



Microsoft Analytics Platform System

POC Module 1 & 2

Course Modules and Objectives

- **Module 1: APS POC Overview**
Establishing Goals And The POC Process
- **Module 2: Planning The APS POC**
Customer Expectations, Scoping & Collecting Artifacts
- **Module 3: POC Execution: Phase I**
POC Preparation & Kickoff, Working With Base Data
- **Module 4: POC Benchmarking in APS**
Tools for Testing and Monitoring, Query Concurrency

Module 1: APS POC Overview

Module 1: Lessons and Overview

- APS POC Overview

Describe the elements of a Successful POC, definition of team and roles

- APS Sales and Delivery Lifecycle

Describe how the APS POC fits into the APS Sales and Delivery lifecycle.

- APS POC Process Overview

Describe the process for scheduling and conducting an APS POC.

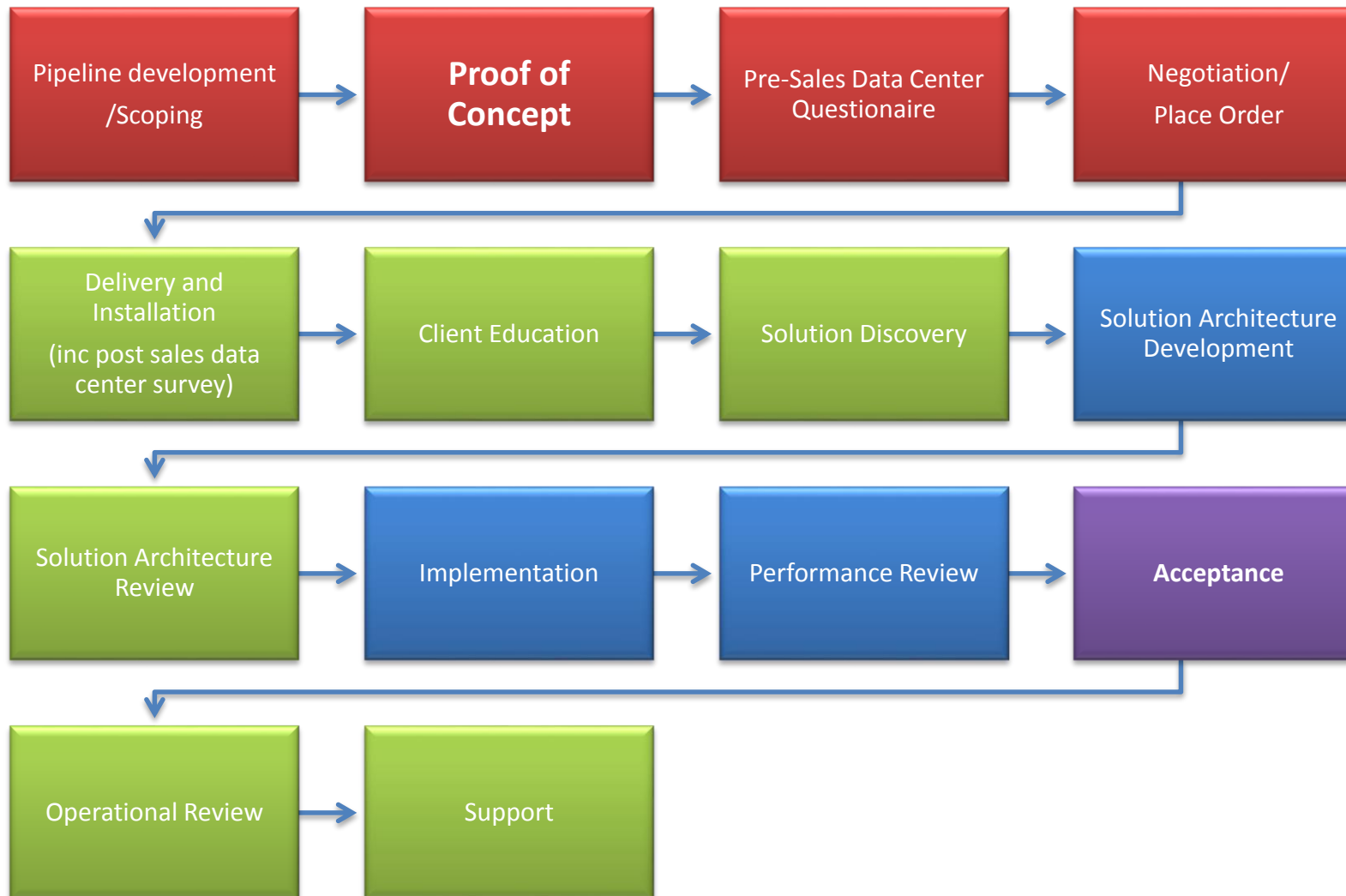
Elements of a Successful POC

- Set up for Success
 - Value Established
 - Proper Resources Allocated and Committed
- Customer Needs Clearly Understood
- Customer Expectations Clearly Set
- POC Scope
 - Clearly Defined Objectives
 - Clearly Defined Execution Process
 - Scope-creep Minimized
- Customer Engaged and Fully Involved
- Regular Progress Updates & Reviews
- Transparency
- Defined Issue Resolution Strategy
 - Process
 - Roles & Responsibilities

