

Introductions

Wednesday, August 2, 2023 10:16 PM

SharePoint (Teams) location: OAS-MRIS Assessment --> Documents--> General -->03 Requirements, Development Specs

Presentation Link: [MRIS Modernization Reference Architecture & Solution Patterns.pptx](#)

Insurance Solutions

Wednesday, March 15, 2023 10:04 PM

https://uhgazure-my.sharepoint.com/:p/g/personal/john_shin_optum_com/EUNdN6SqbG9Oiw7bpkG0aYBafstU8-f3pI3ks2F39IXww

Work Plan

Tuesday, March 21, 2023 10:24 PM

OAS Approach

- OAS will conduct current state assessment
- Provide a high-level time and cost estimates based on prior projects of similar scope, size and complexity
- Provide a CBA report based on industry, organizational and personal experiences
- OAS will adopt a phased approach
- Closely work with the Insurance Solutions (IS) SMEs and Key stakeholders

Assessment Approach

- Identify candidate data platform(s)
- Identify dependencies
- Current toolset
- Identify SMEs (technical and business)
- Identify Key stakeholders
- Collect existing collateral
- Setup meetings and deliver them in advance with proper agenda
- Conduct interview with the SMEs and key stakeholders to understand the nuances
- Understand pain points, vision of the future state
- Collect all data points
- Sort and sift findings
- Conduct analysis
- Create a findings report
- Review it with the SMEs and make necessary tweaks
- Present and deliver the Current state report to IS
- Receive **IS approval sign-off**
- Provide reference costs

Environment Setup

- Make request to setup a Snowflake instance
- Setup development, testing, and production environments
- Identify user personas
- Create a permissions model and make appropriate permission requests
- Stand up a Snowflake instance
- Setup network setup and connectivity
- Setup needed storage account(s)
- Ensure storage accounts have proper permissions
- Setup up the data pipeline (ADF or Kafka)
- Setup monitoring, notifications and logging
- Validate components, processes, and pipelines
- Acquire appropriate approvals, secure implementation team and Set up a Snowflake instance
- Understand the cause creating data access challenges

Parallel to Environment Setup - Roadmap

- A roadmap
- Identify data domains, any dependent processes
- Create a priority based data migration plan
- Create Target/future state Architecture
- Develop user personas, access and governance models
- Establish implementation plans to partner and work with the IS (Doug's team)
- Establish quality and standards up front
- Understand the contacts connection issues and help setup appropriate contacts

Pilot program

- Identify the candidate data domain/data source (*Note: it should be small and simple*)
- Identify test cases
- Test components, connectivity, processes and user permissions using the data workload
- Validate the pipeline and results end-to-end
- Change workload sizes and configurations to see if the system behaves differently
- Make changes, if needed
- Test it again
- Collect and analyze metrics to determine the quality and cost
- Review results with internal SMEs
- Produce the outcome report
- Deliver results to the IS
- Receive **IS approval sign-off**

Development approach

- OAS will adopt a phased approach - platform by platform (domain by domain)
- OAS will use a framework first approach to maximize reusability
- Agile (SCRUM) development methodology will be used for quick feedback and course correction
- At the completion of each phase, a demo will be given, the outcome will be reviewed and accepted by the IS
- Create a POC/pilot project
- Ensure data uniqueness during data ingestion
- Implement Data DevOps

Prerequisites

- All requested collateral is made available at the outset of this endeavor
- SMEs and Key stakeholders are available for meetings and interviews
- Any outstanding questions are responded to within 48 hours
- Access to infrastructure and operations is provided as needed
- Realistic and sanitized data is provided

OAS Team

- Product Owner
- Scrum Master
- Development Teams
 - Business Analyst
 - Architect
 - Technical Lead
 - Azure SMEs
 - Databricks SMEs
 - Data Architect
 - Data Engineer
 - Developers
 - Technical Writer
 - Azure Deployment Engineer(s)

Insurance Solutions Team

- Business SME
- Technology SME
- Project Manager

OAS Team

- Solution Architect - SME (John Shin/Anantha) - PT - provide oversight, guidance, direction and review/feedback, validate approach and outcome
- Delivery Architect (Dave Cheema) - FT - set direction, execution plan, Azure cloud architecture, work closely with MRIS SME, receive, review and analyze current state data platforms documentation, identify gaps and opportunities, incorporate them into future state architecture. Analyze data points, review results with MRIS SME, and create current state assessment report
- Business Analyst (TBD) - FT - will work closely with MRIS SMEs, collect and document findings and will help create necessary collateral
- Cloud Data Architect (junior) (TBD) - FT - help team collect data models, help setup the cloud data platform and structures, document migration, integration, and consumption patterns. Help document findings

MRIS SMEs

- Collect, organize reference material
- Setup technology and business SMEs meetings
- Help us identify data loads
- Provide domain knowledge
- Help resolve issues
- Validate data points collected
- Provide Data Access
- Create data extracts
- Develop user personas, access and governance models
- Provide quality and standards

Program Management & Planning

- Program Status Meetings
 - Daily Scrum
 - Weekly Leadership Status
 - Program Governance Checkpoint
- Program Increment Planning Sessions
 - Identify & Document Features
 - List & describe activities & features (user stories)
 - Enter user stories into Rally
- Program Scope Change Control (As Needed)
 - Potential Change Control for ETL Conversion Tool Costs

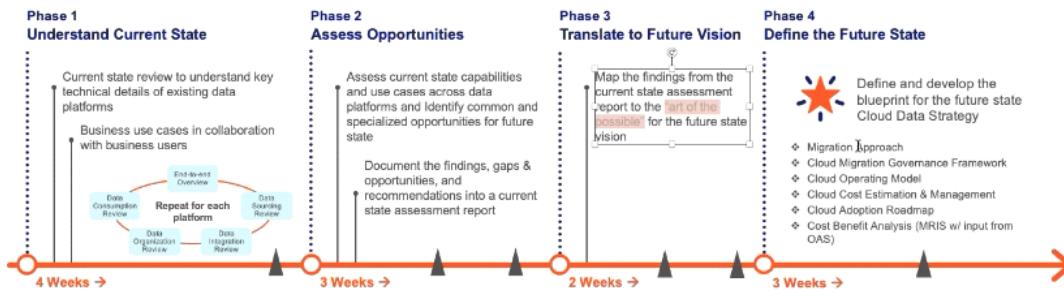
+++++
+++++

| Unique Work Item | SDLC Phase |
|---|--|
| Provide reference costs | Assessment phase |
| Acquire appropriate approvals, secure implementation team and Set up a Snowflake instance | Environment Setup |
| Understand the cause creating data access challenges | Post environment Set |
| Create a POC/pilot project | Initial sprint |
| Create a priority based data migration plan | Parallel to Environment setup |
| Create a roadmap | Parallel to Environment setup |
| Create Target/future state Architecture | Parallel to Environment setup |
| Develop user personas, access and governance models | Parallel to Environment setup |
| Establish implementation relationship and work with the IS | Parallel to Environment setup |
| Establish quality and standards up front | Parallel to Environment setup |
| Identify data domains, any dependent processes, create a priority based data migration plan | Parallel to Environment setup |
| Understand the problem and help setup appropriate contacts | Parallel to Environment setup |
| Ensure data uniqueness during data ingestion | During Data Ingestion |
| Implement Data DevOps | During development, testing and deployment |

| Problem | Solution | Unique Work Items |
|--|--|---|
| Redundancy in data | Ensure data uniqueness | Ensure data uniqueness during data ingestion |
| Not ease to sharing data | Develop user personas, access and governance models | Develop user personas, access and governance models |
| SMART is Kiran's area and are stuck on the connectivity to Snowflake | Understand the cause, create a data access process | Understand the cause, create a data access process |
| Want to create a Target state architecture | Create Target state Architecture | Create Target/future state Architecture |
| Roadmap - what assets could be tackled first and which sequence | Create a roadmap | Create a roadmap |
| Create a focus pilot that could be of value | Create a pilot project | Create a POC/pilot project |
| Automation is expected from the OAS | Implement Data DevOps | Implement Data DevOps |
| How do I get to Snowflake? (Note: Anxious to go to Snowflake) | Future state architecture, proper user access control, permissions and a roadmap | |
| Not familiar with Optum teams | Understand the problem and help setup appropriate contacts | Understand the problem and help setup appropriate contacts |
| What does it take to stand up a Snowflake instance | Acquire appropriate approvals, secure implementation team and Set up a Snowflake instance | Acquire appropriate approvals, secure implementation team and Set up a Snowflake instance |
| Need consultation for how to get there - migrate to Snowflake | Future state architecture, proper user access control, permissions, roadmap, a data migration plan | |
| Need of what pieces to move | Identify data domains, any dependent processes and prioritize | |
| What future state architecture looks like | Create a future state architecture | |
| A roadmap of this journey | A roadmap | |
| They want to do a POC | a POC project | |
| Looking for an implementation partner | Establish implementation relationship and work with the IS | Establish implementation relationship and work with the IS |
| What would be the cost - egress and | Provide reference costs | Provide reference costs |

| | | |
|---|--|---|
| storage cost | | |
| How do we move forward; create a plan together | Create a priority based data migration plan | Identify data domains, any dependent processes, create a priority based data migration plan |
| It should be cleansed and standardized before moving to the platform | Establish quality and standards up front | Establish quality and standards up front |
| Source data flows into the data repository along with proper governance | Establish quality and standards up front + Create a priority based data migration plan | |
| User should read from this data platform | Setup access controls based on user personas | Setup access controls based on user personas |

High-level Work-plan for Execution



Roles & Responsibilities

Insurance Solution

- Providing access, current state artifacts
- Provide key stakeholders availability to participate in the meetings, interviews, workshops as needed (Tech & business)
- SME resources to provide appropriate KT as needed
- SME resources for interim review and feedback

OAS

- Plan agenda for, lead & facilitate meetings / interviews / workshops to understand current state data platforms and related business use cases
- Document key findings from each session and conduct interim review and validation w/ MRIS leadership
- Assessment and identification of gaps & opportunities in current state
- Document gaps and opportunities and conduct interim review and validation w/ MRIS leadership
- Map the findings from the current state assessment report to the **art of the possible** for the future state vision
- Document the future state architecture vision and conduct interim review and validation w/ MRIS leadership
- Define and develop the blueprint for the future state Cloud Data Strategy including the following: Migration Approach, Cloud Migration Governance Framework, Cloud Operating Model, Cloud Cost Estimation & Management, and Cloud Adoption Roadmap
- Provide input from the above for the development of the Cost Benefit Analysis by MRIS leadership

+++++
Parameters to collect to understand the MRIS Data Platform(s) current state: Infrastructure, data sources, data quality, data model(s), data integration methods, consumption patterns, data governance & security practices, languages, technologies, dependencies, workload types, frequencies, and sizes, usage patterns, and related costs

First meeting with Kiran: [Discuss IS Data Warehouse Current State-20230314_110342-Meeting Recording.mp4 \(sharepoint.com\)](#)

Discuss IS Modernization: https://uhgazure-my.sharepoint.com/:v/g/personal/john_shin_optum_com/Ed_rTjLQe51PrGVjWIR955wB8VEkUvgFmTqKmCBhmUdEvw

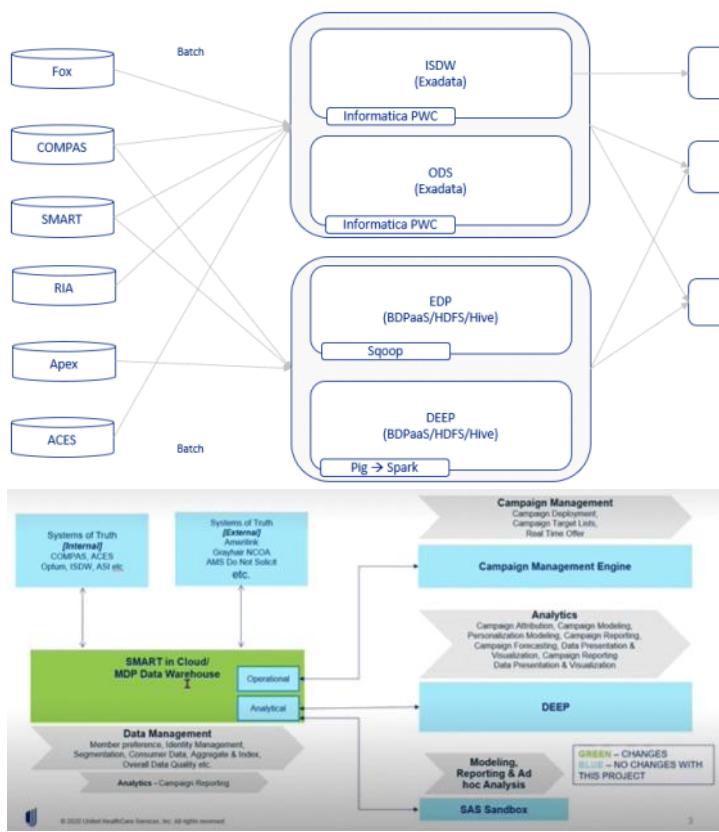
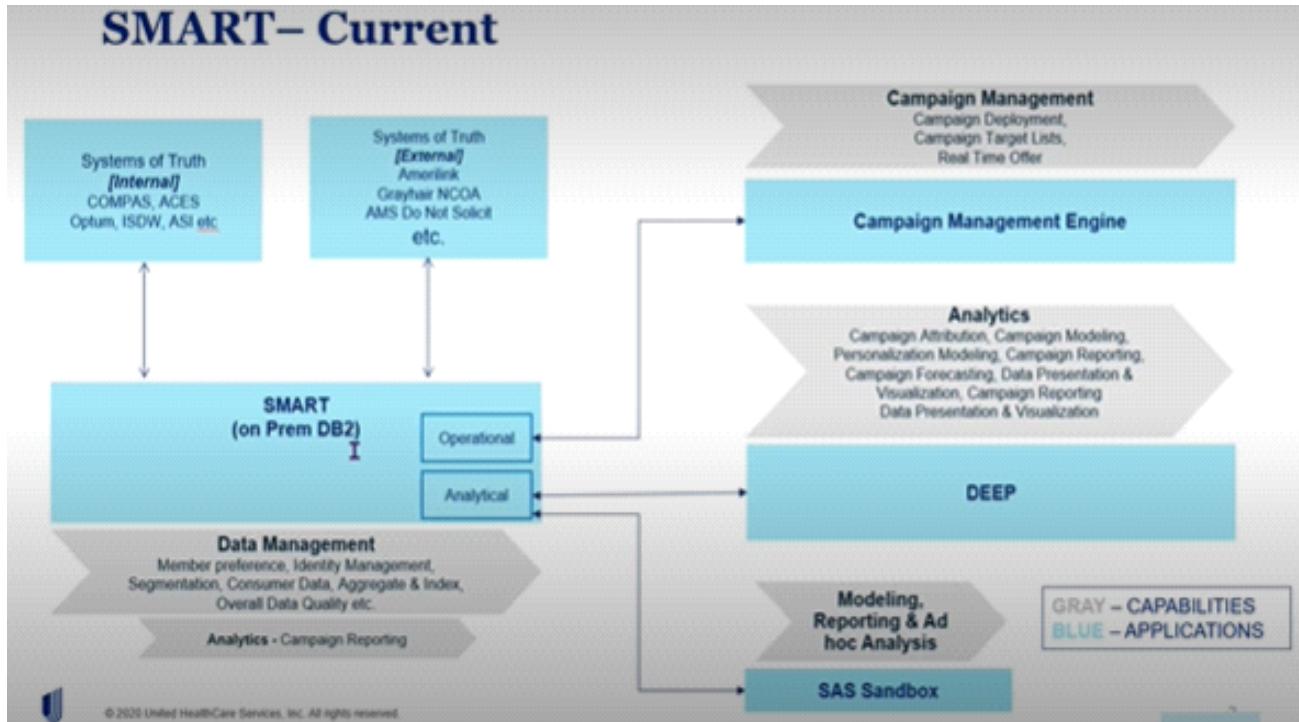
Discuss IS modernization (part 2): https://uhgazure-my.sharepoint.com/:v/g/personal/john_shin_optum_com/EYpVC3mduvpAj82hE5PWk1EBdryTSbz74QBu2_cuZWJ-RA

Meetings

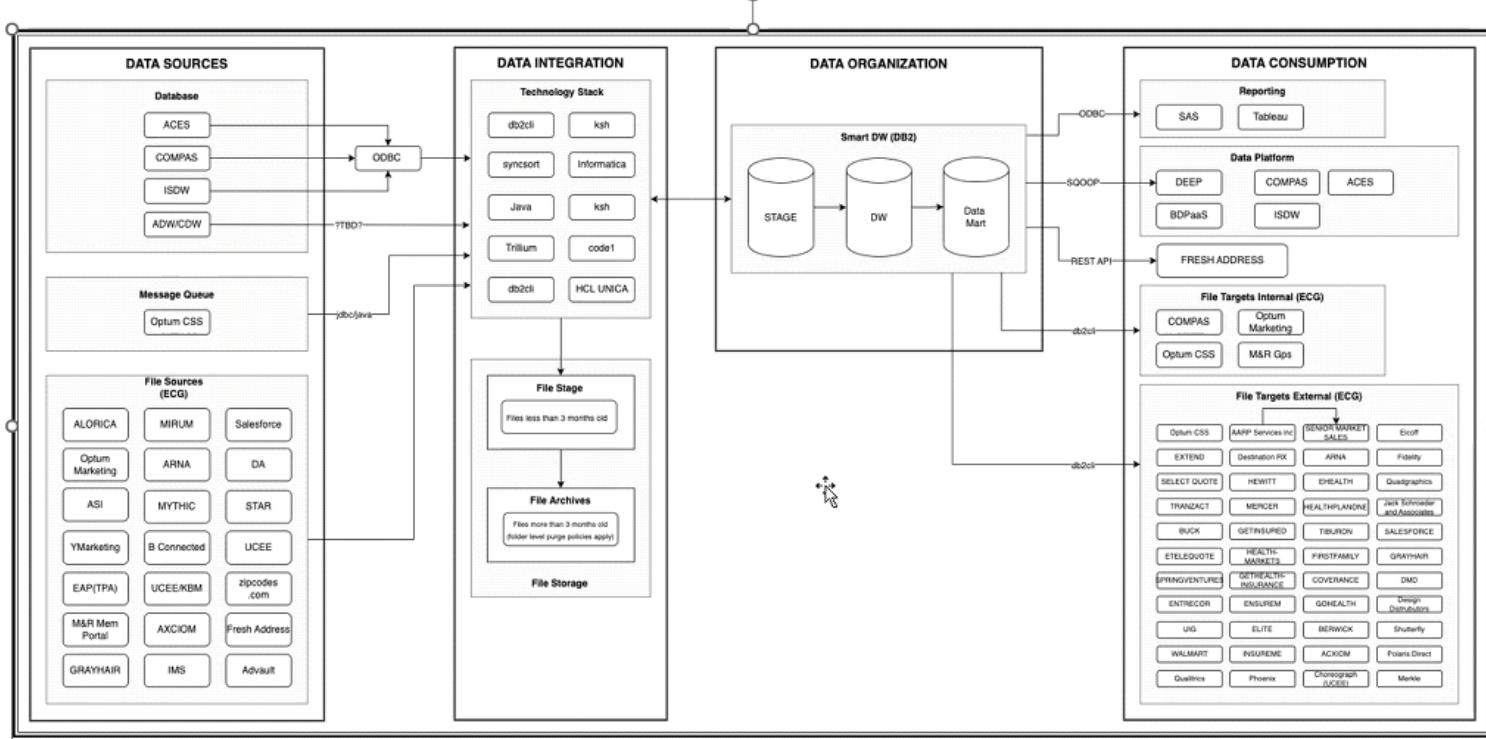
Tuesday, March 21, 2023 5:53 PM

Meeting 03/20/2023 with Kiran & team
Subject: Data modernization

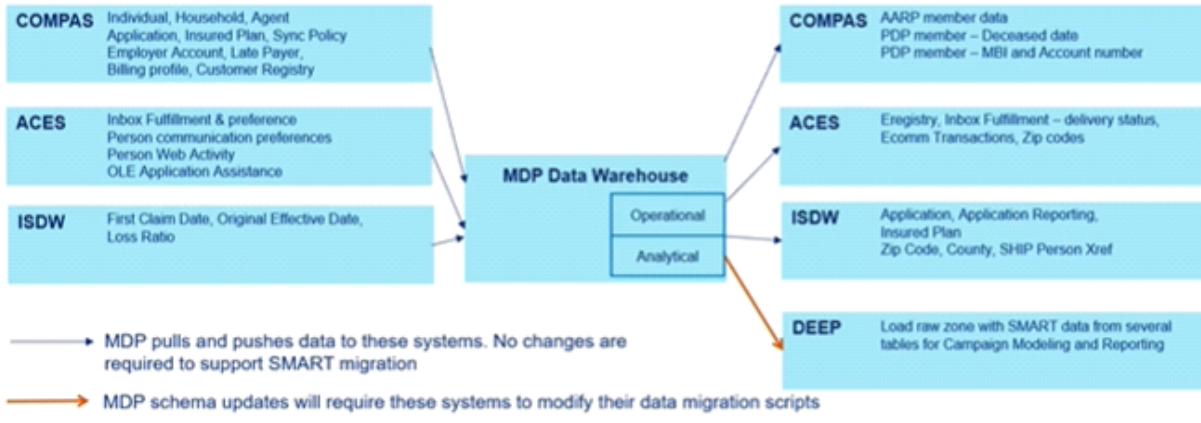
Recording Notes:



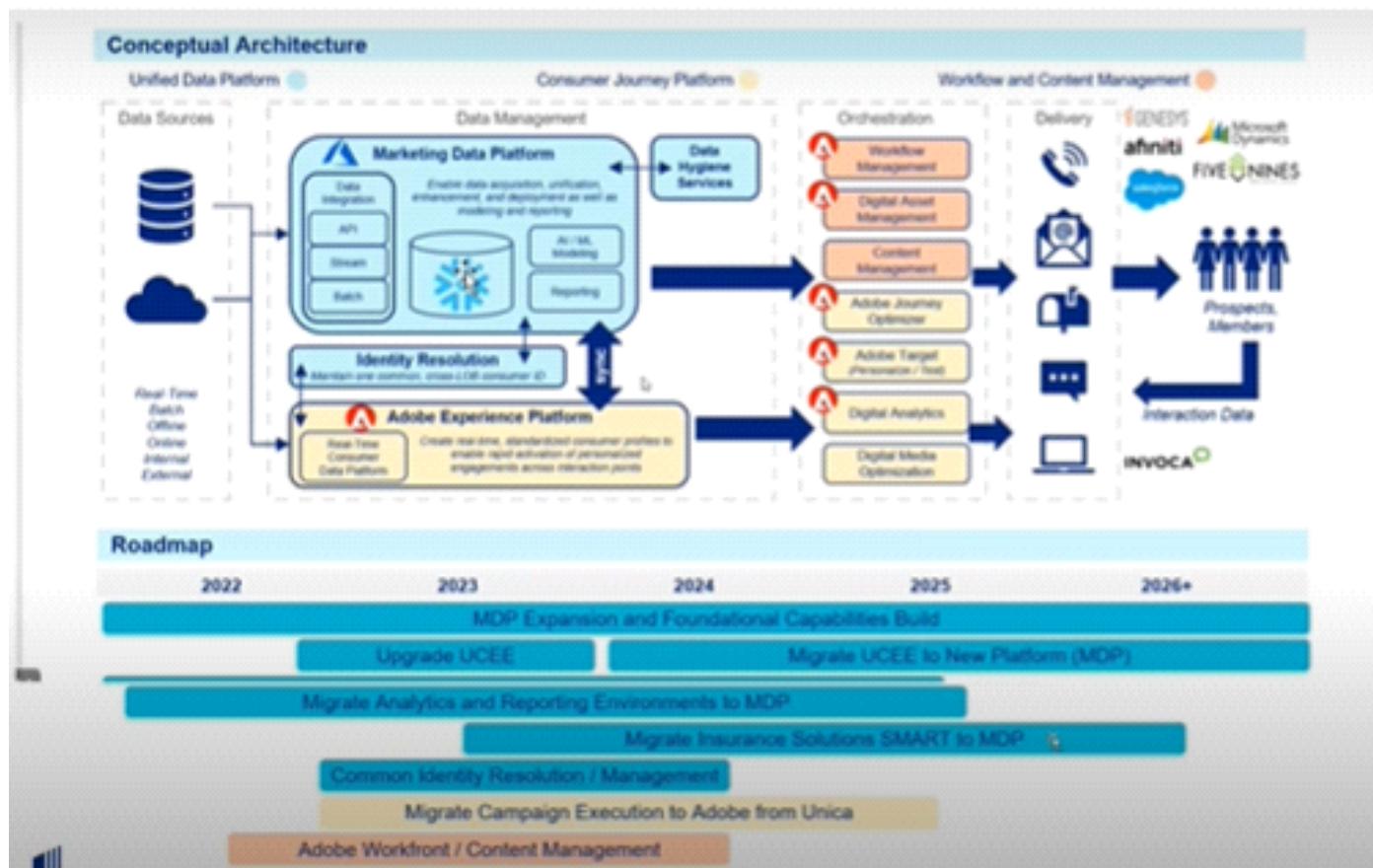
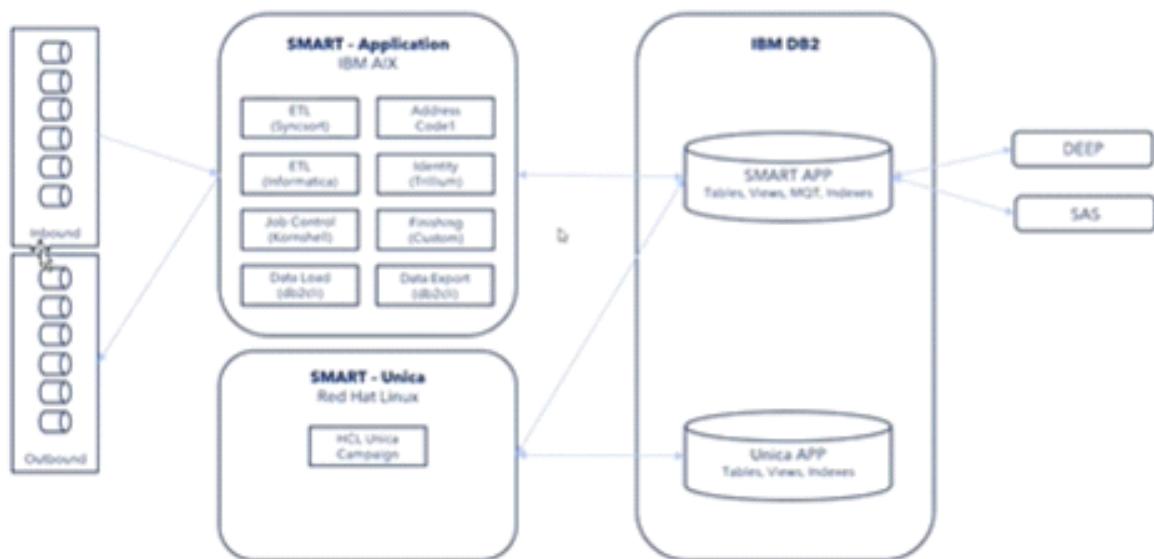
1.1 Current state architecture

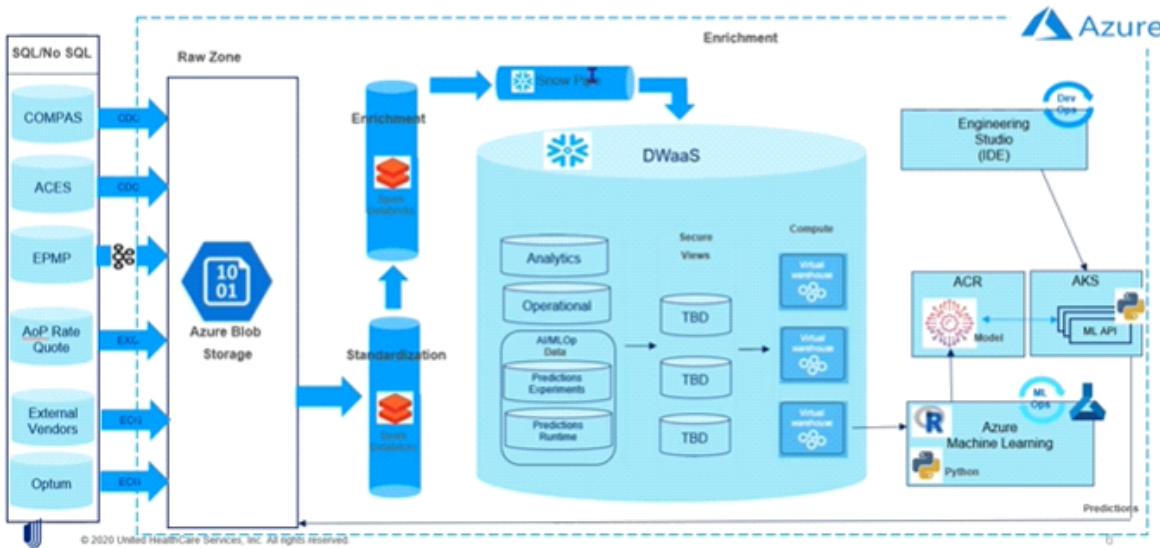


T



SMART – Landscape – Current





SMART is sitting on-prem, DB2

Everything is on-prem, DB2 - has operational data and analytical data

Created some data marts in DB2

Bring data thru ODBC connection or thru files, mainly thru external files that we connect to external sources.

They send us psychographic data, NCOAs, then we have advertising management data that send files of source codes of marketing

Overall, we have over 40 clients that send us files.

Additionally, there are internal systems which are with Oracle, Salesforce database, we bring in data from these sources.

Call center, COMPAS application for policy.

Aces is the digital system has data that is sitting in the data warehouse.

ASI is the AARP clients, we bring in the AARP membership data from ASI, via files.

Aces, COMPAS and ISDW, we have ODBC connection. So we use ETL tool to pull data.

Sometimes push, but mostly pull data. Reason being it sends out marketing material. Both insured members and prospects.

SMART data warehouse is servicing customers looking for additional services, e.g., supplemental insurance plans or indemnityplans; plan request new plans or switch plans.

Reporting tools are sitting on top such as SAS, Tableau. They bring in data to DEEP. DEEP is a Hadoop based platform.

They bring in data from SMART --> DEEP on incremental or full refresh basis in the afternoons.

The system runs almost continuously. Campaign management system, which HCL UNICA tool run on its Linux engine.

It hooks into SMART database and creates files, we don't have a real-time system to pull data; all batch based system.

It sends out these files to our external vendors ,e.g., Salesforce.

Analytics capabilities are on DEEP - DEEP is a custom built platform.

They use R to build analytics. These models were based on SAS.

DEEP is analytic data platform.

DEEP will remain. Currently we are not moving everything from DEEP to Azure. It is a separate team.

Don't know about their timeline, DEEP is on-prem.

DEEP pulls data from MDP data warehouse

DEEP will be heavily involved to verify that all changes work they can read data from all tables

Data is all produced in SMART; DEEP is just a consumer, but they also run models. DEEP is not just SMART data.

Scores are used by the campaigns.

The only thing we're looking to change is whatever is in the DB2 database

Any ETL tools to bring in data or create extracts that is the box that we want to migrate to the cloud.

Data store from SMART and related everything to move to cloud.

Test will stay as is. They send a file via FTP. They have marketing scores data.

SMART is 3 different boxes - one server is DB2 server; one server is application server; these two are on IBM AIX.

3rd server is WebSphere server, on Linux where UNICA application is running.

Database is broken into 2 different schema - one schema for consumer database; UNICA has its own schema and maintain their tables.

SMART App data is about 20 TB.

On the left are apps, old apps, like Syncsort, called in by Korn shell scripts.

Another app is Address cleansing; Address matching app and gender identification;

Trillium also runs on IBM platform.

There is another custom built app called Finishing.

Korn shell scripts create extracts and split into vendor formats.

ETL scripts pull data from SMART database and load data into staging tables.

Some scripts run that take staging data and load into prod tables.

Address cleansing and Address matching apps are crucial.

They are not moving to the cloud; an external vendor will be servicing them.

How do we move data to and from ETLs? Pushing data is easy.

Moving data from DB2 to Snowflake; eventually move ETL to Databricks.

Next phase is to replace address cleansing and Address matching apps.

We want to use common marketing system.

Replace Korn shell scripts with scheduling tools.

UNICA will be replaced with Adobe consumer experience.

UC and SMART have cross capabilities. We try to dedupe.

Migrate data, migrate processes, migrate UNICA.

Roadblocks: Using Databricks is price effective. SMART is a 24/7 system, no downtime. COMPAS is policy system.

Still legacy operational mindset - nightly processing. During daytime, call center come in with heavy user data usage.

Data marts are built incrementally.

Don't have Kafka connect yet.

REST APIs to connect to email vendor.

Planning to create APIs for COMPAS, DOC360 for policy.

We have < 10 Informatics workflows. most of workflows are in DM Express (and SyncSort) (>200, 50-50).

Help needed with:

- AIX DB2 data loading into Snowflake
- Automate script conversion
- Two-way Connectivity between AIX DB2 and Snowflake.

+++++
Cloud Data Strategy

Dinesh's meeting with Liv (03/21/2023)

Subject: Cloud Data Strategy

- Livio wants to have one voice
- Kiran's work is already partially funded, should be part of the initial project
- Liv is Looking for a proposal by Friday

Meeting with Kiran, Doug, and David (03/21/2023)

Subject: Deeper dive into perspective areas

MRIS leadership team: Kiran Padgaonkar, Douglas J Ubele, David R Searfass

Doug support a data warehouse, Operational Data Store, Enterprise Data Platform - a BDPaaS instance

Use Informatica to pull data from various sources and populate data structures, languages are Pig, Spark, Hive tables

On top we have MicroStrategy to produce reports

SAS users do a lot of Data Analytics

Business users use Tableau for reporting

Oracle Exadata, Hadoop with Sqoop and Hive and Pig, Informatica

Everything is on-prem, even RIA

Mostly internal files

Informatica have 1500 jobs and they are staying put for now, they PL/SQLs

It should be good to convert them in procedural languages

Doug's thought: go to Snowflake with Informatica

Want to get Snowflake as quickly as possible, that is where the business value is

RIA is a Data Mart

Structured data is on top right and semi-structured and unstructured data is at the bottom right

SMART is DB2 (on-prem) marketing data, the rest are Oracle databases (on the left top)

Exadata is structured data

They feel trapped - no long-term vision for structured, semi-structured, and unstructured data, e.g., voice recording - will come from Nexidia

They want to perform **sentiment analysis**

DEEP team procures data and put it in reporting, they use Tableau for reporting

First team creates structured data and the second team produce reporting

Source repository is owned and maintained by the Doug's team

Data Regions slide shows their data zones

BI and reporting are Doug and business teams

They are a data provider

Business wants to be independent

Future state should have different zones for different user groups

Want to have a common platform, capable of AI/ML capabilities

Unstructured data is audio files

There is a need for massive inserts and updates in Bulk

Data gets into Data zones via Source onboarding and Source Data Repository

Data Size: 20TB

Pain points

Redundancy in data

Not ease to sharing data

Oracle and Hadoop don't play nice together

Hadoop is terrible for updates and deletes

SMART is Kiran's area and are stuck on the connectivity to Snowflake

Future state

Ease of sharing data

Proper security controls and governance

Want to create a Target state architecture

Roadmap - what assets could be tackled first and which sequence

Create a focus pilot that could be of value

Automation (Data DevOps) is expected from the OAS

On the Doug side - he had shared

Doug - identify Big Data domains, MicroStrategy, SAS user groups do analytics

Oracle Exadata, Spark, Pig

Waiting Verma to send collateral for Kiran's area (Verma had shared his collateral)

Pain points, vision

There is redundancies in the data, not easy to share data

Hadoop is not ideal for sharing

Team does not have comprehensive access to source systems

Doug's ask:

How do I get to Snowflake? (Note: Anxious to go to Snowflake)

Not familiar with Optum teams

What does it take to stand up a Snowflake instance

Need consultation for how to get there - migrate to Snowflake

Need of what pieces to move

What future state architecture looks like

A roadmap of this journey

They want to do a POC

For a separate instance of Snowflake, what would we have to do?

Looking for an implementation partner

How do we get there

What would be the cost - egress and storage cost

How do we move forward; create a plan together

Bring source data into one platform

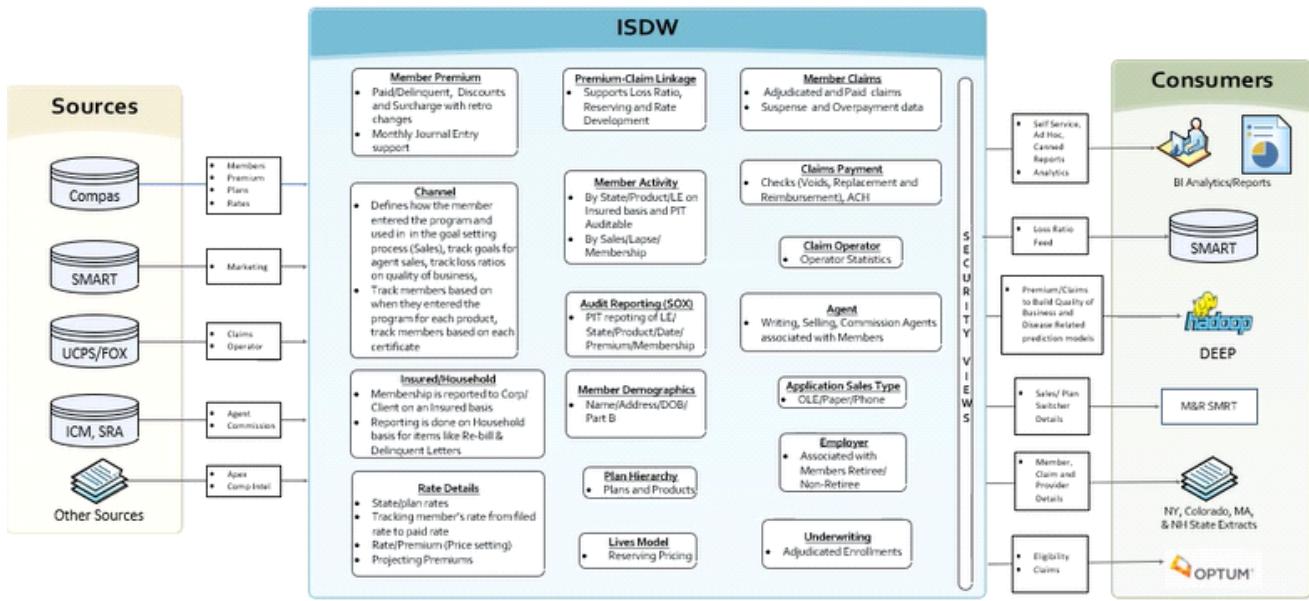
It should be cleansed and standardized before moving to the platform

Source data flows into the data repository along with proper governance

User should read from this data platform

Project Scope (High-level):

| Problem | Solution |
|---|--|
| Pain points | |
| Redundancy in data | Ensure data uniqueness |
| Not ease to sharing data | Develop user personas, access and governance models |
| SMART is Kiran's area and are stuck on the connectivity to Snowflake | Understand the cause, create a data access process |
| Future state | |
| Ease of sharing data | Develop user personas, access and governance models |
| Proper security controls and governance | Develop user personas, access and governance models |
| Want to create a Target state architecture | Create Target state Architecture |
| Roadmap - what assets could be tackled first and which sequence | Create a roadmap from current state to target state |
| Create a focus pilot that could be of value | Create a pilot project |
| Automation is expected from the OAS | Implement Data DevOps |
| How do I get to Snowflake? (Note: Anxious to go to Snowflake) | Future state architecture, proper user access control, permissions and a roadmap |
| Not familiar with Optum teams | Understand the problem and help setup appropriate contacts |
| What does it take to stand up a Snowflake instance | Acquire appropriate approvals, secure implementation team and Set up a Snowflake instance |
| Need consultation for how to get there - migrate to Snowflake | Future state architecture, proper user access control, permissions, roadmap, a data migration plan |
| Need of what pieces to move | Identify data domains, any dependant processes and prioritize |
| What future state architecture looks like | Create a future state architecture |
| A roadmap of this journey | A roadmap |
| They want to do a POC | a POC project |
| For a separate instance of Snowflake, what would we have to do? | Acquire appropriate approvals, secure implementation team and Set up a Snowflake instance |
| Looking for an implementation partner | Establish implementation relationship and work with the client |
| How do we get there | A roadmap |
| What would be the cost - egress and storage cost | Provide reference costs |
| How do we move forward; create a plan together | Create a priority based data migration plan |
| Bring source data into one platform | Create a priority based data migration plan |
| It should be cleansed and standardized before moving to the platform | Establish quality and standards up front |
| Source data flows into the data repository along with proper governance | Establish quality and standards up front + Create a priority based data migration plan |
| User should read from this data platform | Setup access controls based on user personas |



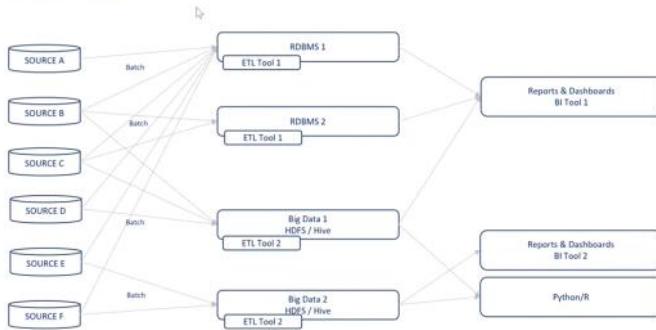
Recording Notes:

RDBMS 1 (Oracle Exadata) - ISDW - Doug's area

RDBMS 2 (ODS, operational) - data store

To big data platforms hosted in BDPaaS, 1 being EDP and the 2nd is DEEP

Current State



Are there overlapping use cases?

