RAPID

Research Analytics Platform for Investment Decisions

MetLife Investment Management

Jan 2021

1 RAPID Proposal Overview

2 POC Use Case Readout

3 Recommendations and Next Steps

1 RAPID Proposal Overview

Investments Research Platform Proposal

Operationalize a technology infrastructure and define an operational framework that will enable investment research to leverage a standard set of tools to perform analytics.

Investments Research Platform Services Goals

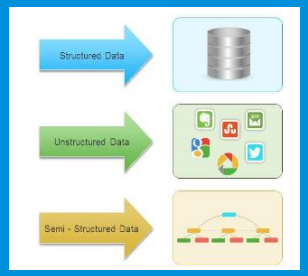
Research Platform Services

- Enable Data Analytics, Insights, Machine Learning, NLP, and AI
- Enable Data Visualization and Presentation
- Support Data transformation and feature engineering
- Expandable Data Storage
- Flexible Data Access

Primary Targeted Users

- Initial roll out to Investments Front Office Analysts
- Investments Technology
- Phased rollout to investments community driven by use cases and user maturity

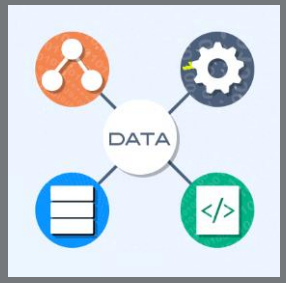
Investments Research Platform Data Scope



Data Types

Increasing Growth

- **Unstructured**: Word, pdf, Web Blogs, Social Media News
- **Semi Structured**: Excel, csv, html, xml, JSON, NoSQL
- **Structured**: Data in sql database table



Data Sources

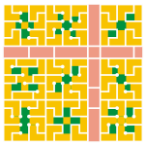
- **Public**: Sec Filing, Public Company Financial
- **Subscriptions**: Intex, Bloomberg, Reuters
- **MIM Proprietary Data**: research publications, international subsidiary holdings
- **Counterparty offerings and data services**: private deal offering, servicer remittance, broker and bank mortgage deal offering

Research Platform Guiding Design Principles

Simplicity: Intuitive ease-to-use platform with **Secured** role-based access

Rapid **Time-to-Market** supported by DevOps

Decentralized Service Model. Self Service where feasible. Investments Business and IT Product Teams have full access to all tools and capabilities



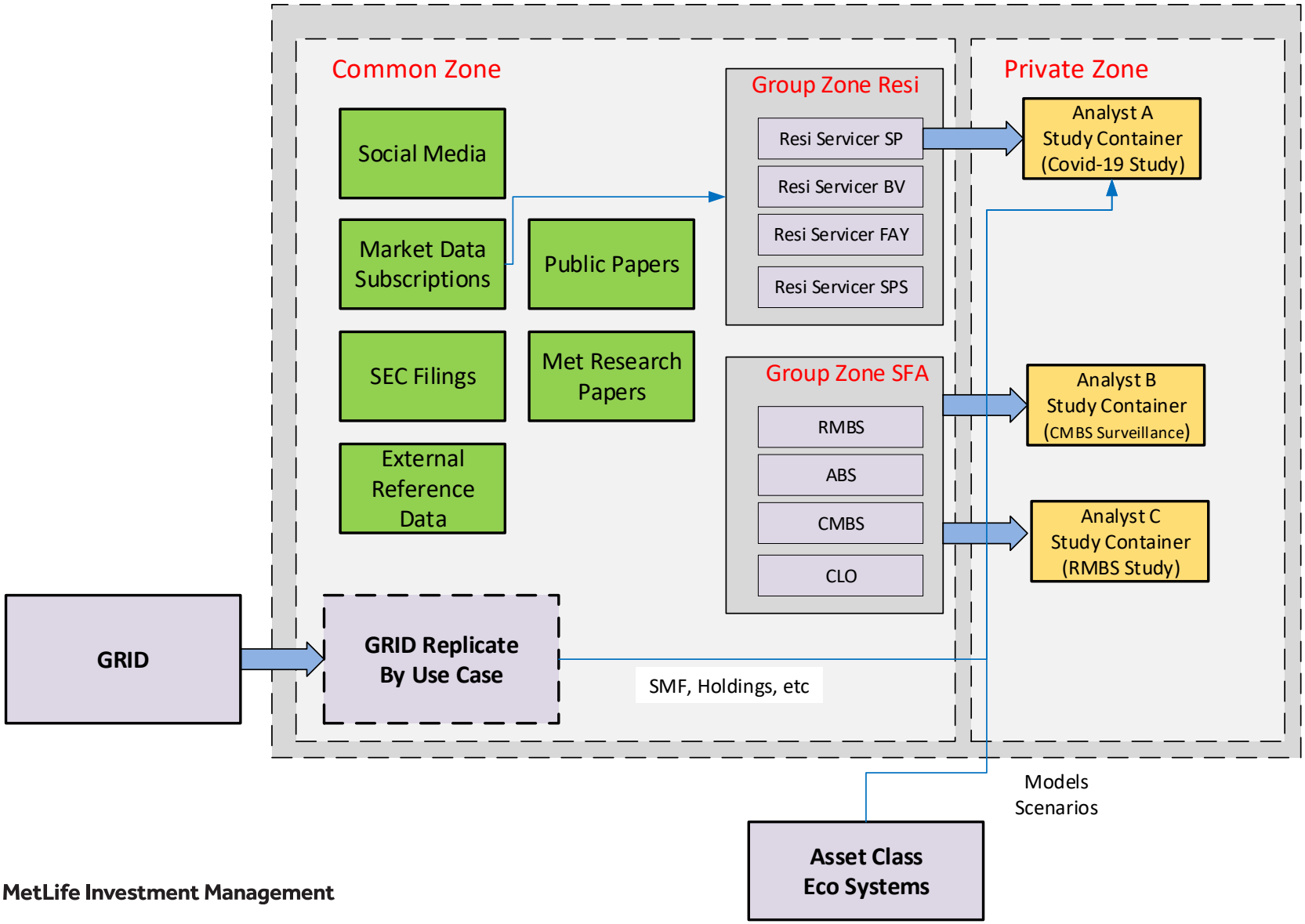
Integrate with Operational Data Lake (**GRID**) where needed

Scalable architecture that is both high-available and fit for multiple-purposes

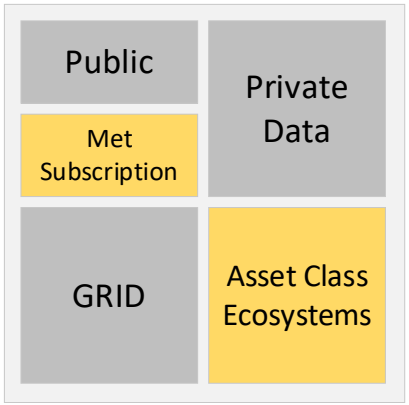
Ease of migration from existing manual / semi manual processes (e.g. excel VB.)

