Azure Data Maturity & Capabilities Assessment / BI Roadmap

Decisive Data
Data Platform Team

DecisiveData

Selected Samples from BI Roadmap Deliverable Deck

Project Overview

Systems Integration + Warehouse Innovation + Data Visualization

Client X needs to consolidate disparate data sources into a single, governed, automated source of the truth. This would speed up development of a broader range of reports, while paving the way for new data sources in the long term.

Key components include an ETL design that accommodates multiple source systems, a conformed data warehouse, a flexible analytic layer for both official and ad-hoc reporting, and a phased roadmap for all the above.

The goal of this discovery project was to determine:

Current BI gaps

Infrastructure Needs

Optimal BI Approach

Optimal Data Warehousing Approach

Tool Recommendations

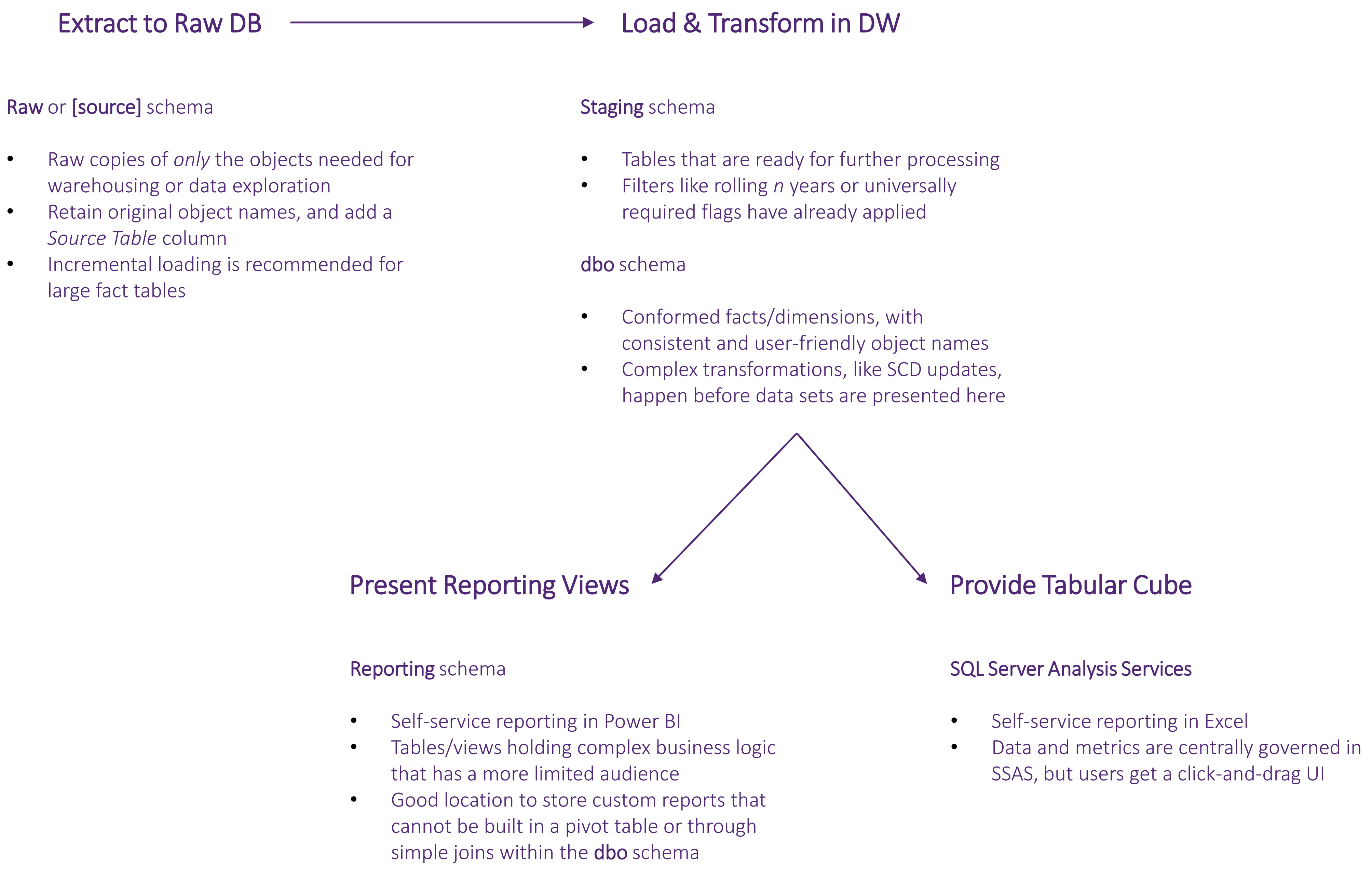
Projected Costs

Projected Schedule

Business Needs & Technical Solutions

Challenge	Solution	Approach
Data Access and Usability	Modern, Conformed Data Warehouse	Offer intuitive data model for analytical use • Exposes relevant and user-friendly tables in a normalized model • Centralizes data governance • Conforms taxonomy for all current and future sources • Single source of record for historical data
Data Freshness and Integration	Robust ETL Methodology	Refresh data warehouse frequently and consistently • Fast and simple method for adding new sources • Granular but actionable ETL logging and alerts • Ability to automate custom scripting (e.g., Python) where needed • Supports SCDs where needed
Change and Version Management	Source Control + Request Tracking Interface	Centralize and create visibility for all DW and report activity • Single repository and mgmt. tool for all code, business logic, and requests • Keeps work in progress readily visible • Preserves code and change history automatically
Flexible Reporting and Data Exploration	Tabular Cube + Excel	Implement and provide training for Power BI and/or Excel • Centralized management of standard reports • Flexible distribution and security models • Self-service analysis through filtering, pivoting, visualizing, and calculating in tool(s) of choice

Data Flow Overview



Extract to Raw DB

Raw or **[source]** schema

- Raw copies of *only* the objects needed for warehousing or data exploration
- Retain original object names, and add a *Source Table* column
- Incremental loading is recommended for large fact tables

Load & Transform in DW

Staging schema

- Tables that are ready for further processing
- Filters like rolling *n* years or universally required flags have already applied

dbo schema

- Conformed facts/dimensions, with consistent and user-friendly object names
- Complex transformations, like SCD updates, happen before data sets are presented here

Present Reporting Views

Reporting schema

- Self-service reporting in Power BI
- Tables/views holding complex business logic that has a more limited audience
- Good location to store custom reports that cannot be built in a pivot table or through simple joins within the **dbo** schema

