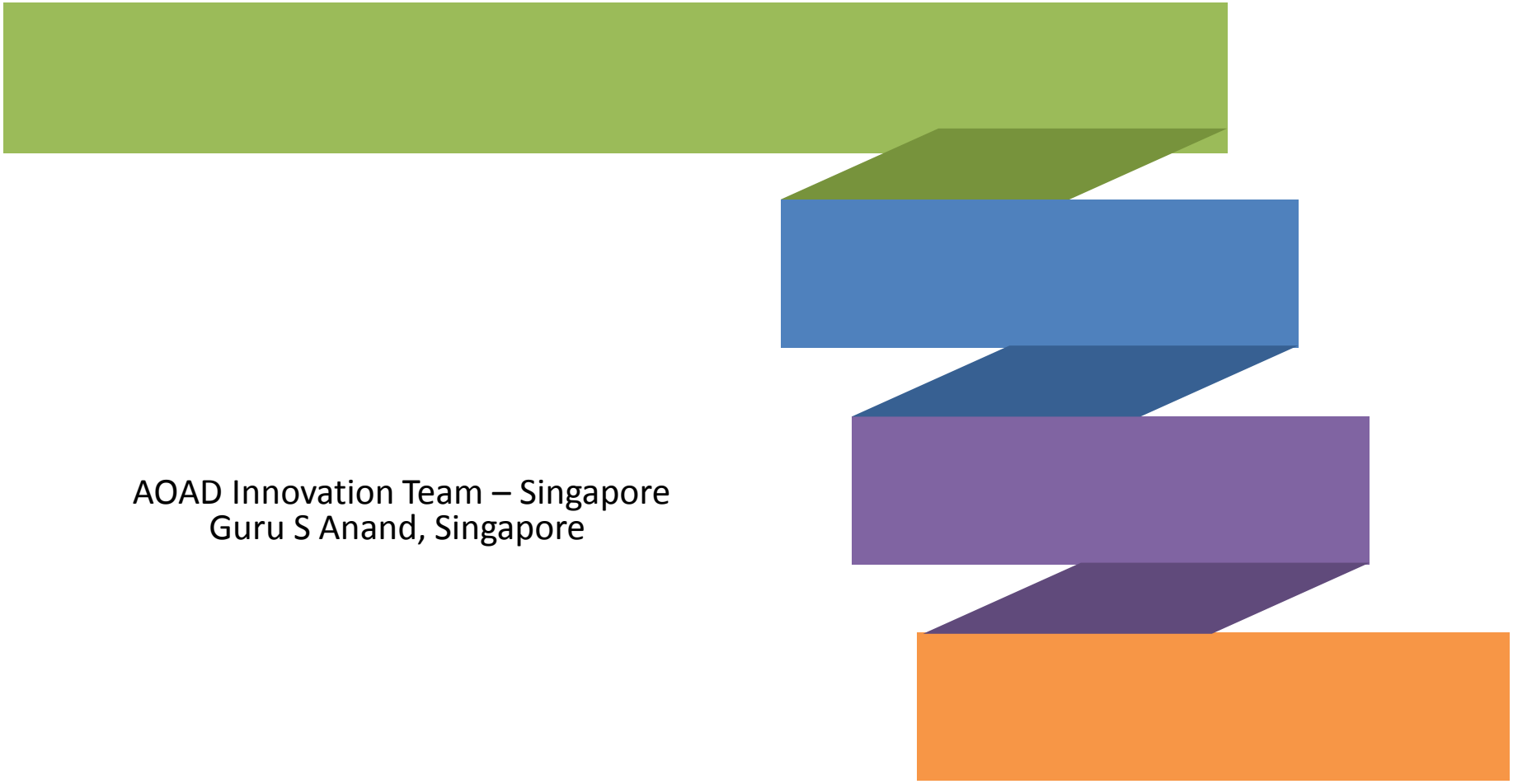
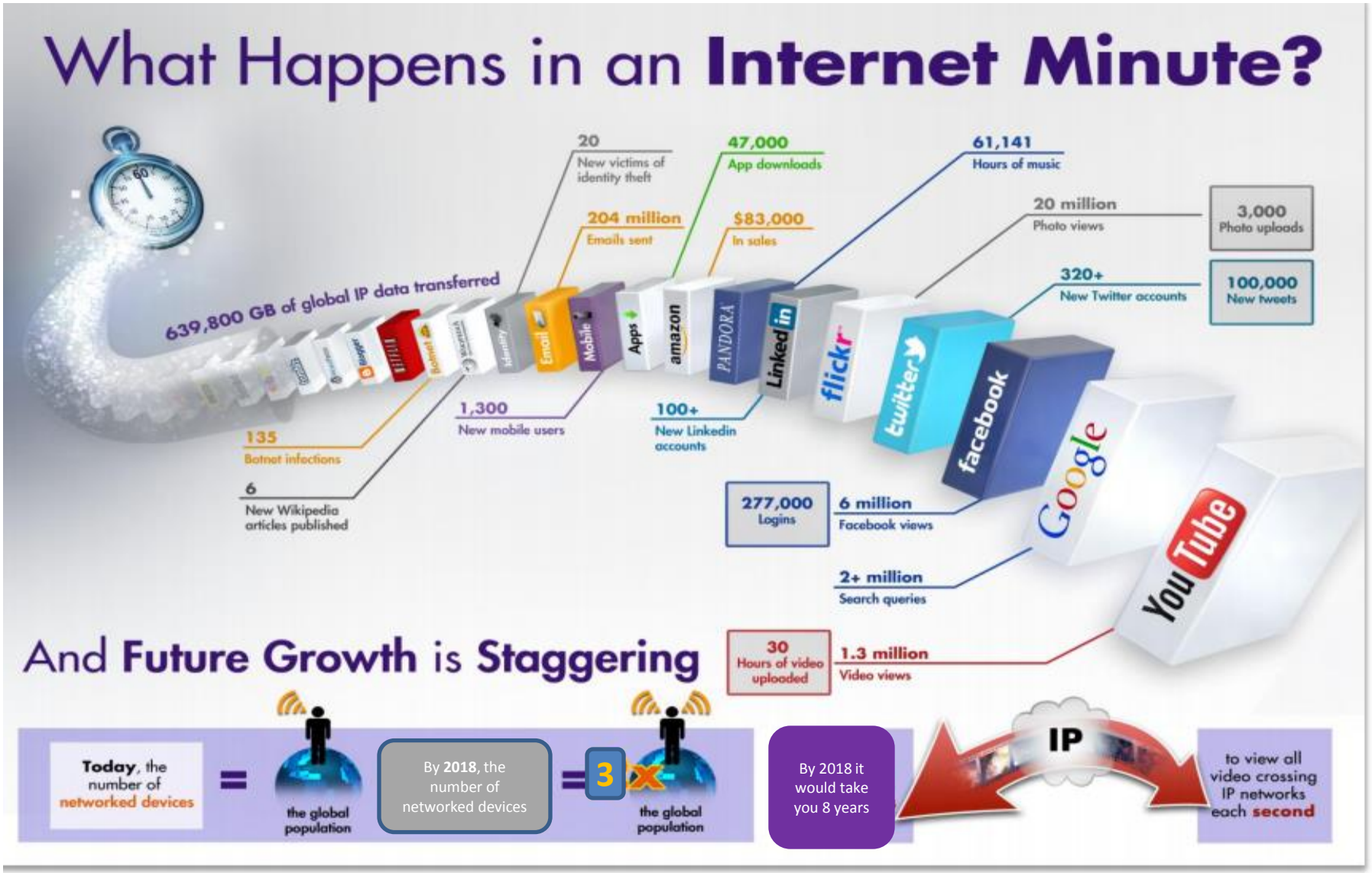
Mizuho OneView
BIGData/DataLake/Unify All Database using one platform
Forecasting/Predicting Customer Stickiness using Machine Learning
POC Results



AOAD Innovation Team – Singapore
Guru S Anand, Singapore

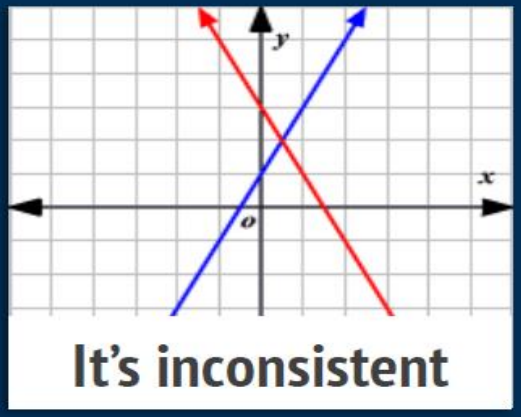
Internet Minute….



The Current Enterprise Problem With Accessing Data

The Current MIZUHO Problem With Accessing Data

Timely Access to Unified Data is a MASSIVE Hurdle to Everything!



Mizuho's current Problem

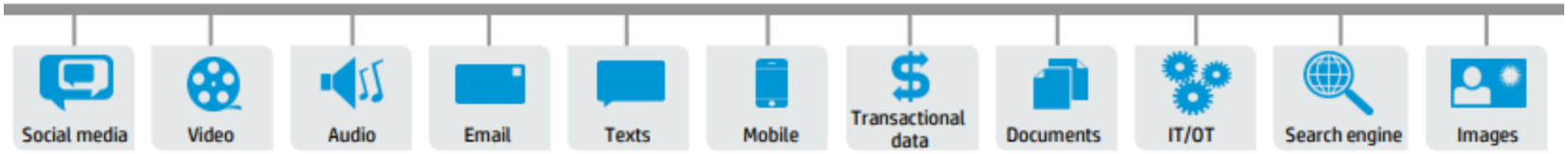
Digital Strategy Foundation - Objective

- BIGData/DataLake – Store All data for various Golden Sources. Structured, Unstrucutred, Real time data from Bloomberg, Reuters etc.
- One Platform – For Data Integration from various Golden Sources, DataLake and to consume for downstream applications, Open APIs

Unify using a single platform

OneView Mizuho

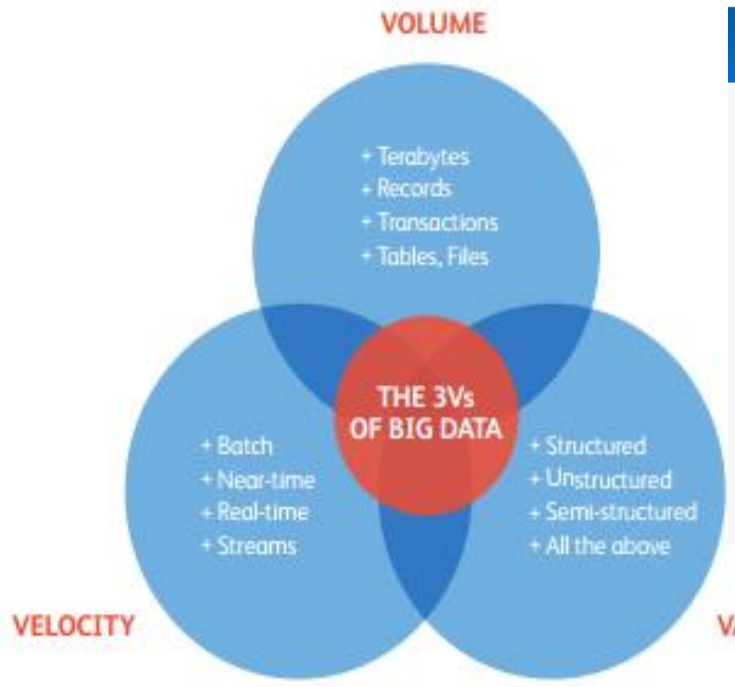
- Create Innovative applications for Forecasting/Predicting for Machine Learning Ready/ Artificial Intelligence Ready