Support **data layers**: private, group, and common data

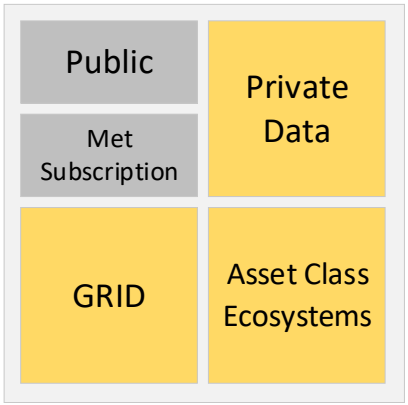
Investments Research Platform Data Zone Diagram



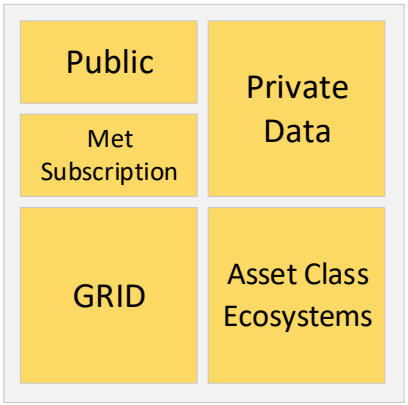
RAPID Use Cases: Data Source Patterns - examples



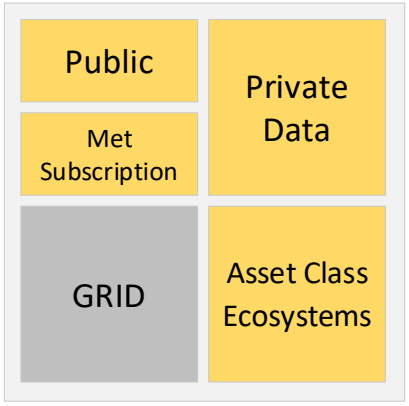
Resi – Covid 19 Exposure Analysis



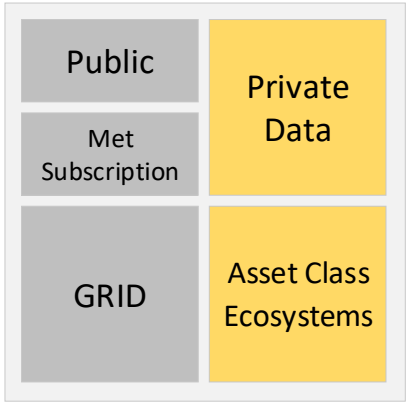
CLO Research
Broker Loan Listing



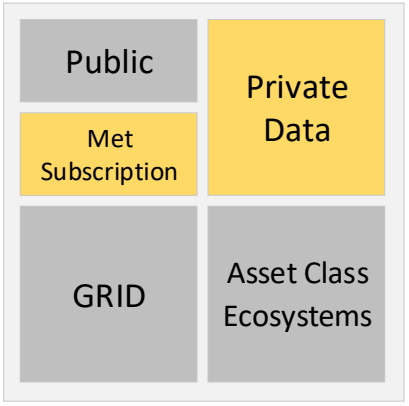
- Public Structured Finance/Kanerai
- ABS – Auto Loan
- Social Media MIM Sentiment



Company Financial Credit Research
CapIQ
Pub/Sub Research Paper



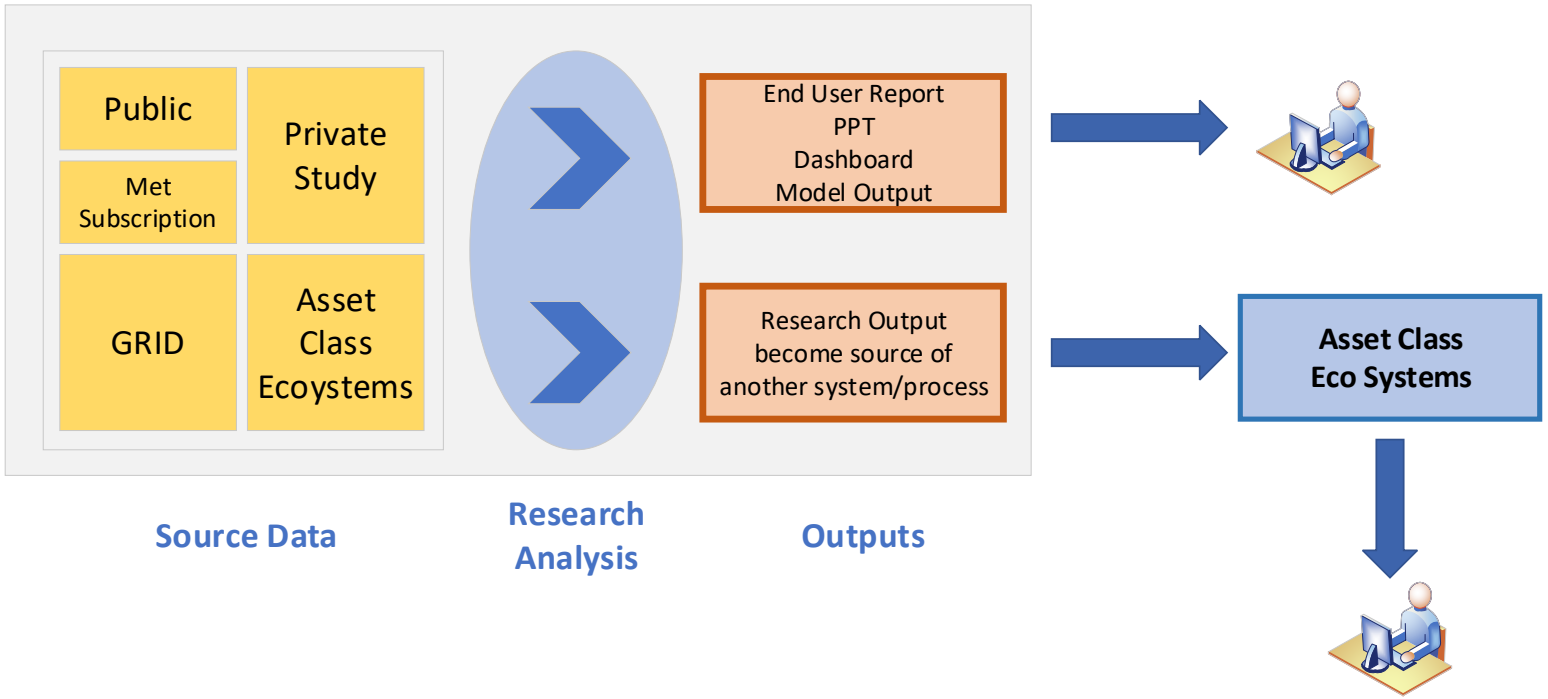
Broker Offering Analysis
Chile Resi



Resi – Single Family Rental (SFR)

Legend: Applicable data source per user case.

RAPID Use Case : 2 Study Output Patterns



Investments Research Platform RACI

	Architecture	Infrastructure Engineering	Data Team	Inv Business	IT Asset Classe Product Teams
Infrastructure and Common Services					
Infrastructure Requirement	C	C	A	I	R
Infrastructure Buildout	C	R	A	I	C
Entitlement Framework Design /Development	C	R	A, R	I	R
Common Technology Module Activation	R	R	A, R	I	C
Governance Framework	R	I	A, R	I	C
Define common and subsription data driven by business use cases	I	I	C	C	A, R
Common Public Data Injection	C	C	A,R	I	C
Shared Subsription Data Injection	C	C	A, R	C	C
Use Cases Implementation Type A - By Business Super Users					
Define Use Case	C		C	A,R	C
Business Engagement	C	I	C	A, R	R
Entitlement Design for business use case	C	I	C	C	A, R
Entitlement Implementation for use case	C	R	I	C	A
Use Case Design	C	I	C	A, R	C
Use Case Implementation	C	I	C	A,R	C
Use Cases Implementation Type B - By Technology Team					
Define Use Case	C		R,C	R,C	A, R
Business Engagement	C	I	R	C	A, R
Entitlement Design for business use case	C	I	C	C	A, R
Entitlement Implementation for use case	C	R	I	C	A
Use Case Design	C	I	C	C	A, R
Use Case Implementation	C	I	R	C	A, R