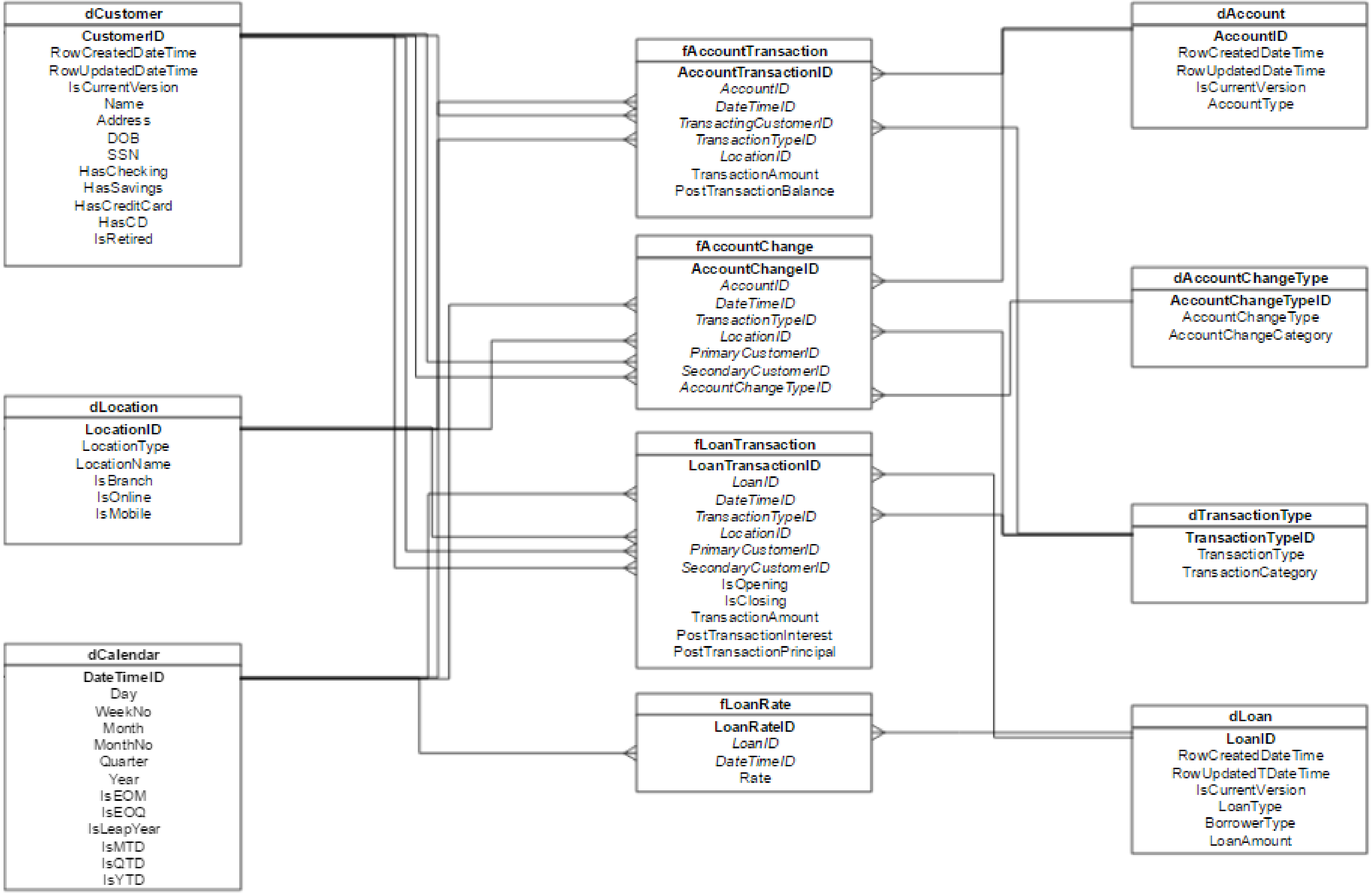
Provide Tabular Cube

SQL Server Analysis Services

- Self-service reporting in Excel
- Data and metrics are centrally governed in SSAS, but users get a click-and-drag UI

Sample Data Model: Foundation

Bold = Primary Key
Italics = Foreign Key



dCustomer

CustomerID
RowCreatedDateTime
RowUpdatedDateTime
IsCurrentVersion
Name
Address
DOB
SSN
HasChecking
HasSavings
HasCreditCard
HasCD
IsRetired

dLocation

LocationID
LocationType
LocationName
IsBranch
IsOnline
IsMobile

dCalendar

DateTimeID
Day
WeekNo
Month
MonthNo
Quarter
Year
IsEOM
IsEOQ
IsLeapYear
IsMTD
IsQTD
IsYTD

fAccountTransaction

AccountTransactionID
AccountID
DateTimeID
TransactingCustomerID
TransactionTypeID
LocationID
TransactionAmount
PostTransactionBalance

fAccountChange

AccountChangeID
AccountID
DateTimeID
TransactionTypeID
LocationID
PrimaryCustomerID
SecondaryCustomerID
AccountChangeTypeID

fLoanTransaction

LoanTransactionID
LoanID
DateTimeID
TransactionTypeID
LocationID
PrimaryCustomerID
SecondaryCustomerID
IsOpening
IsClosing
TransactionAmount
PostTransactionInterest
PostTransactionPrincipal