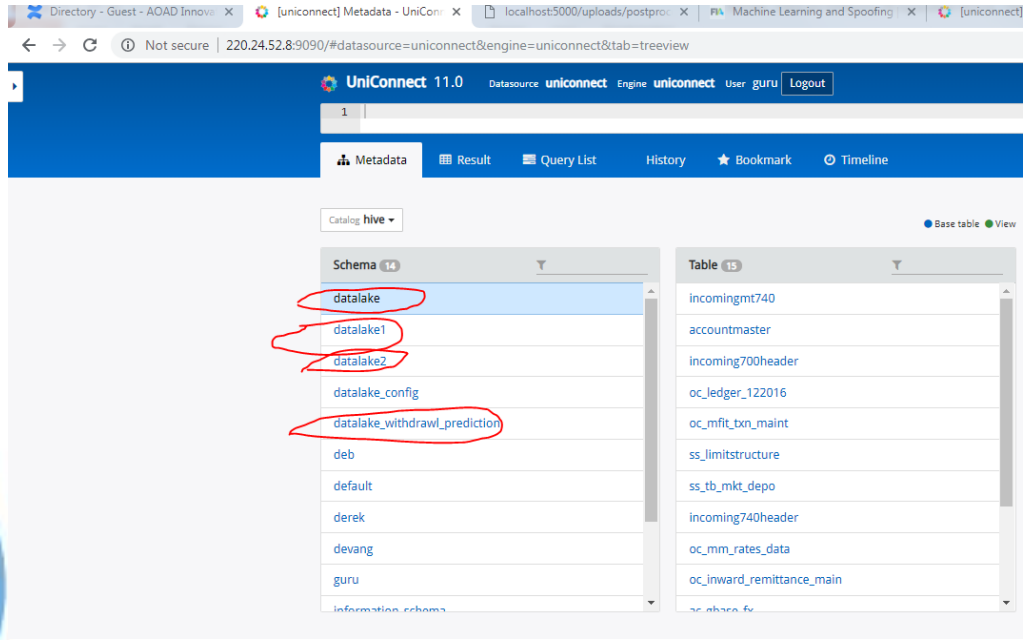


- Data Analytics/ Data Science Ready
- API Ready

DataLake Built in SG Mizuho - POC

BigData - 3 Vs

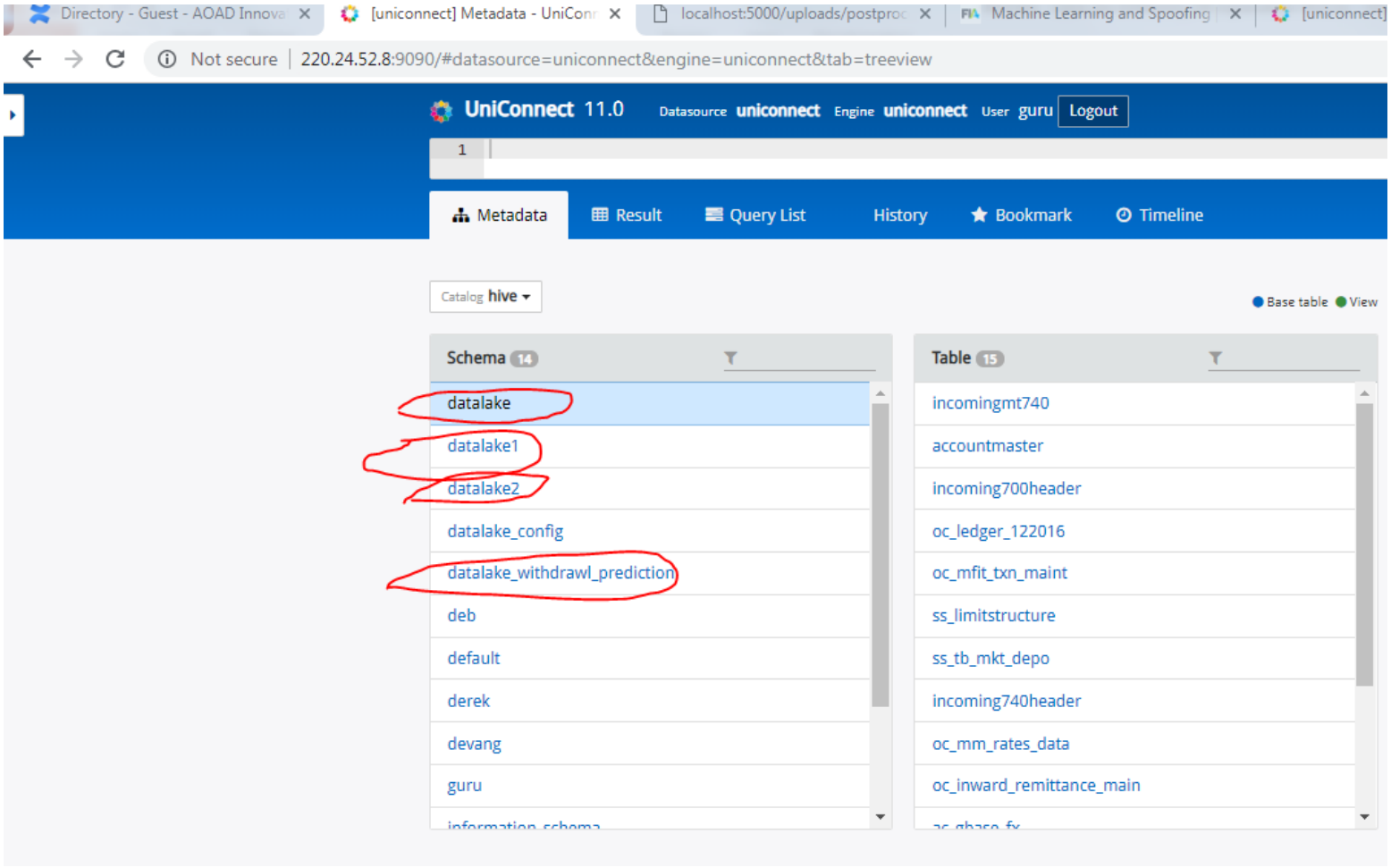




We created a DataLake using Hadoop infrastructure in SG Mizuho

http://220.24.52.8:9090

DataLake Built in SG Mizuho - POC



We created a DataLake using Hadoop infrastructure in SG Mizuho for to use by various department like GTBD, RMD, AOAD, MRDD

http://220.24.52.8:9090

Project Objective

Digital Strategy Foundation

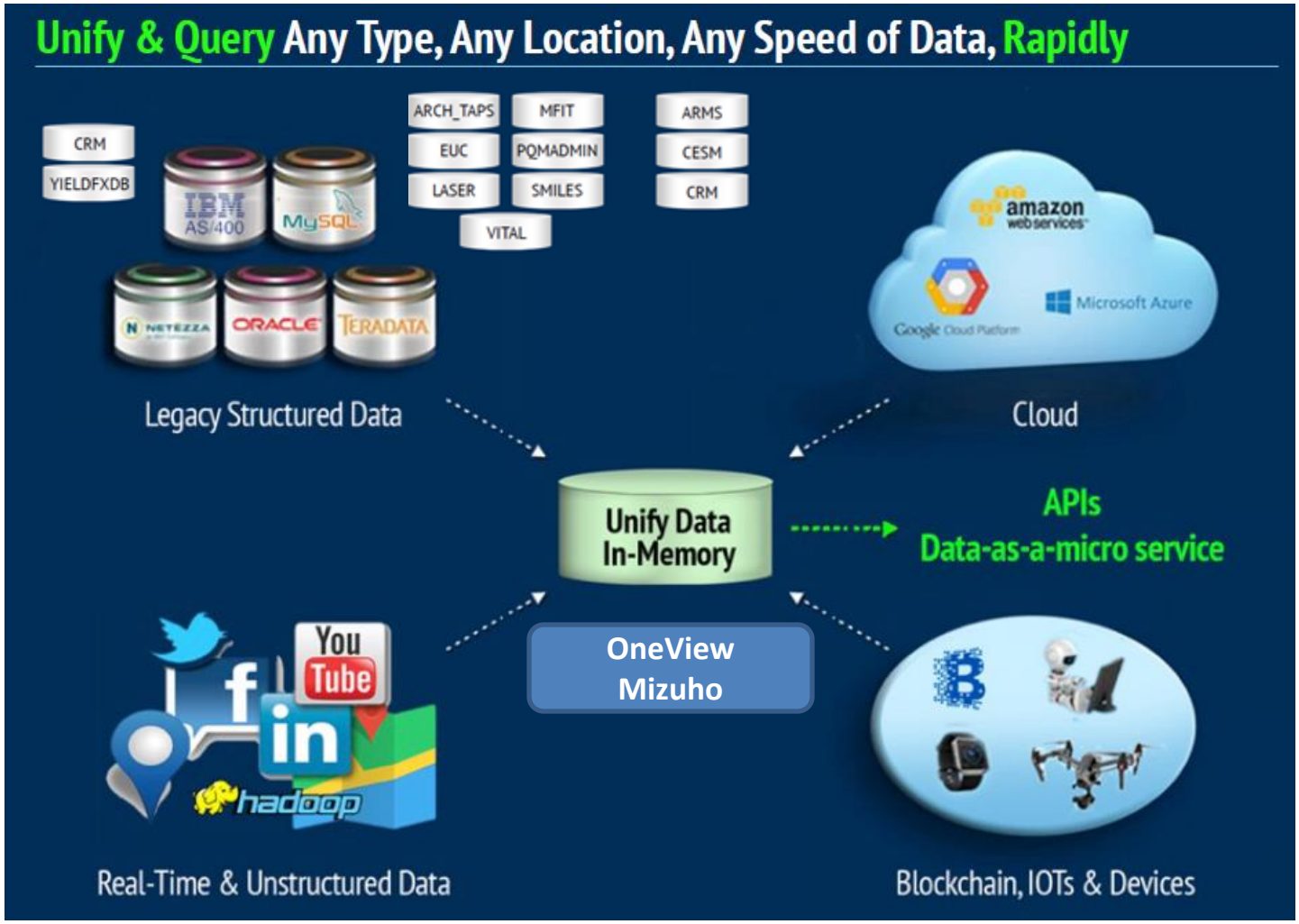
BIGData/DataLake – Store All data for various Golden Sources. Structured, Unstrucutred, Real time data from Bloomberg, Reuters etc.

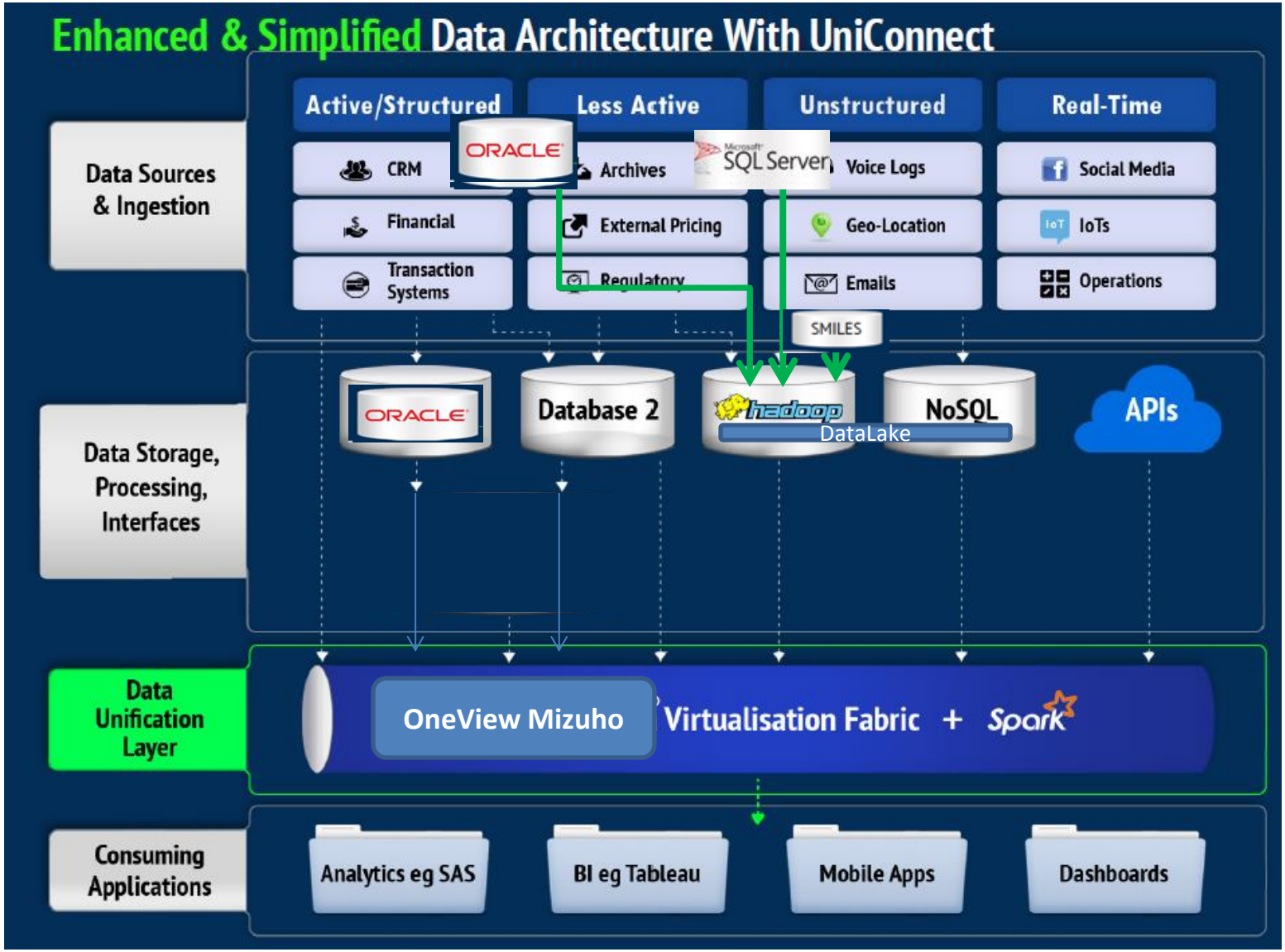
Result - GREEN

- One Platform – For Data Integration from various Golden Sources, DataLake and to consume for downstream applications, Open APIs