A - Accountable R - Responsible C - Consulted I - Informed

2

POC Use Case Readout

POC Goal and Benefits

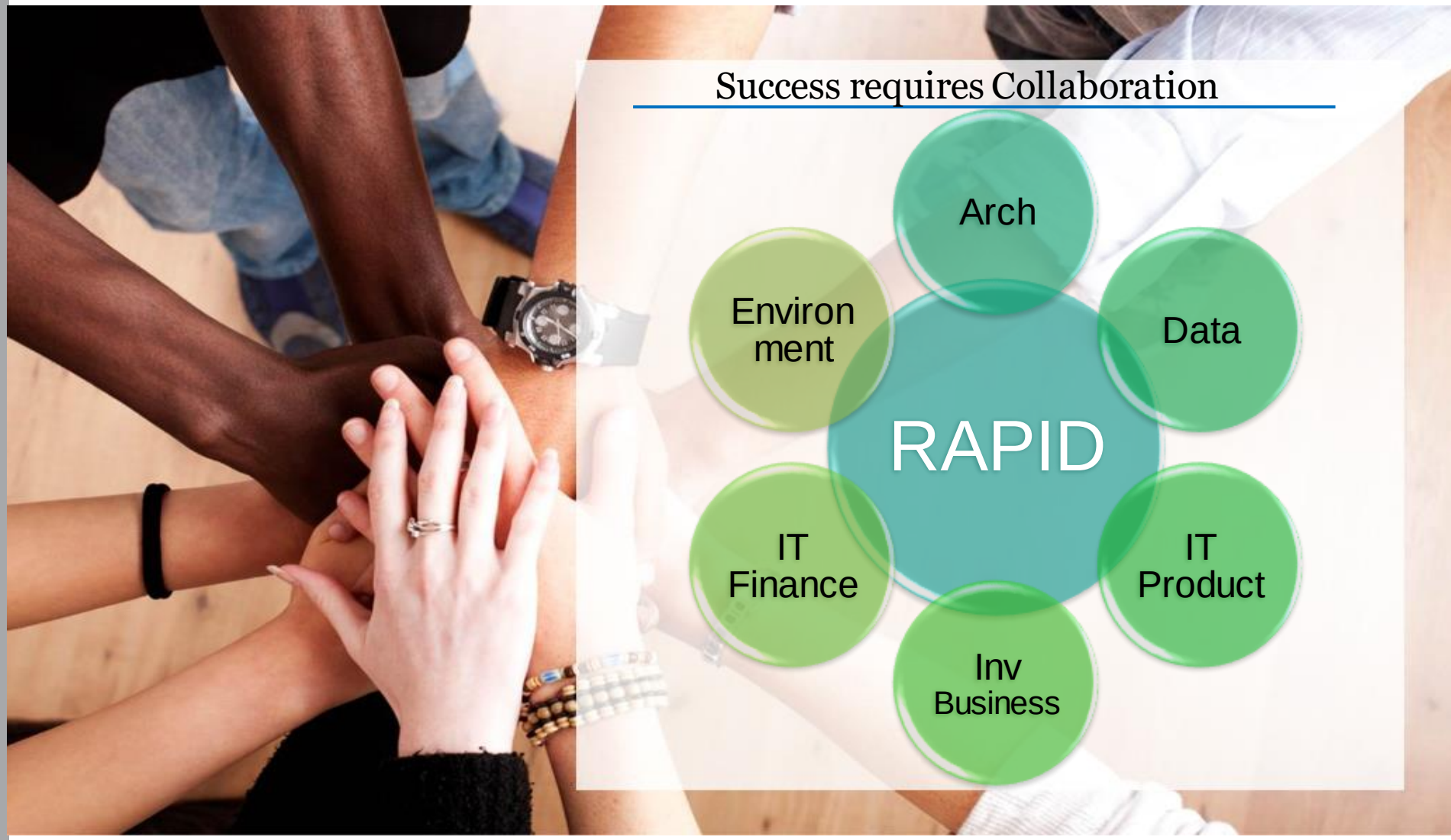
GOAL

Execute a research use case POC on Azure to define technology infrastructure, explore tools, and design operation framework for investment research to optimize analytics.

BENEFIT

- Develop a platform for investment researcher and analyst to land, experiment, analyze, and visualize their data
- Define RAPID design principles
- Define entitlement and deployment strategies
- Gain additional insight on Azure components and capabilities

RAPID POC Team



Use Case : Covid-19 exposure and risk analysis for Resi Loan

Current Process Description

Daily data from 4 residential loan servicers pertaining to Covid-19 impacts to Met resi loan portfolio.

Data includes **forbearance requests** and **workout options**.

Data in Different excel format from each servicer. Some of these files arrive via ftp, others through email.

The current process to compile a single table with all the consolidated information and to provide timely analyses are both extremely manual and very time consuming.

Opportunity w/ RAPID

Opportunity to automate the process in research platform and develop deeper and timely analyses

Opportunity to load files to research platform, make adjustments to standardize, and develop analytics query.

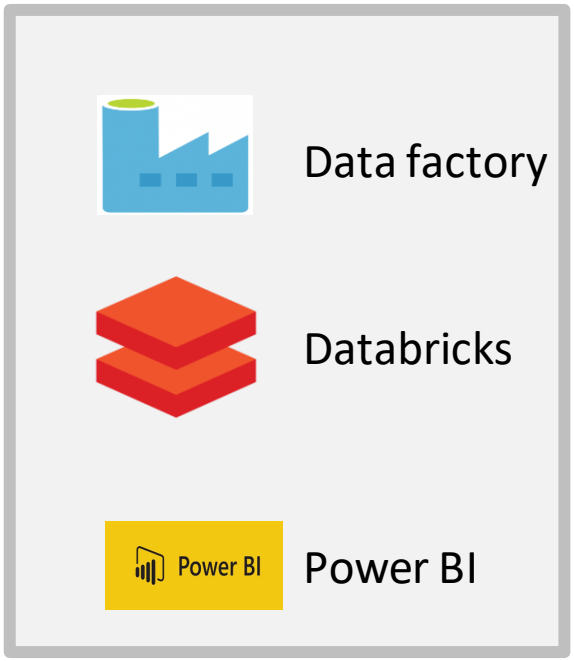
The process could be vastly improved by providing better analytics, more accurate data and more efficient use of time.

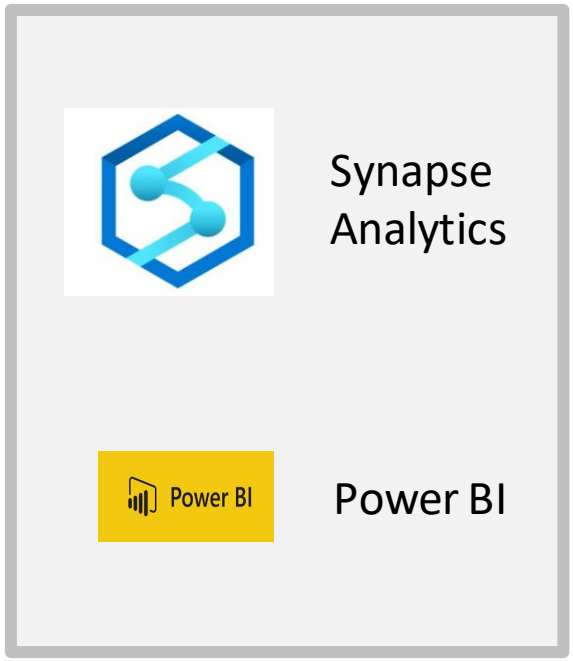
Functional Requirements

- Ability to load multi tab excels with various formats. Join data from different tab or different excel in the lake
- Ability to join holdings data from SPARCS – Resi analytics platform
- Ability to perform data preparation and transformation
- Ability to analyze data into standardize output
- Ability to Visualize data