Is there only a detail level difference, e.g., one detail and the other aggregation?

2 big data tenants hosted in BDPaaS

1 being EDP and the 2nd is DEEP

Business created an instance of their own DEEP.

They built out relational data concepts in there using things like Hive, HQL

It has been brought under Doug's group

We would like to figure out if we migrate these platforms to Snowflake;

Can we re-evaluate the sources of this data that we have and the pipelines that we have, consolidate them

There is no value in taking data from one place and sending it to different platforms

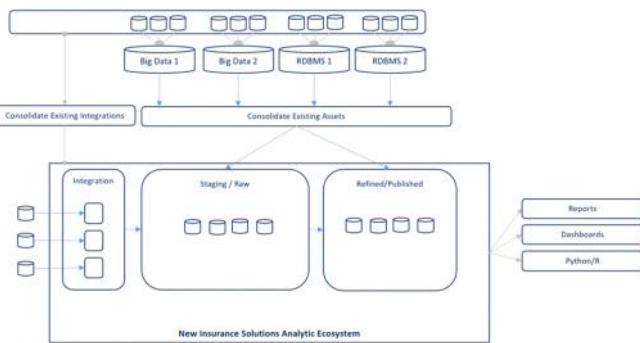
For reporting, we have MicroStrategy, big presence in Tableau. those primary reporting tools

Take a look at the source of these data, and figure out how can we consolidate these assets represented in different technologies

We want to consolidate different pipeline and integrations, so that we have one source for that data and one pipeline for that data;

We might have to serve the Raw data for raw level analytics, create and publish our dimensional data, reconnect reports and dashboards that we have.

Future State – Transition



In Doug's group, there is a big Informatica presence, that is on-prem
Informatica uses CDC, PowerExchange on the source system. How do we leverage that technology in new world?
Are more orgs going to streaming architecture? Is it not necessary?
We have heard about Databricks, Snowpipe. We're not sure how they fit in? How will on-prem Informatica integrate with Azure hosted solution?
For considerations, we're looking to understand, how do we map our existing tools that we have today to the tools that are provided by the DWaaS.

Considerations

- What tooling is supported/offered by Optum DWaaS?
- Are more organizations migrating to streaming architecture?
- Are organizations staying with traditional ETL (Informatica, etc) for their new pipelines?
- Can we integrate on-prem Informatica with Azure hosted Snowflake?
- How are teams approaching their migrations? Lift and shift? Waved migrations?
- How do we estimate costs for the new platform?
- Is there a relation between costs on traditional RDBMS and Snowflake
- What blockers should we be on the lookout for

Kiran - this stuff does not include what Kiran had talked about AIX, on-prem DB2.

It does not include DB2

We are keeping Kiran's effort separate

ISDW is a traditional data warehouse

We introduced the concept of Operational data, which is Raw source data

There is some overlap between RDBMS and big data (Hadoop)

There is no overlap between big data instances

We know our users, there is some concept of personas and they go thru Secure

So there is some overlap between RDBMS' and Hadoop platform

We know our users, they go thru Secure, there is approval process

RDBMS' are heavily regulated data warehouses. there are some outbound extracts

In big data, users experiment, there is overlap in user groups between RDBMS' and Hadoop instances

In the future, we may have to think whether we want to stay with MicroStrategy or think of alternatives?

PowerBI has been discussed, but not currently in use

What are the migration steps

What is our strategic direction and what does the end product look like?

What is the business driver of this migration -

Kiran - we have been discussing for a couple of years. M&R have already created Marketing Data Platform.

MA side has some special contract things and they go after the same customers

Marketing goes to the same org hierarchy

Doug - it is about shareability, more unified data sharing in Raw, bronze, silver gold, etc.

It will ease with development, resource sharing, able to stand up that can house semi-structured, unstructured

Get to a common platform, maybe even leverage resources. Stand up data platform that can handle Structured, semi-structured and unstructured data

Snowflake looks like a great candidate

Kiran - budget is funded from ARB, we already have some money in the budget. We have some flexibility

Cost is another big driver

We want to understand the costing impact; put a pricing picture together will be a big help

Success criteria - Doug - struggling with our head getting around Optum - it is so big.

To know where we're going; what our plan look like. Is this we should be going?

what are the milestone, how do we create the end state vision; what is that vision

David - in Q2, we must have the end state architecture and transition architecture and then pilot program

Kiran - I would love to, we're stuck in the POC mode, in connectivity with AIX to Snowflake, plan is essential

They really want a direct connection from AIX to Snowflake

This is a phased approach

We would like to get away from AIX, maybe to Databricks and Azure native

Kiran's AARP data must be separable?

They should consider a separate Snowflake instance, to get around DWaaS restrictions

Compute is where the cost is, the price model is by the second use; storage cost is almost nothing

David - are there any metrics available?

Kiran's has old tech stack. e.g., DB2, Corn shells, Syncsort; Convert command line to GUI

DS - can work with Doug and Kiran to create the current state architecture

But, would need help to create the future state architecture. Together we build the transition state and architecture

| A | B | C |
|----|---|-----------|
| 1 | Usecases (test cases) for Cloud migration | |
| 2 | | |
| 3 | Source | Target |
| 4 | DB2 | Snowflake |
| 5 | AIX | Snowflake |
| 6 | RHEL (SnowSQL) | Snowflake |
| 7 | Microstrategy | Snowflake |
| 8 | SAS | Snowflake |
| 9 | Tableau | Snowflake |
| 10 | Oracle/Exadata | Snowflake |
| 11 | Oracle/Exadata | SQL |
| 12 | Unica (RHEL) | Snowflake |
| 13 | DEEP (Hadoop/Hive) | Snowflake |
| | | SQL |

++++++

Meeting 03/23/2023 with IS Doug, Kiran, and David

Subject: Align our understanding and expectations

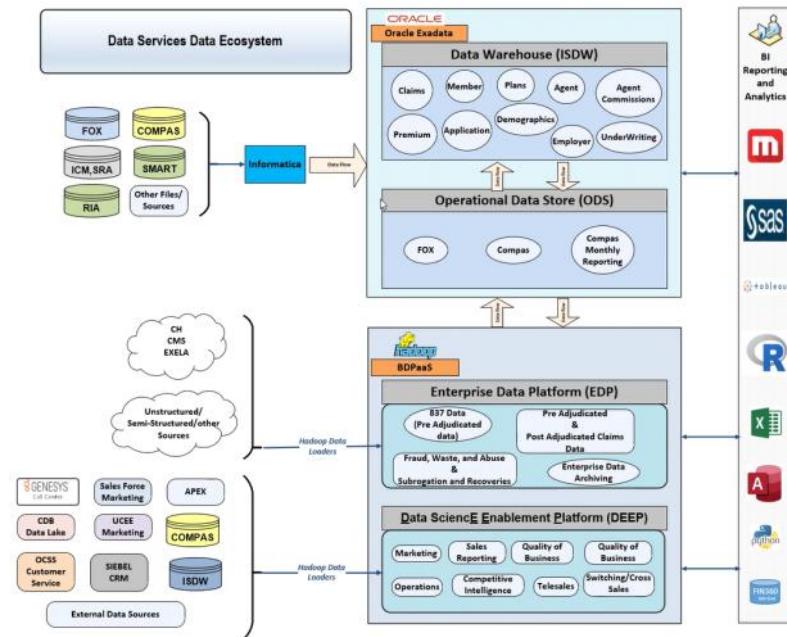
Doug

- Informatica is to be called
- Where ETL occur now? They are familiar with Informatica
- Business benefits of going to the cloud
- Future pain points and value add
- Looking for implementation partner

Kiran

- separability (severability) of Snowflake instance
- consumer (business users) training in migration, adoption step

Meeting recording:



At high-level, we support a data warehouse and operational data store and an enterprise data platform, which is in BDPaaS; we use Informatics to pull data from various sources and we populate these structures

We are ingesting data primarily thru Sqoop, internally using Pig, Spark, Hive tables for deliverables - same concept as relational database on the side we have all our consumers, this team is also under Doug's group.

We have MicroStrategy report that sit on top; there SAS users for analytics, Tableau for reports

Oracle Exadata, Informatica, Hadoop platform, primarily with Sqoop, Pig, Hive tables, and Spark

Pain points - redundancy

Kiran - top left is all Oracle and bottom right is Hadoop. SMART is a marketing database

Everything is on-prem, even RIA is on-prem

Mostly all internal files

Fox - claims, COMPAS - policy admin., SMART - marketing, ICM and SRA are agent based systems, RIA is a data mart that is hanging off Billing and Oracle revenue management, billing installation;

Informatica mapping jobs are about 1500. For now, we stay with Informatica. Lot of Informatica is PL/SQL based, have Informatica wrappers around them; It might be good to push into procedural language, ADF, and Snowflake procs. My thought is to go with Snowflake, Informatica and then convert to procedural language.

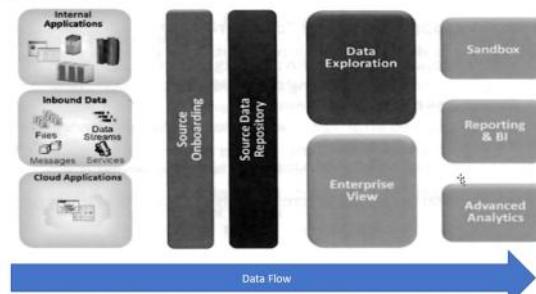
I want to go to Snowflake asap, I see business value there

pain-points - redundancy is there; not ease of sharing data, especially with Hadoop and Oracle - they do play nice together; we will need security

Exadata is structured data, in the upper right corner, bottom right is mostly unstructured data. we don't have a solution that stores structured, semi-structured and unstructured voice recordings for sentiment analysis) data. The voice data comes from Nexidia. That is a big value-add.

DEEP team procure data and put into some sort of structure; Second part is thru reporting - they use Tableau. First team is in Data Services. All teams be able to work in the same, one common, platform, using same languages and able to do the resource shifting and sharing.

Logical Architecture
Data Regions



This shows regions that different folks work in; 1 will be onboarding source data; 2. Source Data Repository (IT only thing); Enterprise View will be the Data warehouse, which will be Doug's zone; then there will be Reporting and BI team (Doug's team and business team), it'll be their region, where they do what needs to be done, they want to be independent; Advanced Analytics is managed by Brian Smith's team; it is the SAS usage, so it'll be another zone specifically for him.

We are a data provider. Different teams be able to use a common platform and be able to do all those capabilities described in the Data Regions diagram. This platform be able to support ML/AI capabilities.

Source repository will be one-to-one sources.

One big opportunity is for users to have access to data in the source systems. It is a real challenge for them - a painful process - to have access to Raw data

There needs to be some governance; mostly folks should read from Enterprise View and the risk would have to be accepted by the business. Data could be cleansed and standardized. They want to understand the cost perspective a little better.

Doug - how do we move forward? Want to see a plan that we work on together; create a roadmap of how we get to Snowflake;

+++++

Meeting 03/31/2023 with MRIS, Durga and Anantha

Subject: Review work plan

Durga and Anantha are quite concerned about the timeline being too aggressive

What tools and technologies?

DU - migrate data and tools?

Map/diagram of the journey, with milestones

milestones for consumption and how different consumers will be able to use it?

what if they don't have those artifacts

A roadmap to future state

Create a template of what we need from them

Plan for risks, update assumptions

Create a high level agenda of what items we need from them for each phase

Miscellaneous

Wednesday, March 15, 2023 10:08 PM

I'm wondering if we should think more broadly with our Snowflake intentions. Could you propose an effort to evaluate the scope, time, cost, CBA and phasing of a Snowflake migration and target architecture. We could include a migration of some asset in the process. Contact points would be Dave Searfass, Doug Ubele, and Kiran Padgaonkar. There are other data stakeholders that would be part of the conversation. Dave can pull them in accordingly.

I feel this may be the best way to get traction on this type of effort.

Our understanding

Insurance Solutions would like to:

- Modernize the data platform
- Adopt Snowflake as the Data Platform on a broader scope
- Request for an effort proposal, which will include scope evaluation, time, cost, CBA
- Create a POC, which will include some assets migration to the Snowflake
- Adopt a phased Snowflake migration approach

Our Approach

- OAS will conduct current state assessment
- Provide a high level time and cost estimates based on prior projects of similar scope, size and complexity
- Provide a CBA report based on industry, organizational and personal experiences
- OAS will adopt a phased approach
- Closely work with SMEs and Key stakeholders

Assessment Approach

- Identify candidate data platform(s)
- Identify SMEs
- Identify Key stakeholders
- Collect existing collateral
- Setup meetings and deliver them in advance
- Conduct interview with the SME to understand the nuances
- Conduct interview with the Key stakeholders to understand their vision of the future state
- Collect all data points
- Sort and sift findings
- Conduct analysis
- Create a findings report
- Review it with the SMEs and make necessary tweaks
- Present and **deliver the Current state report**

Development approach

- After the current state is agreed upon
 - OAS will adopt a phased approach - platform by platform (domain by domain)
 - OAS will use a framework first approach to maximize reusability
 - Agile development methodology will be used for quick feedback and course correction
- At the completion of each phase, the outcome will be reviewed and accepted by the client

Prerequisites

- All requested collateral is made available at the outset of this endeavor
- SMEs and Key stakeholders are available for meetings and interviews
- Any outstanding questions are responded to within 48 hours
- Access to infrastructure and operations is provided as needed
- Realistic and sanitized data is provided

Pain points

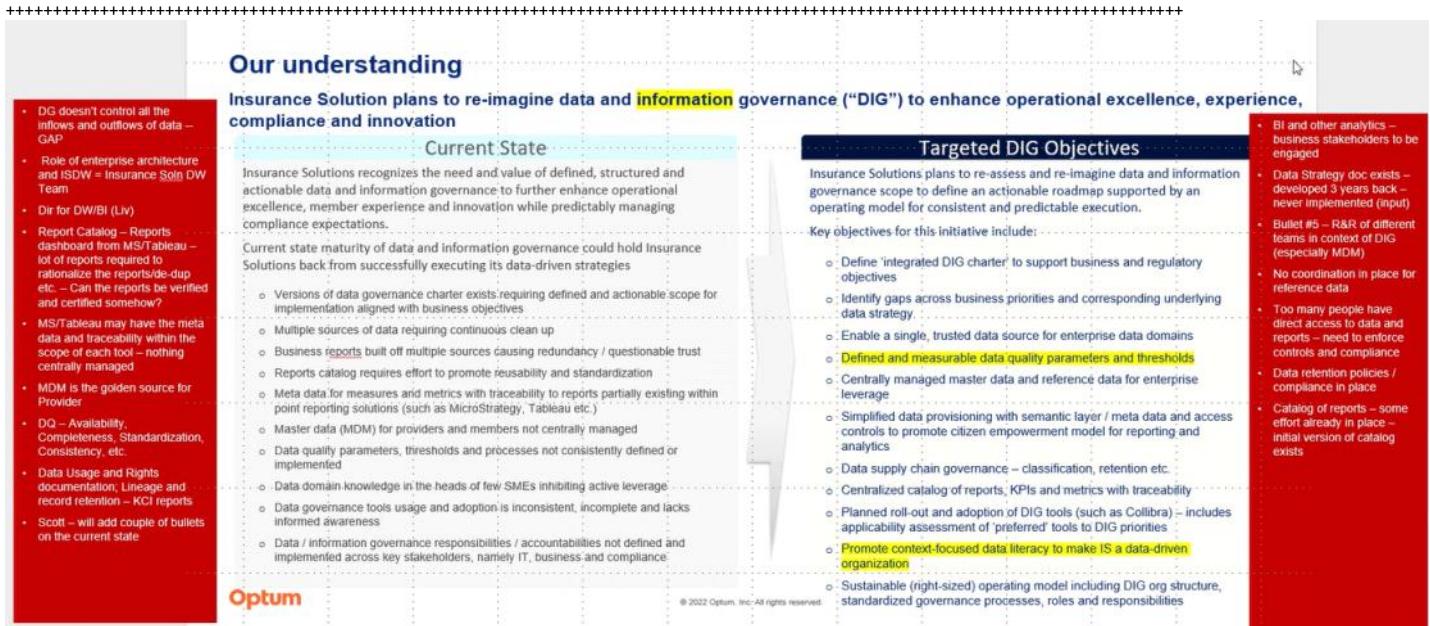
- Redundancy in data
- Not easy to share data
- SMART is Kiran's area and are stuck on the connectivity to Snowflake

Future state

- Ease of sharing data
- Proper security controls and governance
- Want to create a Target state architecture
- Roadmap - what assets could be tackled first and which sequence
- Create a focus pilot that could be of value
- Automation is expected from the OAS

