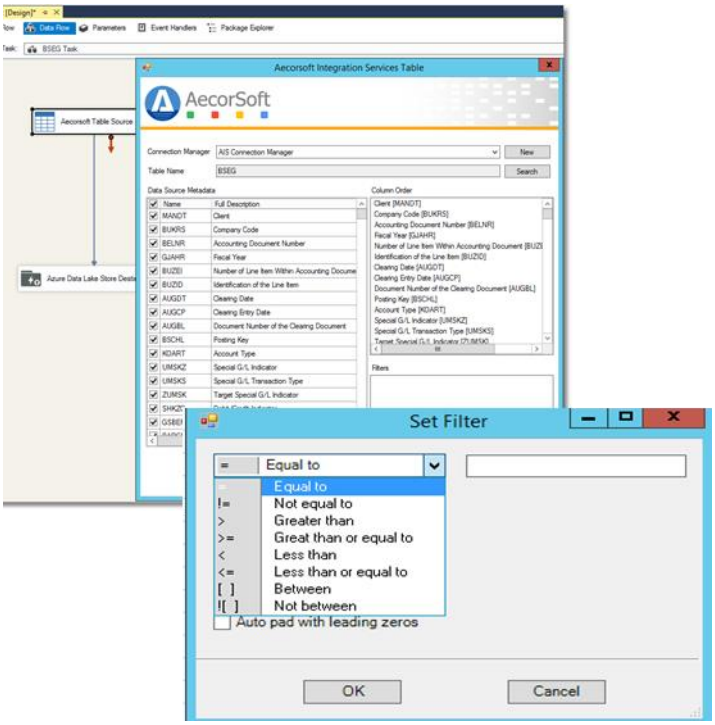
AecorSoft Connection Manager

AecorSoft Connection Manager supports load balancing and single sign-on at SAP ECC/BW's application stack.



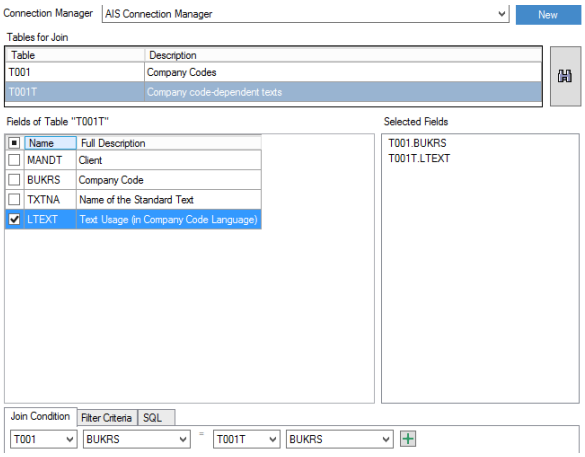
SAP Table / View / CDS View

Through the user friendly interfaces, the metadata of SAP tables, views and CDS Views can be browsed, and filters can be defined. Both static filter values and dynamic filters through SSIS user variable are supported.



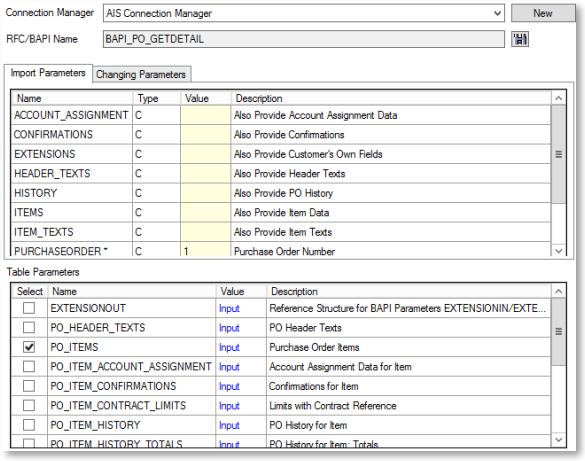
SAP Table Join

Table join conditions and filters can be defined.