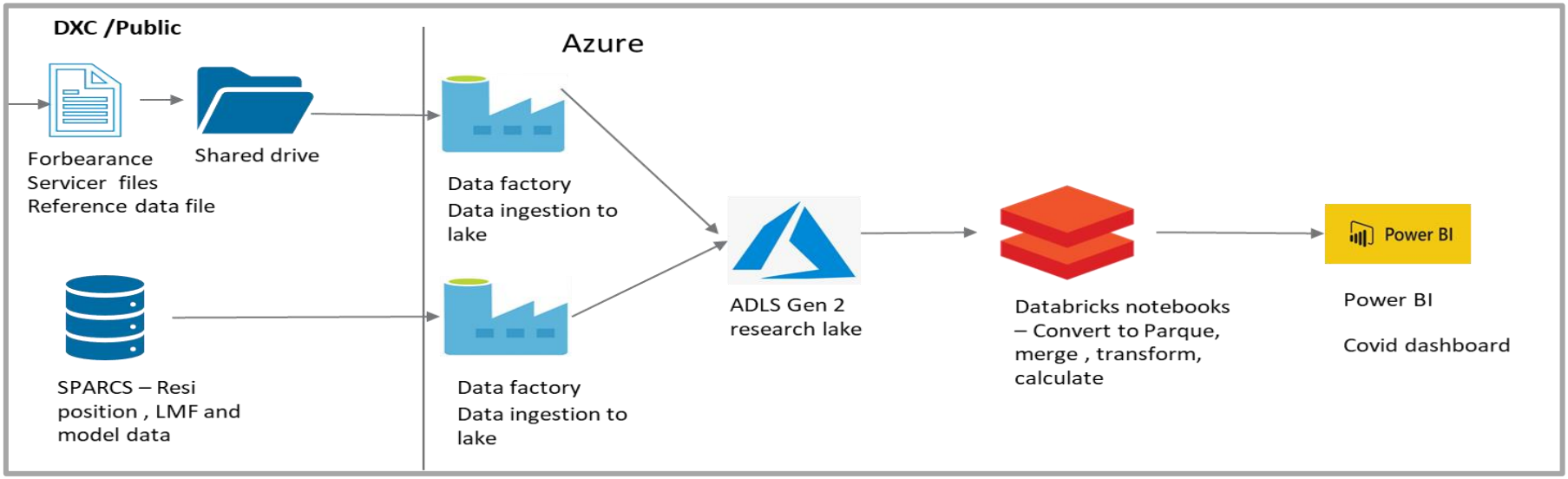
Data Size: 100M to start

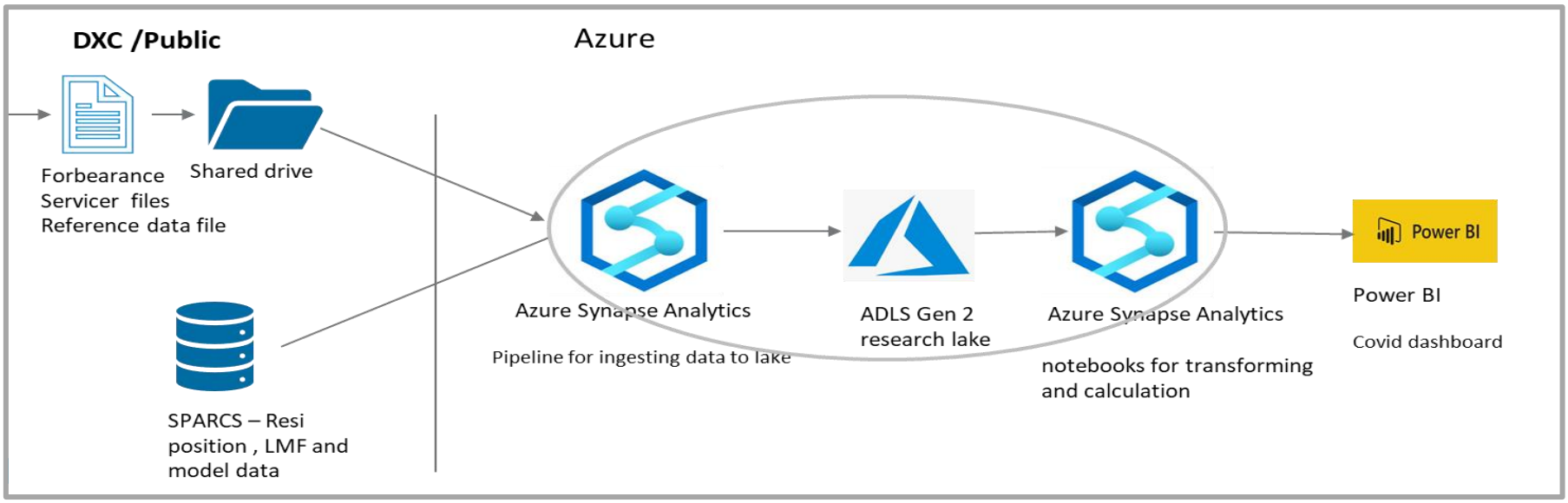
One Use Case, Two Approaches



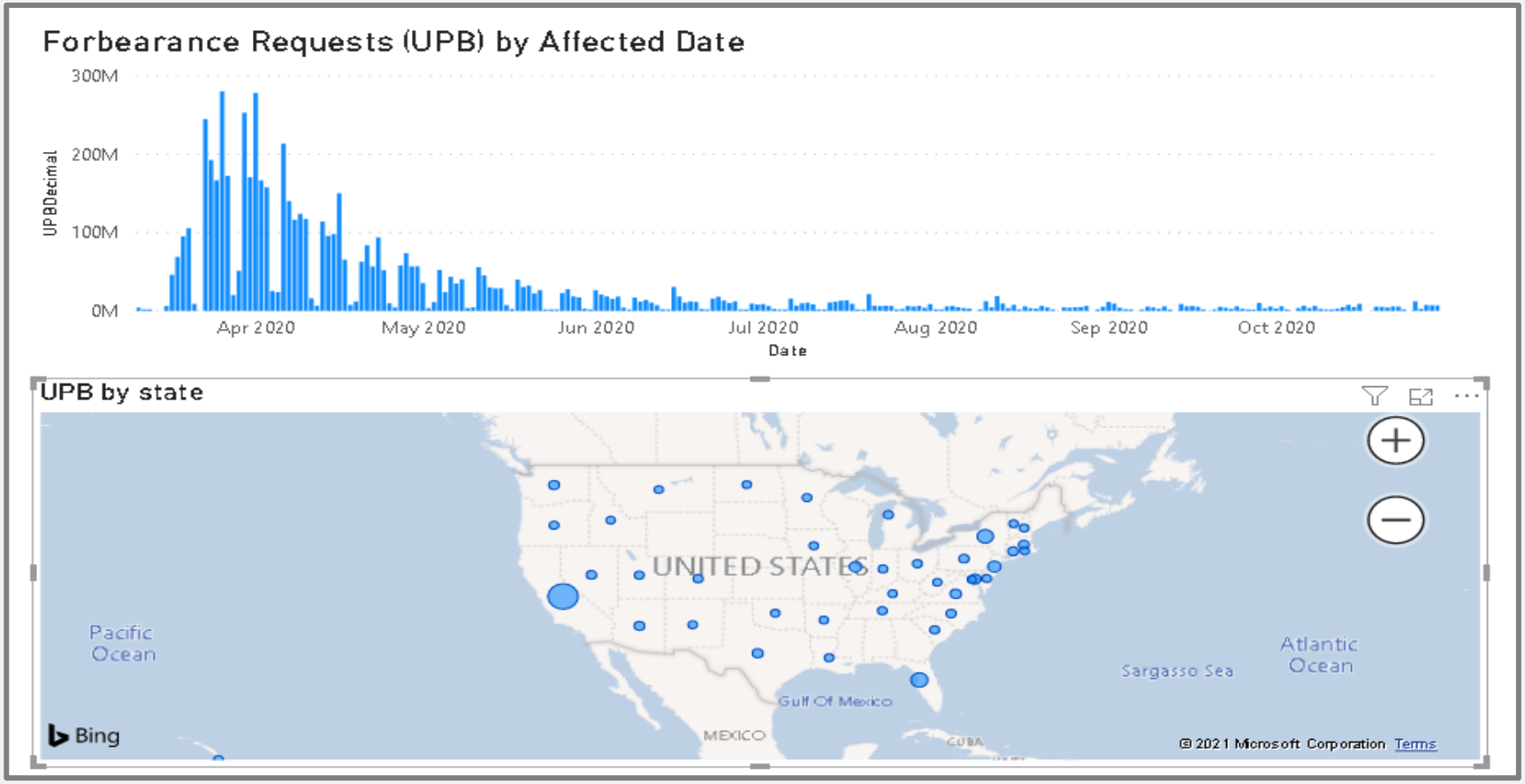


Covid-19 exposure and risk analysis: Implementation Approaches





Covid-19 exposure and risk analysis: Implementation Outcome



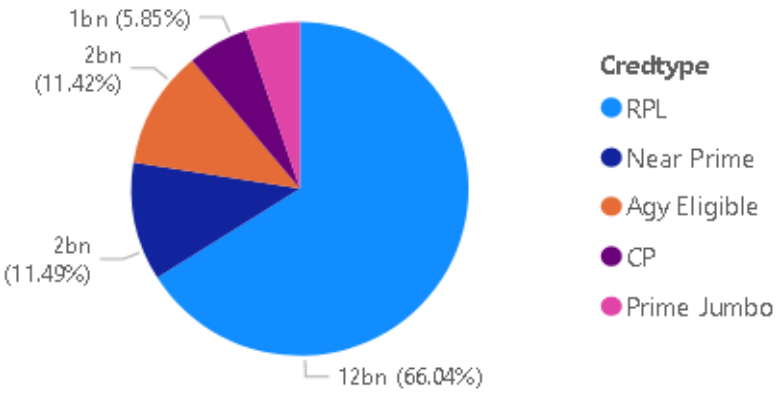
- ✓ Overall process takes **15 minutes** to complete comparing to 4 hours manual effort
- ✓ **No manual intervention** needed to clean up and transform data

Covid-19 exposure and risk analysis: Implementation Outcome

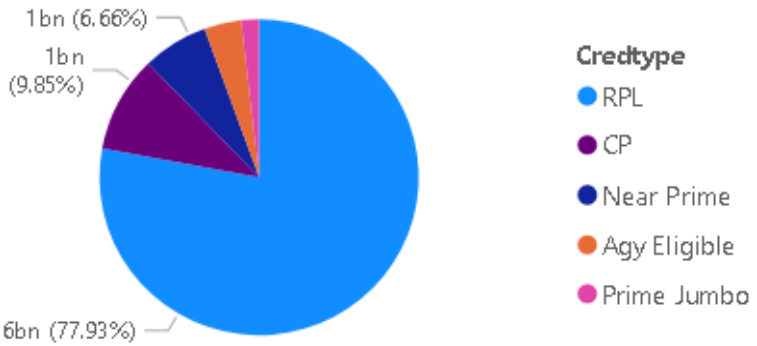
% UPB per state by status

state	Forbearance	Performing
AK	49.35%	50.65%
AR	45.68%	54.32%
AZ	34.20%	65.80%
CA	38.79%	61.21%
CO	22.40%	77.60%
CT	49.62%	50.38%
DC	40.09%	59.91%
DE	40.69%	59.31%
FL	48.46%	51.54%
HI	47.91%	52.09%
IA	45.47%	54.53%
ID	28.61%	71.39%
IL	44.01%	55.99%
IN	43.62%	56.38%
KY	37.72%	62.28%
LA	45.14%	54.86%
MA	39.91%	60.09%
MD	48.99%	51.01%
MI	39.43%	60.57%
MN	34.16%	65.84%
MT	30.07%	69.93%
NC	39.36%	60.64%
Total	**41.86%**	**58.14%**

TOTUPB by Credtype for all the loans



TOTUPB by Credtype for loans in forbearance



UPB (Unpaid Balance) details

Technology Comparison

Feature	Implementation 1 ADF+ Databricks	Implementation 2 Synapse Analytics
Single workspace for Ingestion/analytics and reporting	No	Yes