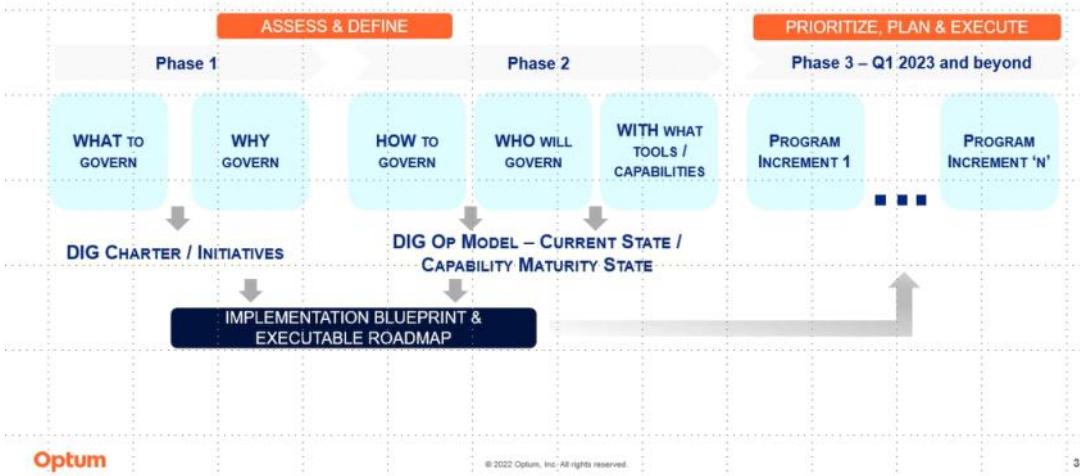
How do I get to Snowflake? (Note: Anxious to go to Snowflake)

- Not familiar with Optum teams
- What does it take to stand up a Snowflake instance
- Need consultation for how to get there - migrate to Snowflake
- Need of what pieces to move
- What future state architecture looks like
- A roadmap of this journey
- They want to do a POC
- For a separate instance of Snowflake, what would we have to do?
- Looking for an implementation partner
- How do we get there
- What would be the cost - egress and storage cost
- How do we move forward; create a plan together
- Bring source data into one platform
- It should be cleansed and standardized before moving to the platform
- Source data flows into the data repository along with proper governance
- User should read from this data platform

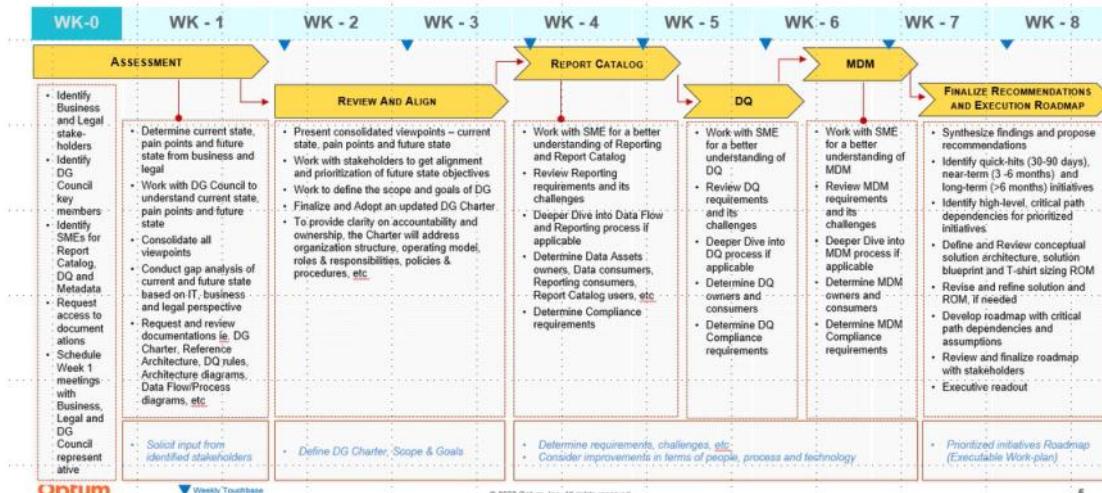


Proposed approach

We propose a simple, phased framework to define the journey roadmap with prioritized and actionable outcomes that can be planned to deliver predictable outcomes in an agile manner



Assess and Define Phase



Assessment Approach

- Identify platform(s), dependencies, and toolset

- Identify Key stakeholders and decision makers
- Collect existing collateral
- Document pain points, vision of the future state
- Create a findings report
- Receive **IS approval sign-off**
- CBA report

Environment Setup

- Setup a Snowflake instance
- Setup up data pipeline (ADF or Kafka)
- Setup dev, test, and run environments
- Create governance model
- Setup monitoring, notifications and logging

Parallel to Environment Setup

- A roadmap
- Create a priority based data migration plan
- Create Target/future state Architecture
- Develop user personas access and governance models
- Establish quality and standards

Pilot program

- Identify the candidate data domain/data source (*Note: it should be small and simple*)
- Identify test cases
- Test connectivity and user permissions
- Collect and analyze metrics
- Produce the outcome report
- Deliver results to the IS

Development approach

- A phased approach - platform by platform
- Reusable framework to maximize reusability
- Agile (SCRUM) development methodology
- Demo the outcome
- Implement Data DevOps

Prerequisites

- All requested collateral is made available at the outset of this endeavor
- SMEs and Key stakeholders are available for meetings and interviews
- Any outstanding questions are responded to within 48 hours
- Access to infrastructure and operations is provided as needed
- Realistic and sanitized data is provided

Data Governance Management

- Data Catalog and Meta Data Management : Data lineage, traceability, source of truths for each domain and scope of metadata management
- Master / reference data management: Master data and reference data acquisition, curation and usage
- Data Quality Management: Data quality parameters and thresholds scope and performance for key entities and attributes
- Data Operations Management: process, controls, backlog for managing data supply chain, acquisition, classification, storage, retention and disposal
- Data Access and Security: Data security and data privacy scope, controls and architecture to support the same
- Information and Insights management (Metrics, Measures, Reports and Dashboards): Governance scope and processes to manage information integrity, usability and reliability – including reusability, redundancy, multiple versions of truth, certification etc.

Data Governance Operating Model

- Operating model in context of roles, governance processes and tools

Work Plan

Ask for executing on the Assessment

- Detailing out the workplan level slide
- Roles & Responsibilities slide

• Current State Assessment Report documenting the following aspects of the current data landscape:

- Key details of existing data platforms including data sources, data quality, data model, data integration methods, data governance & security practices, and technologies & related costs
- Current business use cases being addressed
- Operational model and related costs
- Potential gaps / opportunities existing today for business and operations

Work breakdown

- Existing data platforms
- data sources
- data quality
- data model
- data integration methods
- data governance
 - Management
 - Operating model
- security practices
- technologies
- costs

Assessment Approach

- Assess each data platform, e.g., DEEP, etc.
 - Identify dependencies
- Data Quality
- Data Models
- Current technologies and toolset
- Identify SMEs (technical and business)
- Collect existing collateral

- o Request current data platform documentation
- Setup meetings with proper agenda and set them in advance
 - o Conduct interviews with the SMEs to understand:
 - data quality, data models, data integration methods, data governance management & operating model, security practices, technologies, costs, nuances, pain points, and future state vision
- Collect all data points from the documentation and SME interviews
- Conduct analysis
- Create a findings report
- Review it internally
- Review it with the client SMEs
- Make necessary updates
- Present and deliver the Current state report to IS
- Receive IS approval sign-off

OAS Team roles

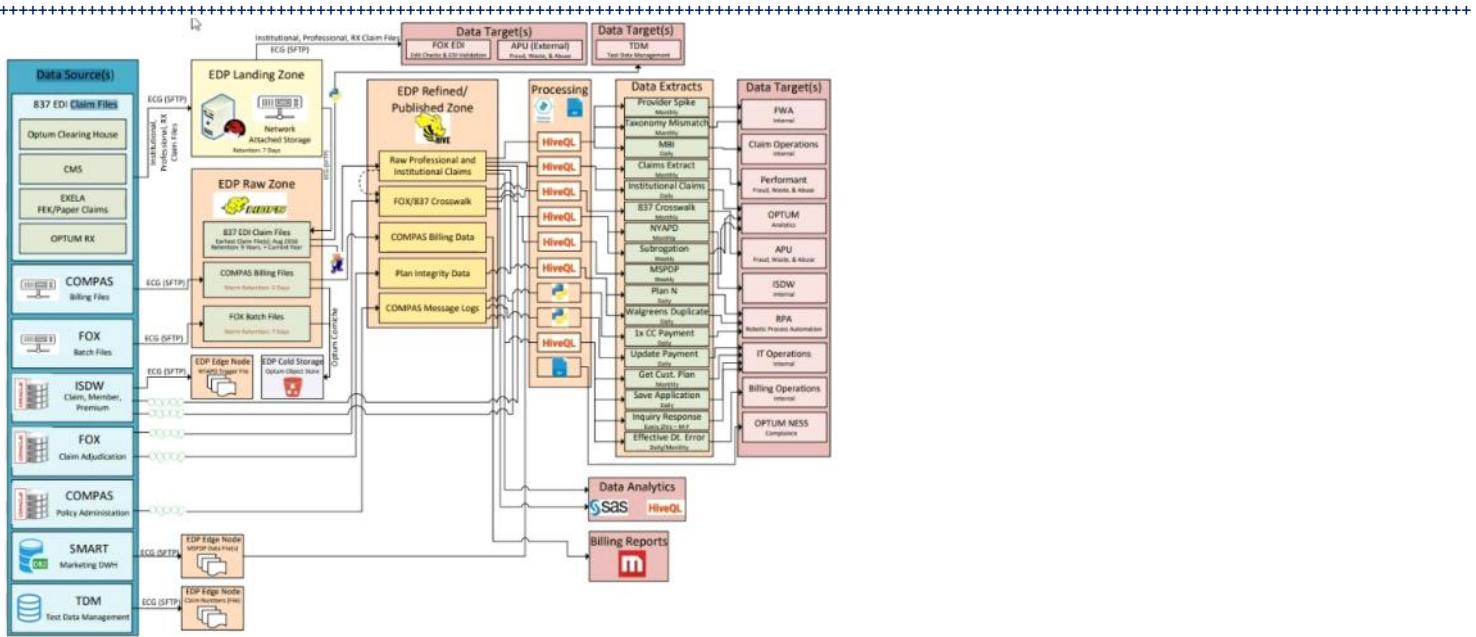
- MRIS Partner/SME
- Delivery Architect
- Solution Architect SME
- Development Teams
 - o Business Analyst
 - o Architect
 - o Cloud Data Architect

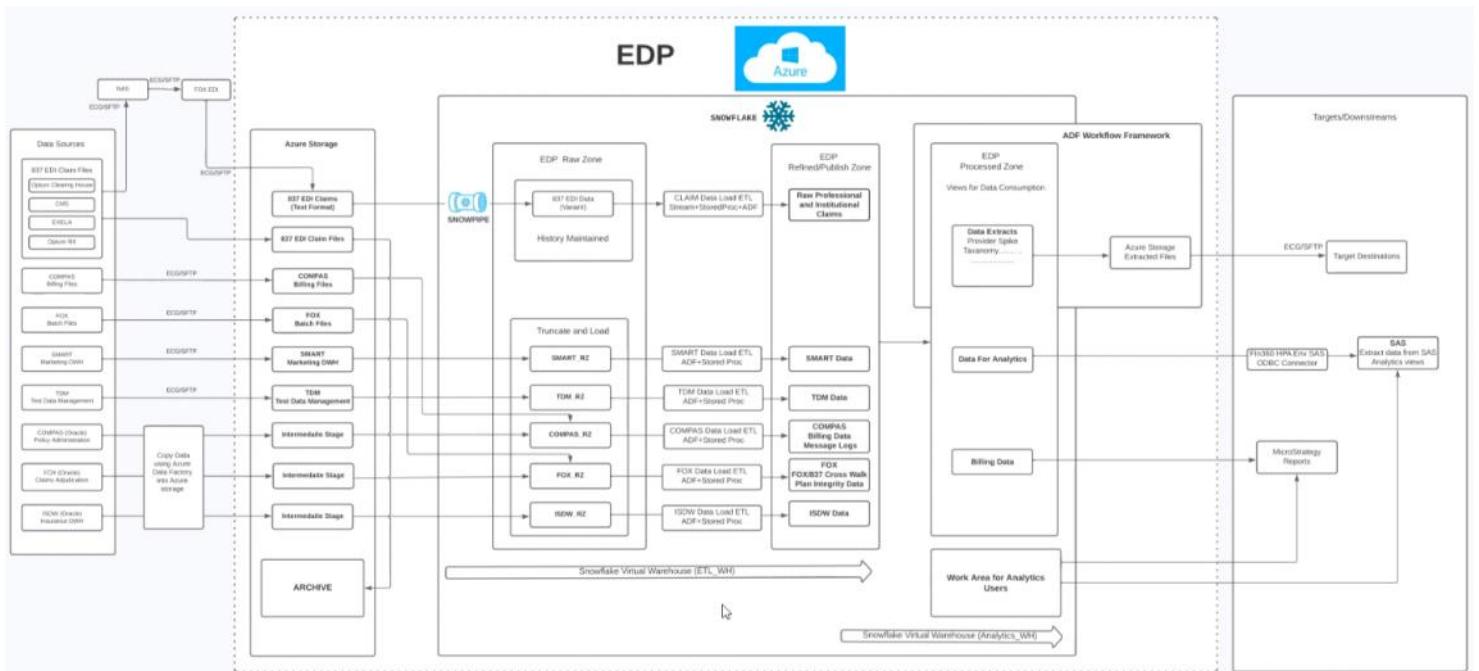
Insurance Solutions Team

- Business SME
- Technology SME

Assumptions

- All requested collateral will be made available at the outset of this endeavor
- SMEs and Key stakeholders will be available for meetings and interviews
- Any outstanding questions will be answered within 48 hours
- Access to infrastructure and operations is available as needed





A comparison of the governance capabilities of Snowflake, Azure Synapse, and Databricks:

| Feature | Snowflake | Azure Synapse | Databricks |
|---------------------|---|---|---|
| Data security | Supports encryption at rest and in transit, role-based access control (RBAC), and audit logging | Supports encryption at rest and in transit, RBAC, audit logging, and data masking | Supports encryption at rest and in transit, RBAC, audit logging, and data lineage |
| Data governance | Provides a centralized catalog for managing data definitions and lineage, as well as a suite of tools for data quality and compliance | Provides a centralized catalog for managing data definitions and lineage, as well as a suite of tools for data quality, compliance, and lineage | Provides a centralized catalog for managing data definitions and lineage, as well as a suite of tools for data quality and compliance |
| Data access control | Supports RBAC and fine-grained access control (FGAC) | Supports RBAC and FGAC | Supports RBAC and FGAC |
| Audit logging | Provides comprehensive audit logging for all data activity | Provides comprehensive audit logging for all data activity | Provides comprehensive audit logging for all data activity |
| Data lineage | Provides lineage tracking for all data flows | Provides lineage tracking for all data flows | Provides lineage tracking for all data flows |

| | | | |
|--------------|---|---|---|
| Data quality | Provides a suite of tools for data quality checking and remediation | Provides a suite of tools for data quality checking and remediation | Provides a suite of tools for data quality checking and remediation |
| Compliance | Supports a wide range of compliance standards, including HIPAA, GDPR, and PCI DSS | Supports a wide range of compliance standards, including HIPAA, GDPR, and PCI DSS | Supports a wide range of compliance standards, including HIPAA, GDPR, and PCI DSS |

Overall, Snowflake, Azure Synapse, and Databricks all offer comprehensive governance capabilities. However, there are some key differences between the three platforms. Snowflake is a pure cloud data warehouse, while Azure Synapse and Databricks are more hybrid platforms that can also be used to process data on-premises. This means that **Snowflake is generally more scalable** and cost-effective for large-scale data warehousing workloads. However, Azure Synapse and Databricks offer more flexibility for hybrid workloads.

In terms of governance capabilities, **Snowflake is a good choice for organizations that need a highly secure and compliant data warehouse**. It offers a wide range of security features, as well as comprehensive audit logging and lineage tracking. **Azure Synapse is a good choice for organizations that need a more flexible governance solution**. It offers a wide range of security features, as well as lineage tracking and data quality tools. Databricks is a good choice for organizations that need a governance solution that integrates with their existing data science and machine learning pipelines. It offers lineage tracking and data quality tools, as well as integration with popular machine learning frameworks.

Ultimately, the best governance platform for your organization will depend on your specific needs and requirements.

Advanced Data Analytics tools:

- Microsoft Power BI:** Power BI is a powerful business intelligence (BI) and analytics platform that can be used to analyze data from a variety of sources, create interactive dashboards and reports, and share insights with others. It is a good choice for businesses of all sizes, and it offers a variety of features that make it well-suited for advanced analytics tasks, such as machine learning and predictive modeling.
- Tableau:** Tableau is another popular BI and analytics platform that is known for its intuitive drag-and-drop interface and powerful data visualization capabilities. It is a good choice for businesses that want to create interactive dashboards and reports that are easy to understand and share. Tableau also offers a variety of features for advanced analytics, such as machine learning and predictive modeling.
- Qlik Sense:** Qlik Sense is a cloud-based BI and analytics platform that is known for its speed and scalability. It is a good choice for businesses that need to analyze large datasets quickly and easily. Qlik Sense also offers a variety of features for advanced analytics, such as machine learning and predictive modeling.
- Looker:** Looker is a BI and analytics platform that is designed for data analysts and data scientists. It offers a variety of features for advanced analytics, such as machine learning, predictive modeling, and natural language processing. Looker is a good choice for businesses that want to empower their data analysts and data scientists to create advanced analytics solutions.
- RapidMiner:** RapidMiner is an open source platform for machine learning and predictive analytics. It is a good choice for businesses that want to build and deploy machine learning models quickly and easily. RapidMiner offers a variety of features for advanced analytics, such as supervised and unsupervised learning, text mining, and image analysis.
- KNIME:** KNIME is another open source platform for machine learning and predictive analytics. It is a good choice for businesses that want to build and deploy machine learning models visually. KNIME offers a variety of features for advanced analytics, such as supervised and unsupervised learning, text mining, and image analysis.

These are just a few of the many great advanced analytics tools available in 2023. The best tool for you will depend on your specific needs and requirements.

Popular and widely used data science tools:

- Python:** Python is the most popular programming language for data science, and for good reason. It is easy to learn and use, yet powerful enough to handle complex data analysis tasks. Python has a wide range of libraries and frameworks for data science, including NumPy, SciPy, Pandas, and Matplotlib.
- R:** R is another popular programming language for data science. It is similar to Python in terms of its capabilities, but it has a different syntax. R is often used for statistical analysis and machine learning tasks.
- Jupyter Notebook:** Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Jupyter Notebooks are a great way to collaborate on data science projects and to document your work.
- Tableau:** Tableau is a data visualization tool that allows you to create interactive dashboards and reports. Tableau is easy to use and can be used to visualize data from a variety of sources.
- Power BI:** Power BI is another popular data visualization tool. It is more powerful than Tableau, but it is also more complex to learn. Power BI is a good choice for data scientists who need to create complex visualizations or who need to integrate data from multiple sources.
- Scikit-learn:** Scikit-learn is a free and open-source machine learning library for Python. It is one of the most popular machine learning libraries in the world, and it is used by data scientists for a variety of tasks, including classification, regression, clustering, and dimensionality reduction.
- TensorFlow:** TensorFlow is a free and open-source software library for machine intelligence. It is used by data scientists for a variety of tasks, including machine learning, natural language processing, and computer vision. TensorFlow is a powerful library, but it can be difficult to learn.
- RapidMiner:** RapidMiner is a commercial data science platform that offers a wide range of features for data preparation, machine learning, and predictive analytics. RapidMiner is a good choice for data scientists who need a comprehensive platform for data science projects.

DataRobot: DataRobot is a cloud-based machine learning platform that automates the machine learning process. DataRobot is a good choice for data scientists who want to build machine learning models without having to write code.