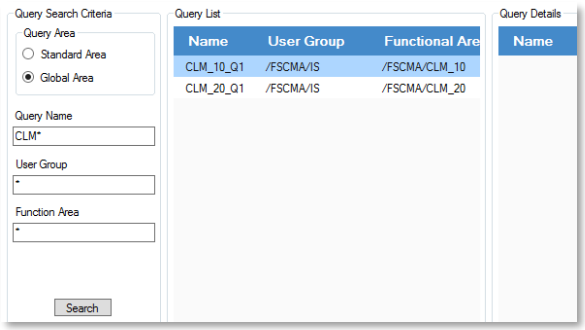
SAP BAPI / RFC

BAPI and RFC functions can be executed, and the result tabular data can be accessed in the data flow.



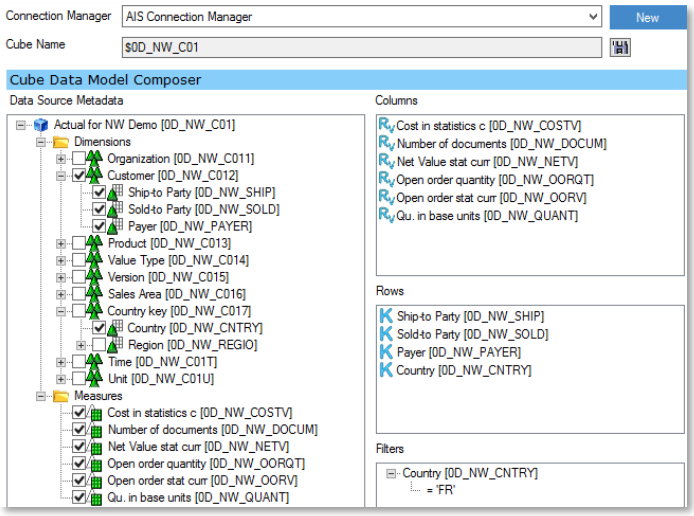
SAP InfoSet Query

SAP InfoSet queries can be browsed and executed easily.



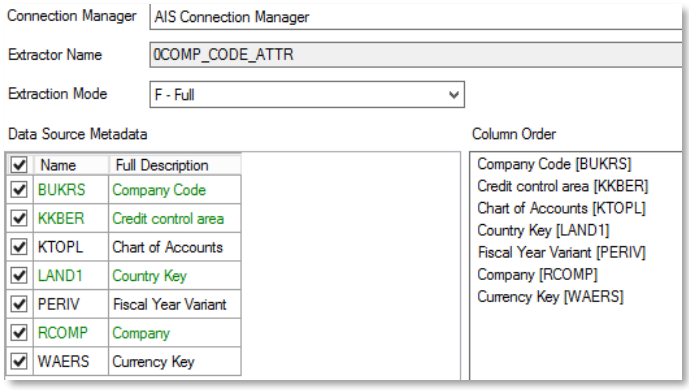
SAP InfoCube / BEx Query

BW InfoCubes and BEx Quries can be browsed through a catalog interface. Dimensions, Key Figures, Hierarchies are supported. Unlike other tools on market, AecorSoft handles cube data extraction package by package with very high performance, instead of loading everything in memory.



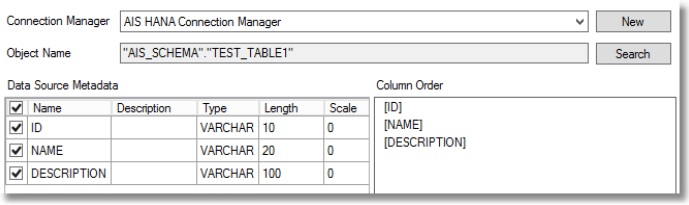
SAP Delta Extractor

Delta queue extractors are completely supported through best-in-class algorithms, with rich delta management features to conduct initialization with transfer, initialization without transfer, full load, ongoing delta, and recovery of recent delta.