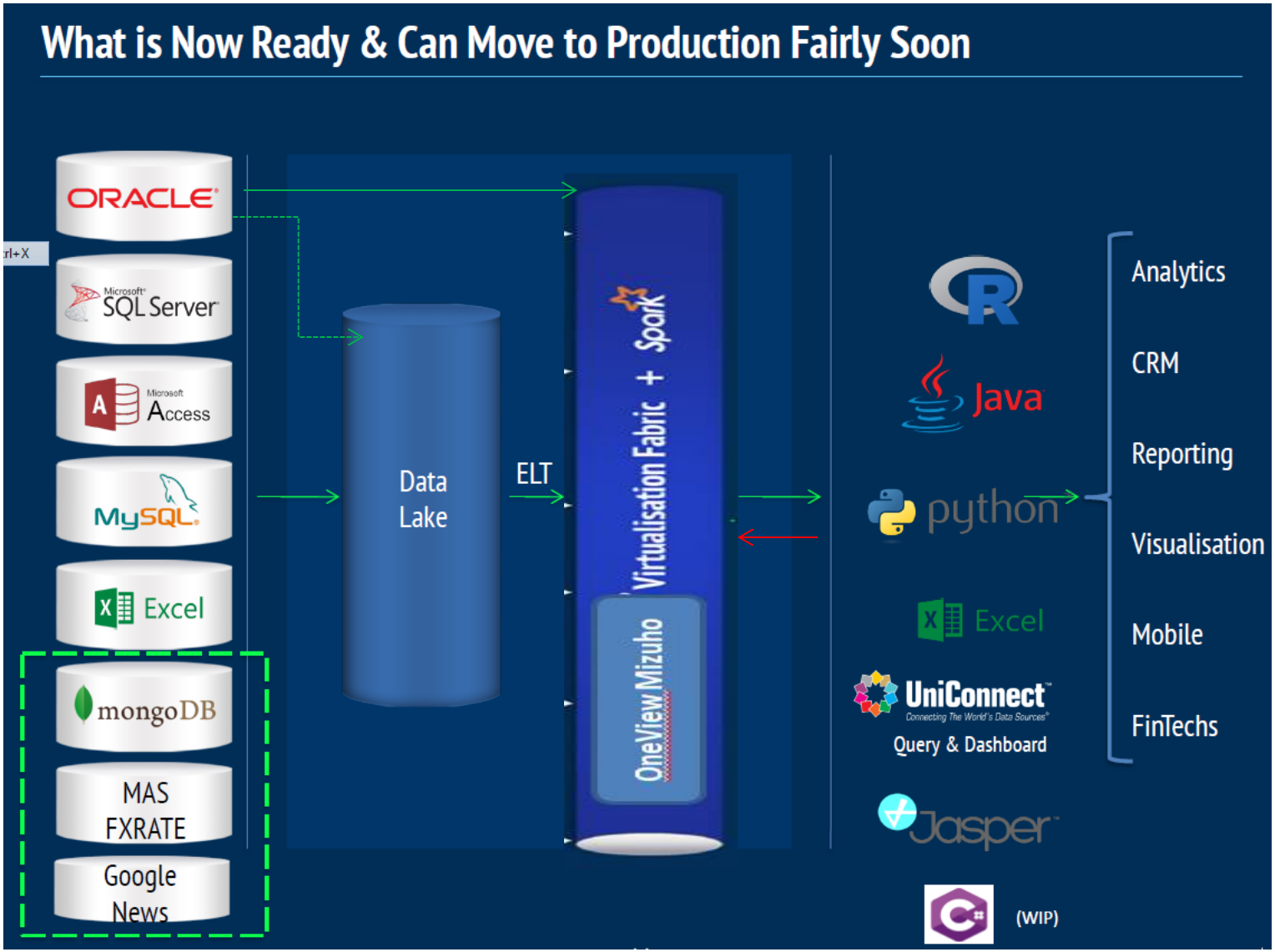
Unify using a single platform

OneView Mizuho


Unify & Query Any Type, Any Location, Any Speed of Data, Rapidly
CRM
YIELDFXDB
ARCH_TAPS
MFIT
EUC
PQMADMIN
LASER
SMILES
VITAL
ARMS
CESM
CRM
IBM AS/400
MySQL
NETEZZA
ORACLE
TERADATA
Legacy Structured Data
amazon web services
Google Cloud Platform
Microsoft Azure
Cloud
Unify Data In-Memory
APIs
Data-as-a-micro service
OneView Mizuho
You Tube
hadoop
Real-Time & Unstructured Data
Blockchain, IOTs & Devices


Enhanced & Simplified Data Architecture With UniConnect
Active/Structured
Less Active
Unstructured
Real-Time
Data Sources & Ingestion
ORACLE
SQL Server
CRM
Archives
Voice Logs
Social Media
Financial
External Pricing
Geo-Location
IoTs
Transaction Systems
Regulatory
Emails
Operations
SMILES
Data Storage, Processing, Interfaces
ORACLE
Database 2
hadoop
DataLake
NoSQL
APIs
Data Unification Layer
OneView Mizuho
Virtualisation Fabric + Spark
Consuming Applications
Analytics eg SAS
BI eg Tableau
Mobile Apps
Dashboards

What is Now Ready & Can Move to Production Fairly Soon



Digital Strategy Foundation

BIGData/DataLake – Store All data for various Golden Sources. Structured, Unstrucutred, Real time data from Bloomberg, Reuters etc.

Unify/In-Memory Database – For Data Integration from various Golden Sources, DataLake and to consume for downstream applications, Open APIs

Unify using a single platform

OneView Mizuho

Result - GREEN

- Forecasting/Predicting for Machine Learning Ready
- Data Analytics Ready

Systems	Version	Link
Operating System	[redacted] RHEL 7+ (64 bit)	NA
Apache Hadoop	2.7.4	http://www-eu.apache.org/dist/hadoop/common/hadoop-2.7.4/hadoop-2.7.4.tar.gz
Apache Hive	2.2.0	http://www-us.apache.org/dist/hive/hive-2.2.0/apache-hive-2.2.0-src.tar.gz
Apache Derby	10.14.1.0	http://www-eu.apache.org/dist//db/derby/db-derby-10.14.1.0/db-derby-10.14.1.0-bin.tar.gz
Apache Spark	2.2.0	http://www-us.apache.org/dist/spark/spark-2.2.0/spark-2.2.0-bin-hadoop2.7.tgz
Apache Kafka	2.11	http://www-eu.apache.org/dist/kafka/1.0.0/kafka_2.11-1.0.0.tgz
Apache Zookeeper	3.4.10	http://www-us.apache.org/dist/zookeeper/zookeeper-3.4.10/zookeeper-3.4.10.tar.gz
Scala	2.11.12	https://downloads.lightbend.com/scala/2.11.12/scala-2.11.12.tgz
Oracle JDK (Recommended Latest one)	JDK1.8.0_151+	http://download.oracle.com/otn-pub/java/jdk/8u151-b12/e758a0de34e24606bca991d704f6dcbf/jdk-8u151-linux-x64.tar.gz
Python	2.7.11 +	https://www.python.org/ftp/python/2.7.11/Python-2.7.11.tgz

Apache

Systems	Version	Link
Apache Hadoop	2.7.4	http://www-eu.apache.org/dist/hadoop/common/hadoop-2.7.4/hadoop-2.7.4.tar.gz
Apache Hive	2.2.0	http://www-us.apache.org/dist/hive/hive-2.2.0/apache-hive-2.2.0-src.tar.gz
Apache Derby	10.14.1.0	http://www-eu.apache.org/dist//db/derby/db-derby-10.14.1.0/db-derby-10.14.1.0-bin.tar.gz
Apache Spark	2.2.0	http://www-us.apache.org/dist/spark/spark-2.2.0/spark-2.2.0-bin-hadoop2.7.tgz
Apache Kafka	2.11	http://www-eu.apache.org/dist/kafka/1.0.0/kafka_2.11-1.0.0.tgz
Apache Zookeeper	3.4.10	http://www-us.apache.org/dist/zookeeper/zookeeper-3.4.10/zookeeper-3.4.10.tar.gz

Server config

Department Wide - 10 TB			
Purpose		**Spec**	**Qty**
Uniconnect Processor	Master Node	DL380 gen10 2P Xeon-Silver 4110 (2.1GHz/8-core), 512GB RAM, 6x1.8TB SAS 10K HDD	1
Uniconnect Processor	Worker Node	DL380 gen10 2P Xeon-Silver 4110 (2.1GHz/8-core), 384GB RAM, 6x1.8TB SAS 10K HDD	4
Hive Datalake		Hadoop Servers, Network & Services	
Enterprise Wide - 50 TB			
Purpose		**Spec**	**Qty**
Uniconnect Processor	Master Node	DL380 gen10 2P Xeon-Silver 4110 (2.1GHz/8-core), 512GB RAM, 6x1.8TB SAS 10K HDD	1
Uniconnect Processor	Worker	DL380 gen10 2P Xeon-Silver 4110 (2.1GHz/8-core), 384GB RAM, 6x1.8TB SAS 10K HDD	10
Hive Datalake		Hadoop Servers, Network & Services	

POC Completed

http://220.24.52.8:9090
http://220.24.52.8:8080/ui
http://spws271:8090/display/AI/Google+News+using+API

Eg: Accessing Multiple DB in a single place

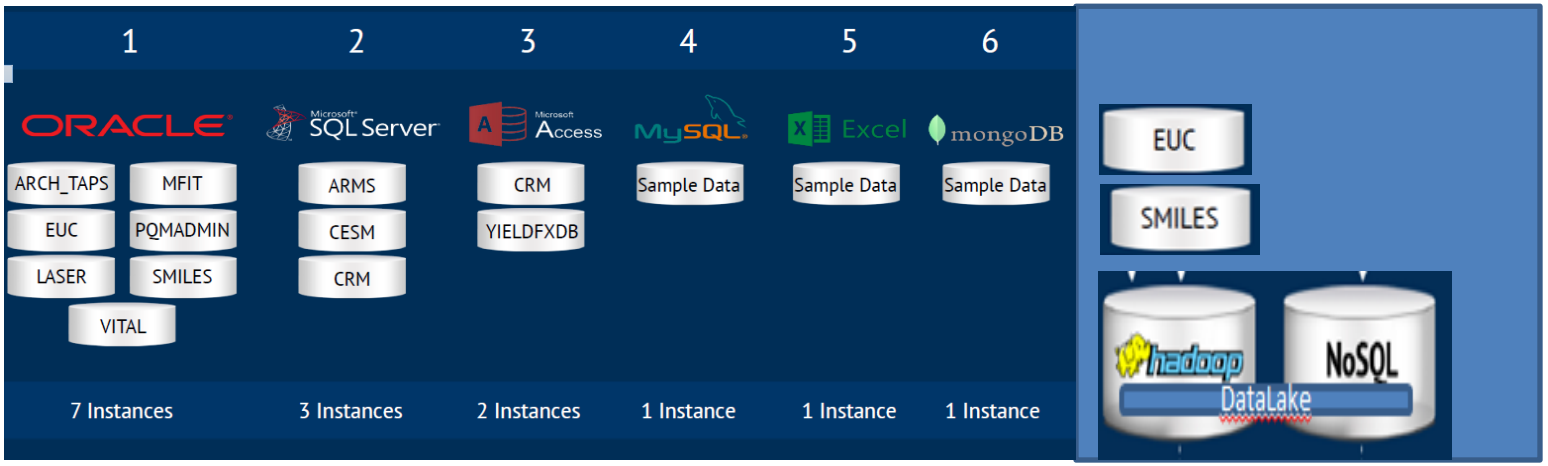
```
SELECT * FROM sqlserver_arms.dbo.tb_mkt_depo SQLSERVER_DEPO INNER JOIN (SELECT * FROM access_yieldfxdb.public.deposits)
ACCESS_DEPO ON SQLSERVER_DEPO.MKT_DATE=ACCESS_DEPO.MKTDATE AND SQLSERVER_DEPO.CCY=ACCESS_DEPO.CURR
AND SQLSERVER_DEPO.PERIOD=ACCESS_DEPO.PERIOD
```