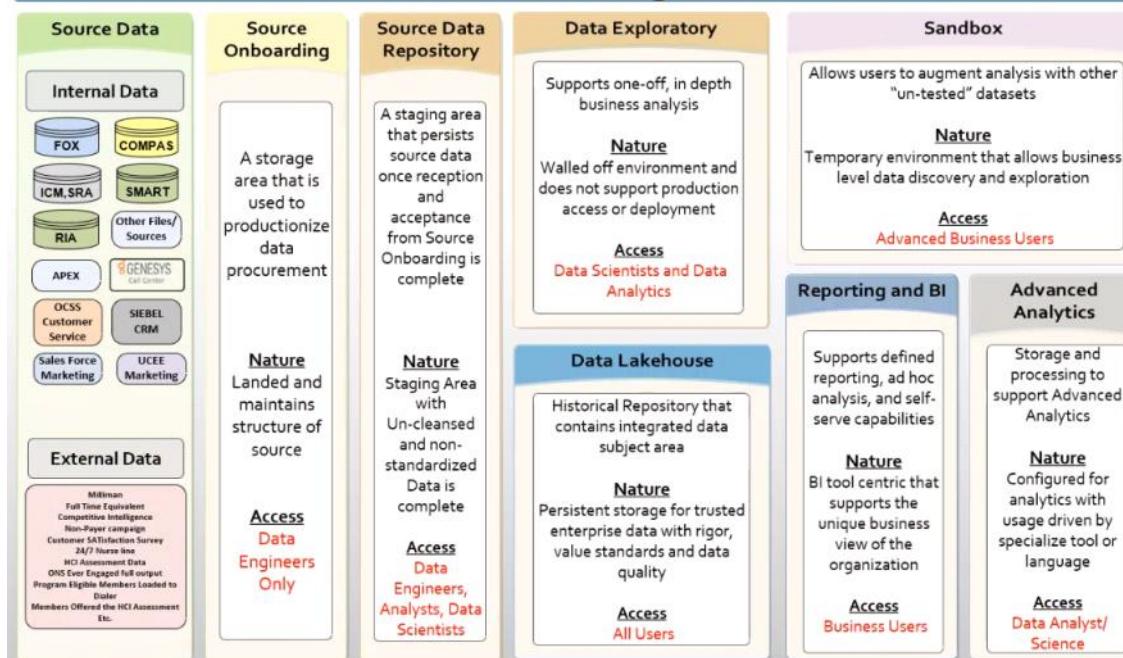
Selected Tools:

Advanced Analytics tools: Tableau, Microsoft Power BI, MicroStrategy, Qlik Sense

Data Science tools: Python, R, Jupyter Notebook, Tableau, Power BI



Cloud Data Regions



+++++
Data Lineage can be done three ways:

Manual - collect data source, destination and transformations

Automated- use a data lineage tool to automatically collect metadata from the data pipeline

Hybrid: combination of manual and automatic methods, can be a good approach

MRIS DQ & Metadata Lineage

Data Lineage

Data Governance - what is done and to achieve what

Integration with other Data Governance functions (Ex. Metadata management)

Map data flow paths describing where the data originated, how it has changed, and its destination

DQ Enablement

DQ Embedment into SDLC / PDLC Processes

Integrate DQ practice, and adoption in Engineering

Integrate technology and process for DQ enablement

Automate adoption monitoring

Enhanced NFR, and architecture guidelines to incorporate DQ /DG into SDLC

Assessment framework to validate eligibility for DQ /DG Monitoring

DQ Enablement KPI

DQ adoption monitoring tracker / dashboard

Metadata Lineage

Business Metadata Lineage Analysis

Process to build business metadata lineage

Buildout metadata lineage

Scratch Pad

Friday, March 24, 2023 10:03 PM

03/24/2023

Assessment of DW, ODS, EDP, DEEP data platforms including:

- Data quality
- Data model
- Data integration methods
- Data governance
 - Management
 - Operating model
- Security practices
- Architecture and technologies
- Costs

+++++

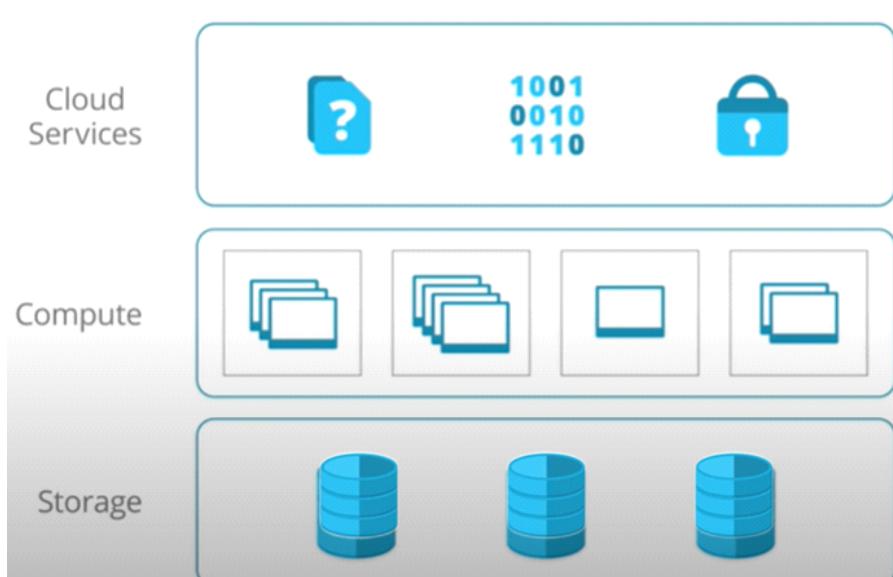
Slide 4 seems redundant to slide 3 and slide 5, move to appendix for now

Slide 5

- Weeks 1 – 3 should be Mobilization timeframe
 - Mobilization activities are before the project clock starts
 - Anything where you have spent time actually analyzing material, reviewing existing architecture, etc. should be after mobilization.
- Weeks 4 and beyond
 - Use a 2-week sprint approach versus showing 1 week blocks
 - Break work down in the lower text box in more detail. Document tasks that will happen for each chevron on the Gantt chart
 - Assessment of each existing asset should result in some perspective on the migration approach and how it potentially fits in to the modernized future state vision. Assessment should give us full understanding of the following and more:
 - If you are intending to use "Assessment of DW, ODS, EDP, DEEP data platforms including:" this text box to detail out specific tasks / activities of the current state assessment, frame it as such.
 - Title it as "Current State Assessment Approach for each Asset as follows:"
 - Don't just list items, change to an action statement with result (outcome)
 - i.e. Assess the quality of data to understand "blah blah blah" and to "blah blah blah" in future state
 - Your chevrons should reflect a little more detail
 - Maybe show the high-level steps of assessment – not sure
 - Future state is way to parallel (start and ends) with current state assessment
 - Findings Analysis... chevron is too generic. Look at slide 3 and call out some of those key things for the future state
- You're plan is way too bland and does not give the person looking at it instant "aha" understanding of how work will be conducted.
- Roles & Responsibilities
 - Call out the roles and responsibilities of MRIS as well and be very specific about what we need them to do

+++++

Snowflake Architecture



Cloud Services



Snowflake - cloud only data warehouse

Key Capabilities

- AWS and Azure Only (As on Date)
- No PK/FK Enforcement (Informational only)
- Snowflake SQL
 - DDL/DML
 - SQL Functions
 - UDF/Stored Procedure (Java Script)
- Views/Materialized Views
- ACID Transactions
- Analytical Aggregations, Windowing, and Hierarchical Queries

Integration Support

- Data Integration Tools (Informatica, Talend, Pentaho etc.)
- Self-service BI Tools (Tableau, QlikView, Spotfire etc.)
- Big Data Tools (Kafka, Spark, Databricks etc.)
- JDBC/ODBC Drivers
- Native Language Connectors (Python, Go, Node.js etc.)
- SQL Interface & Client
 - Snowflake Web Interface
 - Snowflake CLI
 - DBeaver

Unique Offerings

- Scalability (Storage and Compute)
- User Experience (Performance)
 - No Indexing
 - No Performance Tuning
 - No Partitioning
 - No Physical Storage Design
- Tunable Pay per use

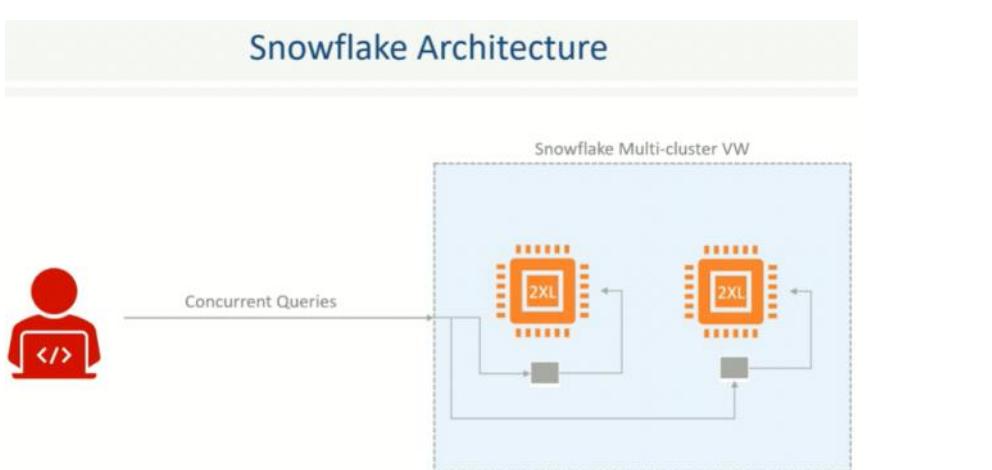
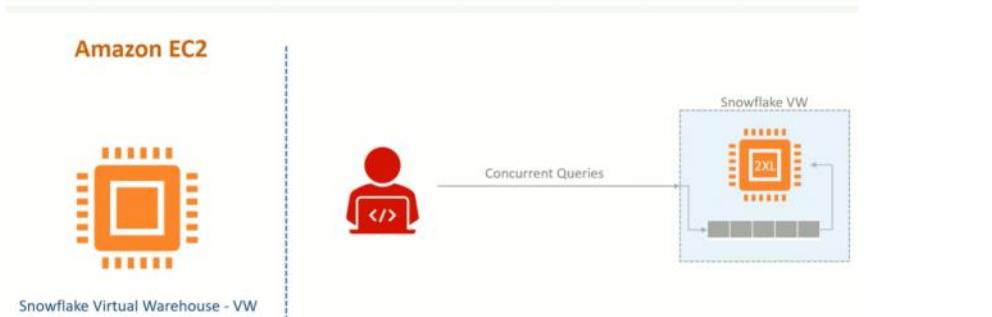
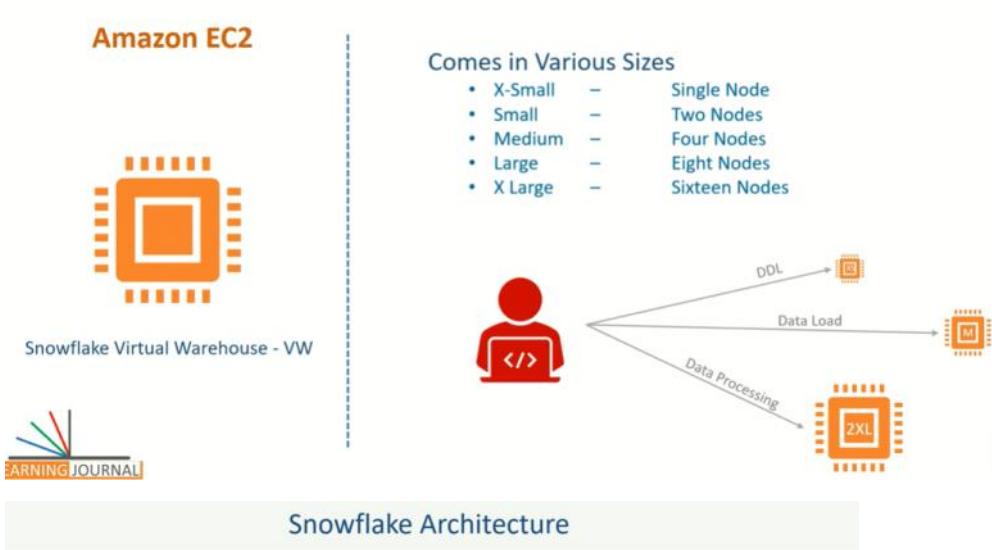
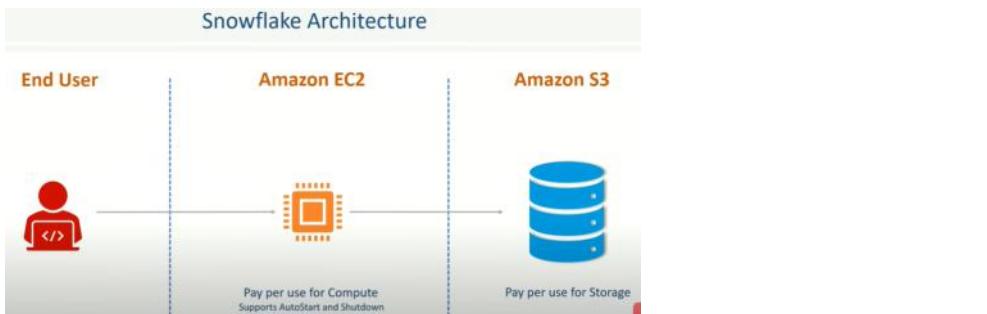
Separation of Storage & Compute

- Database
- Virtual Data Warehouse

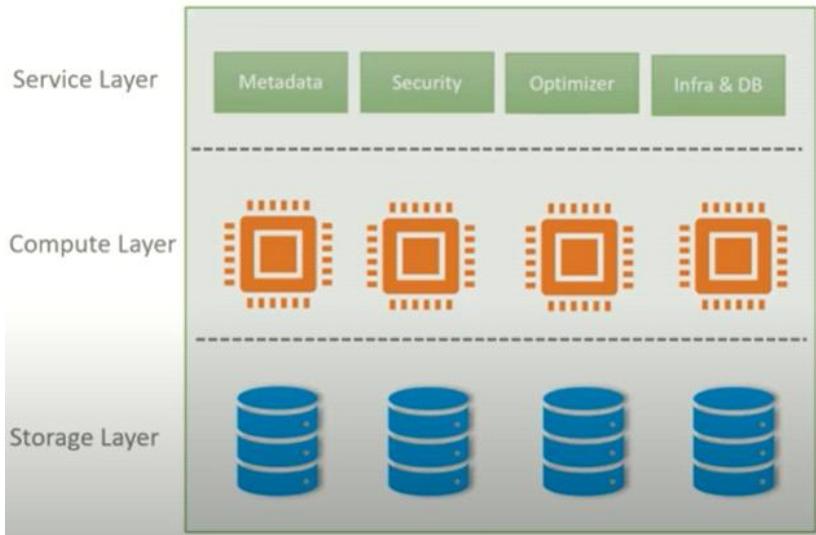
Snowflake Database Layer (using DDL)

- Create Database
- Create Schema
- Create Table
- Load Data in Tables
 - Loading from data files
 - Using Ingestion Tools