fLoanRate

LoanRateID
LoanID
DateTimeID
Rate

dAccount

AccountID
RowCreatedDateTime
RowUpdatedDateTime
IsCurrentVersion
AccountType

dAccountChangeType

AccountChangeTypeID
AccountChangeType
AccountChangeCategory

dTransactionType

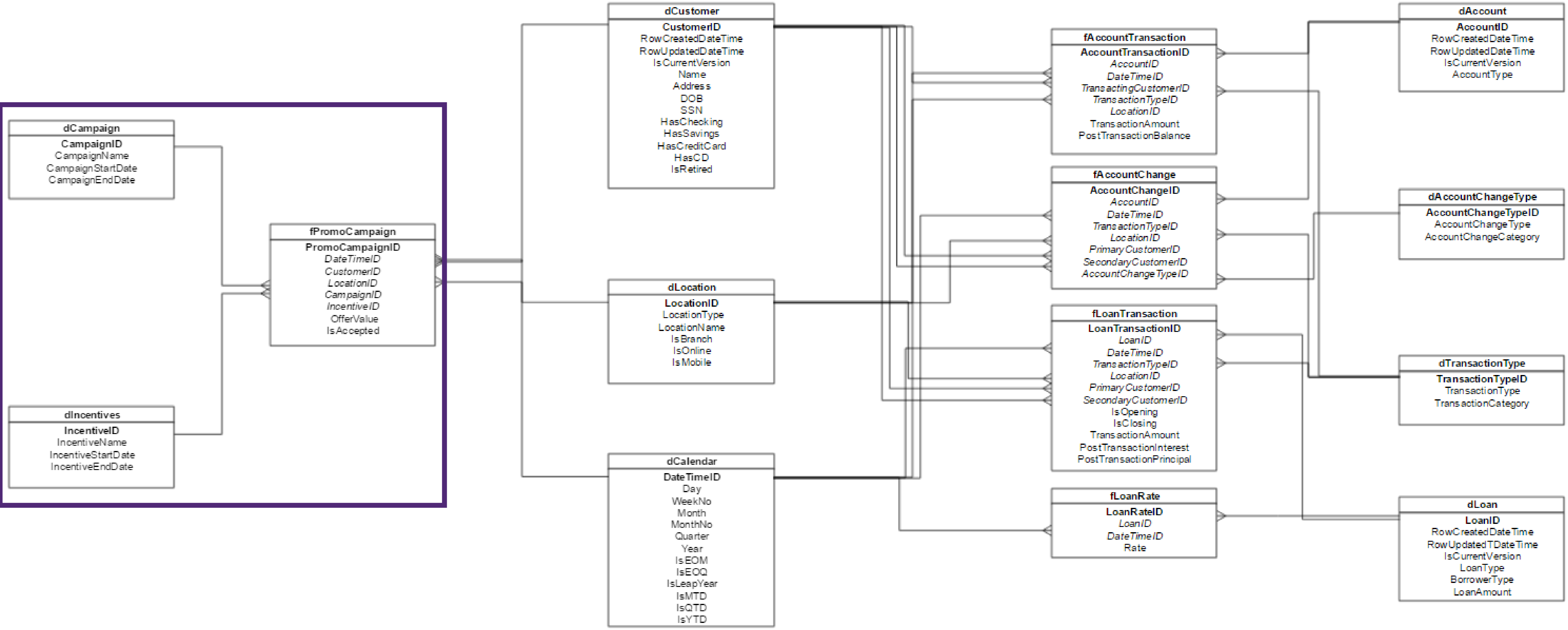
TransactionTypeID
TransactionType
TransactionCategory

dLoan

LoanID
RowCreatedDateTime
RowUpdatedTDateTime
IsCurrentVersion
LoanType
BorrowerType
LoanAmount

Sample Data Model: Expansion

Bold = Primary Key
Italics = Foreign Key



dCampaign
CampaignID
CampaignName
CampaignStartDate
CampaignEndDate

dIncentives
IncentiveID
IncentiveName
IncentiveStartDate
IncentiveEndDate

fPromoCampaign
PromoCampaignID
DateTimeID
CustomerID
LocationID
CampaignID
IncentiveID
OfferValue
IsAccepted

dCustomer
CustomerID
RowCreatedDateTime
RowUpdatedDateTime
IsCurrentVersion
Name
Address
DOB
SSN
HasChecking
HasSavings
HasCreditCard
HasCD
IsRetired

dLocation
LocationID
LocationType
LocationName
IsBranch
IsOnline
IsMobile

dCalendar
DateTimeID
Day
WeekNo
Month
MonthNo
Quarter
Year
IsEOM
IsEOQ
IsLeapYear
IsMTD
IsQTD
IsYTD

fAccountTransaction
AccountTransactionID
AccountID
DateTimeID
TransactingCustomerID
TransactionTypeID
LocationID
TransactionAmount
PostTransactionBalance

fAccountChange
AccountChangeID
AccountID
DateTimeID
TransactionTypeID
LocationID
PrimaryCustomerID
SecondaryCustomerID
AccountChangeTypeID

fLoanTransaction
LoanTransactionID
LoanID
DateTimeID
TransactionTypeID
LocationID
PrimaryCustomerID
SecondaryCustomerID
IsOpening
IsClosing
TransactionAmount
PostTransactionInterest
PostTransactionPrincipal

fLoanRate
LoanRateID
LoanID
DateTimeID
Rate

dAccount
AccountID
RowCreatedDateTime
RowUpdatedDateTime
IsCurrentVersion
AccountType

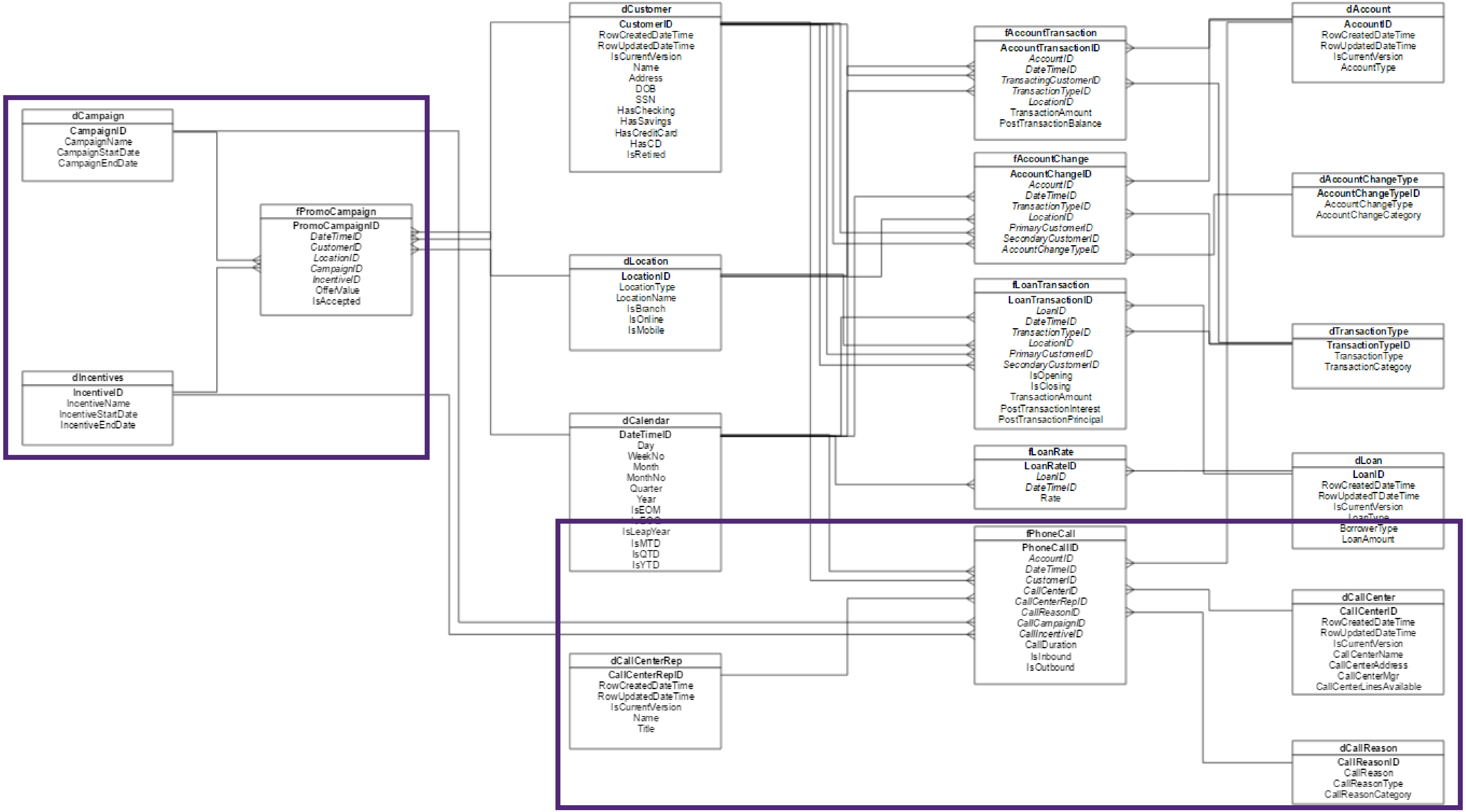
dAccountChangeType
AccountChangeTypeID
AccountChangeType
AccountChangeCategory

dTransactionType
TransactionTypeID
TransactionType
TransactionCategory

dLoan
LoanID
RowCreatedDateTime
RowUpdatedTDateTime
IsCurrentVersion
LoanType
BorrowerType
LoanAmount