SQL on demand pool	No	Yes
GIT Integration	Yes	Yes
CI/CD Support	Yes	Yes
Data warehousing capability	No	Yes
Access Control	Managed through - Azure roles	Managed through - Azure roles , Synapse roles, SQL roles and Git permissions
Use case completion	Successful	Successful
Performance of use case	Comparable	Comparable

Feature	Implementation 1 ADF+ Databricks	Implementation 2 Synapse Analytics
Data ingestion	ADF pipeline	Synapse pipeline
Transformation	Databricks notebook	Synapse Notebook
SPARK version	3.0	2.4
Language support	Python, R, Spark SQL, and Scala	Python, Scala, .NET Core, Java, R, Spark SQL
Magic commands	Fully supported in Databricks	Few not available
Power BI embedded	No	Yes
Embedded charts	Yes	Yes
Support for Integration with Azure Machine learning Workspace	Yes	Yes

Technology Recommendation

- It is recommended to use Synapse workspace for the Research platform: a unified workspace for data preparation, data management, data warehousing and big data analytics, reporting and machine learning, considering support for .NET core, on demand SQL pool which allows complete SQL features and extensive RBAC.

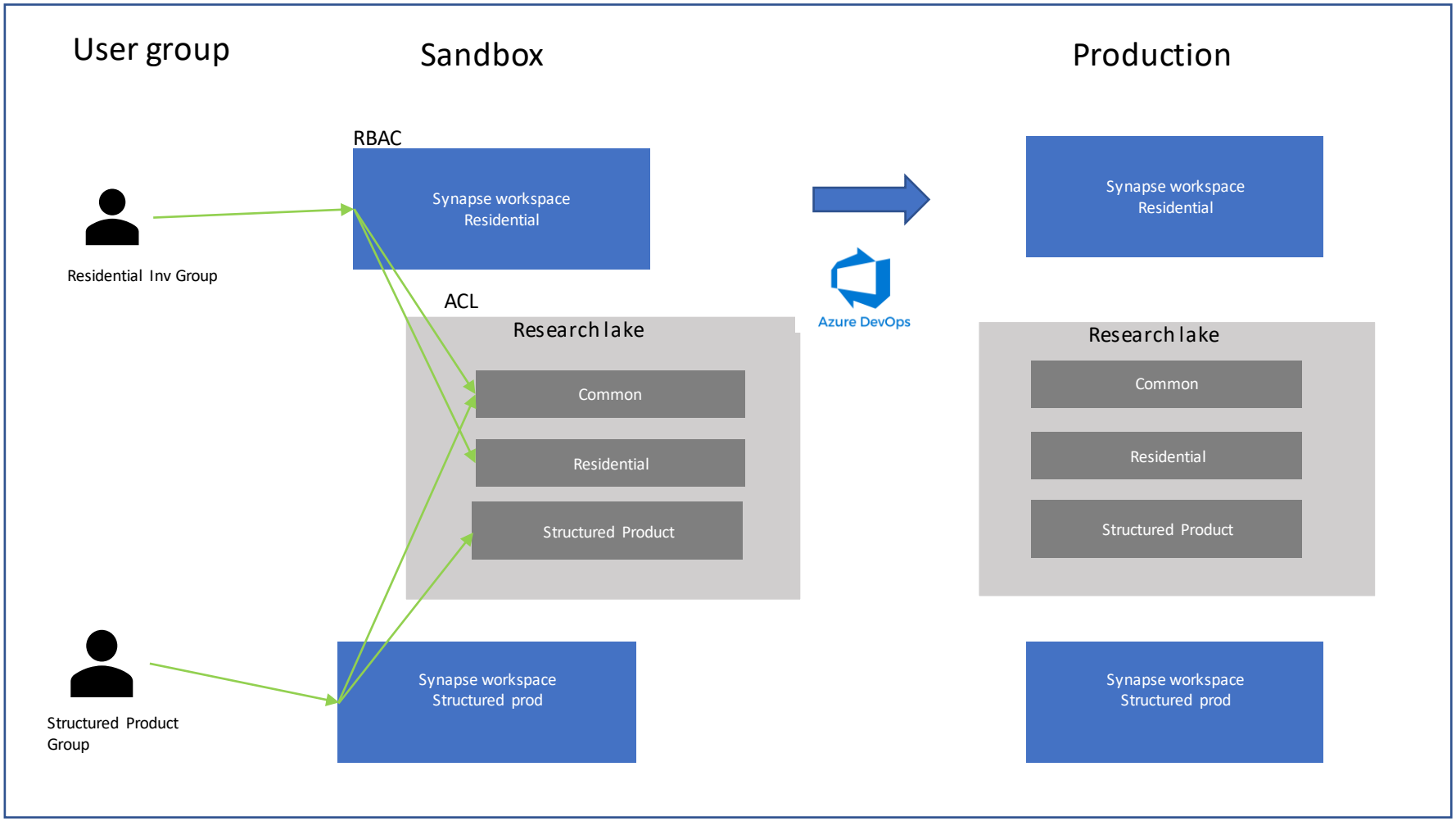
- Databricks will co-exist in RAPID. No additional cost due to pay per use, may be considered in future use cases as certain Databrick library not available in Synapse yet. e.g. DBUtil Library.

