1. **[Canada Life] Meeting June 2020 delivery of all MVP capability promised for the code name Arcadia release.**

Since the Synapse Analytics announcement at end of Q4 2019, MS has been working to assemble enough capability, often leveraging existing capability from other Azure products (e.g. Azure Data Factory), and expose it consistently through the Synpase UX, to provide a reasonably competitive cloud data warehouse offering in the industry. This is the low watermark for a viable product that may be able to meet some of the basic needs of our constituency. If June delivery slides, I’d question whether you’d even have a viable product for our community to make use of, in-year.

**[1. Microsoft viewpoint]**

**1. Summary**

There may be a misconception here that are critical to clarify. The SQL Azure Datawarehouse as well as Azure Data Factory that will support Canada Life’s requirements that we were able to capture were generally available for quite some time and read to use. Microsoft have many customers that have implemented enterprise data warehouses on Azure Synapse today. These customers are also leveraging Azure Data Factory and GA Spark platforms including Azure Databricks and Azure HDInsight

Synapse UX, i that ADD VALUE (i.e. common UI for integration, reporting etc.) comprise of products that are ALSO generally available and ready to use. These products form a solution that currently provide an extremely competitive cloud datawarehouse offering in the industry. Activities such as ingestion of XML can be performed in other Azure Services that are optimized for ingestion and typically cost less than Synapse Analytics.

Please note that Azure Synapse Analytics was previously branded as Azure SQL DataWarehouse (a mature datawarehousing product in the cloud).

**1. Details**

All of the other functionality we are bringing with the Azure Synapse preview (pipelines, ingestion, native Spark, SQL On-demand, native PowerBI) are capabilities that will unify the experiences which no other competitor currently has.  This native integration of the entire data pipeline is so far above and beyond what any other competitors (like SnowFlake) currently offer.  It is important to recognize that these are the capabilities that are still in preview but **our data warehouse solution is GA**, and Snowflake does not have anything close to our vision of a unified analytics platform.

We highly suggest that Canada Life evaluates based on a full offering which includes data integration and big data processing. Many of our competitors do not have a proper roadmap.

Features that are currently Generally Available:

* MPP (Massive Parallel Processing) architecture
* Security (Data encryption at rest, in transit, etc)
* Polybase TSQL queries/CTAS
* Pause and resume/Storage and compute separation
* Columnstore
* Workload classification
* Materialized views
* JSON support
* XML datatypes stored as varchar
* (and many, many more)

1. **[Canada Life] Dynamic, real-time elastic service scalability with guard rails (automatic compute shutdown/start-up, automatic service scaling to accommodate demand).**

Today all instance tuning is manual and applying adjustments involve and instance shutdown/ restart. Because of this, it means that there will be times when were paying for more compute than we’re actually using, as compute doesn’t automatically scale as required and de-scale when not in use. Managing capacity consumption today involves several tools that MS touts as very useful, but adopt an old-style data lake approach, requiring active operational management to juggle (e.g. community compute priority queuing)

**[2. Microsoft viewpoint]**

**2. Summary**

Workload management is a diverse problem with a diverse set of solutions, and the workload isolation in Azure Synapse is a powerful feature that is better suited for handling most types of variability than competitor’s auto-scaling.

The two main points here are:

1. Do we support automatic shutdown/startup/ auto scaling -YES, this is supported. Here are 3 ways it can be implemented:
   1. We have REST API’s, PowerShell and Scripts that support this as well as how to implement it (<https://azure.microsoft.com/en-ca/blog/azuresqldw-cost-savings-with-autoscaler-part-2/>)
   2. This is also on our road map
   3. We have third parties that can support this
2. How to minimize payment on compute and maximize the compute usage when it is allocated i.e. prioritizing workloads – critical to remember:
   1. Competitor auto-scaling is not instant and **evicts the cache, which is a major performance loss**
   2. Workload isolation/prioritization in Synapse does not impact customer workloads or evict the cache and can ensure that your data warehouse is always at 100% utilization, even with constantly variable workloads.  Many customers also appreciate the ability to set a single, defined price point and then manage their workloads within that price point, without ever having to deal with any price unpredictability.