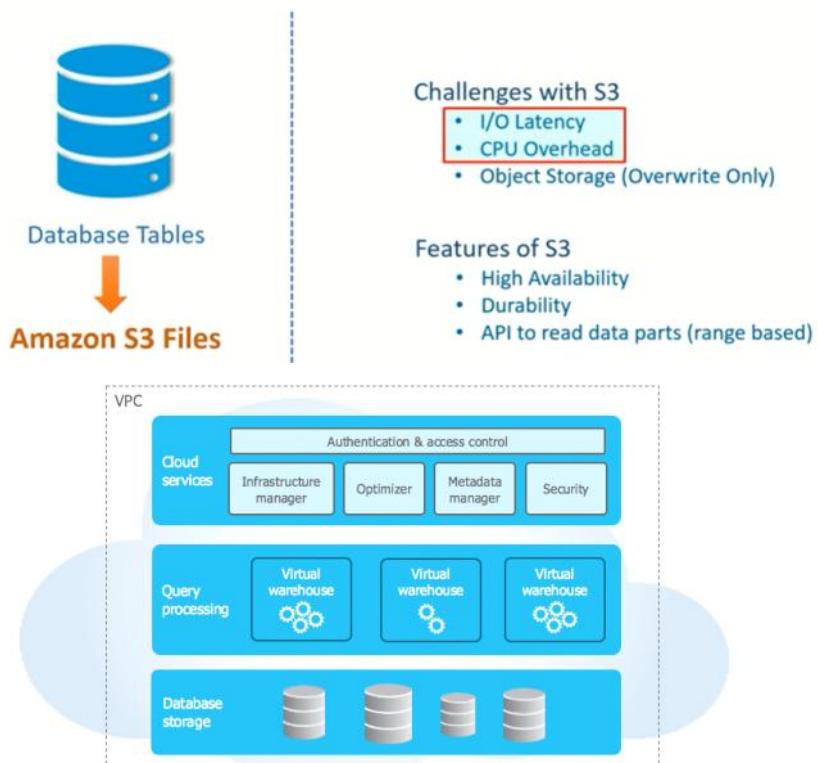
Amazon S3



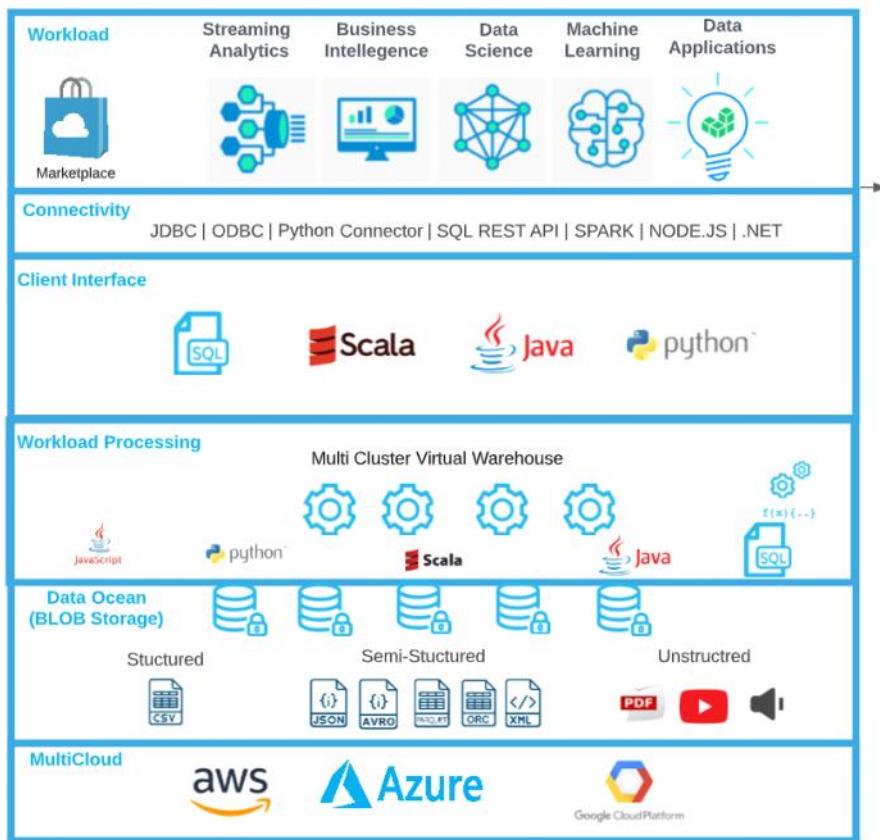
Snowflake Architecture



Snowflake Architecture

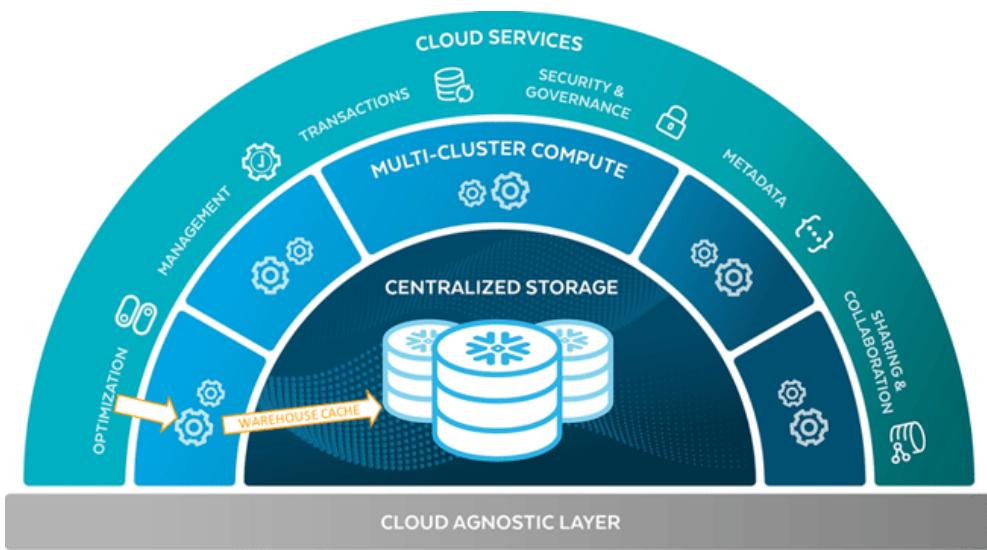
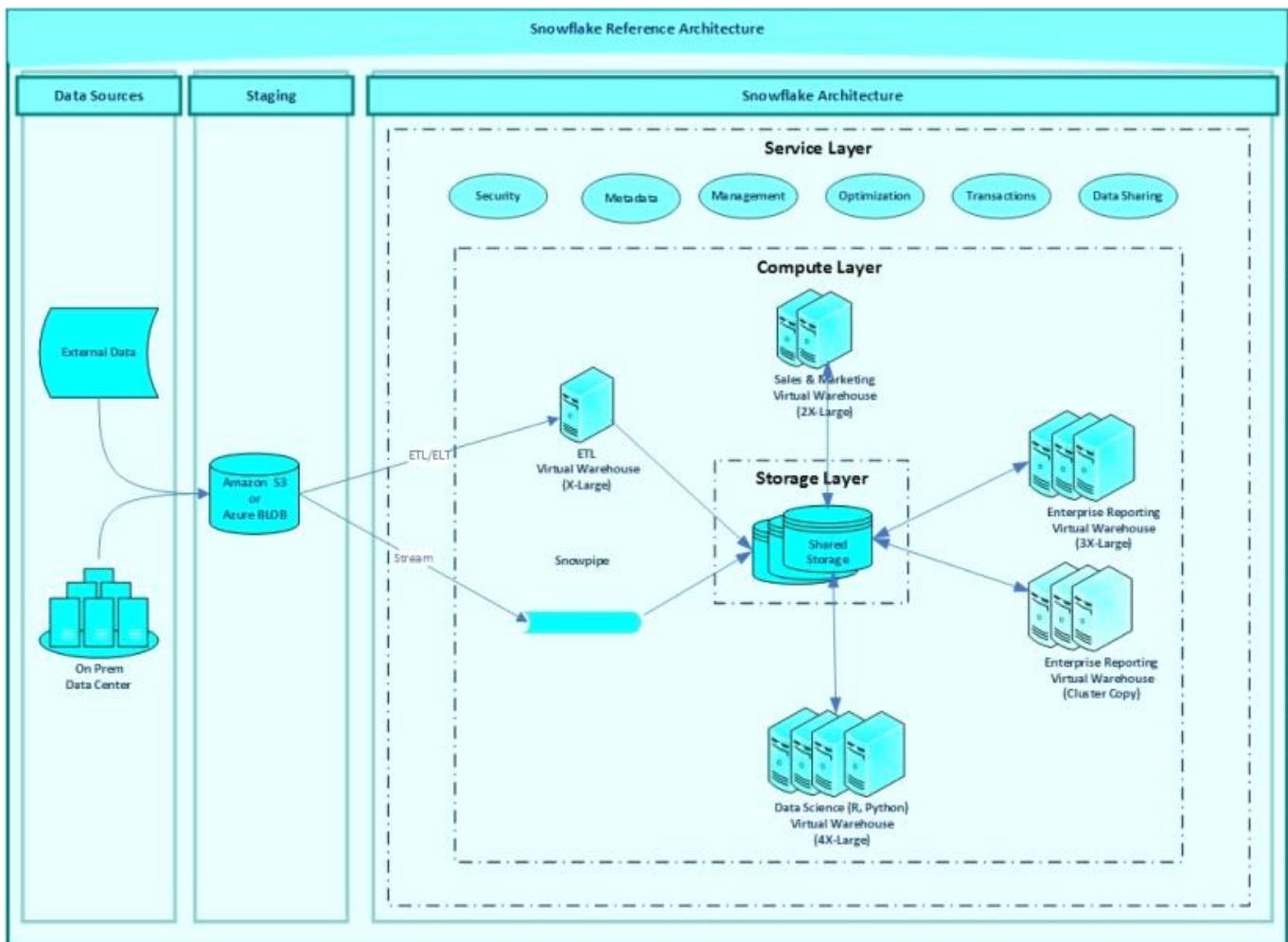


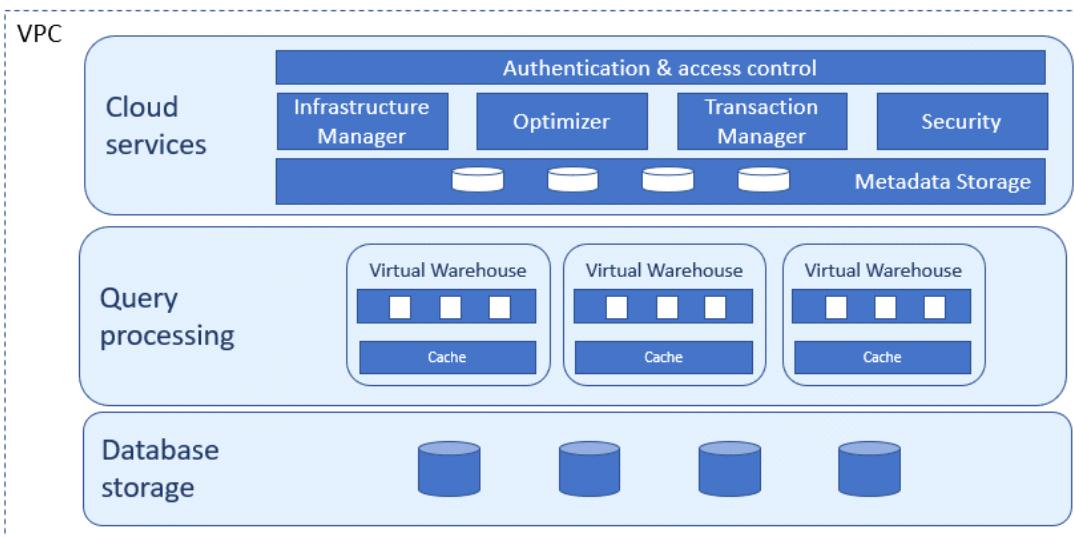
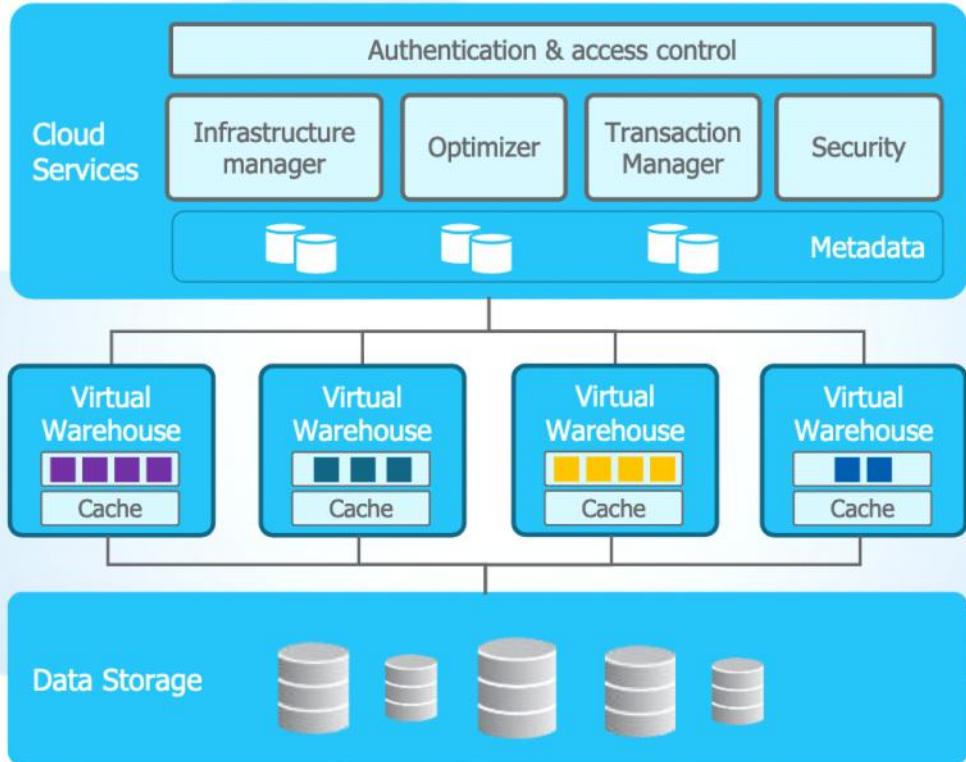
SNOWFLAKE DATA CLOUD ARCHITECTURE

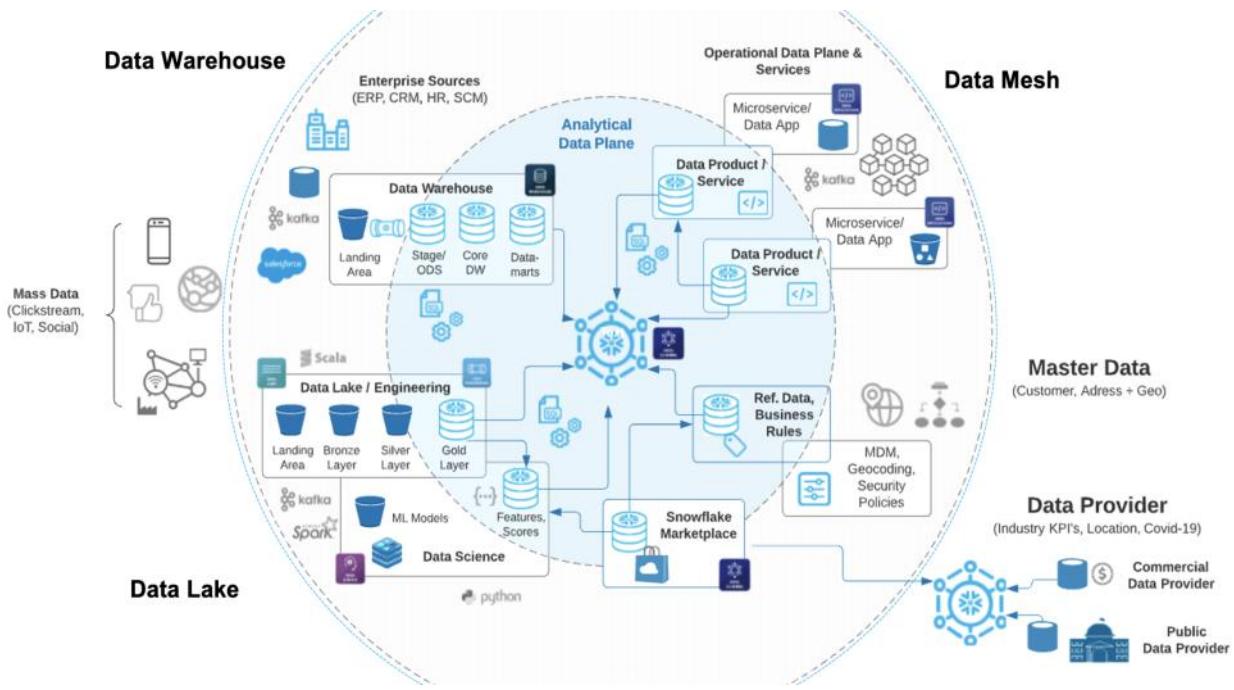


SNOWFLAKE MODERN DATA ARCHITECTURE

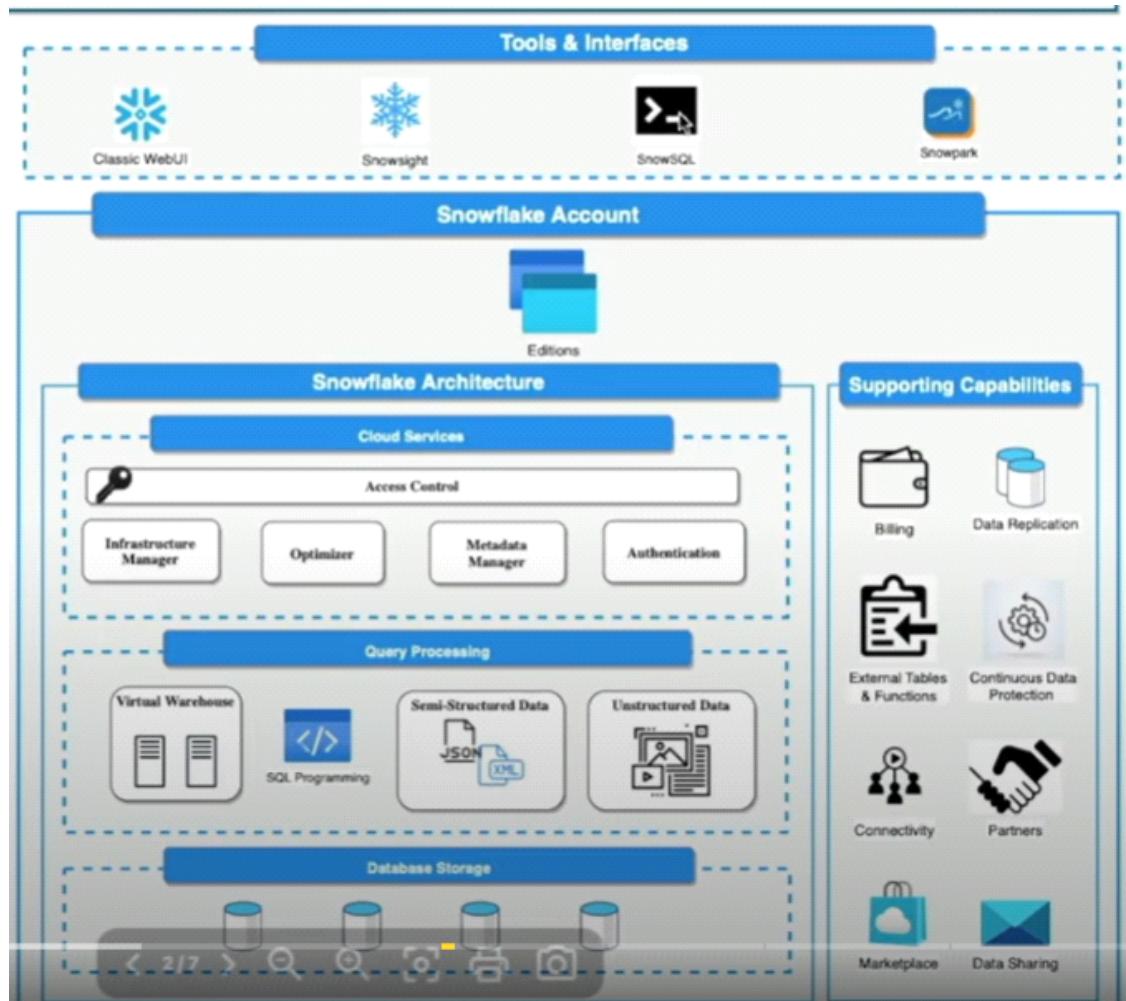


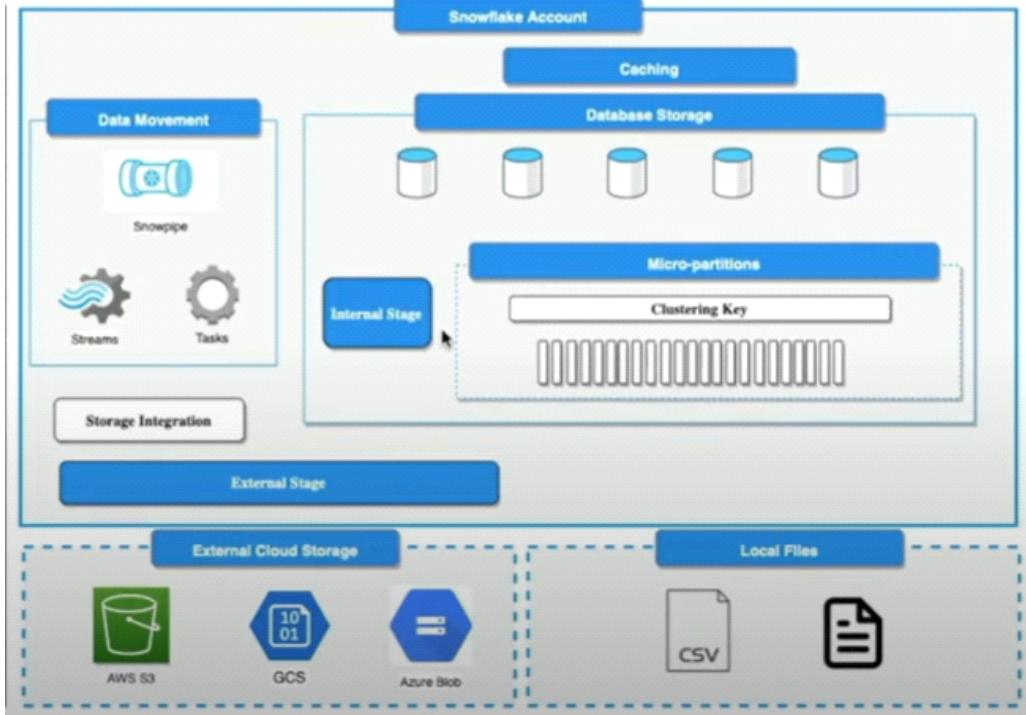






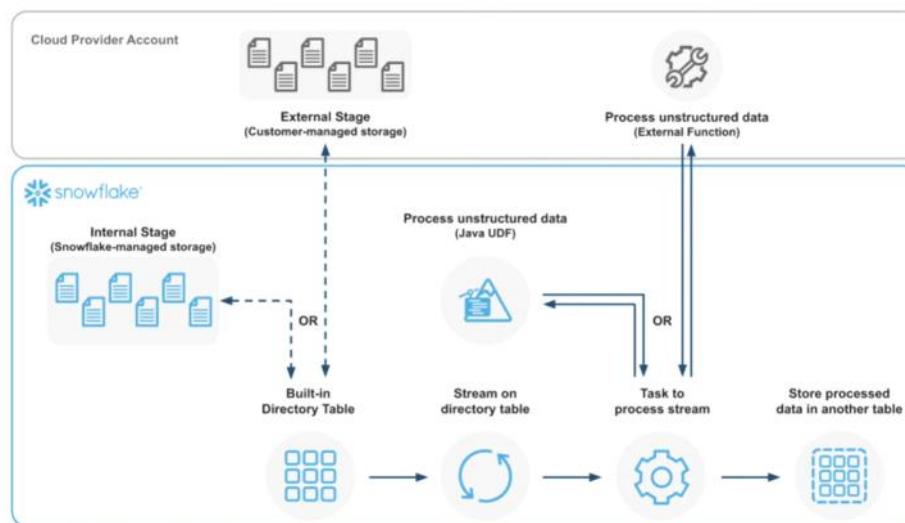
Really good article on Snowflake data architects by Harald Erb: [Snowflake 101 - for Data Architects](#)