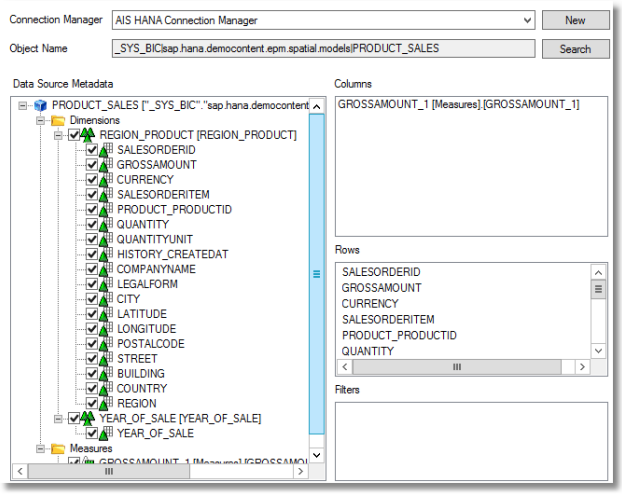
HANA Catalog (Table/View)

HANA tables and view are supported with detailed schema information.



HANA Content (Calc View / Analytic View)

HANA calculated views and analytic views are treated as multi-dimensional.



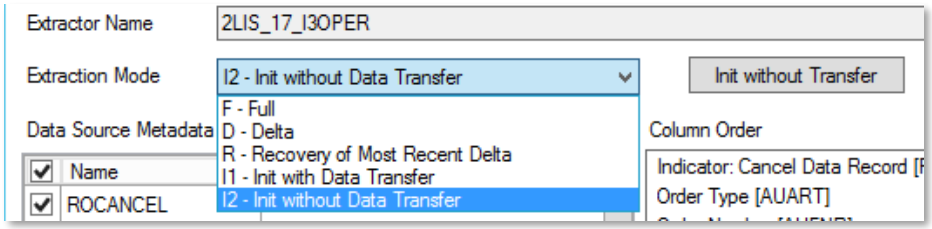
The Journey of Bringing SAP data to Azure: Incremental Delta Extraction

Delta extraction is one of the biggest challenges to almost all SAP centric DW and BI projects. Through AecorSoft, the SAP delta management is significantly simplified into two scenarios:

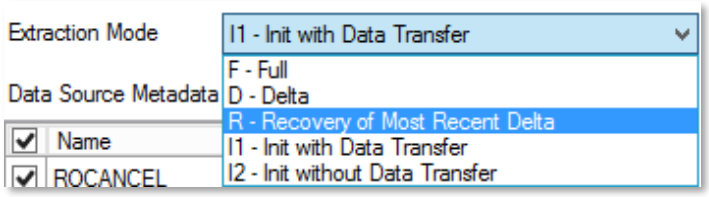
SAP Managed Delta

In this scenario, the inserted/deleted/updated records are actively tracked by SAP itself in the "delta queue". Normally, when standard extractor is available, we want to take advantage of it. "AecorSoft Extractor Source" component is leveraged to communicate with SAP delta queue extractors. The delta capabilities of extractors are intelligently detected to allow data extraction activities like init-with-data-transfer, init-without-data-transfer, and delta-data-recovery.

The intention of init-without-transfer is to set the delta pointer with minimal impact to system downtime. After the delta pointer is reset, all future delta loads will go from that point. BI devops can conduct the necessary full historical data load at a later time.



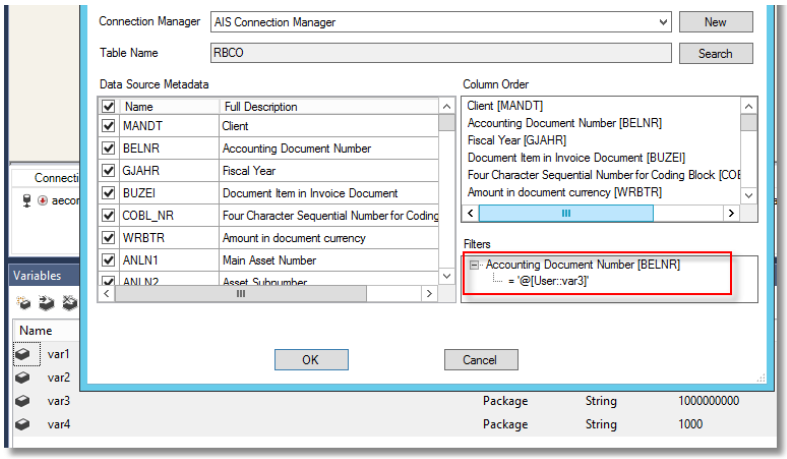
If the delta data package is corrupt at destination, then the most recent delta can be repeated through the option of “Recovery of Most Recent Delta”:



User Managed Delta

When there is no suitable delta extractor available, users can choose to manage delta by themselves. Instead of relying on SAP to track delta pointers, users will need to maintain the delta pointer table, and apply the filter to fetch new or updated records since recent run.