Sample Data Model: Expansion

Bold = Primary Key
Italics = Foreign Key



dCustomer
CustomerID
RowCreatedDateTime
RowUpdatedDateTime
IsCurrentVersion
Name
Address
DOB
SSN
HasChecking
HasSavings
HasCreditCard
HasCD
IsRetired

dCampaign
CampaignID
CampaignName
CampaignStartDate
CampaignEndDate

fPromoCampaign
PromoCampaignID
DateTimeID
CustomerID
LocationID
CampaignID
IncentiveID
OfferValue
IsAccepted

dIncentives
IncentiveID
IncentiveName
IncentiveStartDate
IncentiveEndDate

dLocation
LocationID
LocationType
LocationName
IsBranch
IsOnline
IsMobile

dCalendar
DateTimeID
Day
WeekNo
Month
MonthNo
Quarter
Year
IsEOM
IsEOQ
IsLeapYear
IsMTD
IsQTD
IsYTD

dCallCenterRep
CallCenterRepID
RowCreatedDateTime
RowUpdatedDateTime
IsCurrentVersion
Name
Title

fAccountTransaction
AccountTransactionID
AccountID
DateTimeID
TransactingCustomerID
TransactionTypeID
LocationID
TransactionAmount
PostTransactionBalance

fAccountChange
AccountChangeID
AccountID
DateTimeID
TransactionTypeID
LocationID
PrimaryCustomerID
SecondaryCustomerID
AccountChangeTypeID

fLoanTransaction
LoanTransactionID
LoanID
DateTimeID
TransactionTypeID
LocationID
PrimaryCustomerID
SecondaryCustomerID
IsOpening
IsClosing
TransactionAmount
PostTransactionInterest
PostTransactionPrincipal

fLoanRate
LoanRateID
LoanID
DateTimeID
Rate

fPhoneCall
PhoneCallID
AccountID
DateTimeID
CustomerID
CallCenterID
CallCenterRepID
CallReasonID
CallCampaignID
CallIncentiveID
CallDuration
IsInbound
IsOutbound

dAccount
AccountID
RowCreatedDateTime
RowUpdatedDateTime
IsCurrentVersion
AccountType

dAccountChangeType
AccountChangeTypeID
AccountChangeType
AccountChangeCategory

dTransactionType
TransactionTypeID
TransactionType
TransactionCategory

dLoan
LoanID
RowCreatedDateTime
RowUpdatedTDateTime
IsCurrentVersion
LoanType
BorrowerType
LoanAmount

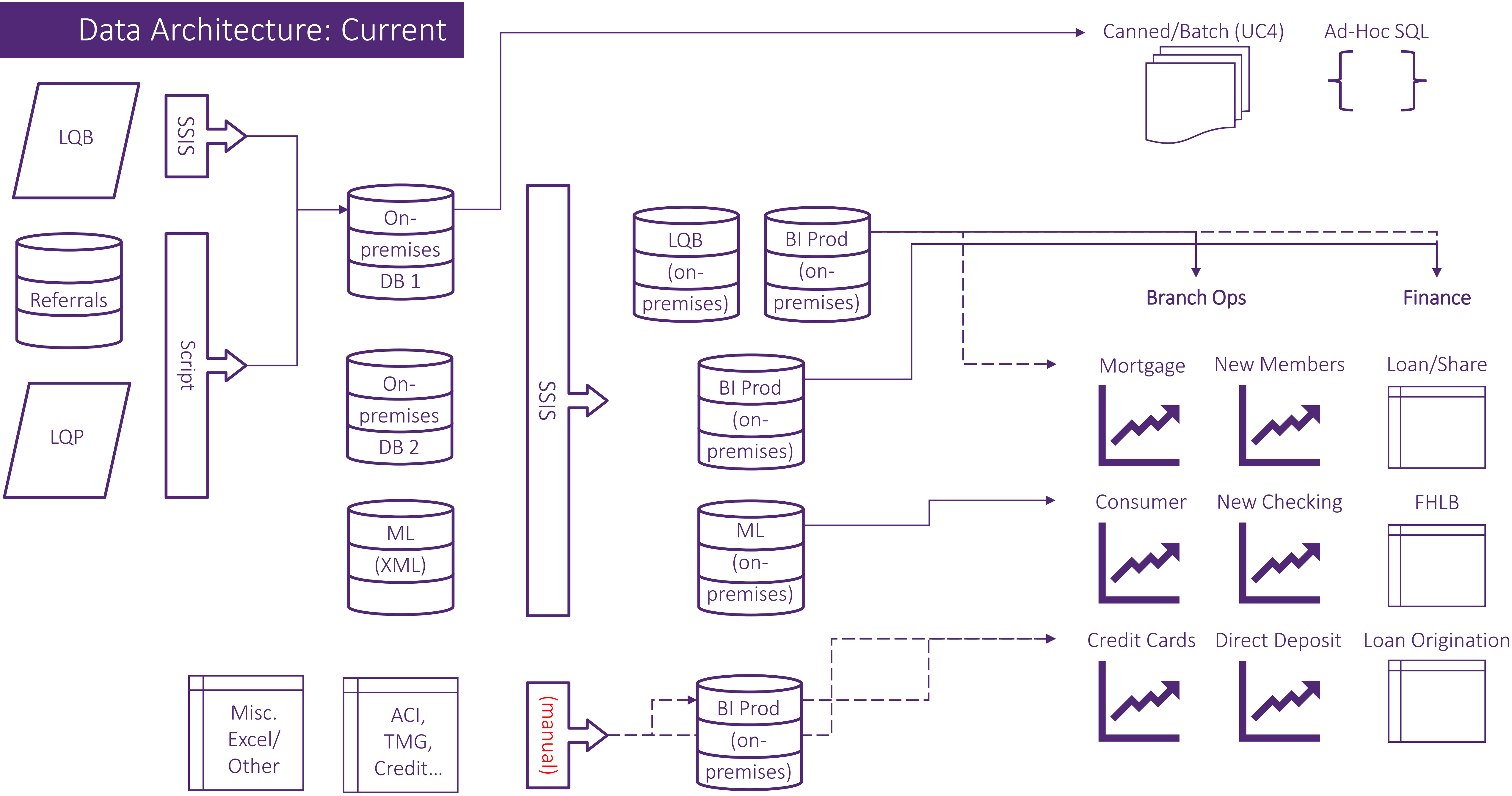
dCallCenter
CallCenterID
RowCreatedDateTime
RowUpdatedDateTime
IsCurrentVersion
CallCenterName
CallCenterAddress
CallCenterMgr
CallCenterLinesAvailable