It is critical to work on the of a use case vs on a feature basis. Canada Life would benefit more from actual business use cases being fulfilled vs a potential feature.

**2. Details**

Workload management is a diverse problem with a diverse set of solutions, and the workload isolation in Azure Synapse is a powerful feature that is better suited for handling most types of variability than competitor’s auto-scaling.  Specifically, auto-scaling is not instant and **evicts the cache, which is a major performance loss**.  If you’re constantly trying to scale say, Snowflake clusters’ up and down to match your workload, your performance will degrade because your cache is constantly evicted.  Workload isolation in Synapse does not impact customer workloads or evict the cache and can ensure that your data warehouse is always at 100% utilization, even with constantly variable workloads.

Workload importance is just another tool in the workload management toolbox that is not offered by other competitors.  With many other customers in the same industry as Canada Life, prioritizing a single workload or user without impacting performance is a critical feature.

Note that unlike the competitors, in order to prioritize workloads Azure Synapse DOES NOT NEED TO:

* spin up a whole new cluster, with a separate endpoint
* That separate cluster would not have to be always on, or the user would face provisioning delays every time they submitted a query.
* separate cluster could not take advantage of the cache in any other clusters, leading to slower performance.

In Synapse, workload prioritization is as simple as just setting the importance in the workload group – no additional overhead, no additional cost.

1. **[Canada Life] Virtual warehouse separation over shared storage.**

Segmented multi-compute accessing a single storage allocation with ACID transaction compliance. Facilitates data sharing with segmented compute (i.e. separate virtual warehouses sharing data) and horizontal scalability, to improve concurrency. Today, instance storage and compute are a linked in a hard 1-1 relationship; not terribly flexible or modern. The only data sharing capability presently is facilitated through asynch snapshot transfer from instance to instance via Azure Data Share – not really data sharing.

**[3. Microsoft viewpoint]**

**3. Summary**

We have seen this question ask a few times and discovered this is a SnowFlake proposed question. The details section will discuss our response. The main thing we also ask customers is instead of asking about the multi compute scenario, to detail the real life use case of what Canada Life’s scenario is with the clear amount of data. We are confident that we can provide a solution architect that can satisfy the needs of Canada Life.

**3. Details**

Snowflake’s ‘horizontal scalability’ in their warehouses may actually be a massive cost. Snowflake **obfuscates** their maximum number of concurrent queries because it allows them to claim unlimited concurrency without customers diving too deep into their numbers.  In a Snowflake warehouse, query performance scales with cluster size (# of servers) while concurrency scales with # of clusters.  **Achieving high concurrency and high query performance can get very expensive very quickly**.

For example, the Gigaom report compares a 15000c Synapse data warehouse (30 nodes, $180/hr) with a 3X-Large Snowflake warehouse (30 nodes, $192/hr at Enterprise Edition).  However, the concurrency for that Snowflake warehouse is only 8 – while that number is flexible, [Snowflake’s documentation](https://nam06.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdocs.snowflake.net%2Fmanuals%2Fsql-reference%2Fparameters.html%23max-concurrency-level&data=02%7C01%7CWinson.Woo%40microsoft.com%7C3a37a2c74ea04be881e208d7d038becb%7C72f988bf86f141af91ab2d7cd011db47%7C1%7C0%7C637206814445293389&sdata=njKBaSsys%2BVQc1uZuOO%2Fq4af7KKngeAru1ePWufhm8Q%3D&reserved=0) encourage customers to only make small, incremental changes, and that they’ve chosen 8 as the recommended values because it balances performance and usage.  Even then, that number doesn’t guarantee 8 concurrent queries as larger or smaller statements may consume many or a fraction of that total.

To scale their concurrency further, the customer would need to provision additional clusters within their warehouse.  However, to ensure query performance, they can’t scale down from 3x-Large, since **scaling out clusters doesn’t improve query perf.** From [Snowflake’s docs](https://nam06.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdocs.snowflake.net%2Fmanuals%2Fuser-guide%2Fwarehouses-multicluster.html%23benefits-of-multi-cluster-warehouses&data=02%7C01%7CWinson.Woo%40microsoft.com%7C3a37a2c74ea04be881e208d7d038becb%7C72f988bf86f141af91ab2d7cd011db47%7C1%7C0%7C637206814445303388&sdata=XSWuMjc7jB7%2F8KMLZkI8fPYUFsN0ZwEW5EkPUsAywBw%3D&reserved=0): “Multi-cluster warehouses are best utilized for scaling resources to improve concurrency for users/queries. They are not as beneficial for improving the performance of slow-running queries or data loading. For these types of operations, ***resizing*** the warehouse provides more benefits.”