Or, dynamic filtering can be achieved through the usage of SSIS user variable.



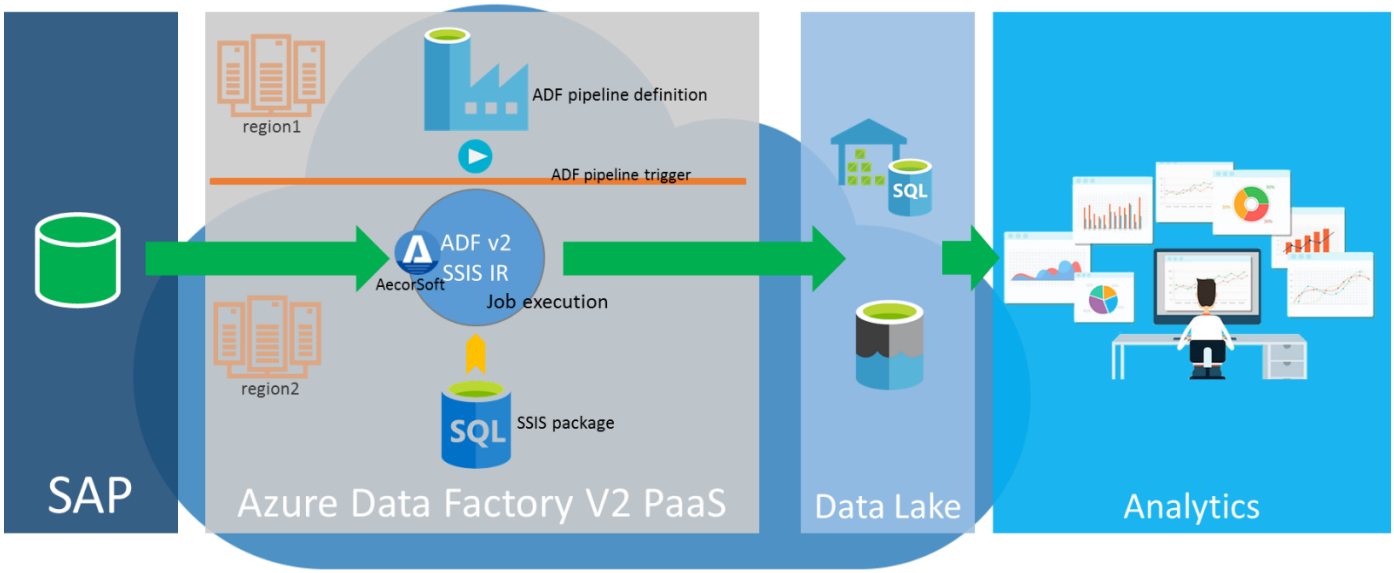
The Journey of Bringing SAP data to Azure: ETL-in-the-Cloud via ADF V2

In order to land the extracted SAP data onto Azure data platform, the Azure destination components from the SSIS Azure Feature Pack can be utilized.

After the on-premise design and development is completed, the SSIS jobs are ready be deployed to on-premise SSIS server or IaaS VM based SSIS server to stream data from SAP to Azure storage.

But how about the idea of ETL-in-the-Cloud? Thanks to the Azure Data Factory V2 SSIS Integration Runtime (ADF V2 SSIS-IR), users can now lift and shift the SSIS solutions to Azure, to enjoy a complete Platform-as-a-Service ETL experience. One immediate advantage is the instant elastic computing, that more nodes can be ramped up immediately to handle peak loads during busy season, and ramped down during regular time to save cost.