dCallReason
CallReasonID
CallReason
CallReasonType
CallReasonCategory

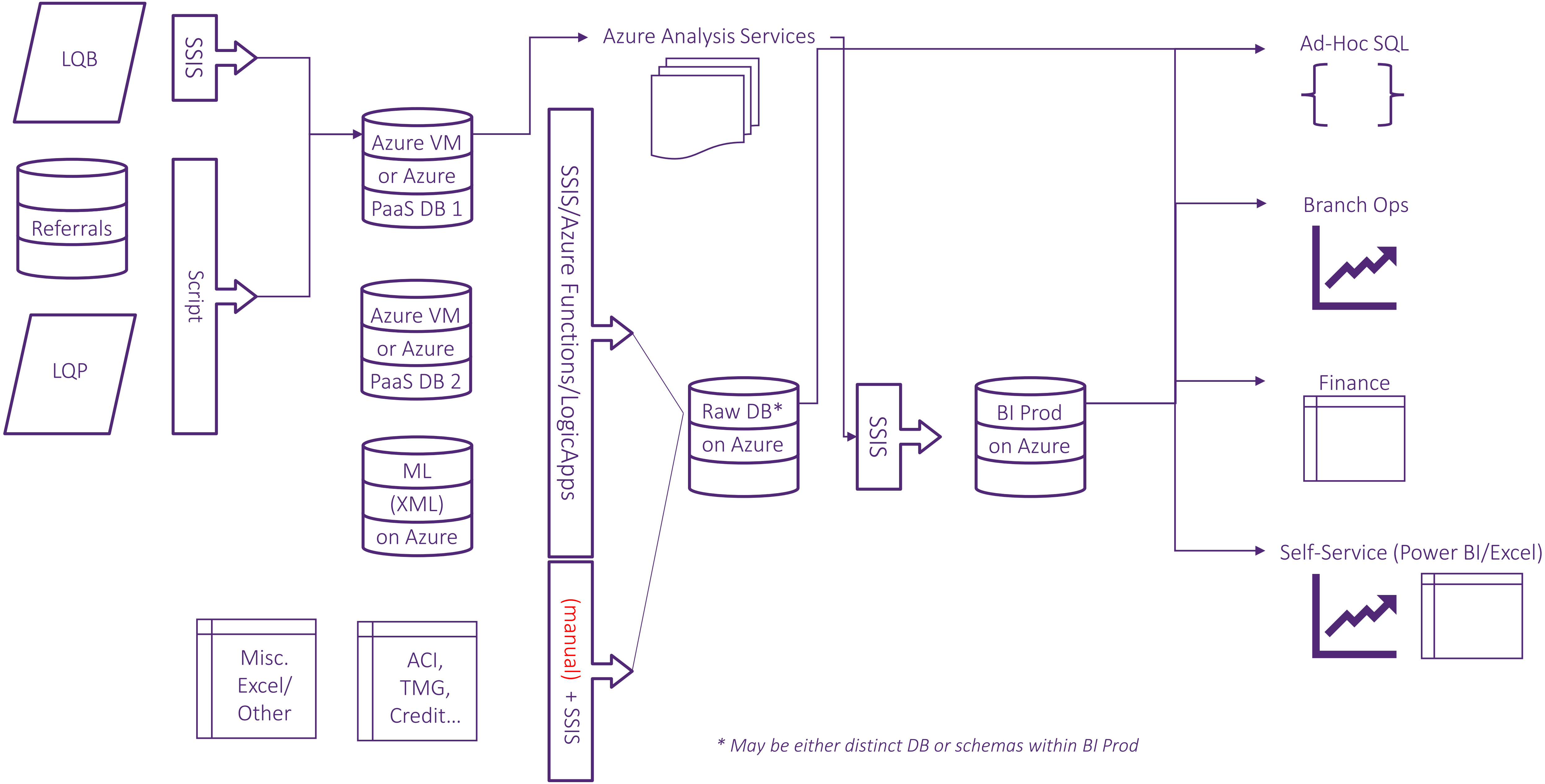
Sample Dimensions, Facts & Metrics for Bus Matrix

Primary Dimensions	Common Measures (Facts)	Metrics	Metric Logic
• Date • Branch • Originating Employee • Org Name • Org Number • Person Number • Account Number • Member Agreement Number • Major Type • Minor Type • Minor Description • Origination Date • Active Date • Credit Limit • Add Date • Date Last Maintained • Current Account Status • Contract Date • Product Group	• Member history • Checking accounts • Direct deposit history • Orig Balance • Refi Amount • Inv Balance • Budget	• New members • New members YTD • New checking accounts • New checks accounts YTD • New direct deposit accounts • New direct deposit accounts YTD • Last year actual loans • Last month actual loans • Current month actual loans • Variance current actual vs. budget • Current month budget • Variance current actual vs. year end budget • Year end budget • Variance current actual vs. last month actual	• (complex; no creation date per se) • Sum of new members for all dates YTD • Distinct count of members with checking account created in last 60 days • Sum of new checking accounts for all dates YTD • Distinct count of members with direct deposit >= $500 into a checking account that *was* new at some point within the last 60 days • Sum of new direct deposit accounts for all dates YTD