Therefore, to achieve higher performance and concurrency, the customer’s costs are now $192/hr x (# of clusters).  Meanwhile, Synapses still provides 128 concurrent queries at $180/hr.  This is without even taking Snowflake’s tiered pricing into consideration, where customers are charged extra for isolated resources and security features, or Azure Synapse’s reserved capacity model to drive costs down even more.

Azure Data Share’s primary advantage over Snowflake is that Azure Data Share enables **open, flexible sharing**.  Snowflake data sharing locks you into Snowflake’s proprietary format, and forces anyone you want to share the data with to also buy Snowflake compute – this is a huge barrier to any partner running any other data warehouse or compute system.  Azure Data Share instead shares the files directly, enabling partners to use their preferred tool to work with the data.  Furthermore, Azure Data Share is also building out in-place sharing as well for multiple database offerings in Azure – as those features move to GA, we will offer a more complete data sharing story, with both in-place and snapshot sharing that **does not lock customers into a proprietary format behind Snowflake’s paywall.**

1. **[Canada Life] Much lower client operational management footprint.**

Automatic service monitoring and adjustment + on-demand elastic horizontal/vertical scaling and shutdown based on instance concurrency + performance. Still have to actively monitor the service (through Azure Monitor), interpret those more traditional infrastructure performance metrics (CPU, memory, I/O, etc.), determine a course of action and then take active manual scaling steps (including service restart) to adjust capacity, in-day, if necessary. Competitors use automatic active monitoring, some ML-powered, and automatically adjust capacity, in real-time, on your behalf.

**[4. Microsoft viewpoint]**

**4. Summary**

This is a similar concern around the multi-cluster discussion.  Workload isolation can help ensure 100% utilization throughout the day at a given scale-factor, and customers still have the ability to scale up or down as needed during down time.  Constantly scaling your data warehouse up and down will tank your performance due to a constantly evicted cache.

The focus should be based on Canada Life’s use cases and the focus on supporting Canada Life’s day today operations. From a client operational management foot print, implementation on continuous integration and security are part of Microsoft’s core offering. Microsoft offers key dev ops, established scenarios around code deployment and support with Azure Synapse Analytics. Many other competitors trail in this space.

1. **[Canada Life] Leading D&A ecosystem vendor integration.**

Lots of integration opportunities with other Microsoft technologies in the D&A space but myopia wrt to leading external D&A ecosystem vendor partner integration (what if we want to augment our tooling with other vendors).  Seems particularly true in the Data Science area.  MS needs to adopt a more open strategy to partner integration in the D&A space – without it you’re left with lowest common denominator ODBC integration; not sufficient.

**[5. Microsoft viewpoint]**

**5. Summary**

This section will require further clarification. Microsoft is known for its integration and data openness. There were questions asked around custom adapters: we had responded with Tableau, Alteryx, QuerySurge and a whole myriad of different connectors. Our partner ecosystem is enormous as they develop custom adapters to our system.

Please note that perhaps Canada Life had been searching for Azure Synapse Analytics which was a rebrand roughly 4 months ago. The previous brand was Azure SQL DataWarehouse and searches against that may yield more results.

We are happy to provide recommended patterns for specific use cases or details on specific partner connectors.

1. **[Canada Life] Data governance.**

Particularly the ability to track technical metadata around lineage/provenance, as data is engineered, curated within the Synapse eco-system. Microsoft has admitted that their current data catalogue offering is very weak. They’ve acquired a data governance vendor, BlueTalon, and are working to reconstitute the offering, integrate purchased capability, and open up the catalogue to ATLAS integration. However, this was still to be a separately licensable offering, apart from Synapse itself, so it’s unclear what level of integration will exist with the initial delivery.