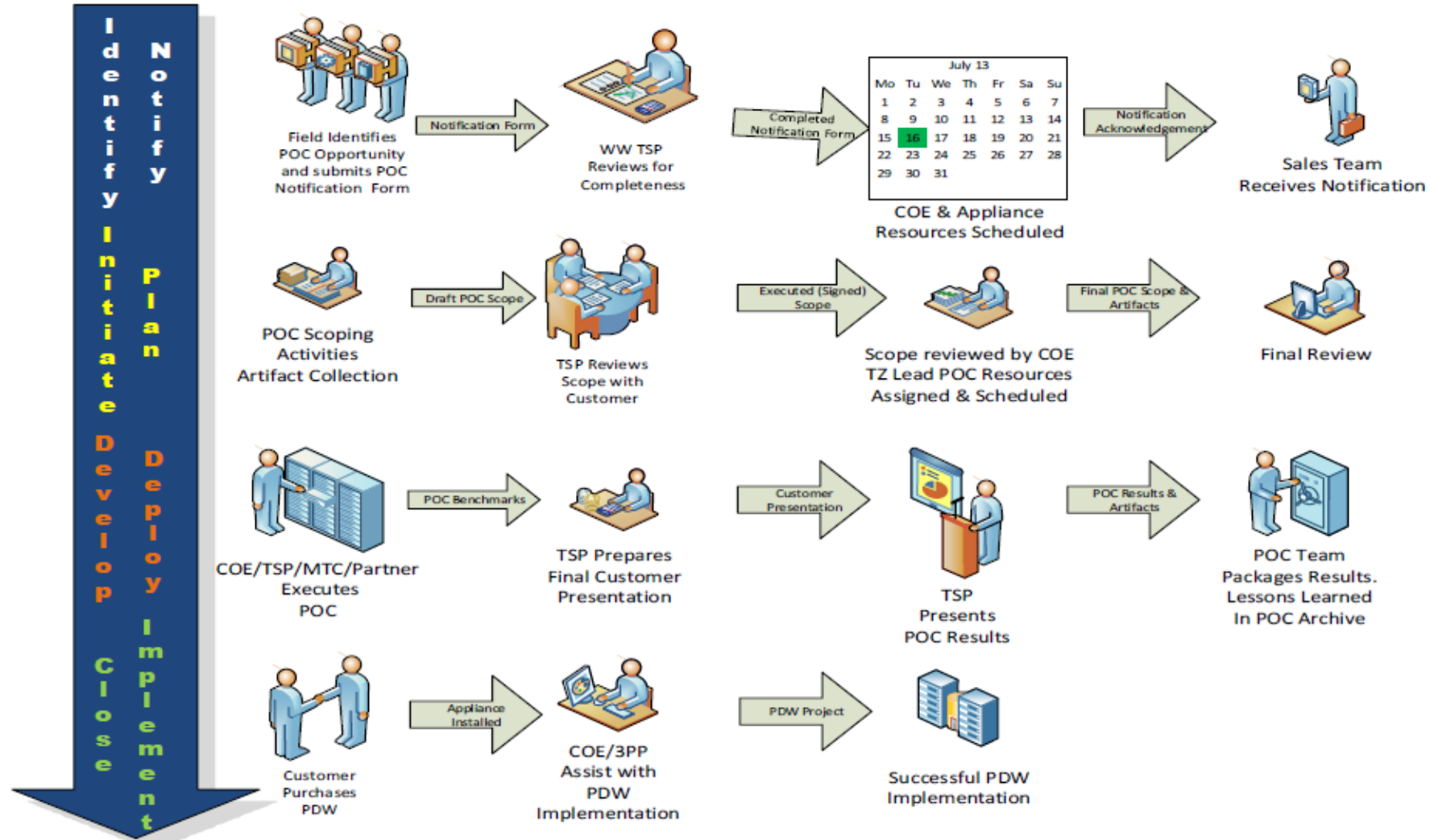
POC Teams and Roles

- Technology Solutions Professional (TSP)
 - Drives POC scope and customer CSF's
 - Monitors POC execution to preempt potential issues
 - Coordinates POC changes between customer and POC Lead
- POC Lead
 - Primary Design Architect
 - Establishes & Drives POC execution plan
 - Reports status regularly
- Client Technical Team
 - Do not bring in client until appliance is set up and the use-case with artifacts has been tested and run at least once
 - Might or might not participate on POC
- Microsoft Support – Premier Field Engineering (PFE) and Dedicated Support Engineering (DSE)
 - Premier Mission Critical Support supports appliance as a production client
 - Any appliance issues are handled by support