Data Architecture: Current

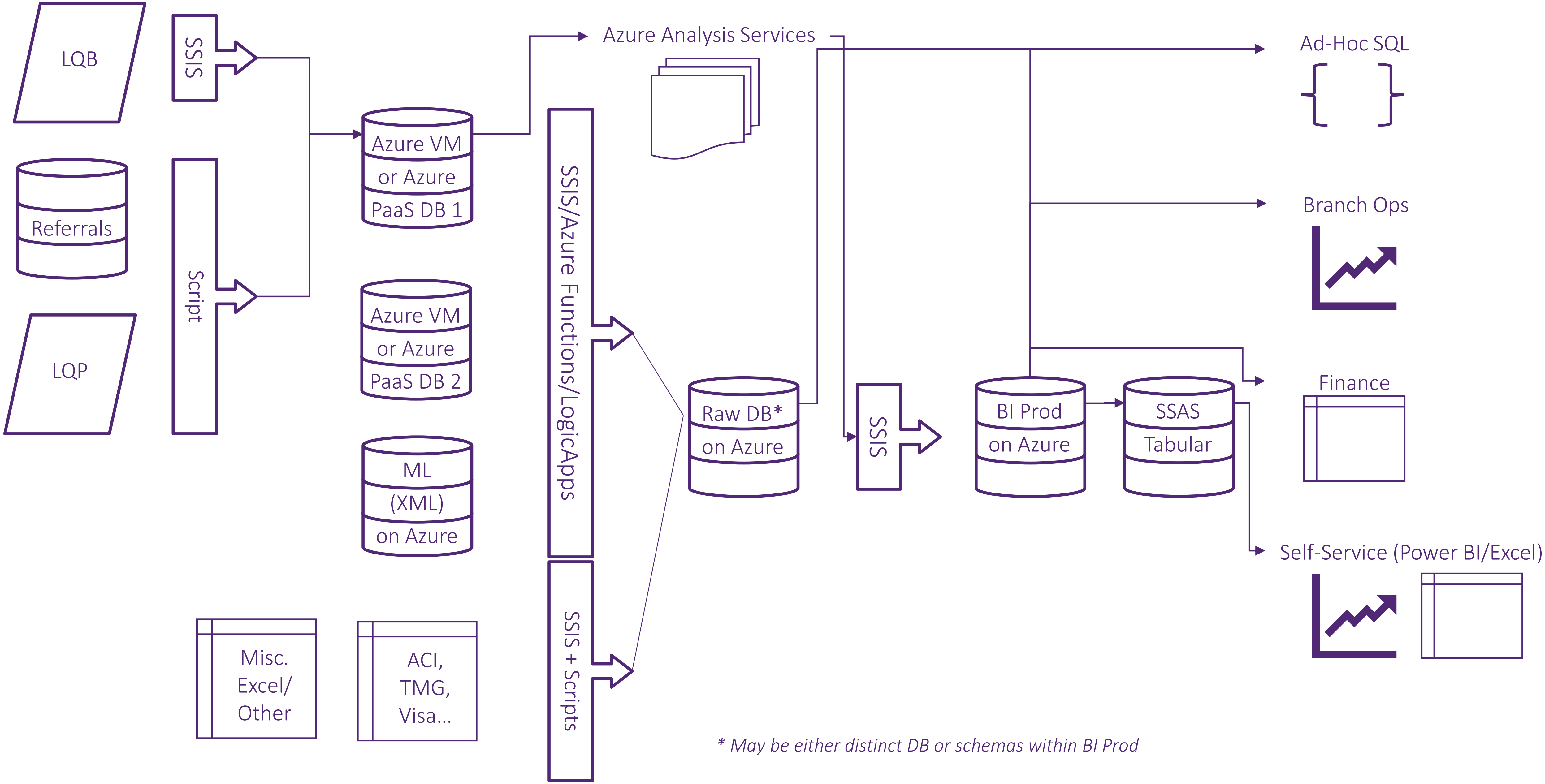


Data Architecture: 6 Months



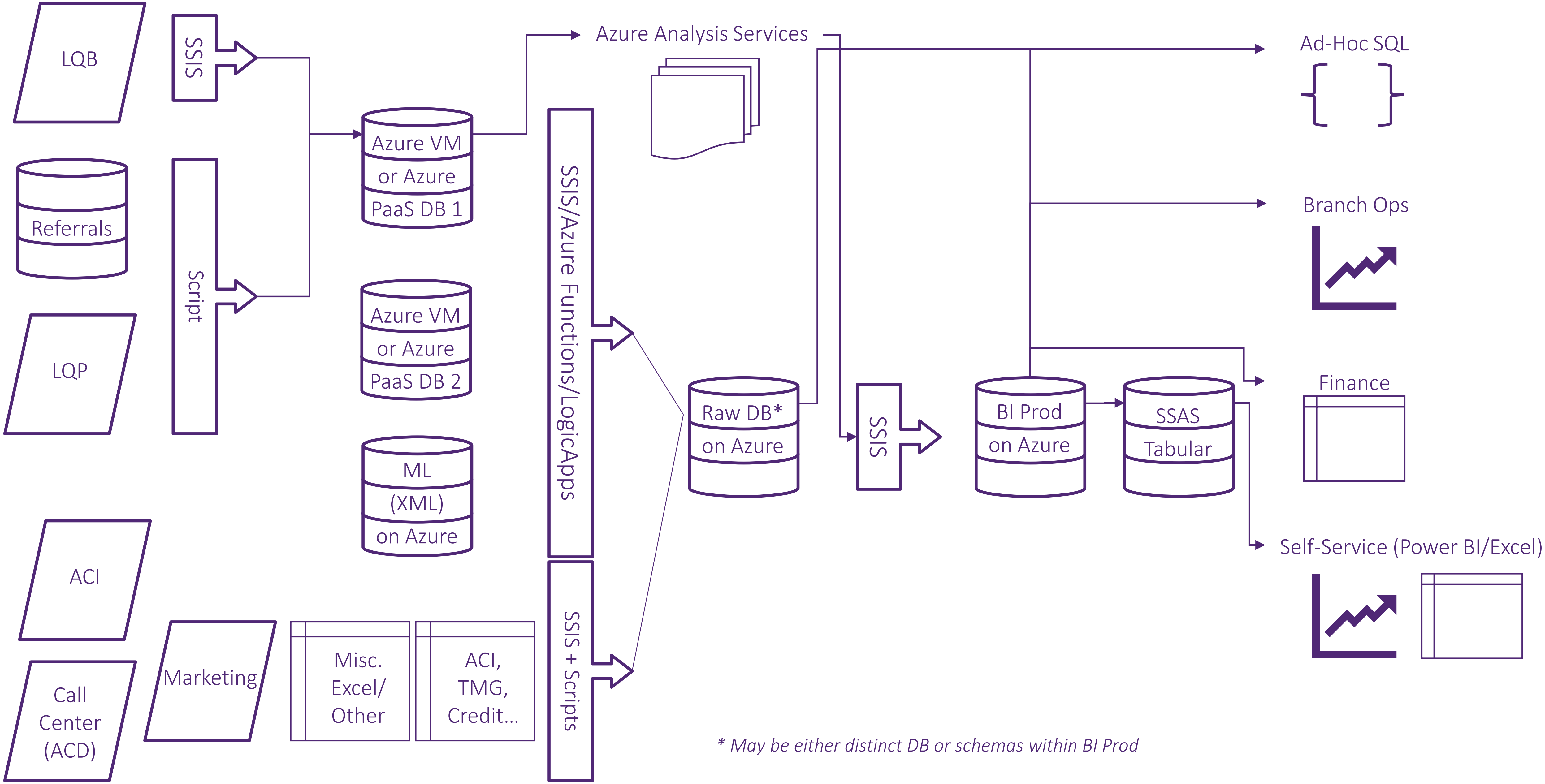
** May be either distinct DB or schemas within BI Prod*