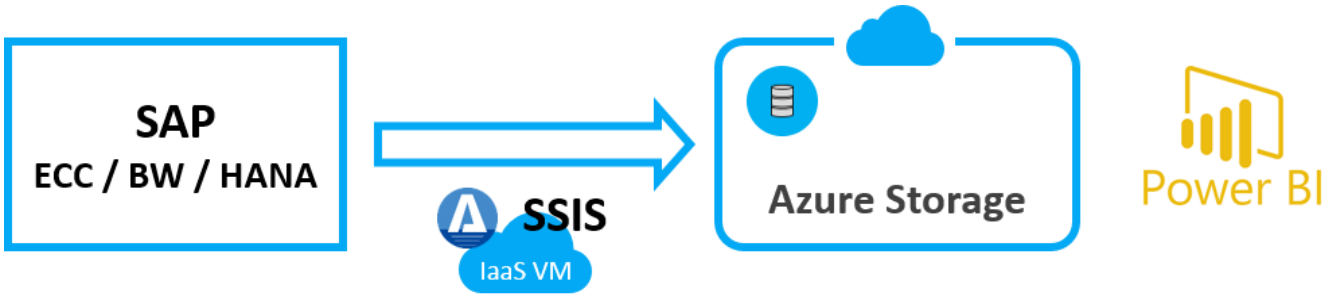
AecorSoft also fully supports ADF V2. The SAP/AecorSoft/ADFv2 reference architecture is:



The Journey of Bringing SAP data to Azure: A Multi-Phase Strategy

Rome is not built in one day. After the Future State Architecture is outlined, a multi-phase roadmap is recommended:

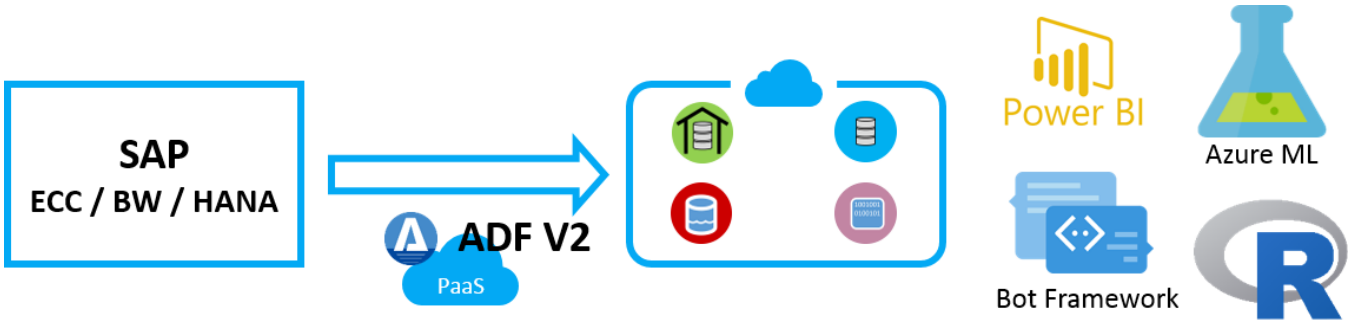
Phase 1: focusing on bringing data out of SAP to realize business value



Highlights of this approach:

- ***Cloud Story***: leveraging Azure IaaS VM to host SSIS server;
- ***Time-to-Market***: quickly delivering SAP data to Azure storage using agile methodology;
- ***Customer Experience***: Modern analytics presented through Power BI for a "quick-win";

Phase 2: marching onto cloud ETL + advanced analytics and insights



Highlights of this approach:

- ***Cloud Story***: deploying SSIS packages to ADF V2 for a complete PaaS ETL experience, with integration to more advanced Azure storage;
- ***Time-to-Market***: continuous shipping possible through cloud-ready technologies and methodologies;
- ***Customer Experience***: More intelligent business insights driven by additional Azure data services, for continuously improved business customer experiences and results.

Conclusion

The power of Azure Data Platform along with the superb data extraction capability of the AecorSoft SAP components lead to an efficient and elegant end-to-end solution, to bring SAP data to Azure securely and productively.

Ultimately, by delivering SAP data to Azure through ADF V2 with high performance and low latency, customers are now able to conduct advanced analytics on the massive SAP data in cloud, and truly take advantage of the elasticity of clout computing to drive efficient business insights.