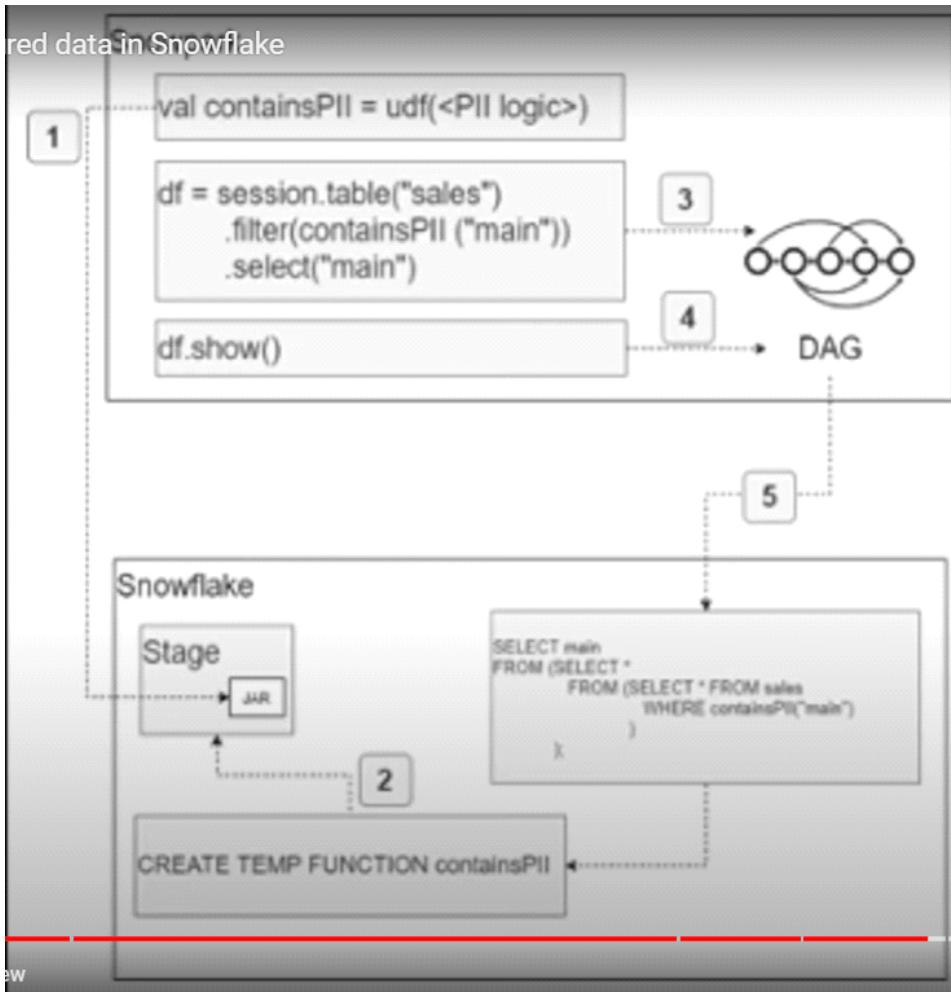
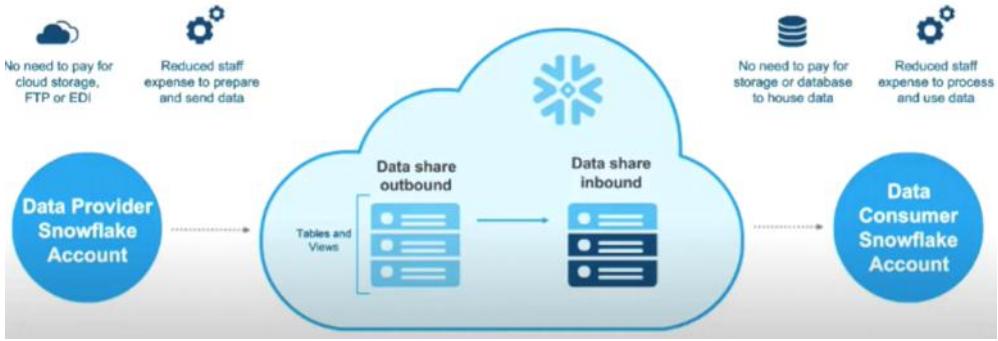


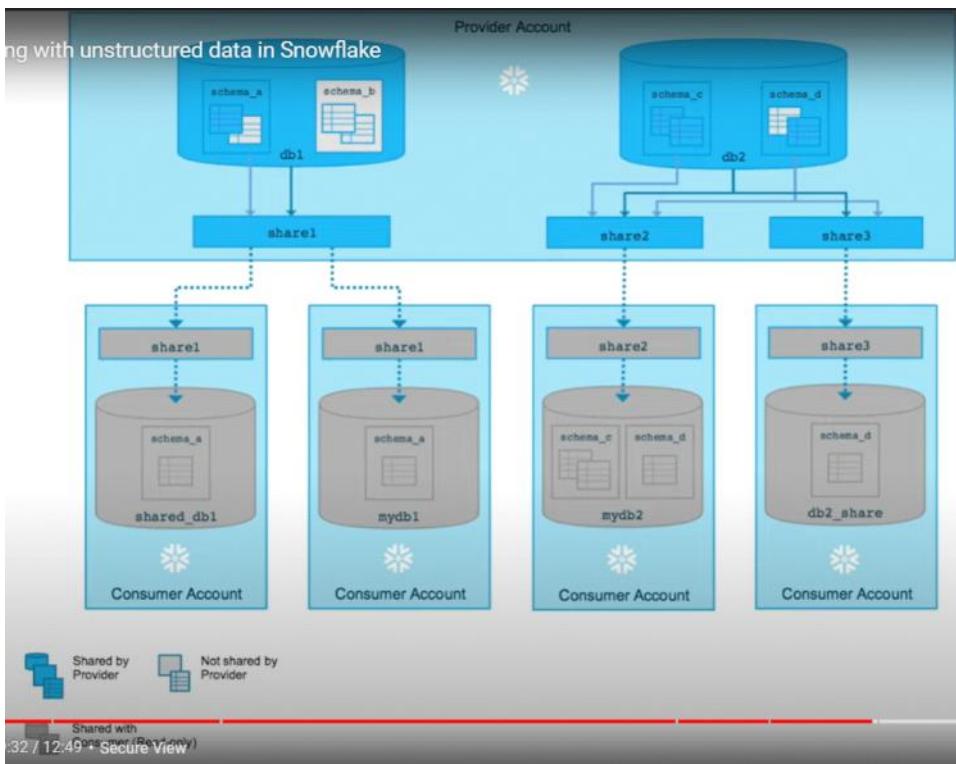


Unstructured data files can be accessed by data consumers via column values of this type in secure views shared by data providers.

- Q. Does snowflake support unstructured data?
- R. Yes, Snowflake does support unstructured data through its support for semi-structured data types such as VARIANT, OBJECT, and ARRAY. These data types allow users to store and query unstructured data in Snowflake. The VARIANT data type is a flexible, semi-structured data type that can store JSON, BSON, Avro, ORC, Parquet, and XML data. It allows for nested and repeated structures within a single VARIANT column, which makes it suitable for storing unstructured data. The OBJECT data type is similar to the VARIANT data type, but it is designed for storing data that has a well-defined schema. It can store data in key-value pairs and allows for nested and repeated structures. The ARRAY data type is used for storing ordered lists of values of any type, including semi-structured data types. Overall, Snowflake's support for semi-structured data types allows users to store and query unstructured data alongside structured data, making it a versatile platform for data warehousing and analytics.







1. Data Ingestion Layer: tools and technologies for data ingestion, extraction, transformation, and data loading
2. Data Storage Layer: storing data in a structured or unstructured format, include storage types such as, databases, data lakes, or data warehouses.
3. Data Processing Layer: tools and technologies for data cleaning, data modeling, data analysis, and machine learning.
4. Data Delivery Layer: Delivers data to end-users or downstream applications, using delivery mechanisms, such as APIs, dashboards, or reports, etc.

Assessment includes series of meetings to get full picture of end-to-end high-level view of the current state. The work plan is as follows:

Day 1 - Overview and sourcing sessions to understand how data sourcing and acquisition occur; challenges faced and opportunities discovered; tools and technologies for data ingestion, extraction, transformation, and data loading

Details: Request MRIS SMEs to go over the overview of this effort. Ask about the data sources (internal and external), any transformation/filtering; governance model, security practices, who are the users, how they interact with the data platform; data sources, types, sizes and frequency

Day 2 - Integration - how do migrations and integrations occur

Details: Ask MRIS SMEs how data is loaded, integrated and merged? What are the load patterns, e.g., streaming, batch, event based. Which processes and tools used; any challenges and opportunities? How many integrations? Issues, challenges, nuances, suggestions?

Day 3 - Organization of data - how is the data organized, what are the data models, data schema? Any issues/challenges/desires?

Details: Ask MRIS SMEs how the data is currently organized, what are the data models and data schema. What consumer personas are using it; what types of demands are on the system, e.g., Raw, aggregated, APIs, Advanced analytics, AI/ML, etc.

Day 4 - consumption patterns - how is data consumed, who is consuming it, how often it is consumed, how much is consumed? What are the current delivery channels, challenges, pain-points, etc.

Day 5 - Consolidate and review documentation

Activities: Consolidate, review and document all data points and documentation received

Repeat the same process for each data platform (4 weeks)

Week 5 - Consolidate and review documentation

Activities: Consolidate all data points and documentation provided from all data platforms; get clarifications/issue resolutions on outstanding issues. Conduct necessary research, and create the current state assessment report; review it internally and with the MRIS SMEs; get MRIS acceptance

Week 6 - start with the future state vision

Create future state architecture design. Detail out architecture layers, e.g., data ingestion, transformation, organization, delivery, etc. Factor in the desired features and current pain-points; map MRIS requirements against the future state architect vision. Conduct gap analysis, develop mitigation strategy; conduct necessary research for reference material to ensure future state architecture vision meets/exceeds expectations; document how those features and pain-points are addressed in the future state architecture.

Week 7 - Migration approach

Activities: Create cloud migration playbook to be used iteratively including potentially automated conversion tools

Request MRIS SME to identify and prioritize existing data sources to be migrated to Azure based Snowflake data platform. Ask for testing

validation criteria. Existing migration patterns, processes, and tools.

Divide up data sources into manageable workloads and migrate iteratively. Workloads may be by data type, target types, ingestion type, consumption type and frequency type. Request MRIS SMEs to identify ETL conversions/transformations

Create an adoption plan, that may require training and switching business users to new platform and switching off existing platform

Week 8 - Cloud migration governance framework

Activities: Define roles and responsibilities of different personnel involved in the migration process, e.g., cloud architect, cloud data architects, business analysts, security personnel, and operations staff.

Request MRIS SMEs to validate depicted roles and responsibilities. Get compliance, security and performance requirements. Collect performance metrics

Identify events such as, security, compliance, and performance, that may adversely affect the migration plan, assess risks, and devise plans to mitigate them.

Since there is compliance involved, develop policies and procedures to ensure that the migration adheres to regulations and standards.

Develop procedure(s) to manage cost associated with migration, such as, infrastructure, license fees, 3rd party tools/services, etc.

Create a process to manage change, e.g., testing, validation, rollbacks

Define the performance monitoring and management process to ensure established performance thresholds are met

Create security policies and controls to protect data assets against internal and external threats.

Review the framework with MRIS (business and technical) SMEs.

Week 9 - Cloud operating model

Activities: To appropriately manage and effectively operate data cloud services, OAS will recommend the following set of processes, tools, and best practices:

Create a plan which ensures that MRIS's objectives are aligned with the cloud operating model. Define the scope of the cloud environment, identification of needed services and technologies and an effective governance model to manage cloud resources.

Data DevOps are an essential component of the cloud operating model to deliver and manage cloud services at scale. OAS will suggest agile development practices, CI/CD pipelines and infrastructure as code processes to automate deployments and cloud resources management.

To effectively manage cloud environment, tools such as, logging, monitoring, security, compliance, automation, and cost management are required. This will help in optimization of cloud resources, contain costs, and ensure that the environment is compliant, safe, and secure.

To be successful in the cloud will need alignment across MRIS to ensure that everyone is marching toward the same target, including creating unambiguous roles and responsibilities, setting up cross functional teams, emphasis on collaboration and culture of innovation.

Week 10 - Cloud cost estimation and management

Activities: Based on OAS prior experiences, we will provide cost estimates similar services in the cloud. In addition, we will also provide cost management, tools, techniques and best practices

Week 11 - Create Cost Benefit Analysis report

Activities: The CBA report will be based on the current costs received from the MRIS SMEs and the estimated costs of similar services running in the cloud

Week 12 - Final delivery

Activities: The assessment report will be finalized that details the current state, identifies gaps, issues, opportunities and recommendations. A future state architecture vision will be finalized that aligns with MRIS desires and objectives. A CBA report will be delivered.

+++++
Meetings are to understand, sourcing, integration, organization and consumption

MRIS walks through of what they have

OAS will review meetings notes, documentation provided and conduct research

MRIS will talk about the architecture layers of the architecture landscape

Day 1 - Overview and sourcing sessions - how they sourcing and acquire data, what are the challenges and opportunities

Day 2 - Integration - how is data integrated and merged, what are the patterns and what are the workload sizes

Day 3 - Organization of data - how is the data organized, what are the data models, data schema? Any issues/challenges with it

Day 4 - consumption patterns - who is data consumed, who is consuming it, how often it is consumed. what are the challenges, pain-points, etc.

Day 5 - Consolidate and review documentation

Repeat the same process for all four data platforms

Week 5 - Consolidate, review and get clarifications and clear view of current state

Week 6 - start with the future state vision

Review documents to firm up our understanding

overview

sourcing

integration

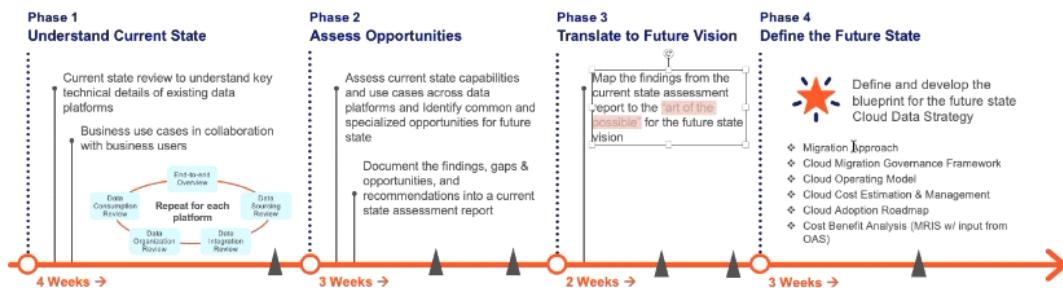
organization of data

consumption

understand and consolidate

+++++

High-level Work-plan for Execution



Roles & Responsibilities

Insurance Solution

- Providing access, current state artifacts
- Making right resources available to participate in the meetings, interviews, workshops as needed (tech & business)
- SME resources to provide appropriate KT as needed
- SME resources for interim review and feedback

OAS

- Plan agenda for, lead & facilitate meetings / interviews / workshops to understand current state data platforms and related business use cases
- Document key findings from each session and conduct interim review and validation w/ MRIS leadership
- Assessment and identification of gaps & opportunities in current state
- Document gaps and opportunities and conduct interim review and validation w/ MRIS leadership
- Map the findings from the current state assessment report to the "art of the possible" for the future state vision
- Document the future state architecture vision and conduct interim review and validation w/ MRIS leadership
- Define and develop the blueprint for the future state Cloud Data Strategy including the following: Migration Approach, Cloud Migration Governance Framework, Cloud Operating Model, Cloud Cost Estimation & Management, and Cloud Adoption Roadmap
- Provide input from the above for the development of the Cost Benefit Analysis by MRIS leadership

First meeting with Kiran: [Discuss IS Data Warehouse Current State-20230314_110342-Meeting Recording.mp4 \(sharepoint.com\)](https://sharepoint.com/:v/g/personal/john_shin_optum_com/Ed_rTjLQe51PrGVjWIR955wB8VEkUvgFmTqKmCBhmUdEvv)

Discuss IS Modernization: https://uhgazure-my.sharepoint.com/:v/g/personal/john_shin_optum_com/Ed_rTjLQe51PrGVjWIR955wB8VEkUvgFmTqKmCBhmUdEvv

Discuss IS modernization (part 2): https://uhgazure-my.sharepoint.com/:v/g/personal/john_shin_optum_com/EYpVC3mduvpAj82hE5PWk1EBdryTSbz74QBu2_cuZWJ-RA

Discuss IS modernization (part 3) on 03/31/2023: [Review Cloud Data Strategy Proposal Draft-20230331_120516-Meeting Recording.mp4 \(sharepoint.com\)](https://sharepoint.com/:v/g/personal/john_shin_optum_com/Review_Cloud_Data_Strategy_Proposal_Draft-20230331_120516-Meeting_Recording.mp4 (sharepoint.com))

[9:22 AM] Shin, John S

slide 6 seems redundant now so move to back please

[9:22 AM] Shin, John S

also, slide 7 isn't accurate in timelines

[9:23 AM] Shin, John S

week 6 isn't start of future state

[9:23 AM] Shin, John S

please take a look at slide 4 again and align

Proposed Approach

| Project Phases | WK 1 | WK 2 | WK 3 | WK 4 | WK 5 | WK 6 | WK 7 | WK 8 | WK 9 | WK 10 | revised |
|----------------|--|--|--|--|--|------|------|-----------------------|----------------------|-------|---------|
| | Mobilize and Plan | Assess and Analyze | | | Synthesize and Align (across multiple sourcing strategies) | | | Confirm and Recommend | Finalize and Debrief | | |
| Activities | | | | | | | | | | | |
| | <ul style="list-style-type: none"> • Mobilize team • Reference documentation for UX Use Cases • Information models: current state data source and client roll-out plan • Confirm internal data sources, data source SMEs and project managers, review schedule, data sources, review meetings • Identify external data sources and data vendors used by Optum Insight and Change Health | <ul style="list-style-type: none"> • Review current state solution; chairside shadowing of advocacy and self-service patient experience scope. • Review current state data sources; data availability; source data owners and data integration architecture • Review prioritized experience use cases with corresponding data needs • Review 2024 / 2025 client roll-out plan; identify external carriers in scope; covered lives and employer group's contract duration w/ carriers • Execute viability assessment of securing member consent for non-UHC carriers (as applicable under 21st Century Cures Act) w/ legal and 'selected' employer groups • Execute external data source research (in-scope data aggregators and data vendors) for in-scope use cases and quality data and experience gaps. | <ul style="list-style-type: none"> • Member consent - identify next steps for broader employer group population implementation approach (assuming this is a viable option) • External carrier engagement: compare data completeness across shortlisted external vendor sources to identify final list; identify data contracting requirements, lead time and estimated cost • Carrier strategy - Carrier specific data gaps and carrier specific data acquisition strategies per employer group generation • Consolidate experience impact for unresolved data gaps across acquisition strategies • Identify 'Plan B' (impact on people, process, technology) to mitigate the risk for unresolved data gaps | <ul style="list-style-type: none"> • Member consent – review and finalize recommendations • External data sources – review and confirm member consent; vendor's scope and identify next steps to initiate contracting • Carrier sourcing – review and finalize recommendations • Unresolved data gaps: mitigation execution plan | <ul style="list-style-type: none"> • Develop overall roadmap for various solutions • Develop executive debrief material • Review Roadmap and executive debrief material with the project sponsor • Executive readout | | | | | | |
| Deliverables | <ul style="list-style-type: none"> • Data sources scope confirmed • Define plan for Assess and Analyze • Define success criteria | <ul style="list-style-type: none"> • External data source mapping / data gaps / impacted experiences • Employer group specific strategies – Carrier penetration heat map • Member consent viability | <ul style="list-style-type: none"> • Unresolved data gaps with mitigation plan • Member consent implementation / integration plan • External sources engagement / integration plan • Direct carrier sourcing / integration plan | <ul style="list-style-type: none"> • Data sourcing recommendations and acquisition plan • Unresolved gaps mitigation plan, Change management scope | <ul style="list-style-type: none"> • Data Acquisition RoadMap | | | | | | |
| Optum | | | | | | | | | | | |

Notes from meeting with Dinesh 04/07/2023

Four environments in scope - call it out somewhere

Get rid of Day by day, day by day, setting us up for failure out of the gate.

Need an assessment framework that captures all the variables; a high level blueprint and say here are the things that we would be assessing.

if we can show the client of framework of assessment
then it is easier to say, hey, Mr. client, this is what we would be assessing. These are the parameters of assessment.