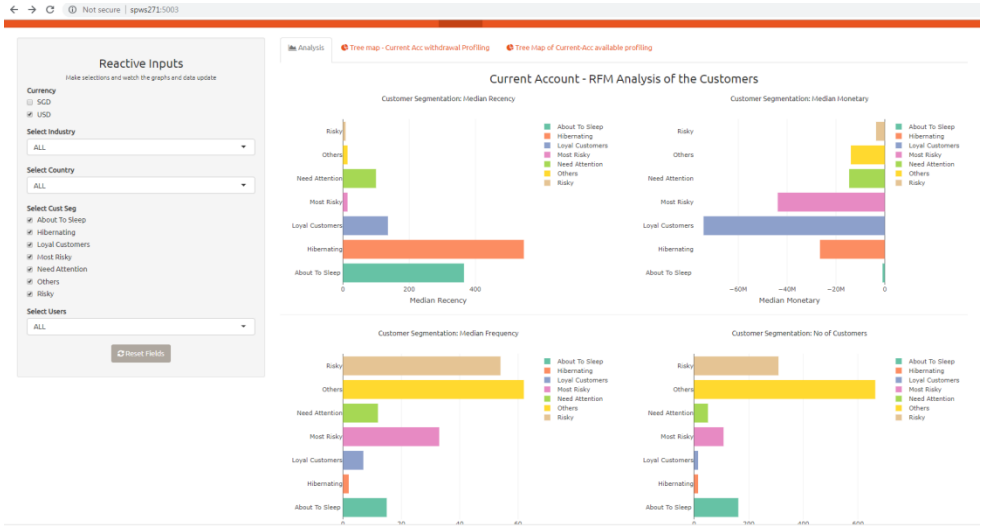
Result - GREEN

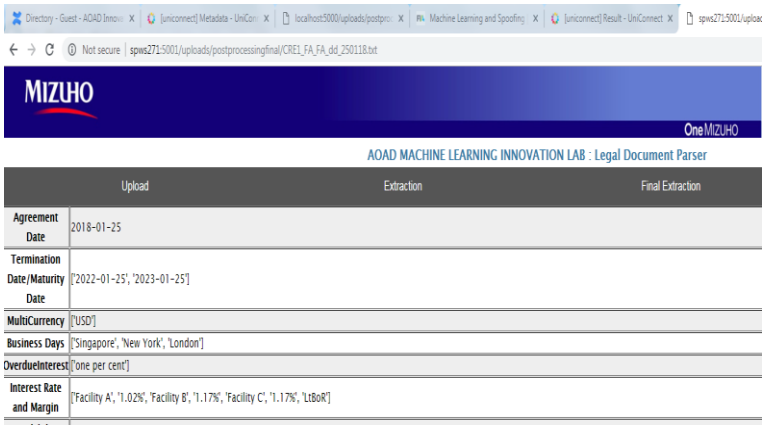
POC - Technical Results

S. No	Test cases - Activity	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
	30-Jul-18										
1	Tested Hive/Datalake connectivity by executing queries	GREEN									
2	Tested Oracle connectivity by executing queries (Different subnet)	GREEN									
3	Issue with MySQL connectivity at database side (Different subnet) - ID permission issue	AMBER	AMBER	GREEN							
4	Issue with SqlServer connectivity (Password must be changed) - ID permission issue	AMBER	GREEN								
5	Installed query executer, but was having issuefor loading CSS and JS (refereeing online links). Resolved and will deploy again it on tomorrow.	RED	GREEN								
6	Trying to setup Python libraries to access UniConnect via Python code.	AMBER	AMBER	In Progress	In Progress	GREEN					
7	We had issue with internet proxy setup in Unix.	RED	RED	GREEN							
	31-Jul-18										
8	Installed ScaleETL		AMBER	GREEN							
9	Installed UniConnect Dashboard		AMBER	GREEN							
10	Installed reconfigured query executor (Testing yet to be done) - **Waiting for Mizuho to install Java 8 -32bit**		AMBER	GREEN							
11	Tested executing query via JAVA sample code.		GREEN								
12	Offloaded sample tables in Hive by CTAS on SqlServer		GREEN								
13	**203 Million of records**, creation time around 25 Minutes. (ORC format)		GREEN								
14	**28 Million of records,** creation time around 16 Minutes. (ORC format)		GREEN								
15	Configured extra database connectors for Oracle and SqlServer		GREEN								
16	Installed Mongo DB on local and tested.		GREEN								
18	Mizuho Mongo DB is not connecting. Permisison issue		AMBER	GREEN							
19	Not able to run Python code in the server due to missing libraries - Planning to download Python libraries separately		AMBER	In Progress	In Progress	GREEN					
	01-Aug-18										
20	Installed and configured Apache Kafka.			GREEN							
21	Tested ScaleETL on Google news API and dumped data into Kafka. Fetched by UniConnect.			GREEN							
22	Tested and configured access control for 3 dummy users.			GREEN							
	02-Aug-18										
23	Security authorisation - Matrix tested				GREEN						
24	Introduced the Uniconnect Web Ui to RMD team				GREEN						
25	Tested Access DB connectivity				GREEN						
26	Deployed and configured Jasper server				GREEN						
27	Jasper studio Report generation is pending.				In Progress	In Progress	GREEN				
28	WebUI - Latest binaries deployed				GREEN						
29	Java code to connect to Uniconnect provided for verification				In Progress	In Progress	GREEN				
30	Trying to access **Uniconnect Web UI URL in TK office.** TK office is having issue in connecting to this URL. Working with TK team to resolve this issue.				In Progress	In Progress					
	03-Aug-18										
31	Demo provided to Mizuho team on work accomplished till now. Briefing to Key Stakeholders					GREEN					
32	Tested CSV files upload via CSV connector					GREEN					
33	Able to source Google News data into Uniconnect. Used Open Google API					GREEN					
34	Transformation of Google data are in progress.					In Progress	GREEN				
35	Able to use Python code to connect and query Uniconnect data					GREEN					
	06-Aug-18										
36	Email SMTP setup completed. Able to send test mail on Mizuho network						GREEN				
37	ScalETL and Kafka setup tested						GREEN				
38	Able to source and transform MAS FX Daily rates from open source API and push to hive						GREEN				
39	Able to deploy and run locally created Jasper report on Mizuho						GREEN				
40	Tested basic connectivity with Postgres SQL.						GREEN				
41	Basic data offload using 2 tables from different source hive has been tested. Able to send						GREEN				

Leverage DataLake and Built Applications







POC Business use cases

POC using the Oneview Mizuho and Datalake

- Current Account – Cash Flow Forecasting/Prediction using Machine Learning, Analysis and Customer stickiness
    1. Customer profiling
    2. Customer segmentation by Industries
    3. Forecasting
    4. Prediction
- Time Deposit – Cash Flow Forecasting/Prediction using Machine Learning, Analysis and Customer stickiness
    1. Customer profiling
    2. Customer segmentation by Industries
    3. Forecasting
    4. Prediction
- Legal Document extraction
- MRDD – Remittance Analysis
- API Ready

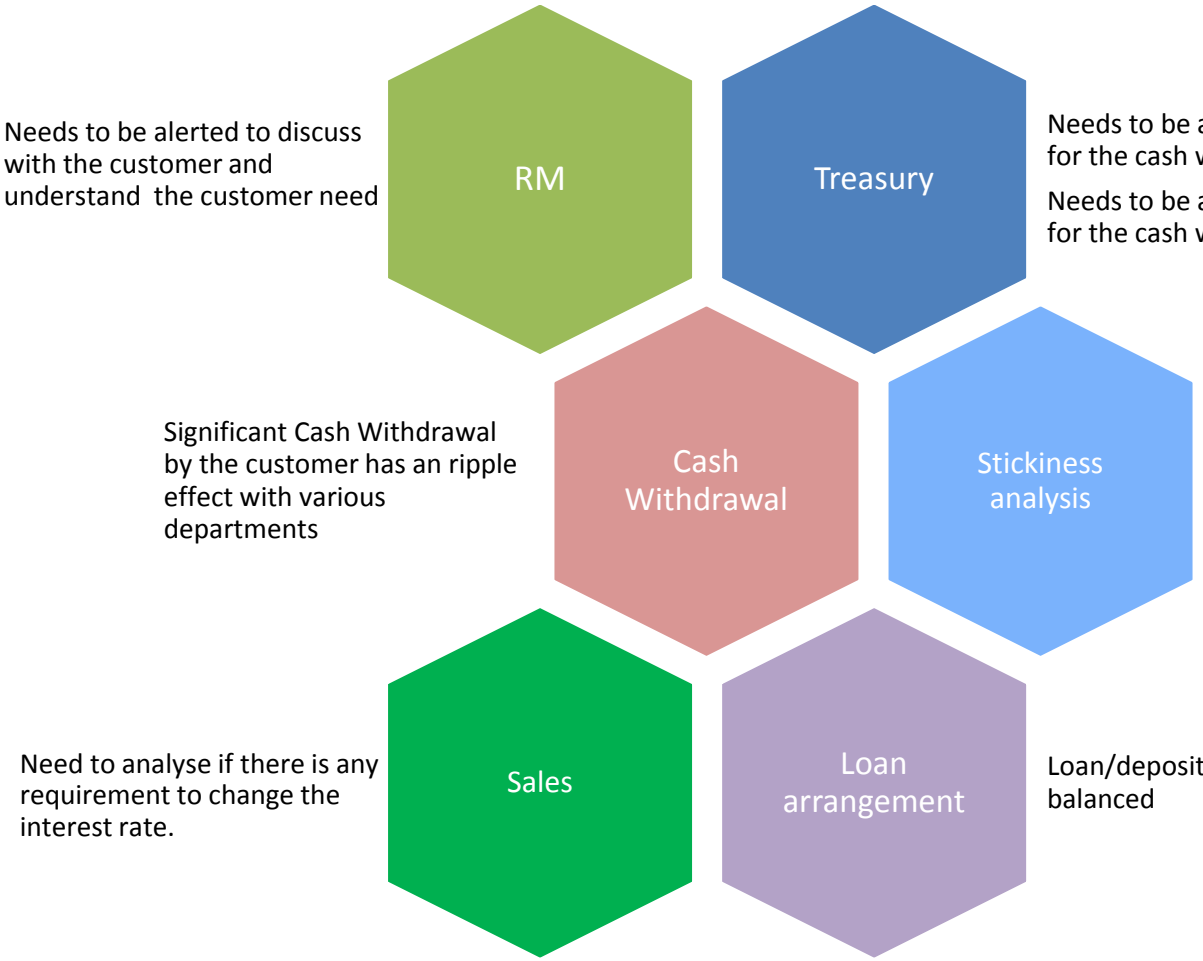
POC Business use cases

POC using the Oneview Mizuho and Datalake – **Use case 1**

- Current Account – Cash Flow Forecasting/Prediction using Machine Learning, Analysis and Customer stickiness
    1. Customer profiling
    2. Customer segmentation by Industries
    3. Forecasting
    4. Prediction

Objective – Customer cashflow Forecasting/Prediction

Predictive Analysis



Needs to be alerted to discuss with the customer and understand the customer need

Needs to be alerted to arrange for the cash withdrew

Needs to be alerted to arrange for the cash withdrew

Significant Cash Withdrawal by the customer has an ripple effect with various departments

Identify sticky customers.
Analyse top depositors stickiness

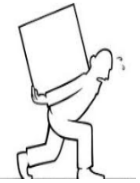
Need to analyse if there is any requirement to change the interest rate.

Loan/deposit has to be balanced

Retain the customers

Cashflow forecasting is still a

- cumbersome,
- manual, and
- spreadsheet-based process



Involving many people from across the organization

Resulting in monthly or quarterly, rather than weekly, updates.

Cash flow forecasting can be an important function to predict large deposit withdrawals; to analyse top depositors stickiness level for stable source, etc. In Mizuho we are doing this analysis at crude level. No back-testing.

Cashflow Withdrawal Prediction Benefits

Retain the customer

Early Alert

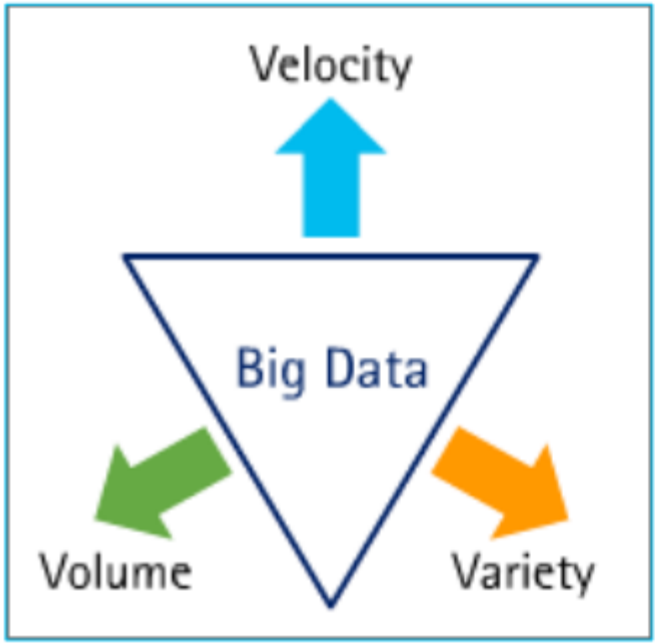
Identify trends / customer segmentation

Planning for liquidity / Asset Liability Mgmt

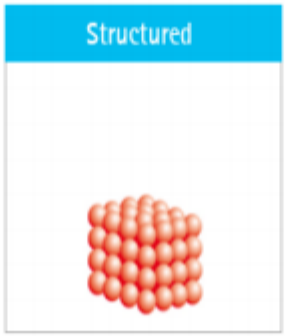
Why we need BigData?