- Synapse workspace has GIT integration as well as CI/CD support with Azure DevOps. A DevOps strategy must be defined which provides flexibility to researchers as well as enforce governance.

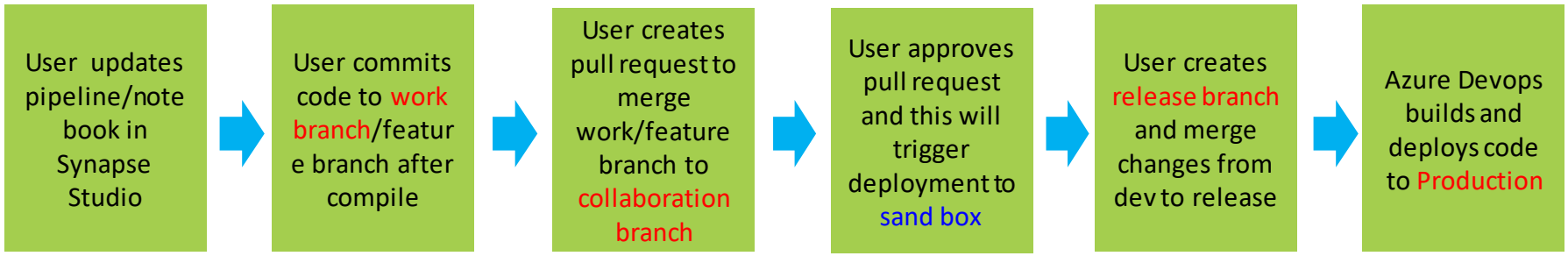
Infrastructure Cost

2020	2021
$ 20,894*	**$86,450**
• December 2020 FY Run Rate. The total RAPID December cost multiplied by 12 months. Dec cost $1,741 • **2020 actual FY cost: $4,056** • One POC Environment for 3 Month	• Planned 2021 FY spend for 2 environments • 2021 RAPID Growth Room: 207%

Entitlement and Deployment 1/2



Entitlement and Deployment 2/2



3

Recommendations and Next Steps

Recommended Technology Stack for Initial Rollout of RAPID

Platform Foundation	• Azure Data Lake (Storage Accounts. Block Blob Storage) • Self Hosted integration runtime (IR) Windows Virtual Machine • *Azure SQL Database
ETL and Analytics Tools	• Data Factory: Data Injection. Pipeline Service and transformation • Databricks / SPARK: Data Engineering (ETL, Analytical, and ML) • Synapse Analytics Workspace: Pipeline. Data Querying from IRDL and ecosystems. ETL, Analytics, Machine Learning • Data Visualization: Power BI • *Machine Learning and NLP Package: R, Python, and SAS, • *Financial Modeling Packages: Fincad, MetLab, Loan Models.
Environments	Two Environments: Sand Box and PROD

* Not included in POC use case. Required for initial rollout for subsequent use cases.
** Only MetLife Technology Stack approved tools are allowed to access RAPID

Next Steps

- GARB and IT Risk Review for moving on to IMPT phase
- Train business initially focusing on Resi team. Business to gain hands on experience
- Train Inv IT
- Obtain and implement Business and Technology Feedbacks
- Implement additional 'POC' user cases across LOB
- Define support model. Move from POC mode to Initial Rollout of RAPID
- Engage Data Scientist for model migrating
- Engage DevOps