**[6. Microsoft viewpoint]**

**6. Summary**

Please note that the Azure Data Catalog does fall under a different team. Similar to Azure Synapse Analytics, there are features that are mature and GA. We have not heard of the use case requirement from Canada Life and can happily engage the right team to review this offering as we have done with the Synapse Team.

While this may be billed as a different product, our customers will not be paying for a datawarehouse service to perform meta data management.

As a point of clarification, this product will not be licensed but as with most cloud offering billed by use. As part of the briefing, we are happy to explain the level of integration and delivery. There is currently little to no other offerings in the industry that will provide the capabilities that this new Azure Data Catalog can provide.

1. **[Canada Life] Data marketplace capability.**

Ability to easily share data between internal constituencies and securely, with qualified external partners (e.g Group Sponsors). Ability to easily incorporate and leverage useful external data sources (StatsCan).  Azure Data Share asynch snapshot sharing is not sufficient, would be challenging for federated, non-technical, citizen data science use, and would require similar integration target effort as a B2B feed.

**[7. Microsoft viewpoint]**

**7. Summary**

In all fairness, we had spent a total of 5 mins on Azure Data Share which is a different product than Azure Synapse. We can provide more details on these as well as its main feature set offers unprecedented flexible and secure sharing in B2B scenarios.

We urge you to provide scenarios and use cases on what data sharing capabilities are required. Many of the open data set capabilities that require publishing and sharing are simple solution architectures that we have done many times before. With Financial type organizations integrations to equities, government publications and other types of open streams we have been performing for decades.

**7. Details**

The following provides a higher level of detail around some of the capabilities of Azure Data Share.

* Azure Data Share can be used in a number of different industries. For example, a retailer may want to share recent point of sales data with their suppliers. Using Azure Data Share, a retailer can set up a data share containing point of sales data for all of their suppliers and share sales on an hourly or daily basis.
* Azure Data Share can also be used to establish a data marketplace for a specific industry. For example, a government or a research institution that regularly shares anonymized data about population growth with third parties.
* Another use case for Azure Data Share is establishing a data consortium. For example, a number of different research institutions can share data with a single trusted body. Data is analyzed, aggregated or processed using Azure analytics tools and then shared with interested parties.

Azure Data Share enables data providers to:

* Share data from the list of [supported data stores](https://docs.microsoft.com/en-us/azure/data-share/supported-data-stores) with customers and partners outside of your organization
* Keep track of who you have shared your data with
* Choice of snapshot or in-place sharing
* How frequently your data consumers are receiving updates to your data
* Allow your customers to pull the latest version of your data as needed, or allow them to automatically receive incremental changes to your data at an interval defined by you

Azure Data Share enables data consumers to:

* View a description of the type of data being shared
* View terms of use for the data
* Accept or reject an Azure Data Share invitation
* Accept data shared with you into a [supported data store](https://docs.microsoft.com/en-us/azure/data-share/supported-data-stores).
* Trigger a full or incremental snapshot of a Data Share that an organization has shared with you
* Subscribe to a data share to receive the latest copy of the data through incremental snapshot

All key capabilities listed above are supported through the Azure portal or via REST APIs. For more details on using Azure Data Share through REST APIs, check out our reference documentation.

**Security**

Azure Data Share leverages the underlying security that Azure offers to protect data at rest and in transit. Data is encrypted at rest, where supported by the underlying data store. Data is also encrypted in transit. Metadata about a data share is also encrypted at rest and in transit.

Access controls can be set on the Azure Data Share resource level to ensure it is accessed by those that are authorized.

Azure Data Share leverages Managed Identities for Azure Resources (previously known as MSIs) for automatic identity management in Azure Active Directory. Managed identities for Azure Resources are leveraged for access to the data stores that are being used for data sharing. There is no exchange of credentials between a data provider and a data consumer. For more information, refer to the [Managed Identities for Azure Resources page](https://docs.microsoft.com/azure/active-directory/managed-identities-azure-resources/services-support-managed-identities).