Big data in general has context in three Vs:

- Sheer quantity of data
- Speed with which data is produced, processed, and digested
- Diversity of sources inside and outside



The different types of data that contribute to this are:



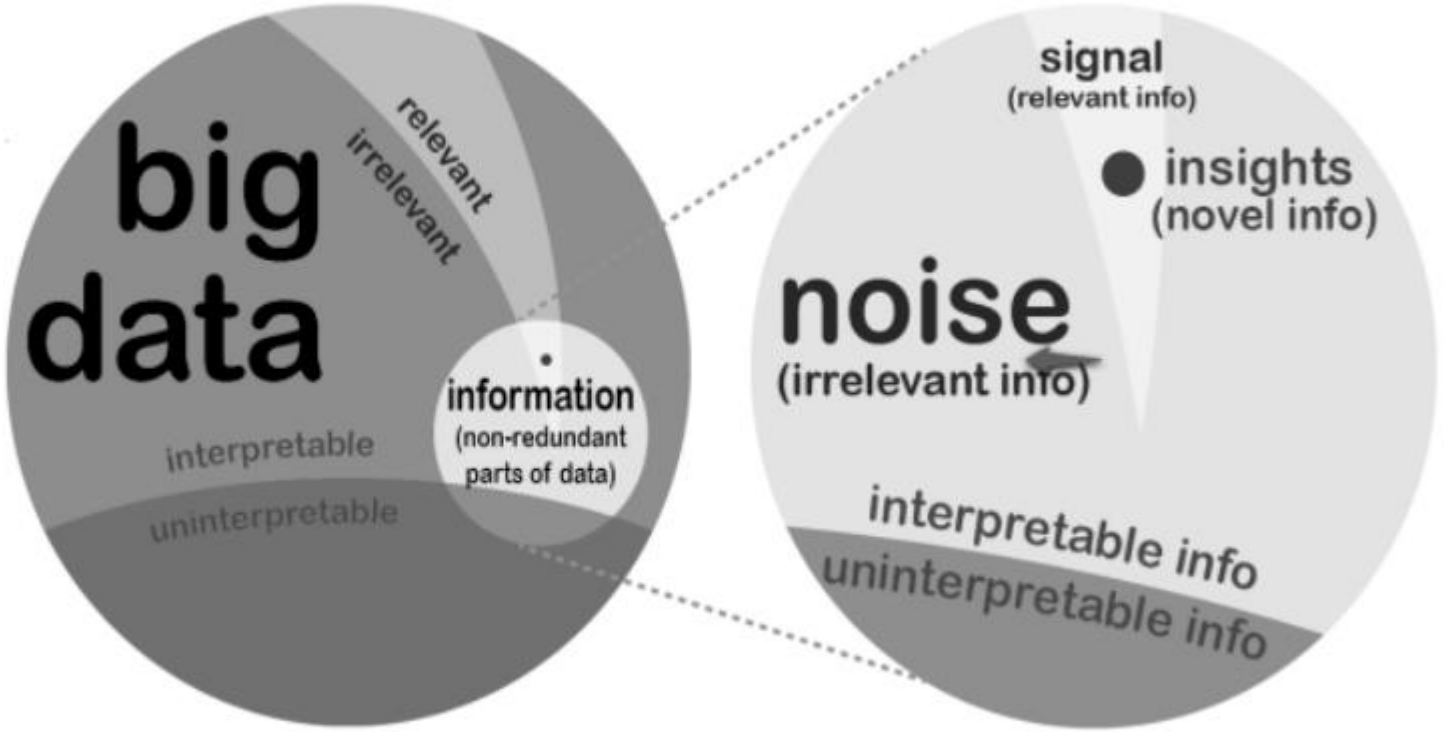
- Fields/Tables /Columns
- RDBMS/Spreadsheet

- Markers/Tags to separate elements
- XML/HTML

- No fields/attributes
- Free form text (email body, notes, articles)
- Audio, video, and image

Separating the signal from the noise[1] becomes really relevant

to



1 http://techcrunch.com/2012/11/25/the-big-data-fallacy-data-≠-information-≠-insights/

Some Parameters outside our Bank for prediction considered in the POC for to have more accuracy

- Structured Data
  - Historical daily **SGX** price from Yahoo finance
  - Historical daily **Nasdaq** Exchange Price from Yahoo finance
  - Historical daily **S&P** Exchange Price from Yahoo finance
  - Historical daily **Nikkie** Price from Yahoo finance
  - Customer Equity price from **Bloomberg** terminal for two customers
- Unstructured Data
  - **News** related to the particular customer OG30 from newsapi.com using API call from python and **Kafka(hadoop)**
  - **Top 20 news** headline around the worldwide from kaggle.

Our Bank – Transaction data

- Transaction Data
- Processed approximately **20 million** records from Hive(Hadoop - BigData)

Not secure | spws271:5003

Current Account

Reactive Inputs

Make selections and watch the graphs and data update

Currency
- [ ] SGD
- [x] USD

Select Industry
ALL

Select Country
ALL

Select Cust Seg
- [x] About To Sleep
- [x] Hibernating
- [x] Loyal Customers
- [x] Most Risky
- [x] Need Attention
- [x] Others
- [x] Risky

Select Users
ALL

Reset Fields

Analysis | Tree map - Current Acc withdrawal Profiling | Tree Map of Current-Acc available profiling

Current Account - RFM Analysis of the Customers

Not secure | spws271:5003

MIZUHO SG - AOAD INNOVATION Machine Learning - Customer Stickiness(Current Account) Dashboard | Analytics | Forecasting

Reactive Inputs

Make selections and watch the graphs and data update

Currency

- [ ] SGD
- [x] USD

Select Industry

ALL

Select Country

ALL

Select Cust Seg

- [x] About To Sleep
- [x] Hibernating
- [x] Loyal Customers
- [x] Most Risky
- [x] Need Attention
- [x] Others
- [x] Risky

Select Users

ALL

Reset Fields

Analysis | Tree map - Current Acc withdrawal Profiling | Tree Map of Current-Acc available profiling

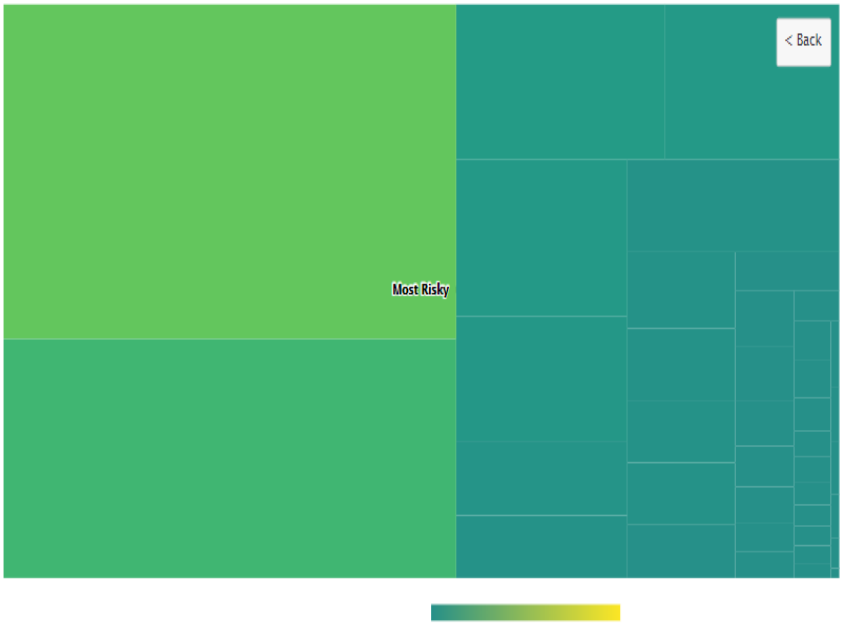
Select Measure

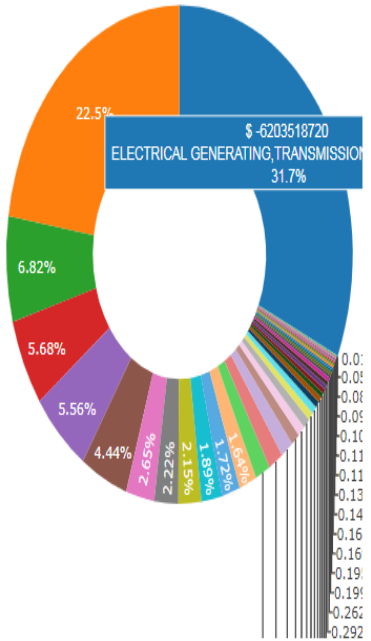
Cust_segment

Select Measure

industry_name

Customer TreeMap Explorer





Machine Learning - Forecasting

MIZUHO SG - AOAD INNOVATION Machine Learning - Customer Stickiness(Current Account) Dashboard | Analytics | Forecasting

Forecasting Inputs

Make selections and watch the graphs and data update

Date input: yyyy-mm-dd

2016-02-04

Select Model

tslm.basic

Select Currency

SGD

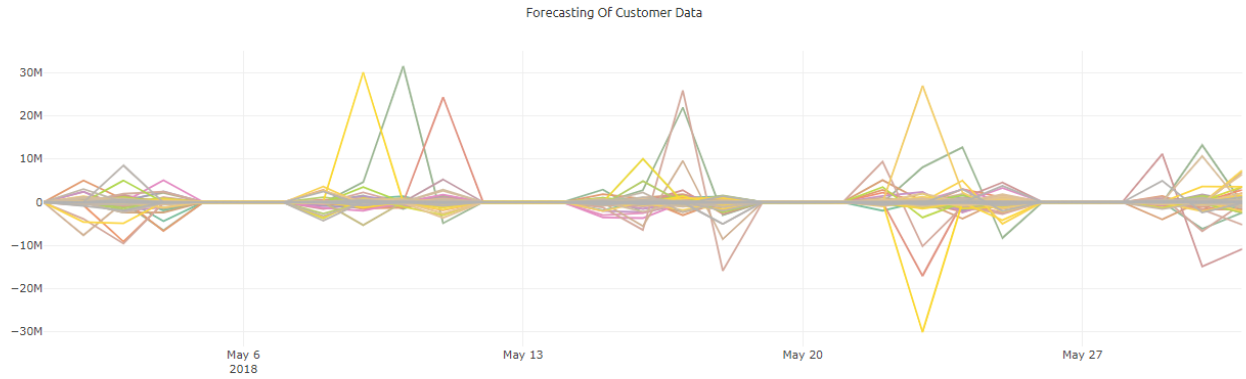
Select Account

CURRENT ACCOUNT

Forecast/Prediction

* Please Note it will take some time to load the data. The data is dynamically computed on Hadoop. After computation, forecasting model is ran on each of the customers individually

Forecasting Overall | Forecast Top5



Show 10 entries Search:

Date	AA17	AA81	AB72	AB86	AD11	AD50	AD95	AD96	
2018-05-01	-58.6420545746396	-1214.32343499209	8.46548956661144	-0.0288924558587482	-4030.49759229543	-1.43017656500807	-2.38362760834671	-28.6444622792939	-7767
2018-05-02	-58.642054574638	-1214.32343499196	8.46548956661871	-0.0288924558587481	-8293.99759229547	-1.43017656500803	-2.38362760834679	714.355537720704	2469
2018-05-03	-58.64205457464	-100109.323434992	8.46548956660195	-0.0288924558587482	-19619.4975922955	-1.43017656500806	-2.383627608347	-2492.1444622793	1795
2018-05-04	-58.6420545746085	-7244.32343499198	8.46548956644822	-0.0288924558587484	-5750.4975922954	-1.43017656500808	-2.38362760834672	1318.8555377207	-1916
2018-05-05	0	0	0	0	0	0	0	0	
2018-05-06	0	0	0	0	0	0	0	0	

POC Business use cases

POC using the Oneview Mizuho and Datalake – Use case 2

- Time Deposit – Cash Flow Forecasting/Prediction using Machine Learning, Analysis and Customer stickiness
    1. Customer profiling
    2. Customer segmentation by Industries
    3. Forecasting
    4. Prediction

MIZUHO SG - AOAD INNOVATION Machine Learning - Customer Stickiness(Time Deposit) Dashboard | Analytics | Forecasting