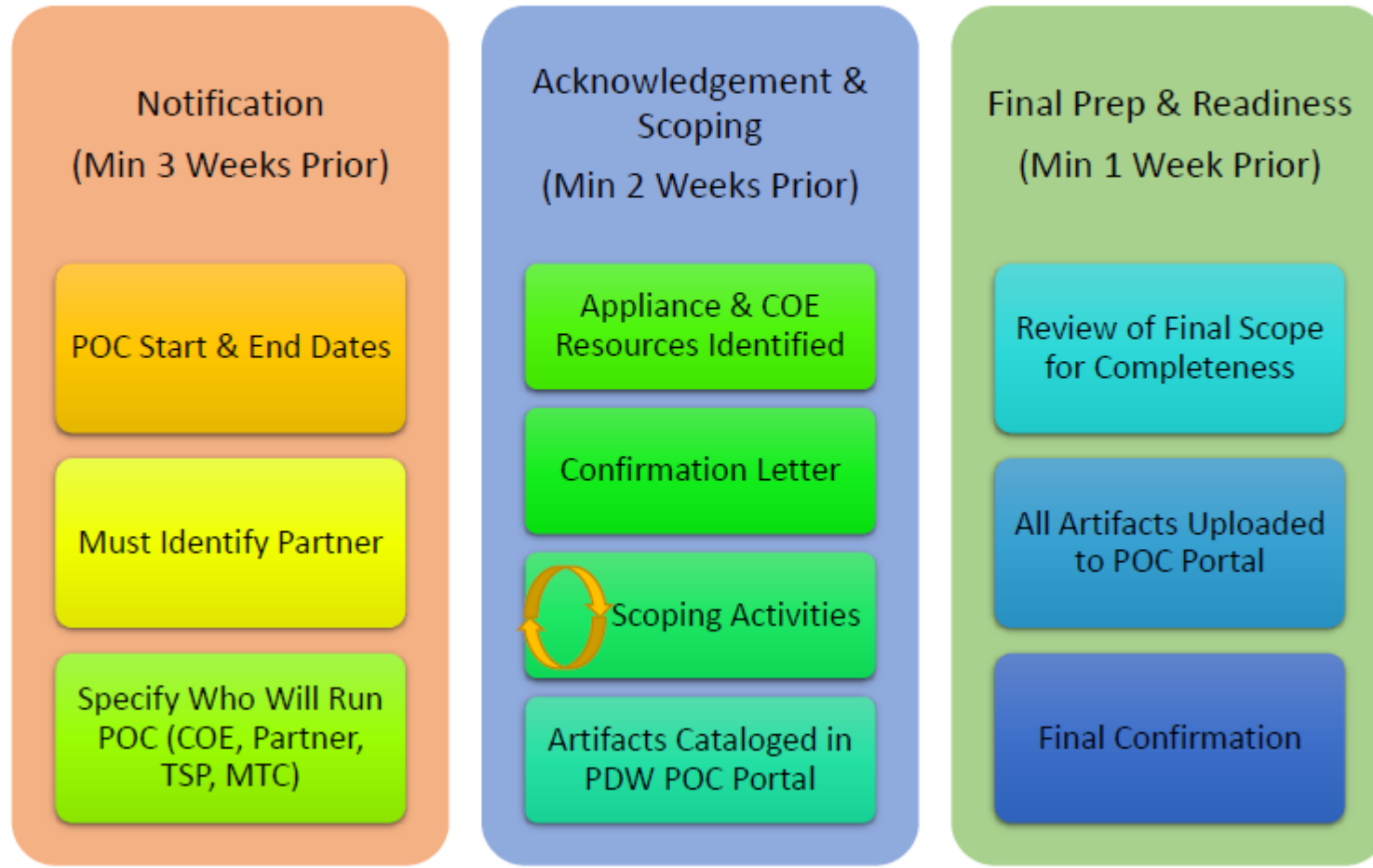
APS Sales and Delivery Lifecycle



PDW POC Process



POC Notification Process



Questions

Module Summary

In this module, you learned...

- The goals, particularities and context of a APS POC
- The steps required to deliver a APS POC
- The general roles and responsibilities of the APS POC team

Module 2: Planning the APS POC

Course Module and Objectives

- Module 1: APS POC Overview
Establishing Goals And The POC Process
- **Module 2: Planning The APS POC**
Customer Expectations, Scoping & Collecting Artifacts
- Module 3: POC Execution: Phase I
POC Preparation & Kickoff, Working With Base Data
- Module 4: POC Benchmarking in APS
Tools for Testing and Monitoring, Query Concurrency

Module 2: Lessons and Overview

- POC Goals

Understanding customer requirements & setting proper expectations

- Developing a Proper POC Scope

Determining CSF's and POC Success Criteria

- Developing the POC Work plan

Building a realistic POC project schedule and data strategy

- Gathering Customer Artifacts

Understand best practices to gather, analyze & organize POC Artifacts

Goals of the APS POC

Conversation Points and Critical Success Factors

- Current DW Design & Technology
- Future Capacity
- Load Performance
- Cost of Ownership
- High Availability
- DW Maintenance
- Disaster Recovery
- Smaller Batch Windows
- Ease of Use
- Mixed Query Workloads
- Query Performance
- Concurrency

Getting the Customer Committed to POC Process

Step	Task Description	Roles			Estimated Task Timing									
		Client	POC Lead	TSP	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6				
1	Initial Planning Discussion	X	X	X										
2	Initial Inventory													
	Database DDL	X	X	X										
	Sample Data	X	X	X										
	Sample Queries	X	X	X										
	Create Sample Database		X	X										
	Load Sample Data		X	X										
	Run Sample Queries		X	X										
3	Generate Scoping Document		X											
4	Physical DB Design & Review	X	X	X										
5	Review Scoping & Sign-off	X	X	X										
6	Primary Data Acquisition													
	Formally Approve Sample Data	X	X	X										
	Extract Data/FWD NAS Device	X												