1. **[Canada Life] Lack of MS training / education / community offerings.**

Our federated constituencies are going to need access to training and self-service Q&A support. MS-moderated/curated user community supports are practically non-existent, both for the current offering and its pre-cursor, Azure Data Warehouse. We asked about training opportunities, but never received information back.  We investigated self-service education – it’s pretty sparse (< 20 use case-focused YouTube videos). Active community support was practically non-existent on Stack Overflow; we found a few Q&A threads. May actually be a harbinger of very low actual industry adoption of the toolset.

**[8. Microsoft viewpoint]**

**8. Summary**

Microsoft has been a longtime leader and creator of content and communities to help customers and partners get educated and drive value out of their Microsoft investments. The list below identifies clearly that there is no lack of documentation, self-paced learning, community support and momentum that would guide Canada Life through their journey. Whether it be through Microsoft Learn, Docs, GitHub, Ignite, Build and Inspire events and Tech Community or the thriving global partner populus, the Azure Synapse community continues to flourish representing a healthy ecosystem being established. As well, we want to highlight the specific investment that Microsoft is making into Canada Life specifically on upskilling employees to help accelerate their overall Azure journey through our Technical Skills for Business program. The following is a very abbreviated list of the many resources available.

# Enterprise Skilling Initiative Program

This program from Microsoft is dedicated to strategic customers. The aim is to introduce, plan and implement a customized Azure learning plan for Canada Life. Skills are approached in 3 phases:

* 1. Fundamentals
  2. Role Based Skilling and Training
  3. Specialty Courses

The contents of a Skilling Plan include:

* 1. Access to Local Events
  2. Digital Skilling including Online Resources and Online training sessions
  3. Instructor Lead Training
  4. Certification Vouchers to validate the training

The program with the majority funded by Microsoft, including vouchers and is not one time. Once a certain threshold of Cl employees obtain certification, the entire program is re-issued again. This program has already been introduced and operationalized to Canada Life with Hye Chong Li being the Program Coordinator and Mike Brown as the Exec Lead. A 50 person Azure Fundamentals training class and testing session was scheduled for March 13 but was postponed due to the impact of COVID-19. An online equivalent will be provided and attendees can take their certification tests at home. Role based skilling has yet to be defined by CL but can result in a program leading employees to attain Azure Data Engineer Certification which includes extensive training for Azure SQL Data Warehouse and Azure Synapse.

The strong mix of self-paced learning, communities, certification tracks, global partner ecosystem, live in person events, vast Microsoft resources and investments in Canada Life reflects the commitment we have to the platform, our customers and the ecosystem.

# Public (Microsoft)

**Microsoft Docs** – Including reference architectures, performance tuning, managing/monitoring, security and loading.

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/>

**Microsoft Learn** – self paced learning for Synapse as part of broader Data Engineer and Data Architect Learning paths.

<https://docs.microsoft.com/en-us/learn/browse/?roles=data-engineer%2Csolution-architect&products=azure-synapse-analytics>

**Microsoft Whitepapers**

<https://azure.microsoft.com/en-us/resources/whitepapers/search/?service=synapse-analytics&type=WhitePaperResource>

**Microsoft Resources**

Canada Life has an extensive team from Microsoft to help support their journey. As an earmarked Strategic Customer, Canada Life also has access to resources that other Microsoft customers may not have access to including Cloud Solution Architects, Global Black Belts (field specialists with direct connection to Engineering) and Solution Specialists. As well, Canada Life has had direct engagement with Synapse Product Engineering on Synapse first at the Executive Briefing in January, 2020 and then in direct meetings with Engineering during a visit to Canada.

**Additional Resources** – Including customer references, technical documentation and guidance on how to start

<https://azure.microsoft.com/en-in/services/synapse-analytics/resources/>

# Community & Events

**Microsoft Tech Communities** – Azure Synapse, part of a 113K Azure user community, with dozens of blog posts with thousands of views

<https://techcommunity.microsoft.com/t5/azure-synapse-analytics/bg-p/AzureSynapseAnalyticsBlog>

**28 GitHub Repos on Synapse** including:

Azure Synapse Toolbox

<https://github.com/microsoft/Azure_Synapse_Toolbox>

Azure Synapse Labs on GitHub (using National Oceanic and Atmospheric Organization, NOAA)