Reactive Inputs

Make selections and watch the graphs and data update

Currency

- [x] SGD
- [ ] USD

Select Industry

ALL

Select Country

ALL

Select Cust Seg

- [x] About To Sleep
- [x] Hibernating
- [x] Loyal Customers
- [x] Most Risky
- [x] Need Attention
- [x] Others
- [x] Risky

Select Users

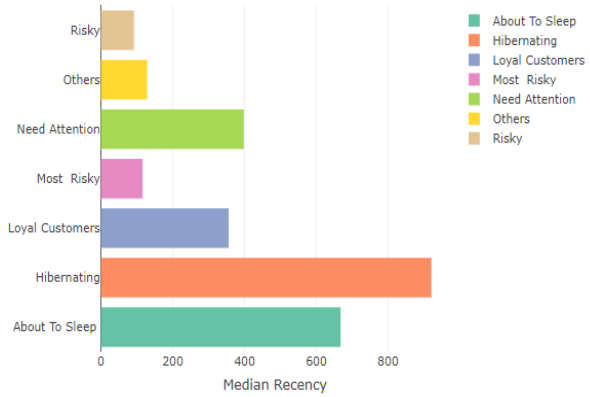
ALL

Reset Fields

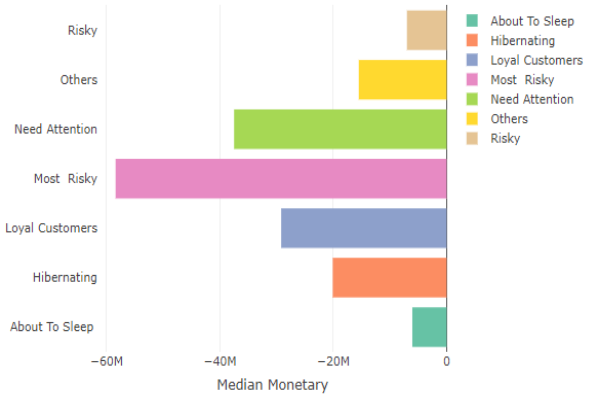
Analysis | Tree map - Deposit withdrawal Profiling | Tree Map of Deposit available profiling

Time Deposit - RFM Analysis of the Customer Base

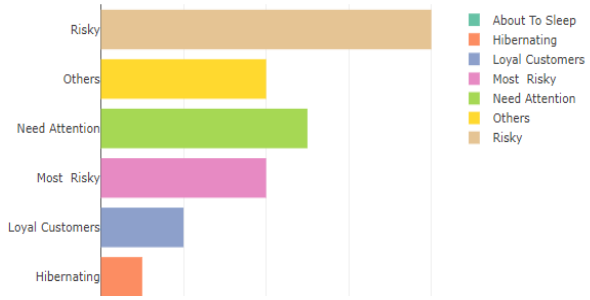
Customer Segmentation: Median Recency



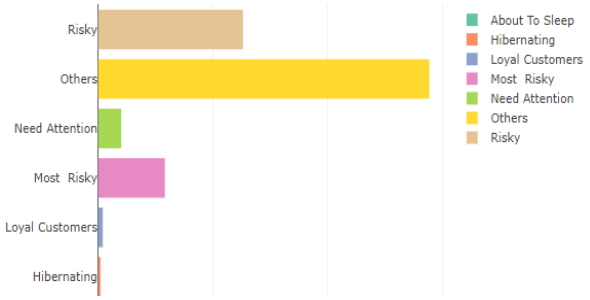
Customer Segmentation: Median Monetary



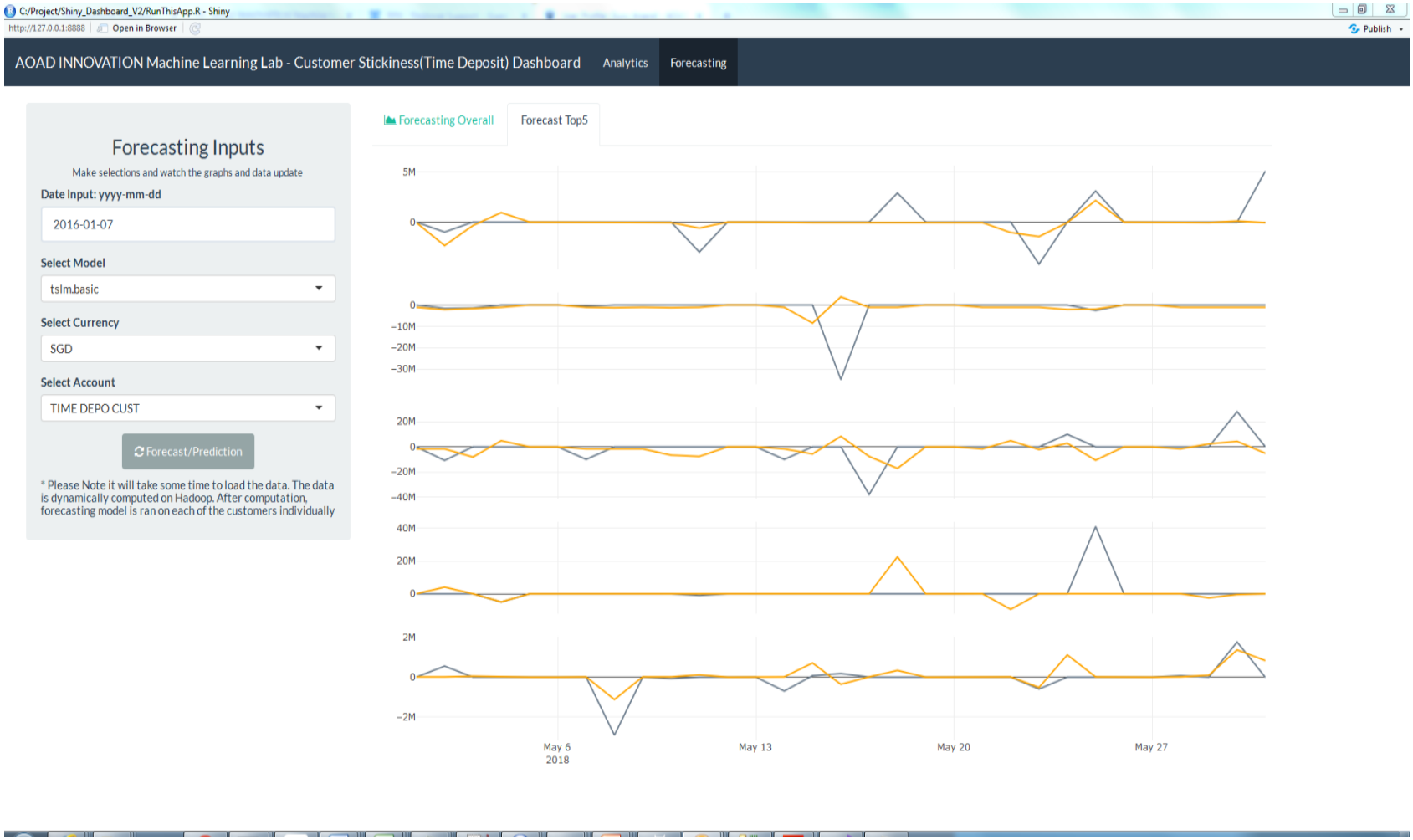
Customer Segmentation: Median Frequency



Customer Segmentation: No of Customers



Machine Learning – Prediction for one month



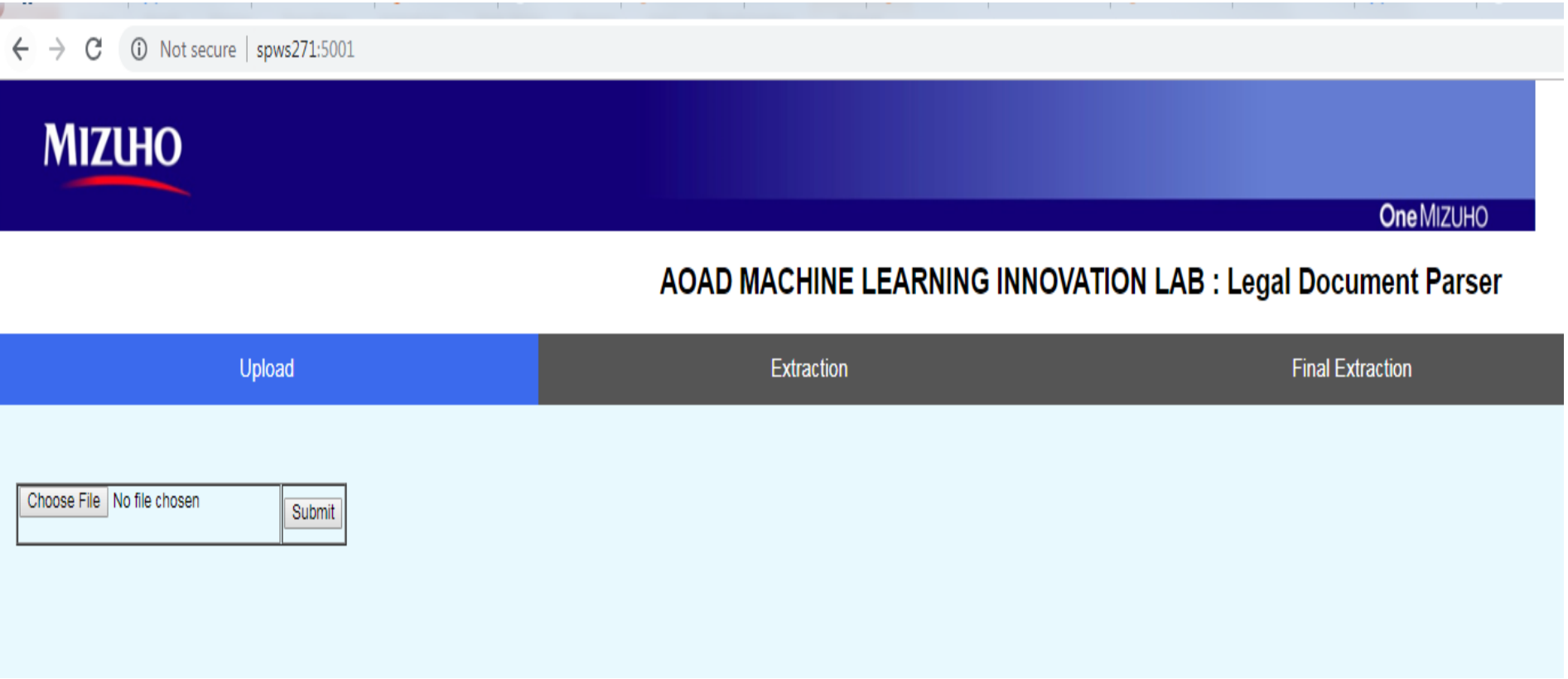
POC Business use cases

POC using the Oneview Mizuho and Datalake – Use case 3

- Legal Document extraction

Legal Document extraction using NLP

http://spws271:5001/



Legal Document extraction using NLP

http://spws271:5001/

Upload	Extraction	Final Extraction
Agreement Date	2018-01-25	
Termination Date/Maturity Date	['2022-01-25', '2023-01-25']	
MultiCurrency	['USD']	
Business Days	['Singapore', 'New York', 'London']	
OverdueInterest	['one per cent']	
Interest Rate and Margin	['Facility A', '1.02%', 'Facility B', '1.17%', 'Facility C', '1.17%', 'LtBoR']	
Availability Period Processed	['2018-04-25', '2022-12-25']	
Repayment Processed	Bullet	
Borrower	Dated 25 January 2018 AMONG -. HUDSON DC ASSETS LLC, NAVARRO DC ASSETS LLC, ETOWAH DC ASSETS LLC, REDWOOD DC ASSETS LLC Collectively, as the Borrower MIZUHO BANK, LTD.	