Data Architecture: 1 Year



** May be either distinct DB or schemas within BI Prod*

Data Architecture: 2+ Years



** May be either distinct DB or schemas within BI Prod*

Comparison of ETL Approaches

***Bolded** information is especially relevant to Client X

	Cost	Maintenance	Performance	Usage	Flexibility	Features
Azure (specifically, Azure Data Factory)	• Relatively low compared to other options	• **Minimal maintenance**	• Build faster with the leading cloud platform	• Use any development tool or language	• **High (AutoScaling, etc.)** • **Easy integration with Power BI**	• **Virtual servers via "Virtual Machines"** • **Backend process logic via "Event Grid"** • **Scalability via "AutoScaling"** • **Many storage and backup/recovery options** • Orchestrates and automates movement and transformation of data from various sources
SQL Custom Code (Table ETL Framework)	• **Low** - does not require an additional application	• Less maintenance required • Fairly strong T-SQL knowledge required	• Slow processing for larger data sets	• Any SQL developer can use it (don't need knowledge of 3rd party tools) • Fairly fast to develop	• **Moderate** • Does not require Microsoft products / autonomy from vendors • Flexibility to integrate with other products (such as SSIS and Azure Data Factory)	• **No additional tools needed** • Flexibility to integrate with other products (such as SSIS and Azure Data Factory) • Superior logging capabilities • Most code is generated automatically • Good control flow and branching features (notifications, error message emails) • **Easy to incorporate new data sets**
SQL Server Integration Services (SSIS)	• **None** – included with SQL Server	• Environment management is more complex • **Tedious to implement and manage**	• **Faster performance for large data sets** • More processing power than SQL (run parallel processes, larger batch loads than SQL)	• Allows a less technical developer to define ETL process (drag-and-drop-features)	• **High** • Dependent on Microsoft products	• **Included in SQL Server license** • Can pair with custom SQL ETL Framework • Extensive control flow and branching features (notifications, error message emails)

Recommendations

ETL Approach

Azure Data Factory

Client X needs a robust ETL platform to serve as a foundation for reporting today and a conduit for new data in the future. Azure Data Factory accomplishes this.

Infrastructure Requirements

SQL Server + Analysis Services

We recommend that Client X use its existing SQL Server Standard license as the foundation of its data warehouse. A single server will suffice for now, but multiple servers may be prudent once the DW reaches enterprise scale. SSAS is highly recommended if Client X will make significant use of tabular-compatible tools like Excel. An SSAS tabular model is not strictly necessary, but will streamline data governance and security for self-service.

Visualization Software

Power BI

Power BI offers the greatest visual flexibility currently available. We recommend that Client X a) continue using it for visualization and b) identify potential "power users" to train internally in the next 6-12 months.

Change & Request Mgmt

Visual Studio + Visual Studio Team Services/Other

The desktop Visual Studio IDE is optimal for tabular cube development, and works smoothly with SQL Server database projects, for both Git and TFS repos. Visual Studio Team Services (formerly Visual Studio Online) is one of many choices for web-based project management/task tracking, but we recommend it for use alongside Visual Studio. As a simple and low- or no-cost option, Git and Trello may work well, albeit through a less polished UI.