7	Prepare/Verify Appliance		X											
	Download Data Onto Load Server													
	Verify Data Integrity		X											
8	Initial Data Load													
	Load Data		X											
	Compile Load Benchmarks		X											
	Publish Load Benchmarks	X	X	X										
9	Run Baseline Queries													
	Execute Queries													
	Compile Query Timings		X											
	Publish Query Timings	X	X											
10	Data Expansion													
11	Re-Run Baseline Queries													
	Execute Queries													
	Compile Query Timings		X											
	Publish Query Timings	X	X											
12	Concurrency Testing													
	Prep Extracts for Incr Loads		X											
	Prepare Concurrency Matrix		X											
	Execute Concurrency Tests													
	Compile Concurrency Timing		X											
	Publish Concurrency Timing	X	X											
13	Miscellaneous POC Proofs													
	Backup/Restore													
	Failover													
	Partition Split/Merge													
	ETL/ELT													
	SSIS/SSAS/SSRS Integration													
	Table Maint (CTAS) Operations													
14	Final PDW Benchmark Reviews	X	X											



PDW POC Process Template

The POC Scoping Document – Overview and Contents

- “Living Document” – Keep it Updated!
- Brief Company Overview and Key Background
- Critical Success Factors
 - Business Needs
 - Use-Cases
- Success Criteria (How We Will Win)
 - SLA's
 - Key Proof Points (Stated Empirically)
- Describes Existing Environment
- POC Environment
- POC Execution Plan
 - Roles/Responsibilities
 - Dates
 - Deliverables
- Key Milestones
- Appendices (Supporting Docs)

The thumbnails show the following content:

- Top Left:** Title page for 'A. Datum Corporation PDW POC Scope Document' dated October 13, 2011.
- Top Middle:** Table of Contents with sections like Project Overview, Success Criteria, etc.
- Top Right:** Project Overview table with columns for Project Name, Status, and Date.
- Bottom Left:** Project Charter section with a diagram showing the project flow.
- Bottom Middle:** Project Charter section with a diagram showing the project flow.
- Bottom Right:** Project Charter section with a diagram showing the project flow.