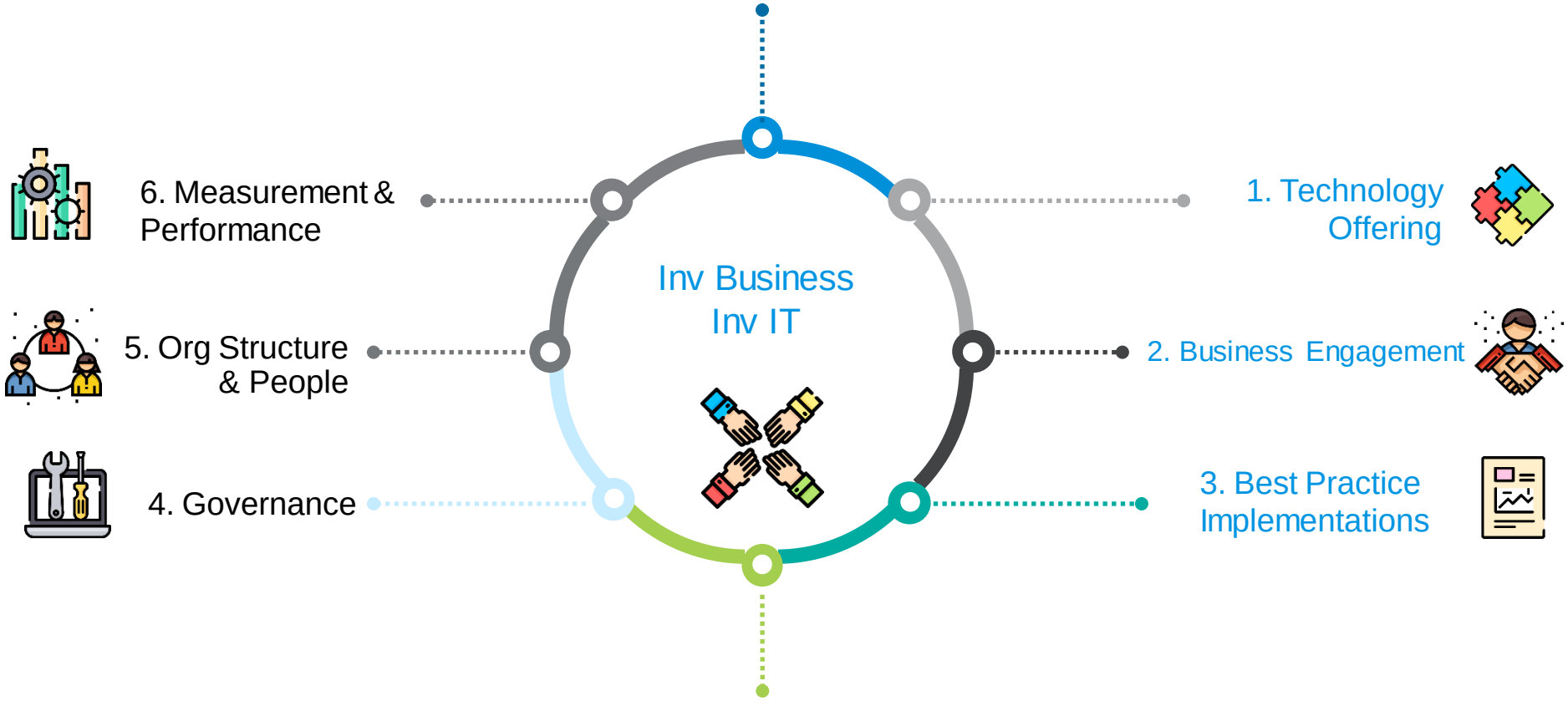
Appendix

Thoughts on Data Governance and Policy Guidelines*

- Data Quality – Responsibility of data owner
- Data cataloging and Metadata definition and maintenance - Responsibility of data owner
  - Technical: source, structure
  - Operational: lineage, quality, profile, version, accuracy
  - Business: usage, distribution
- Defines Data cardinality: **how it relates to other data**
- Data Owner: Individual analyst owns private data, group owner owns group data, Data Gov team owns common data.
- Raw data may not be normalized.
- **Research code management** may follow different path as standardized DevOps Process. Researcher owns their own code on Azure/TFS. Research code promotions might follow a different pipeline.
- Data backup: Daily backup. Special needs might require more frequent backup
- Computing Power: Expandable on demand
- Interaction with Operational Data Lake (GRID): Leverage data in GRID.
  - Develop interface with GRID
  - Retrieve GRID data into research lake to be combined with research data for analytics
  - Where applicable, develop a process to productionize research data into GRID
- Define usage thresholds for budgetary management
- Implement data retention, data refresh, and disposition policy

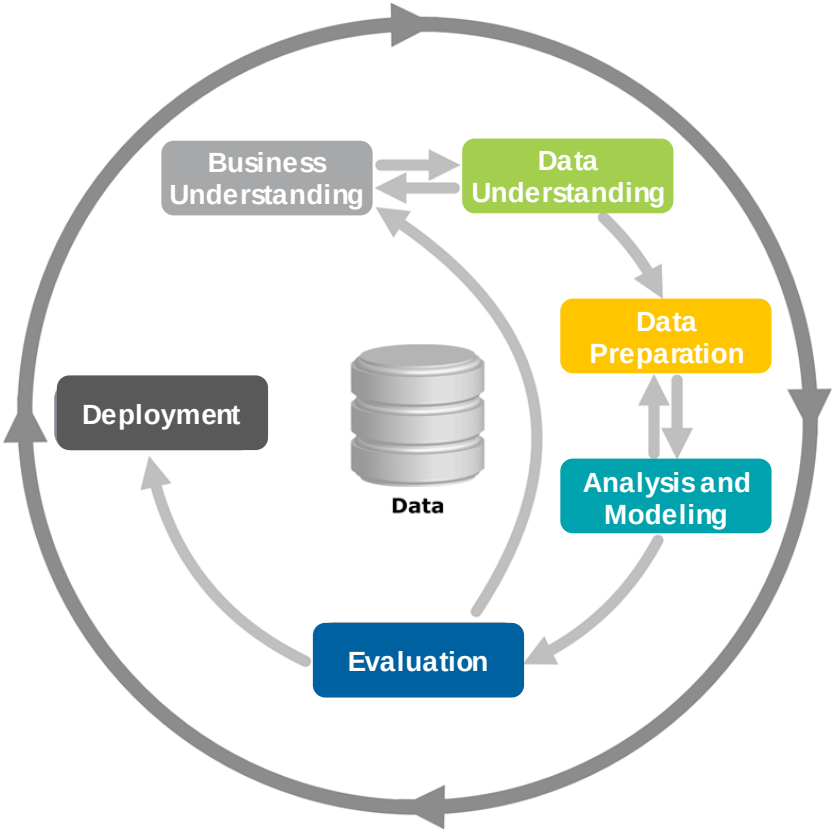
* Needs to be reviewed by Investments' Data Governance Office

Investments Research Platform Operating Model



Critical Success Factors –
Strong Partnership and Technology Enablement

Implementation Phases



- **Business Understanding** – project objectives and requirements;
- **Data Understanding** – data collection & understanding; define data quality
- **Data Preparation** – attribute selection, transformation and cleaning
- **Analysis and Modeling** – selecting and applying modeling techniques (i.e. Financial Engineering, Nearest Neighbors, Predictive Modeling, etc)
- **Evaluation** – review steps executed to construct the model; ensure achieves business objective
- **Deployment** – report generation and/or production deployment

Public Structured Finance Analytics Team CLO Use Cases:

Scope Description

Use case A: Cashflow
1. Load Intex cashflow to the platform
2. Load holding data from Aladdin to the platform
3. Calculate the discounted cashflow outside platform using the data from the platform

Use case B: Cashflow aggregation
1. Load Intex cashflow to the platform
2. Load holding data from Aladdin to the platform
3. Aggregate the data in the platform and provide aggregated data as output for processing outside platform

Technology Requirement

- Must compatible with SQL Server
- Must compatible or easily transferable with spreadsheet macro or VB
- Code management should be from the SFA repo.
- Data should be shareable within the group

Public Structured Finance RMBS Use Case:

Scope Description

Use case A: RMBS New Issues

1. Inject new RMBS deals from excel into SQL Server
2. Data enrichment and feature engineering in SQL
3. Run whole loan model
4. Export the model result into excel for reporting

Use case B: Existing RMBS Monthly Remittance

1. Load monthly remittance from Intex into SQL database
2. Data enrichment and feature engineering in SQL
3. Run whole loan model
4. Export the model result into excel for reporting – the output of the whole loan model is summarized into a single line

Technology Requirements

- Must compatible with SQL Server
- Must compatible or easily transferable with spreadsheet macro or VB
- Performance is key. Be able to index on both row and column at query time

Data Size: 1TB – 4TB

- RMBS: 600 GB - 3 TB
- CMBS: 100 GB - 200 GB
- ABS: 100 GB - 200 GB

Growth rates 200 GB per year

Public Structured Finance CMBS Use Case:

Scope Description

Use case A: Same as RMBS Use Case
Use case B: Data Visualization. Needed in Oct and Nov 2020
Use case C: Advanced Analytics and Machine Learning
1. Sentiment Analysis for Servicer Comments using NLP

Technology Requirements

- Must compatible with SQL Server
- Must compatible or easily transferable with spreadsheet macro or VB
- Performance is key. Be able to index on both row and column at query time
- NLP Tools
- Visualization Tools: e.g. Power BI