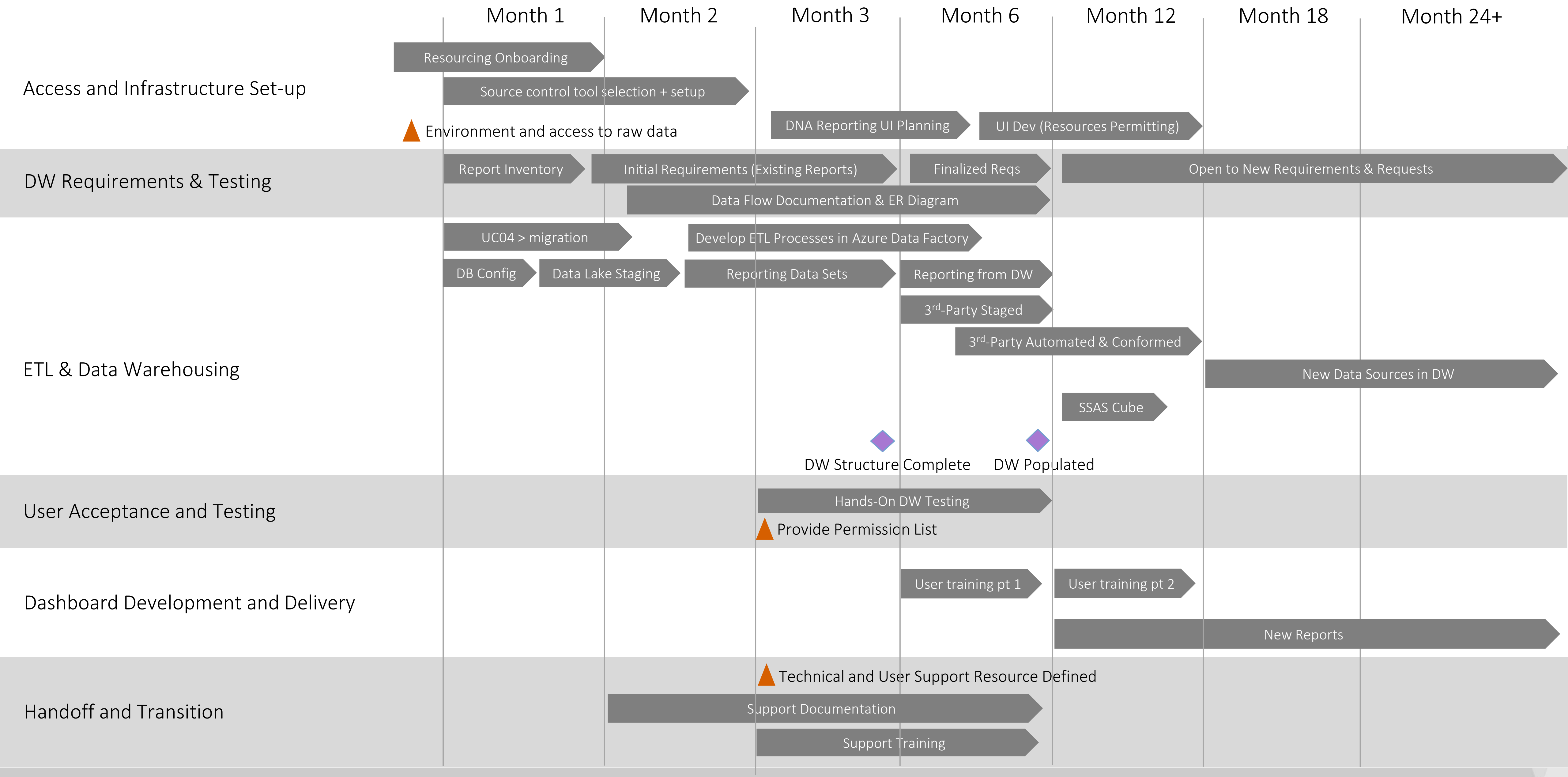
Project Phases

Phase	Time From Start	Milestones
1	3 months	• DW requirements and bus matrix are taking shape, and DW is in active development • Source control + change mgmt. tool is in use • UC4-automated reports migrated to scheduled Azure Data Factory workflows • All objects that support existing reports are staged in the raw DB/schemas • Final reporting data sets are consolidated into biprod, and it has been communicated that all reports should connect to the reporting schema • Where feasible, raw third-party sources (e.g., credit card data) are staged via Azure Data Factory
2	6 months	• DW requirements are solidified and initial DW iteration is in production • Reporting data sets are being rebuilt from conformed DW • Users have received basic Power BI/Excel training for self-service using DW • Third-party sources previously staged have been conformed and pushed to DW • DNA reporting UI has been scoped out, and development begins (if resources permit)
3	1 year	• Tabular cube is in production • Users have received additional training for self-service using cube • Third-party sources that were not feasible in Azure Data Factory have been automated and conformed through custom scripts
4	2+ years	• New sources (e.g., call center, ACI, marketing) are available in the conformed DW

Approximate Project Timeline

◆ DD Deliverable

▲ Client Deliverable



Month 1 | Month 2 | Month 3 | Month 6 | Month 12 | Month 18 | Month 24+

Access and Infrastructure Set-up

- Resourcing Onboarding
- Source control tool selection + setup
- ▲ Environment and access to raw data
- DNA Reporting UI Planning
- UI Dev (Resources Permitting)

DW Requirements & Testing

- Report Inventory
- Initial Requirements (Existing Reports)
- Finalized Reqs
- Open to New Requirements & Requests
- Data Flow Documentation & ER Diagram

ETL & Data Warehousing

- UC04 > migration
- Develop ETL Processes in Azure Data Factory
- DB Config
- Data Lake Staging
- Reporting Data Sets
- Reporting from DW
- 3rd-Party Staged
- 3rd-Party Automated & Conformed
- New Data Sources in DW
- SSAS Cube
- ◆ DW Structure Complete
- ◆ DW Populated

User Acceptance and Testing

- Hands-On DW Testing
- ▲ Provide Permission List

Dashboard Development and Delivery

- User training pt 1
- User training pt 2
- New Reports

Handoff and Transition

- ▲ Technical and User Support Resource Defined
- Support Documentation
- Support Training

RACI Matrix

Responsible Persons whose contributions result in completion of a task

Accountable Members whose approval is required to complete an activity

Consulted Individuals who play an indirect role by contributing their knowledge

Informed Little to no direct involvement, but must be made aware of activities

	CIO	VP of IT	PM	Data Lead	Data SMEs	BI Lead	BI Dev	Finance Lead	Retail Lead
Sponsorship and high-level advocacy	R	C	C	I					
Software & Hardware	I	R	C	R		C			
General oversight	C	C	R	R		R			
Requirements and definitions	I	C	I	R	C	R	C	C	C
Report inventory maintenance			I	I	C	C		R	R
Report creation				R	R	R	C	A	A
Report validation	I	I	I	C	C	C	C	RA	RA
DW architecture	I	A	I	R	C	R	C	I	I
DW implementation	I	I	C	R	C	C	R		
DW maintenance				R	C	C	C		
Data validation (upstream)			I	R	C	RA	R	C	C
Data governance	I	C	C	C	I	R	C	C	C
Security	A	A	C	R	C	R	C	C	C
Data flow documentation - initial			I	RA	R	RA	R	RA	RA
Data flow documentation - expansion and auditing				C	I	RA	C	RA	RA
Change and version control	I	I	C	RA	C	RA	R	R	R
End-user DW documentation - initial			A	AC	C	RA	C	I	I
End-user DW documentation - expansion and auditing			C	AC	C	RA	C	I	I

Applicable BI Skills & Gaps

Skill	Availability	Comments
BI/DW expertise	Contract	Most expertise currently from contractors. Employees are learning but have limited time. Training highly recommended. We do not provide general BI/DW training, but do give extensive training on managing all solutions we work on.
MSBI stack expertise	N/A	SQL + Azure Data Factory is sufficient for now, but MSBI expertise will be needed for tabular cube development and, potentially, Excel report development. Training highly recommended. As above, DD does not do general product training but does offer practical training on any MSBI solution we work on.
Azure Data Factory development	Contract + In-House	Both contract and in-house resources, but time is limited.
T-SQL development	In-House	Strong knowledge within Applications team, again subject to limited time.
Visualization and dashboard development	Contract	Existing resources are enough, but user training is recommend once self-service becomes available. DD provides both general product training and solution-specific training as needed.

Solution Risks & Mitigation Steps

Risk	Likely Impact	Comments & Mitigation Steps
Resource limitations	Very High	• If the RACI matrix is realistic, then the Data Lead and BI Lead roles will be greatly overburdened • For this timeline, we recommend the equivalent of a) one full-time business analyst (with strong technical background) and b) one part- or full-time BI developer with Azure Data Factory and MSBI experience
Lack of consistent code and change mgmt. process	Very High	• After implementing whichever source control + change mgmt. tool Client X prefers, use it universally • Maintaining *zero* deviation is absolutely critical for success!
Inconsistent requirements gathering practices	High	• Challenging to "translate" between business and data requirements • [Client Name 1] and [Client Name 2] have templates and practices that are excellent (we recommend using them universally) but foreign to Applications team • Important to identify one person to oversee *all* requirements gathering and to train Applications team on methodology
Demand for "dashboard-first" approach	High	• Communicate that slower report development now (1-2 quarters including reqs gathering) means faster, more valuable reporting in the long run • Emphasize benefits of self-service • Keep progress/statuses highly visible
Gap between internal technical skills and internal BI/DW knowledge	Moderate	• Encourage training (whether external or on-site) in DW methodology for Applications team • Adhere to a bus matrix as a bridge between data and business SMEs
Lack of familiarity or clarity discourages end users	Moderate	• Document with painstaking clarity from day one per standards of [Client Name 1] and [Client Name 2] • Provide internal training on documentation methodology (*especially* critical if no additional resources can be hired) • Vet documentation with end users • Consider brief end-user training sessions, especially if/when SSAS is live
Slow DW delivery due to scope creep	Moderate	• Start small, scale gradually, and validate as you build • Err on the side of more, rather than less, time for requirements gathering

Decisive Data