Meeting sessions (one 2-hour session in the morning and one 2-hour session in the evening)

What are we going to target to capture in those discovery meetings?

It can be in the appendix to say here are the 20 parameters that would be included in our assessment.

We should add a slide to enumerate on what will be covering as part of the overview session, what are the inputs that we're looking for? What are we trying to understand, etcetera for each of those sessions; as a secondary slide that we will have in the appendix to refer to as needed; refer to Dinesh's slide on how to present this information. This slide is saying what we are going to do.

Frame it into tasks and activities in a somewhat abbreviated manner, without losing the intent and context.

Two weeks, three weeks, four weeks. This is basically kind of like a three-week window. We are saying we are going to do the assess and analyze in the these three weeks. Here are all the things these are going to be, the outputs that we are going to create. So the deliverables need to be created, work products. So my assumption is if I go back to your this.

Assessment report - Is that the only output that would get created?

Overview Day two, break it into working sessions to say we anticipate 8 working sessions for each source system. Here's the breakdown. Whatever you are covering in the working session, that should map to all the assessment parameters.

Look across people, process technology

Emails

Sunday, April 9, 2023 3:26 PM

Re: MRIS Cloud Data Strategy - Need to build TWO Core Collaterals to sell the deal



Nand, Durga

To: Malhotra, Dinesh; Cheema, Dave; Shin, John S
Cc: Ramakrishnan, Anantha

Retention Policy: UHGInbox (90 days)

Reply Reply All Forward Print ...

Sun 4/9/2023 3:03 PM

Expires 7/8/2023

Internal

Start your reply all with: Ok, sounds good. Ok, thank you. Thanks. I will take a look. Feedback

Dinesh,

Usage of "cloud day" word muddies the scope and responsibility and this is not an industry standard term either. Since John, Dave, and you have been in meeting, I am at loss here. Pl clearly define what deliverables would come from cloud side and which one from data and some may be common that we can talk later. I would review and come back to you, I have not seen this information anywhere.

Thanks

Durga

Get [Outlook for iOS](#)

From: Malhotra, Dinesh <dinesh.malhotra@optum.com>

Sent: Sunday, April 9, 2023 3:10:26 PM

To: Nand, Durga <durga.nand@optum.com>; Cheema, Dave <dave.cheema@optum.com>; Shin, John S <john.shin@optum.com>

Cc: Ramakrishnan, Anantha <anantha.ramakrishnan@optum.com>

Subject: RE: MRIS Cloud Data Strategy - Need to build TWO Core Collaterals to sell the deal

Durga and team –

This is a Cloud Data Strategy project – with a primary focus on data estate running on legacy infrastructure primarily and how that data estate need to be strategized to be solutioned, architected, migrated to cloud infrastructure and application stack with the go-forward operational and governance considerations going forward.

Analytics, reporting etc. – need to be addressed from business continuity perspective – and not as an Analytics Application strategy project.

I have had client conversations and have an understanding of Liv's expectations. John (and Dave) has already socialized the scope and outcomes and the client leaders are on board with the directional approach / outcomes.

My guidance to the team –

- Let's get grounded on the scope of the assessment – and the scope of recommendations – and use assessment blueprint and recommendations framework to finalize what we would like to commit to deliver.
- The project will be strategized as ONE project with different workstreams as needed – operating under ONE, singular solution direction and project delivery ownership.
- Let us not second guess amongst ourselves about the scope of the project.

Best,

Dinesh Malhotra

M: +1 773-398-7713

From: Nand, Durga <durga.nand@optum.com>

Sent: Sunday, April 9, 2023 12:33 PM

To: Malhotra, Dinesh <dinesh.malhotra@optum.com>; Cheema, Dave <dave.cheema@optum.com>; Shin, John S <john.shin@optum.com>

Subject: RE: MRIS Cloud Data Strategy - Need to build TWO Core Collaterals to sell the deal

Hi Dinesh,

I have not put together a framework, when we will have engagement, we will put one.

On the other topics, my suggestions are as below:

- There are two streams of work – Cloud Platform Engineering (Assessment/ Design/ Stand up / Migration) and Analytics Applications (Data warehouse type applications) migration framework and in my opinion we are conflating two. These are completely different skill sets and stream of work.
- Let us separate it out and have separate responsibility. I am not sure if Analytics Applications Migration Assessment Framework exists – that need to be developed as well. Cloud Platform assessment should be my responsibility and Analytics application John/ Dave / Anantha.
- Let us confirm from Ins. Soln if they have made an assessment of

- type applications) migration framework and in my opinion we are conflating two. These are completely different skill sets and stream of work.
- Let us separate it out and have separate responsibility. I am not sure if Analytics Applications Migration Assessment Framework exists – that need to be developed as well. Cloud Platform assessment should be my responsibility and Analytics application John/ Dave / Anantha.
 - Let us confirm from Ins. Soln if they have made an assessment of
 - They can or cannot use HCP
 - If not, have they done any work in this area or what is their thinking
 - Let us go from there.

With Regards

Durga Nand | Optum
 Optum Advisory Services
 T +1 (952) 324-4007 (O)

From: Malhotra, Dinesh <dinesh.malhotra@optum.com>
Sent: Saturday, April 8, 2023 8:15 PM
To: Nand, Durga <durga.nand@optum.com>; Cheema, Dave <dave.cheema@optum.com>; Shin, John S <john.shin@optum.com>
Subject: MRIS Cloud Data Strategy - Need to build TWO Core Collaterals to sell the deal
Importance: High

Durga – thanks for sending the governance slide

- Do you have an assessment framework for Cloud Migration Assessment that you may have put together as you built multiple collaterals over last 12 months? That could be perhaps tailored for MRIS

All – the snippets Durga has shared below need to be bucketed under different categories so that these can be methodically assessed as a part of a formal assessment.

- We will need a structure to categorize our collective random thoughts and build a formal collateral that can be used for positioning client conversations. This should become our formal GTM collateral as Optum's / OAS's Cloud Data Strategy Assessment Blueprint
- We also need to define the framework (or Table of Contents) for the Future State Cloud Data Strategy Recommendations.
 - You all have good experience in various aspects of cloud and cloud data strategy, architecture, migration, delivery, and post migration aspects. Please put yourself in the client's shoes and challenge yourself what should the final assessment deliverable include – and let's create a Table of Contents for the Future State Cloud Data Strategy Deliverable to deliver to the client

Best,
Dinesh Malhotra
 M: +1 773-398-7713

From: Nand, Durga <durga.nand@optum.com>
Sent: Saturday, April 8, 2023 3:55 PM
To: Cheema, Dave <dave.cheema@optum.com>; Shin, John S <john.shin@optum.com>
Cc: Malhotra, Dinesh <dinesh.malhotra@optum.com>
Subject: Material from yesterday's discussion

Dave,
 Material from yesterday is attached.

John and Dinesh,
 The key questions to answer before we move forward is to understand their platform strategy

- Do they have one
- What is it – Hybrid Cloud / Public Cloud
- Which public cloud they have chosen
- Cloud management strategy
- Cloud compute strategy
- Cloud Storage management
- IaaS vs Paas, Network,
- CI/CD & DevOps
- Infrastructure as code maturity
- Cloud brokering
- Cloud security
- Integration Services – Streaming, Kafka, catalogs,
- Cloud native development strategy / capacity / frameworks
- Cloud financial management – tagging, Budgeting, chargeback

- Cloud security
- Integration Services – Streaming, Kafka, catalogs,
- Cloud native development strategy / capacity / frameworks
- Cloud financial management – tagging, Budgeting, chargeback
- Scaling plan
- Functional components
 - Automation
 - Data Management
 - Portals

These are the high level one's that we would need to understand or define before we move one way or another.

With Regards

Durga Nand | Optum
 Optum Advisory Services
 T +1 (952) 324-4007 (O)

FP Re-imagine User Experience - high-level plan.xlsx

 Malhotra, Dinesh
 To: Cheema, Dave; Shin, John S
 Cc: Ramakrishnan, Anantha; Nand, Durga; Koneti, Rajesh; Malhotra, Dinesh
 Retention Policy: UHGIinbox (90 days) Expires: 7/8/2023


 Reply  Reply All  Forward  ...

Sun 4/9/2023 2:27 PM

 FP Re-imagine User Experience - high-level plan.xlsx

22 KB

Here is how a construct like this can be approached to critically think and plan for projects like the one we are targeting for MRIS:

1. Define end outcomes – scope out all elements that would need to be delivered to client as Final Deliverable
2. Using the scope of end outcomes and the Final Deliverable, identify all assessment activities to be undertaken and interim work products to be created that would lead to the synthesis of current state findings to develop Final Deliverables
 - a. For the scope of assessment activities:
 - i. Identify who would have primary and secondary ownership
 - ii. Identify the pre-requisites for each of the activity
 - iii. Key skills to be staffed on the OAS team – and who would own which activities / work products
3. Use the detailed breakdown of #2 above to come up with sequenced flow of activities and tasks plotted against weekly / daily plan
 - a. Use this weekly / daily plan to assess reasonableness of assessment scope of activities (or Work Breakdown Structure)
 - b. Identify which activities / tasks will be at risk for execution and how to mitigate those execution risks
4. Use #2 and #3 to verify your staffing plan – and identify what specific roles need to be full time and what roles would be part time and when they would need to be engaged

Hope this helps develop your own mental model for planning and strategizing the proposals for projects like MRIS Cloud Data Strategy and other projects.

Let me know if you have any questions or need any additional input / clarification. Should you have any suggestion to further expand on this critical thinking framework, do not hesitate to share your feedback.

Best,
 Dinesh

RE: Insurance Solutions - Cloud Data Strategy Proposal - Review and Refine

 Malhotra, Dinesh
 To: Cheema, Dave; Shin, John S; Nand, Durga; Ramakrishnan, Anantha
 Retention Policy: UHGIinbox (90 days) Expires: 7/8/2023


 Reply  Reply All  Forward  ...

Sun 4/9/2023 3:05 PM

Anantha and I will be in a full day client working session

Dave, Durga and John – You please go ahead and review and refine the Assessment Blueprint (h/l scope, activities and work products) and the Final Deliverable scope (what all should be targeted for the Final Deliverable).

- You don't have to wait till Wednesday to do this.

I will make time on Thursday to review our finalized approach, scope, activities and final deliverable outline

Best,
Dinesh Malhotra
 M: +1 773-398-7713

+++++

[Yesterday 10:16 PM] Malhotra, Dinesh

+++++

[Yesterday 10:16 PM] Malhotra, Dinesh

I took a quick look - overall it comes across more coherent - for the last three phases, I would encourage you to challenge yourself with the 'why' for each of the bullet to confirm the intent, timing and the required pre-requisites

[Yesterday 10:18 PM] Malhotra, Dinesh

We still need to define an Assessment Blueprint and the Final Deliverable Strawman. Assessment Blueprint should include all the topics / subject areas (including the ones Durga sent in the email) that need to be systematically addressed during the Discovery phase - and all need these need to tick and tie into the various report-out topics included in the Final Deliverable Strawman

[Yesterday 10:19 PM] Malhotra, Dinesh

Please put together a strawman for the Final Deliverable - what all topics we need to be able to provide recommendations to the client that can be cleanly mapped on a roadmap with methodical and logical sequencing of various recommendations

[Yesterday 10:21 PM] Malhotra, Dinesh

Let's start with a 'Table of Contents' of Topics and Sub-Topics that should be addressed in the Final Deliverable - that would help assess what all need to be assessed during the Discovery phase - that would also help estimate the rough order of magnitude for schedule and effort and staffing - and accordingly we can objectively and convincingly work with the client whether this work can be done in 8 weeks, 10 weeks or 12 weeks

[Yesterday 10:21 PM] Malhotra, Dinesh

Without having some handle on a granular game plan - it will be difficult to objectively defend the schedule and the price of the project

[Yesterday 10:22 PM] Malhotra, Dinesh

I will share these messages with John as well so that he too is in the loop - as we bring our critical thinking together

Big Data Comparison - Snowflake, Synapse, Databricks

Sunday, July 30, 2023 11:14 PM

Sure, here is a comparison of Snowflake, Synapse, and Databricks:

| Feature | Snowflake | Synapse | Databricks |
|--------------|---|--|---|
| Pricing | Pay-as-you-go, with a minimum of 60 seconds of usage | Pay-as-you-go, with a minimum of 12 hours of usage | Serverless, with a minimum of 100 milliseconds of usage |
| Performance | High performance, with low latency | High performance, with low latency | High performance, with low latency |
| Scalability | Highly scalable, with the ability to handle petabytes of data | Highly scalable, with the ability to handle petabytes of data | Highly scalable, with the ability to handle petabytes of data |
| Features | Supports a wide range of features, including ACID transactions, data sharing, and data encryption | Supports a wide range of features, including ACID transactions, data sharing, and data encryption | Supports a wide range of features, including ACID transactions, data sharing, data encryption, and machine learning |
| Ease of use | Easy to use, with a simple user interface | Easy to use, with a simple user interface | More complex to use, as it requires knowledge of Spark |
| Integrations | Integrates with a wide range of applications, including BI tools, data visualization tools, and machine learning frameworks | Integrates with a wide range of applications, including BI tools, data visualization tools, and machine learning frameworks | Integrates with a wide range of applications, including BI tools, data visualization tools, and machine learning frameworks |
| Use cases | Ideal for traditional business intelligence and analytics workloads | Ideal for a wide range of workloads, including traditional business intelligence and analytics, machine learning, and data streaming | Ideal for machine learning and data engineering workloads |

As you can see, all three platforms offer high performance, scalability, and a wide range of features. However, they each have their own strengths and weaknesses. Snowflake is a good choice for traditional business intelligence and analytics workloads, while Synapse is a good choice for a wide range of workloads. Databricks is a good choice for machine learning and data engineering workloads.

TCO - Snowflake, Synapse, Databricks

Sunday, July 30, 2023 11:15 PM

The total cost of ownership (TCO) for Snowflake, Synapse, and Databricks can vary depending on a number of factors, including the size and complexity of your data, the type of workloads you run, and the pricing plans you choose. However, a study by Gigaom found that Azure Synapse had the lowest TCO of the three platforms, followed by AWS and Snowflake.

Here is a more detailed breakdown of the TCO for each platform:

- Azure Synapse: Synapse is a unified analytics platform that combines the best of data warehouses and data lakes. It is a good choice for organizations that need a scalable and flexible platform for a variety of workloads. The TCO for Synapse is lower than Snowflake and AWS because it offers a pay-as-you-go pricing model and does not require any upfront costs.
- AWS: AWS is a popular cloud platform that offers a wide range of services, including data warehousing and analytics. The TCO for AWS can be higher than Synapse because it requires more upfront costs and has a more complex pricing structure. However, AWS offers a wider range of features and flexibility than Synapse.
- Snowflake: Snowflake is a cloud-based data warehouse that is known for its high performance and scalability. The TCO for Snowflake can be higher than Synapse and AWS because it does not offer a pay-as-you-go pricing model. However, Snowflake is a good choice for organizations that need a high-performance data warehouse for mission-critical workloads.

Ultimately, the best platform for you will depend on your specific needs and requirements. If you are looking for a scalable and flexible platform with a low TCO, then Azure Synapse is a good option. If you need a wider range of features and flexibility, then AWS is a good choice. And if you need a high-performance data warehouse for mission-critical workloads, then Snowflake is a good option.

Here are some additional factors to consider when evaluating the TCO of these platforms:

- Data storage: The amount of data you need to store will affect the TCO of all three platforms. Synapse and AWS offer more storage options than Snowflake, so they may be a better choice if you have a large amount of data.
- Compute: The amount of compute power you need will also affect the TCO of all three platforms. Synapse and AWS offer more compute options than Snowflake, so they may be a better choice if you need to run complex workloads.
- Support: The level of support you need will also affect the TCO of all three platforms. Synapse and AWS offer more support options than Snowflake, so they may be a better choice if you need help with troubleshooting or optimization.

TCO Databricks & Snowflake

Sunday, July 30, 2023 11:16 PM

The total cost of ownership (TCO) for Databricks can vary depending on a number of factors, including the size and complexity of your workloads, the cloud provider you use, and the features you need. However, in general, Databricks is a cost-effective platform for big data processing.

One of the main benefits of Databricks is that it can help you to reduce your infrastructure costs. Databricks is a fully managed platform, which means that you don't have to worry about provisioning or managing servers. This can save you a significant amount of money, especially if you're running large-scale workloads.

Databricks can also help you to reduce your operational costs. The platform includes a number of features that can help you to automate tasks, such as data loading and processing. This can free up your IT team to focus on other tasks, such as developing new applications.

In addition to the infrastructure and operational costs, you also need to consider the cost of the Databricks licenses. The licenses are priced based on the number of active users and the amount of compute resources you use. However, Databricks offers a number of pricing options, so you can choose the one that best fits your needs.

Overall, the TCO for Databricks can be significantly lower than the TCO for other big data platforms. However, it's important to carefully evaluate your needs before you make a decision.

Here are some additional factors to consider when evaluating the TCO of Databricks:

- The type of workloads you run. Databricks is well-suited for a variety of workloads, including data warehousing, machine learning, and data science. However, some workloads may be more cost-effective on other platforms.
- The size of your data. Databricks can handle large datasets, but the cost of storing and processing data can increase as the size of your data grows.
- The cloud provider you use. The cost of Databricks can vary depending on the cloud provider you use. For example, Databricks is generally more expensive on AWS than it is on Azure.
- The features you need. Databricks offers a wide range of features, but not all features are necessary for every organization. If you only need a subset of the features, you can save money by choosing a lower-priced plan.

If you're considering using Databricks, it's important to carefully evaluate your needs and the TCO of the platform. By doing so, you can ensure that you're making the best decision for your organization.

+++++

The total cost of ownership (TCO) for a Snowflake data warehouse can vary depending on a number of factors, including the size of the data warehouse, the number of users, and the level of performance required. However, in general, Snowflake is a cost-effective option for data warehousing.

Here are some of the factors that can affect the TCO of a Snowflake data warehouse:

- **Storage:** Snowflake charges for storage on a per-byte basis. The cost of storage will vary depending on the type of data being stored, the compression ratio, and the region where the data is stored.
- **Compute:** Snowflake charges for compute on a per-second basis. The cost of compute will vary depending on the size of the virtual warehouse, the number of concurrent users, and the level of performance required.
- **Networking:** Snowflake charges for networking on a per-byte basis. The cost of networking will vary depending on the amount of data being transferred and the distance between the data warehouse and the users.
- **Other costs:** In addition to storage, compute, and networking, there are a few other costs associated with a Snowflake data warehouse, such as licensing fees, support fees, and data transfer fees.

Overall, the TCO of a Snowflake data warehouse is typically lower than the TCO of a traditional on-premises data warehouse. This is because Snowflake eliminates the need to purchase and maintain hardware, software, and staff. Additionally, Snowflake's pay-as-you-go pricing model can help to reduce costs during periods of low usage.

Here is a table that summarizes the TCO of a Snowflake data warehouse compared to a traditional on-premises data warehouse:

| Factor | Snowflake | Traditional on-premises |
|--------------------|------------|-------------------------|
| Storage | Per-byte | Per-byte |
| Compute | Per-second | Per-second |
| Networking | Per-byte | Per-byte |
| Licensing fees | None | Yes |
| Support fees | Yes | Yes |
| Data transfer fees | Yes | Yes |
| Total cost | Lower | Higher |

As you can see, the TCO of a Snowflake data warehouse is typically lower than the TCO of a traditional on-premises data warehouse. This is because Snowflake eliminates the need to purchase and maintain hardware, software, and staff. Additionally, Snowflake's pay-as-you-go pricing model can help to reduce costs during periods of low usage.

If you are considering a new data warehouse, I recommend that you evaluate Snowflake as an option. Snowflake is a cost-effective, scalable, and reliable solution that can help you to improve your data analytics capabilities.