PDW POC
Scoping Template

Critical Success Factors

What are Good CSF's?

- Driven by Business Objectives
- Clear/Succinct
- Measurable/Valuable

Examples

- Improved analytics efficiency will contribute 10% to bottom line.
- Reduced COE will reduce overall expenses by 7%
- Replacing more expensive EDW solution will remove 35% from our IT budget.
- Centralized/Consolidated EDW will increase overall productivity by 20%
- Less expensive environment to meet government regulations and archive data for 7 years reduces expenses by 8%
- Ability to quickly adapt to variable data/computing capacity will save 14% of variable expenses.

POC Success Criteria

What are Good POC Success Criteria?

- Specific Proof Points
- Highlights APS Strengths
- Differentiate Against Competitor(s)
- Outline of key POC Tasks
- Reasonable

Examples

- Overall data Load times must be reduced by 20%
- SSAS Cube XYZ refresh must improve by 15%
- Key analytic query performance must improve by 10X.
- System workload/throughput must improve by 25%
- Full Database Backup must be reduced by 3 hours.

Limiting the POC Scope

- Differentiate Between A POC And An Implementation
- Distinguish Between ETL/ELT And Data Loading
- Limit Conversion/Refactoring Of SP's To No More Than 5
- Consider Logistics Of 3rd Party Integration – Try To Eliminate
 - Licensing/Installation/Setup
 - SME For Tool Usage
 - Platform Differences
 - Adds Minimum Of 5 Days To POC
- SSIS/SSAS/SSRS Integrations Add 3-5 Days To POC
- Keep Number Of Tables < 50
 - Data Must Be Validated Against Each One
- Reasonable Number Of Queries – < 30 Recommended
 - Small/Medium/Large
 - Representative
- Failover Is “Software Only”

Onsite POC Considerations

Avoid Onsite POC's !!

... But if client is qualified and Onsite is unavoidable ...

- Plan for Onsite Survey (Who will perform it?)
 - Where will appliance be received?
 - Power/Thermal requirements
 - Where will appliance live?
 - Where will packing materials be installed?
- Travel logistics/costs of Systems Engineer
- Who pays for transport costs?
- VPN connectivity for support?
- Security clearance for POC team?

POC Data Strategy: Logistics

- How will data files be supplied/transported?
- What are the file formats?
- Delimited? Zipped? SQL Server Backup? (not a good idea!)
 - Make sure you get the schema
- How will we handle HADOOP Clusters?
- What are the obfuscation (Anonymization) requirements?
- How long will it take to perform extracts?
 - Impact on Production System
 - Time to upload files to transport drives
- What is the initial DB size versus final DB size?
- What are the plans to incrementally load additional data?
- What is the data load frequency (near real-time, daily, hourly, etc.)
- What are the data load volumes (Initial, Incremental)

POC Data Strategy: Expansion

- “Standard” Data Expansion Using INSERT/UPDATE Should Be Used Judiciously In APS
 - Serialized Operation
 - Logged Operation
 - Roll-back
- Always Try To Utilize CTAS
 - Parallelized
 - Minimal Logging
 - Minimal TEMPDB Use.
- Common Expansion Scenarios
 - See Word Document



**PDW Data
xpansion Strategie**

Collecting Customer-Provided “Standard” Artifacts

- Database DDL
 - Table Definitions
 - Partition Definitions
 - PK/FK Definitions
 - Indexes (Clustering, Non-Clustering)
 - UDT's/UDF's
- Queries
 - Small/Medium/Large
 - Representative Workloads
 - Serve as “Baseline” throughout POC
- Sample Data
 - Flat files (Resist DB Backups)
 - Need sample data for ALL tables used in POC
 - Determine proper column delimiters before client performs extract