Data Size: Included in RMBS Use Case

Azure PaaS RACI

Azure PaaS RACI Chart

Functions	Task	Product Owner	AD Teams	Application Management Team	ITRS & Compliance	Architecture Team	- Investments IT Eng. / Env Team	Cloud Ops Team	Microsoft	Finance	Investments IT Management	Comments
Application Development	Requirement Analysis	A	R	C		R					I	
	Functional Design	A	R	C		C					I	
	Technical Design	A	R	C		C					I	
	Development	A	R	C		I					I	
	Build & Support Continuous Integration & Testing for Azure applications	A	R	C	A	C					I	
Application Testing	Test Environment Availability	A	I	C		I					I	
	Unit Testing	A	I	C		I					I	
	System Testing	A	I	C		I					I	
	UAT Testing	A	I	C		I					I	
	Deployment	A	I	R		I					I	
Application Management	UAT Monitoring	A	I	R		I					I	
	Production Parallel Support	A	I	R		I					I	
	Production Support	C	I	A,R		I	R	R	C		I	
	Platform issues / Service Outage	I	I	A		I	R		C		I	
	Incident and Problem Management	A	I	R		I					I	
	Define azure service changes and processes for release management	C	I	I		C	A,R		R		I	
	Azure Application Service Monitoring	A	I	R		I	A		R		I	
Infrastructure /Platform Support	Azure environment provisioning & Decommissioning	C	I	C	I	I	R	R	C	I	I	
	Azure Security, Monitoring and Configuration	I	I	I	A,R	I	A,R	R*	C	I	I	***1**
	Azure Network/Environment Security (vnet)	C	I	C	A,R	I	A,R	R	C	I	I	
	Express Route		I		A		A,R	R	AT&T	I	I	
	NSG (Network Security Group)		I		A		A,R	R	C	I	I	
	Azure Firewall		I		A		A,R	R	C	I	I	
	Azure Infrastructure (Including Virtual Machines) & Service Monitoring (CloudOps/BizCloud	I	I	C,I	A,C	I	A	R*	C	I	I	*2
	Security Monitoring	I	I	I	A,C	I	A	R*	C	I	I	
	Co-ordination with Microsoft to resolve Azure platform issues	I	I	C,I		I	A	R	R	I	I	
	Data BackUp and Restore	I	I	C,I		I	A	R	R	I	I	
	DR Testing	I	I	A,R	C,I	I	R	R	C	I	I	
Azure PaaS Service	CosmosDB	I	I	I	C,I		R		A,R	I	I	
	DataBricks	I	I	I	C,I		R		A,R	I	I	
	ADLS	I	I	I	C,I		R		A,R	I	I	
	Service Bus	I	I	I	C,I		R		A,R	I	I	
	Event Hub	I	I	I	C,I	C	R		A,R	I	I	
	Azure Data Factory	I	I	I	C,I		R		A,R	I	I	
	SQL Datawarehouse	I	I	I	C,I		R		A,R	I	I	
	ASE	I	I	I	C,I		R		A,R	I	I	
	SQL DB	I	I	I	C,I		R		A,R	I	I	
	Azure Function* (Applied to PaaS stack)	I	I	I	C,I		R		A,R	I	I	
	Web Apps	I	I	I	C,I		R		A,R	I	I	
	App Gateway	I	I	I	C,I		R		A,R	I	I	
Status & Metrics Reporting	Tracking progress of tasks	A	I	C		I	A,R				I	
	Identify and communicate Issues, Risk and dependencies related to tasks	A	I	C		I	A,R				I	
	Tracking the overall project plan and providing appropriate status	A,R					C				I	

Other Investments Use Cases:

Research Data Lakes Use Case:	Description	Key Feasures	Business
A strategic solution for Structured Finance Public Sector Research and Modeling Capabilities	**Data Sources:** Structured Finance Publis Sector: CMBS, RMBS, ABS, and CLOs. Load research data from various data sources with various formats including holdings, collateral, bond/tranche, ratings actions, and sector specific additional data. **Processes:** Data quality controls, check, and data engineering Data summarization Run analytical model to produce expected losses / coverages Summarize collateral strats using different attributes Calculated an automated metric that allows that Bonds to be classified as Performing normal, performing outlier – bad, performing outlier good. Credit designation. **Outputs:** Flagging of outlier bonds for review Summary Reports indicating/depicting the healthy state of the non-outliers Trends being observed in underlying collateral – period over period (monthly/quarterly/semi-annual/annual) Presentation/Generation of pre-canned reports for distribution	- Automated Data Retrival from various sources - Data cracker and data engineering - Run Analytics Model - Reporting and visualization	Harish Yerapathi
Resi - Covid-19 Related Data for exposure and risk analysis	**Daily data from servicers, each in different format to performance s**tructured as well as ad hoc data analysis for resi portfolio risk/exposure. Example data fields: **Called In**: This field represents a borrower that has called in(**or ask**ed for help by other channel), some borrowers call in and eventually do not elect forbearance. **Affected:** Whether the loan is in Forbearance (FB) **As of Date:** When the data is up to and including. Each servicer reports data at different periods. At this stage only SPS and SHP are reporting daily **Next Due Date**: When the next payment of the loan is due **Last Pay:** When the borrower made their last payment. **Prior ACH**: Whether the borrower had an ACH prior to asking for FB **Current ACH:** Whether the is still currently on ACH **Channel:** What channel the borrower used to request FB	- Data clean up and normalization - Large Data Volumn - Data Analytics - Data Visualization	Colin Robina / Charles Donninger
Single Family Rental (SFR) – Resi	Consolidate quarterly Single Family Rental (SFR) loan level data in excel, build timeseries reports for exposure analysis	- Timeseries report -Flexibility -Simplicity -Business Self Managed	Colin Robina/Grafas, Stephen
Chile Resi Loan	Load Chile Resi loan data extracted from local system to Research Data Lakes (RDL), enabling US resi research team to analyze the data	- Data Integration - Data Analytics - ETL - Run Resi Model	Colin Robina/Razeen
Agency loan level data from Fannie Mae and Freddie Mac	Download Single Family Loan Level Dataset from Fannie Mae and Freddie Mac. Analysis data ready for user query	- Data reporting and aggregation	Jason Valentino
Sec company 10K and 10Q	Download IPO Company Financial Report from public domain SEC Securities and Exchange Commission (Domastic/International) Company Financial Data to be utilized for credit research	- Automatic data retrieve from Public Site - ETL - Data Analytics and Reporting	Alternative Source for CapIQ
BRS CLO Data	Consolidate and aggregate monthly CLO Deal from Aladdin and display in Power BI for reporting and visualization	- ETL - Data Analytics and Visualization	Sean Lyng
Prosper Loan Listing and Mothnly data	Load Prosper Loan Listing data and monthly data for data analysis - ad hoc need to explore investment opportunities (Monthly Data 16 GB, Listing 4GB)	- ETL - Undefined data format - Run Model	Colin Robina /Wang, Charles
ABS- Auto Loan:	Daily/Weekly download from SEC site for new and existing ABS with underlying loans. For Data analysis and valuation. Estimation 254GB per year, it may grow up to 1TB in 5 years. Auto loan duration usually finished in 5 years	- Auto retrieve data structured data from public site. - Data Analytics - Visualization	Loritta Cheng

Azure Component Primer

COMPONENT	CAPABILITY
Azure Data Lake	Azure Data Lake store is a hyper-scale repository for big data analytics workloads and a Hadoop Distributed File System (HDFS) for the cloud. Allows unstructured and structured data in their native formats.
Azure SQL Data Warehouse	SQL Data Warehouse is a cloud-based Enterprise Data Warehouse (EDW) that uses Massively Parallel Processing (MPP) to quickly run complex queries across petabytes of data. Use SQL Data Warehouse as a key component of a big data solution.
Azure HD Insights HIVE	Hive allows you to project structure on largely unstructured data.
Azure Data Factory	The Azure Data Factory (ADF) is a service that enables developers to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation across disparate sources. It is a platform somewhat like SSIS in the cloud to manage the data you have both on-prem and in the cloud.
ADF + Azure Function	Azure Functions is a serverless compute service that enables you to run code on-demand without having to explicitly provision or manage infrastructure. Use Azure Functions to run a script or piece of code in response to a variety of events.
ADF + SSIS	Microsoft SSIS (SQL Server Integration Services) is an enterprise data integration, data transformation and data migration tool built into Microsoft's SQL Server database
Application Service Environment (ASE)	An App Service Environment is a fully isolated and dedicated environment for securely running Azure App Service apps at high scale, including Web Apps, Mobile Apps, and API Apps

Azure Component Primer (continued)

COMPONENT	CAPABILITY
Azure Service Bus	Microsoft Azure Service Bus is a fully managed enterprise integration message broker. Service Bus is most commonly used to decouple applications and services from each other, and is a reliable and secure platform for asynchronous data and state transfer.
CosmosDB	Microsoft's proprietary globally-distributed, multi-model database service "for managing data at planet-scale". It is schema-agnostic, horizontally scalable and generally classified as a NoSQL database
Azure Analysis Services	Azure Analysis Services is a fully managed platform as a service (PaaS) that provides enterprise-grade data models in the cloud. Use advanced mashup and modeling features to combine data from multiple data sources, define metrics, and secure your data in a single, trusted tabular semantic data model.
Power BI	Power BI is a business analytics solution that lets you visualize your data and share insights across your organization, or embed them in your app or website. Connect to hundreds of data sources and bring your data to life with live dashboards and reports.