<https://github.com/TechDataGeek/AzureSynapseHOL>

**Microsoft Ignite 2019**

This is the event where we announced Synapse Analytics. There were over 700 individual sessions that made references to Synapse with 10 specific breakout sessions dedicated to Synapse including the main Keynote by Satya Nadella making the initial announcement. [Link](https://myignite.techcommunity.microsoft.com/sessions?t=%257B%2522from%2522%253A%25222019-11-03T08%253A00%253A00-05%253A00%2522%252C%2522to%2522%253A%25222019-11-08T19%253A00%253A00-05%253A00%2522%257D&g=%255B%2522on-demand%2522%255D&q=synapse&s=%257B%2522name%2522%253A%2522translate.refine.label.sort.relevance%2522%252C%2522type%2522%253A0%257D) is for the 10 specific Synapse sessions, completely viewable on demand.

In addition, Azure Synapse Analytics embraces T-SQL as the standard language. What this means is that everyone in the SQL ecosystem can immediately be productive on Synapse. Microsoft has invested decades into developing the community and ecosystem for SQL.

There are many more, this is only a quick two-page copy of a web search. The following are just some of the trainings available:

**YouTube (Azure Synapse)**

<https://www.youtube.com/results?search_query=azure+synapse>

**YouTube (Azure SQL Datawarehouse)**

<https://www.youtube.com/results?search_query=azure+sql+data+warehouse>

# Microsoft and Third-Party Learning Providers

**Azure Synapse On Demand Virtual Event (Microsoft)**

<https://info.microsoft.com/Analytics-in-Azure-virtual-event-Accelerate-Time-to-Insight-with-Azure-Synapse-Analytics-On-Demand-Registration.html>

**Azure Hands On Labs (Microsoft):** In person events happening across the globe, including several in Canada. Note these will be rescheduled due to ongoing COVID-19 crisis

<https://www.microsoft.com/en-ca/sites/azurehands-onlabs/?wt.mc_id=AID2478210_QSG_381386&ocid=AID2478210_QSG_381386>

**Cloud Academy**: [Cloud Academy Link](https://cloudacademy.com/course/design-document-data-flows-azure/azure-sql-data-warehouse/)

**Pluralsight**: [Plurasight Link](https://www.pluralsight.com/courses/azure-sql-data-warehouse-first-look)

**EdX**: [EDX Link](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=14&cad=rja&uact=8&ved=2ahUKEwir-cu_l7boAhWMbc0KHfkVBY4QFjANegQIARAB&url=https%3A%2F%2Fwww.edx.org%2Fcourse%2Fdelivering-a-data-warehouse-in-the-cloud&usg=AOvVaw2-V0PiXl1_w860YwtoYgls)

**Udemy**: [Udemy Link](https://www.udemy.com/course/the-complete-sql-bootcamp/?gclid=CjwKCAjwguzzBRBiEiwAgU0FTw6Lq1zyWzuZZGsaCaH_1D1hoakx5HVAs_jg9ABpJvl2OcIWr8LIpBoCXjcQAvD_BwE&matchtype=e&utm_campaign=SQL_CA&utm_content=deal4584&utm_medium=udemyads&utm_source=adwords&utm_term=_._ag_62726991097_._ad_340861214140_._kw_azure+sql+data+warehouse+training_._de_c_._dm__._pl__._ti_aud-632926098907%3Akwd-506608979234_._li_9000784_._pd__._)

**LinkedIn Learning**: <https://www.linkedin.com/learning/search?keywords=azure%20sql%20data%20warehouse&u=3322>

1. **Product pricing, particular in the area of storage**

out of line with competitors and even MS Azure similar storage pricing. MS has committed to closing that gap and has offered to price specially, in the interim, for us.

**[10. Microsoft viewpoint]**

Azure Synapse offers a number of programmatic offerings, like reserved capacity discounts to drive down costs. When comparing data warehouse costs, compute costs quickly out-scale storage costs, so Canada Life could potentially see greater savings on Synapse because of our superior price/performance for compute.

In short, Microsoft is willing to work with Canada Life in good faith to establish pricing for Azure Synapse that is mutually agreeable and aligned with Canada Life’s specific use cases.