POC Business use cases

POC using the Oneview Mizuho and Datalake – Use case 4

- MRDD – Remittance Data Analysis

Remittance Data (MRDD)

Remittance Data (MRDD)

Remittance Data (TRICS by MRDD)

- Able to upload **15 million** remittances to Data Lake in a few minutes
- It covers Apr 2014 to May 2018
- Total of 18 trillion USD notional remittance, where Mizuho is one party
- Ran **complex** query and found **4.8 million** remittances, where the FX is **not** done by Mizuho over the period of 5 years.*Rough estimate
- Query took 20 seconds
- 'Potential' FX business opportunity of notional USD 1.5 trillion over 5 years
- ***Rough estimate condition examples where FX business is likely not captured:
  - Sending non-SGD from Singapore, but the remitting bank is not Mizuho

POC Business use cases

POC using the Oneview Mizuho and Datalake – Use case 5

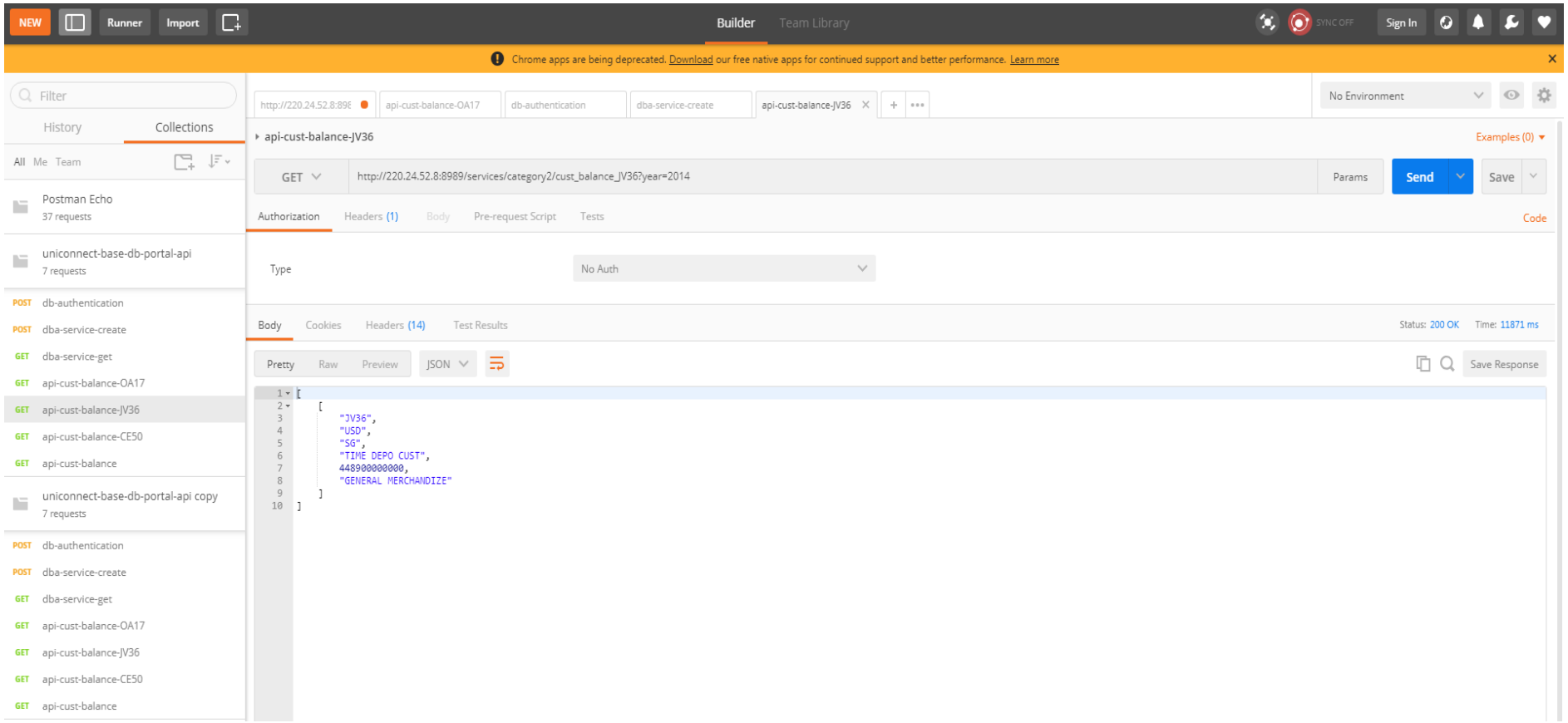
- API in Mizuho
- Internal API and
- External API

OneView Mizuho Features

PRODUCT FEATURES	ASTER DATA	DENODO	ABINITIO QUERYIT	CIRRO	TAMR	OneView Mizuho
Data Is Copied To Platform	Y	Y	Y	Y	Y	N (In-Memory)
Analytics Eco-system On Top	N	N	N	N	Y	Y (with Spark)
Deployment	On Premises	On Premises	On Premises	On Premises	On Premises/Cloud	On Premises/Cloud
Data Cleaning Enrichment	N	N	Y	N	Y	Y (Partial, with SQL)
Volume / CPU Based Licensing	Y	Y	Y	Y	Y	Free To Grow
Scalability	Fixed	Fixed	Fixed	Fixed	Fixed	Auto Elastic
Can Develop Connector API	N	N	N	N	N	Y
JDBC/ODBC	Y	Y	N	Y	N	Y
Learning Curve	Y	Y	Y	Y	Y	N (Use SQL)
Security / Admin / Audit	Y	Y	Y	Y	Y	Y

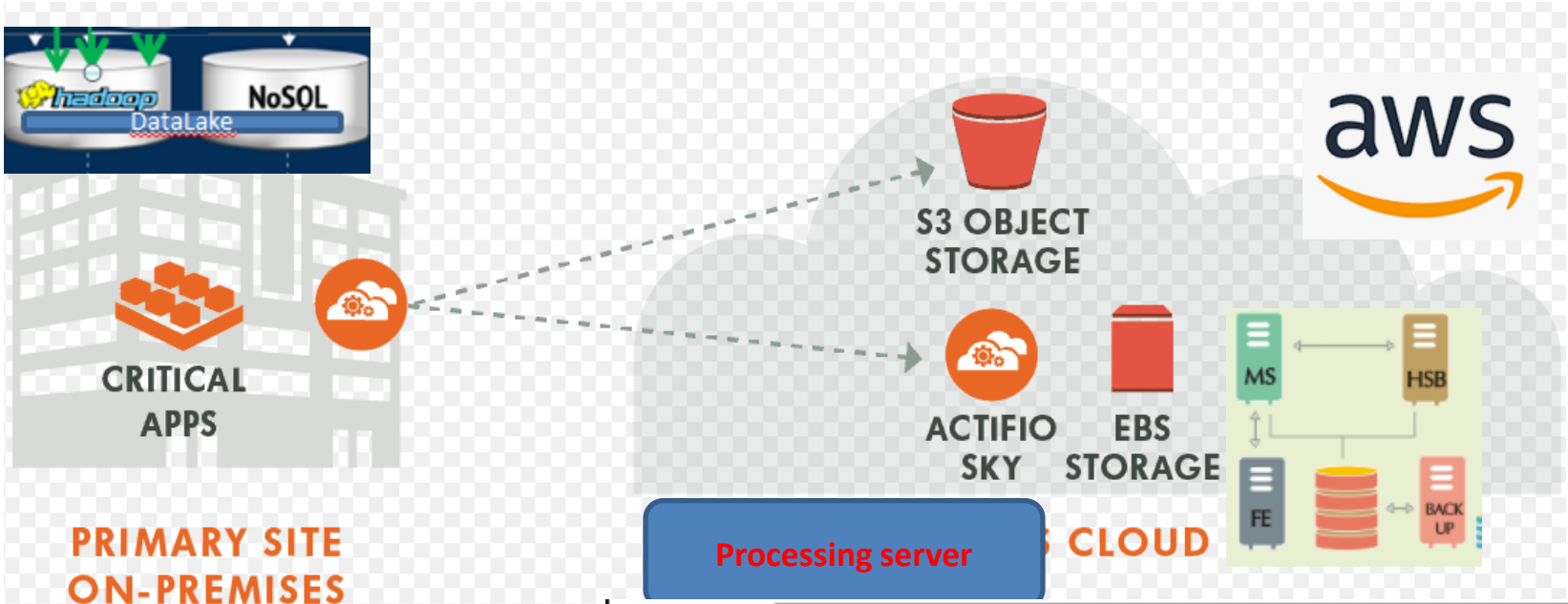
API Ready

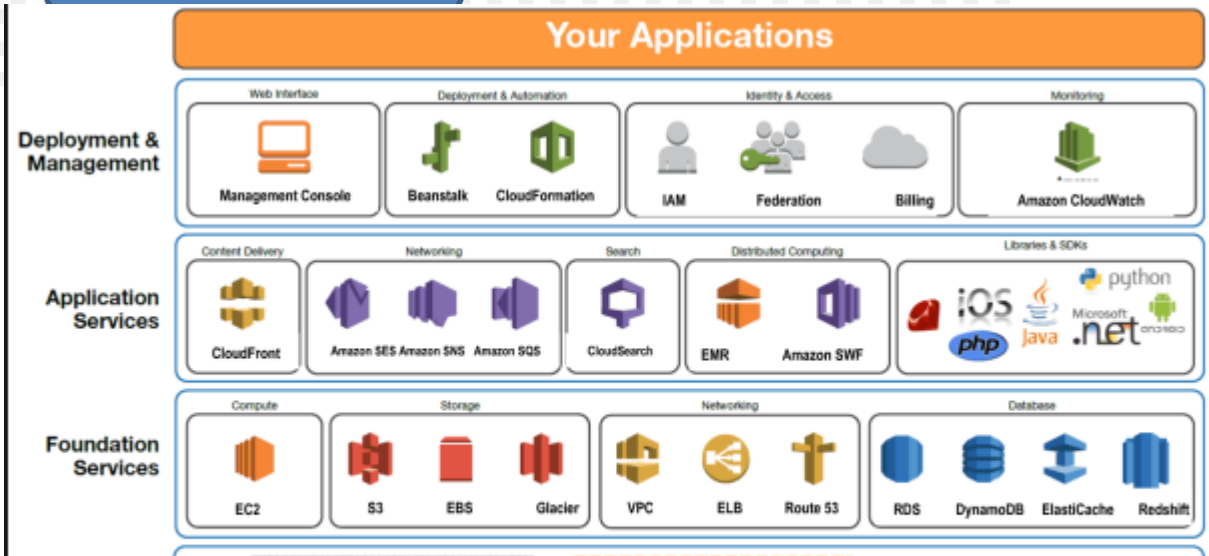
APIs


NEW
Runner
Import
Builder
Team Library
SYNC OFF
Sign In
Chrome apps are being deprecated. Download our free native apps for continued support and better performance. Learn more
Filter
History
Collections
All Me Team
Postman Echo
37 requests
uniconnect-base-db-portal-api
7 requests
POST db-authentication
POST dba-service-create
GET dba-service-get
GET api-cust-balance-OA17
GET api-cust-balance-JV36
GET api-cust-balance-CE50
GET api-cust-balance
uniconnect-base-db-portal-api copy
7 requests
POST db-authentication
POST dba-service-create
GET dba-service-get
GET api-cust-balance-OA17
GET api-cust-balance-JV36
GET api-cust-balance-CE50
GET api-cust-balance
http://220.24.52.8:89
api-cust-balance-OA17
db-authentication
dba-service-create
api-cust-balance-JV36
No Environment
api-cust-balance-JV36
Examples (0)
GET
http://220.24.52.8:8989/services/category2/cust_balance_JV36?year=2014
Params
Send
Save
Authorization
Headers (1)
Body
Pre-request Script
Tests
Code
Type
No Auth
Body
Cookies
Headers (14)
Test Results
Status: 200 OK
Time: 11871 ms
Pretty
Raw
Preview
JSON
Save Response
[
[
"JV36",
"USD",
"SG",
"TIME DEPO CUST",
448900000000,
"GENERAL MERCHANDIZE"
]
]