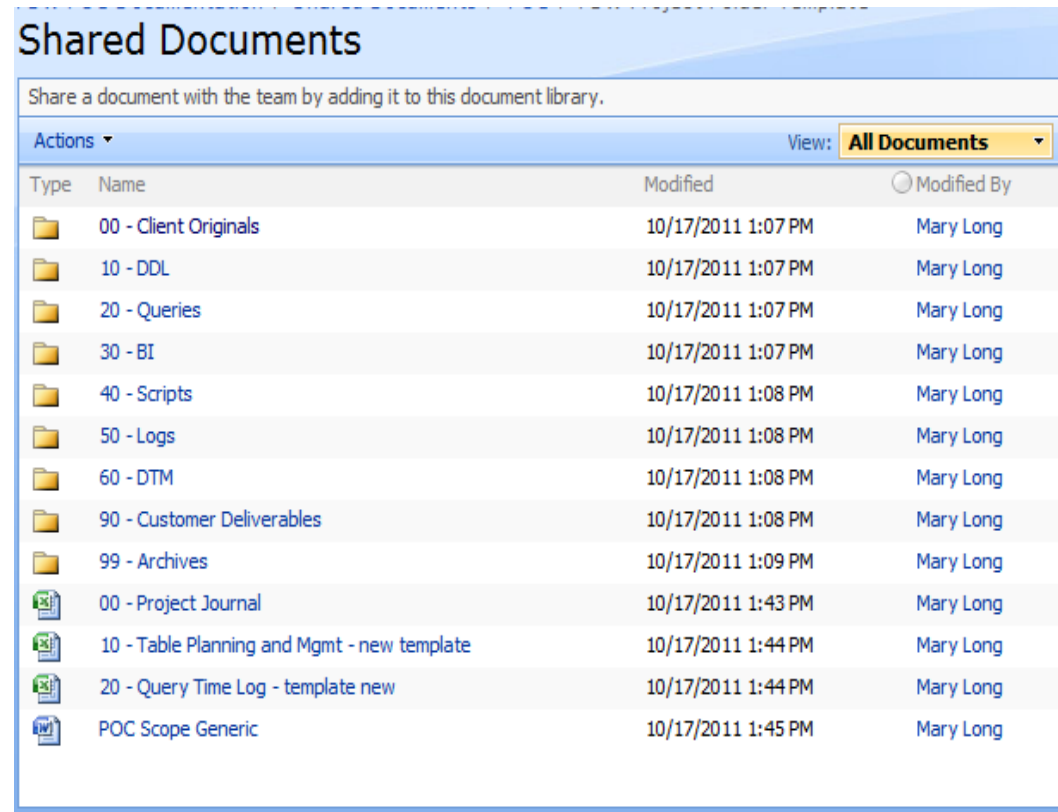
Collecting Customer-Provided “Extended” Artifacts

- SSIS/SSAS/SSRS Packages
 - Package Definitions
 - Stored Procedures
 - Cube RI Requirements
- ETL Tools
- Third Party BI Tools
 - Business Objects
 - Cognos
 - MicroStrategy
 - How will you analyze these components

The APS SME SharePoint Site

It is recommended that a SharePoint site be established to organize and hold documents and results

- Captures all Customer Artifacts
- Organized consistently across all APSPOC's
- Should be refreshed regularly
- Documentation Templates:
- POC Journal
- POC Process & Scoping Document
- Table Planning & Management
- Benchmark Results
- Sample Customer Presentation



The screenshot shows a SharePoint 'Shared Documents' library. At the top, there is a header 'Shared Documents' and a sub-header 'Share a document with the team by adding it to this document library.' Below this, there is an 'Actions' dropdown menu and a 'View: All Documents' dropdown menu. The main content is a table with four columns: 'Type', 'Name', 'Modified', and 'Modified By'. The table lists various folders and documents, all of which were modified by 'Mary Long' on '10/17/2011'.

Type	Name	Modified	Modified By
Folder	00 - Client Originals	10/17/2011 1:07 PM	Mary Long
Folder	10 - DDL	10/17/2011 1:07 PM	Mary Long
Folder	20 - Queries	10/17/2011 1:07 PM	Mary Long
Folder	30 - BI	10/17/2011 1:07 PM	Mary Long
Folder	40 - Scripts	10/17/2011 1:08 PM	Mary Long
Folder	50 - Logs	10/17/2011 1:08 PM	Mary Long
Folder	60 - DTM	10/17/2011 1:08 PM	Mary Long
Folder	90 - Customer Deliverables	10/17/2011 1:08 PM	Mary Long
Folder	99 - Archives	10/17/2011 1:09 PM	Mary Long
Document	00 - Project Journal	10/17/2011 1:43 PM	Mary Long
Document	10 - Table Planning and Mgmt - new template	10/17/2011 1:44 PM	Mary Long
Document	20 - Query Time Log - template new	10/17/2011 1:44 PM	Mary Long
Document	POC Scope Generic	10/17/2011 1:45 PM	Mary Long