Future Plan with AWS

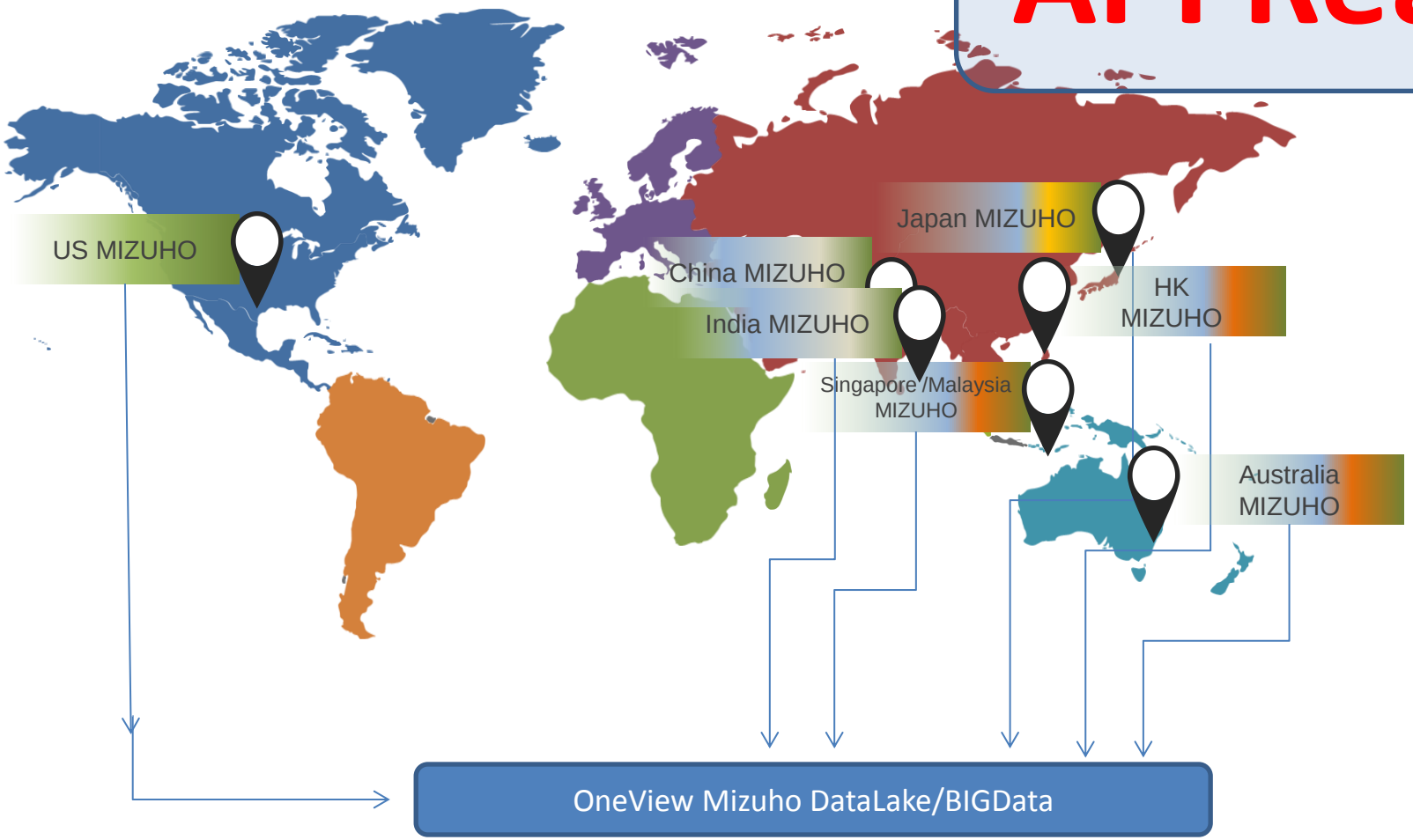
Data in house and Processing server in Cloud for the applications that we built





Future - DataLake/In-memory Database

API Ready



Case Study

Percipient
Connecting the world's data

MIZUHO

Corporate Time Deposit Customer Profiling and Analysis of Customer Stickiness

By achieving integrated access to multiple sources of data, a Percipient and HPE Proof of Concept demonstrated how Mizuho Singapore's Innovation team could forecast and predict customer cash flow and improve stickiness

"We are very pleased with the insights and forecasting we have been able to achieve, underpinned by the UniConnect platform. We believe the solutions that Percipient has introduced to us will substantially lift our analytical capabilities and correspondingly, our productivity, in today's hyper- competitive corporate banking sector."

Guru S. Anand,
Vice President from the Innovation Team,
Asia and Oceania Department,
Mizuho Bank Singapore.

MIZUHO

Executive Summary

The Business Need:

- Track behaviour patterns within Time Deposit customer segments
- Incorporate market news and events into customer analytics
- Forecast potential for customer attrition

The Tech Challenge:

- Aggregate and store fragmented customer data
- Ingest third party data in real time
- Make available business-specific information

The Solution:

- Unify the bank's multiple databases
- Deploy a firm level Datalake platform to host logical Datalake/Datamarts
- Integrate with the bank's analytics and machine learning code

Manual processes

SG Mizuho's Asia and Oceania Administration Department - Innovation team tracks and forecasts customer cash flows in order to predict the likelihood of large deposit withdrawals and customer stickiness. This helps drive early RM alerts, identification of customer behaviour trends, customer segmentation and liquidity management.

However, these analytics exercises are currently performed by manually extracting and transforming data sourced from a variety of internal databases, a process that is cumbersome, slow and resource-intensive. Factoring in external data adds further complexity and as a result, is not routinely done. As a result, forecasts can only be updated monthly or quarterly, rather than on a more impactful weekly cycle. It is also impossible to continuously back-test and refine the forecast model.

New data elements

To address these challenges, Mizuho engaged integration software provider, Percipient, for a trial of the latter's flagship UniConnect platform. A key requirement for this Proof Of Concept (POC) was for structured and unstructured data to be delivered to a single end-point, paving the way for easy discovery and consumption by various business, technology and analytics teams.

The external data accessed by Uniconnect included:

- SGX prices from Yahoo finance
- Nasdaq Exchange Prices from Yahoo finance

Percipient

MIZUHO

- S&P Exchange Prices from Yahoo finance
- Nikkei Prices from Yahoo finance
- Customer Equity price from Bloomberg terminal
- News items from news api.com using API call from python and Kafka (Hadoop)
- Top 100 news headlines worldwide from the website Kaggle.

This was unified with internal data comprising millions of rows of transaction data stored over the last five years in a variety of repositories and formats, including:

- Oracle
- SQL Server

Next Gen Solution

Based on the above specifications, Percipient proposed a solution comprising a number of elements central to meeting the bank's needs.

Connectors
UniConnect Connectors deployed in order to automatically ingest data from the multiple sources outlined above.

Data Lake
A Hadoop datalake was built on premise. UniConnect 's engine was used to move new and existing data to the datalake.

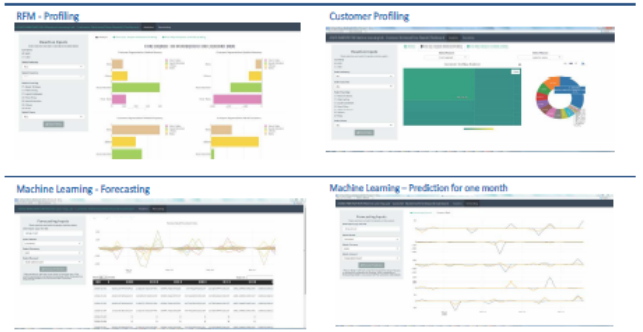
Transformation
Approx 30 million customer records were transformed virtually to align with the bank's Python-based forecast models.

Consumption
Data could be discovered and queried in seconds or micro-seconds as a single view using UniConnect's SQL interface

To provide hardware support for this POC deployment of the UniConnect and datalake platforms, Percipient and Mizuho partnered with leading IT infrastructure and services company, HPE. Based on the SLAs defined for the trial, HPE were able to supply a pair of DL380 servers for the Uniconnect and Hadoop Datalake processing.

Analytical Outcomes

Leveraging UniConnect's integration capabilities, Mizuho's SG Asia and Oceania Department's Innovation team built dashboards as shown below:



Analysis of the customer profiling and stickiness patterns allowed the team to uncover a number of important stickiness indicators. Stickiness was determined to be a stable customer variable that relates closely to withdrawals. These results can be used to drive new analytical models and customer retention strategies.

Guru S. Anand, Vice President from the Innovation Team, Asia and Oceania Department, Mizuho Bank Singapore, said, "We are very pleased with the insights and forecasting we have been able to achieve, underpinned by the UniConnect platform. We believe the solutions that Percipient has introduced to us will substantially lift our analytical capabilities and correspondingly, our productivity, in today's hyper-competitive corporate banking sector."

The full Mizuho Bank Case Sudy document is available on www.percipientcx.com

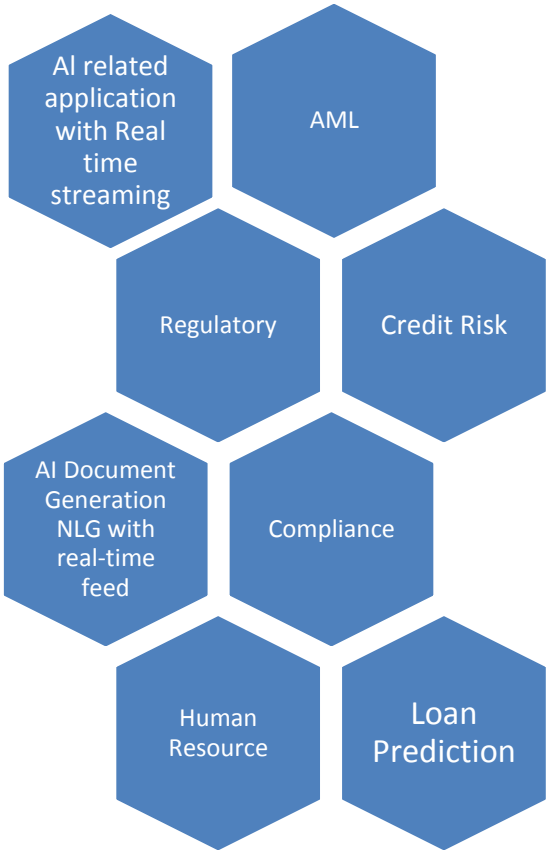
Additional info

POC - DBs

Data type	Database	Data accessible	Description	Section
Structured data	CDM ODBC	Customer profile and credit facilities	customer profile, financial data, customer rating, credit facilities, protection, facility rating (outstanding & limit), write-off & reserve etc.	GTBD
	Hyperion	Profit & exposure	data from the profit and exposure management systems (PEGASYS, PEGASYS+), and the customer relationship management system (IGCIS) etc.	GTBD
	Gbase EUC	Previous business day's data of Gbase	data storage system by Head Office, contains numerous tables and the table names are self-explanatory (e.g. LOANS, REMITTANCE etc.)	AOAD,GTBD
	Gbase for other branches	data of Gbase for other branches	data storage system by Head Office, contains numerous tables and the table names are self-explanatory (e.g. LOANS, REMITTANCE etc.)	AOAD,MRDD
	SMILES	Singapore branch's transaction processing systems	branch's integrated transaction processing systems which cover business areas listed below and having screens for end users. Remittance (inward, outward, in-house transfer), GIRO, Loan, Cheque issuances and clearing, Deposit, FX and money market, Certificate of balance, eStatement and eAdvice, Trade processing (import, export, LC etc.), Gbase reports on the screen	AOAD,GTBD
	SMILES Web	Singapore branch's transaction processing systems	branch's integrated transaction processing systems which cover business areas listed below and having screens for end users. Remittance (inward, outward, in-house transfer), GIRO, Loan, Cheque issuances and clearing, Deposit, FX and money market, Certificate of balance, eStatement and eAdvice, Trade processing (import, export, LC etc.), Gbase reports on the screen	AOAD
	TAPS	Singapore branch's transaction processing systems	branch's integrated transaction processing systems which cover business areas listed below and having screens for end users. Remittance (inward, outward, in-house transfer), GIRO, Loan, Cheque issuances and clearing, Deposit, FX and money market, Certificate of balance, eStatement and eAdvice, Trade processing (import, export, LC etc.), Gbase reports on the screen	AOAD
	LASER	Singapore branch's small scale data warehouse	branch's data storage system that holds historical data of Gbase EUC (not all) and branch's transaction processing systems (SMILE & TAPS). Main usages of this system are regulatory reporting and maintaining historical data.	GTBD
	External data providers (e.g. Bloomberg, Reuters etc.)	Complement internal data	customer equity info, macroeconomic indices, financial market indices etc.	AOAD,GTBD
	IGSIS / IIMI	Customer info	customers' profile and performance data	RMD
	DEVON	Customer FX transactions	data source for customers' FX transactions executed by AOTD	GTBD
	CES	Data for MAS	Data for upcoming MAS revised return 610.	RMD
	VITAL	Front office for Treasuray	Front office for Treasuray for Hanoi Branch	MRDD
Semi-structured data	Social media & Websites	Opinions about customers and corporate actions	news, public consensus, corporate actions, company announcements etc.	AOAD,GTBD
Unstructured	N.A.	Communication with customers and internal customer/industry analysis	business memos, credit applications/ review, legal agreements, communication with customers by email and phone	AOAD,GTBD

Other Use cases in Corporate Banking

Fraud Detection and Analysis	Risk Management	Customer Insights/Management
Micro targeting	Trade/Payments Analytics	Long-Term Storage & Analytics
New Opportunities	Reputational Risk	Marketing Campaigns



Facts

- We are building a foundation here for BigData/DataAnalytics/Machine Learning platform.
- BigData/DataLake is one of the fundamental blocks for Digital Strategy/Innovation.
- We are not implementing a solution per say like what other teams are doing in a fragmented fashion for other specific use cases.
- We are building a BigData/DataLake for all the departments/sections/regions/countries usage(Global usage) storing Terabytes/Petabytes of structured/unstructured/semi-structured data to move from traditional computational ecosystem to **predictive computational** ecosystem.
- Also the use cases used in POC are some very few samples only and not the actual implementation.
- We have lot of use cases from other departments like RMD, MRDD etc. Even their use use cases will be tested during POC which is not documented.
- Moreover the use cases are not limited in the BIGData/DataLake solution. It is unlimited which will evolve as we move forward from various departments/sections/regions/globally.
- **Other department stakeholders like GTBD, MRDD, RMD** expressed their interest to participate in the conference call to explain the importance of this BigData/DataLake project in their respective departments if necessary.

Questions