Validating/Collating/Cataloging DDL

- DDL Validation
 - Use conversion /migration utilities if available
 - Consider candidate distribution keys from PK/FK/CI/Index Specifications
 - Identify unsupported datatypes (See User manual) and resolve with customer
 - Identify UDT's/UDF's and resolve
 - Ensure all table definitions are present
 - Normalize Data Types: Validate data type consistency between tables
- DDL Collation/Cataloging
 - Remove extraneous objects reducing table definitions to Tables & Distribution
 - Optimize column sizes if appropriate (e.g. reduce VARCHAR (2000) to VARCHAR (256))
 - DMS performance greatly affected by oversized columns
 - Normalize data types
 - Record Rowcounts (Initial/expanded, etc.)
 - Develop *initial* distribution model based on sample data
 - Consolidate DDL into a single definition.
 - Consider generating DB Views

Validating/Collating/Cataloging Queries

- Query Validation
 - Ensure you have entire set of queries needed
 - Check Query structures and determine potential issues
 - Unrestricted Queries
 - Outer Joins
 - Joins/Distribution strategy
 - Potential Skew on Join
 - Capture DSQL Plans
 - Double-check for data inconsistencies on join columns
- Query Collation/Cataloging
 - Enumerate Queries Q01, Q02, etc.
 - Insert (OPTION (LABEL = 'QXX – Query Description') where QXX is the query number into each query to identify in Admin Console
 - Consider modifying to use DB Views
 - Store DSQL plans as "initial DSQL"

Working with Sample Data

- Ensure Consistency Between Data And Column Definitions
- Request A “Representative” Data Sample Ideally To Cover Unique Data Characteristics
- Request That Extracted Sample Data Is Relationally Integrated
(Don't Just Simply SELECT TOP 100 * From Each Table)
- Perform Cardinality Checks On Candidate Columns To Determine Initial Distribution Strategy
- Observe And Record Compression Ratio's For Large Distributed Tables.

Table Planning and Management

- Record All POC Tables
- Pre-determine Compression Ratios
- Perform Cardinality checks on candidate columns to determine initial distribution strategy
- Observe and record compression ratio's for large distributed tables.

Table	Table Type	Distribute	Partition	Clustered Index	Rows Loaded	Actual Table Size (MB)	Compression Ratio
DimDate	Replicated			DateKey	11,880	0.60	0.00
DimDepartmentGroup	Replicated			DepartmentGroupKey	70	0.20	0.00
DimEmployee	Replicated			EmployeeKey	2,960	0.60	0.00
DimGeography	Replicated			GeographyKey	6,550	0.40	0.00
DimOrganization	Replicated			OrganizationKey	140	0.20	0.00
DimProduct	Replicated			ProductKey	6,060	2.70	0.00
FactCallCenter	Distributed	FactCallCenterID		DateKey	113	1.20	0.00
FactCurrencyRate	Replicated			CurrencyKey	142,640	3.30	0.00
FactFinance	Distributed	FinanceKey		DateKey	39,466	2.40	0.00
FactInternetSales	Distributed	SalesOrderNumber		OrderDateKey	60,222	4.00	0.00
FactInternetSalesReason	Distributed	SalesOrderNumber		SalesOrderNumber	64,359	1.60	0.00
FactResellerSales	Distributed	SalesOrderNumber	OrderDateKey	OrderDateKey	182,754,354	9,127.50	3.40
FactSalesQuota	Replicated			EmployeeKey	1,630	0.20	0.00
FactSurveyResponse	Distributed	CustomerKey		DateKey	2,757	1.60	0.00
FactWholesale	Distributed	SalesOrderNumber		StartDate	0	0.00	0.00
ProspectiveBuyer	Replicated			ProspectiveBuyerKey	20,590	2.70	0.00

Managing the Scope of an APS POC

What if client makes additional requests during the POC?

- Goal is to stay within the constraints of the scoping document
- Always ask yourself “How will this help us win?”
- Determine risk of changing the project plan
- Establish the priority of the change request
 - Review the Scoping Document and discuss what can be modified/removed to accommodate change
 - Tell customer you’ll confer with TSP
 - Let the TSP handle the request
- Consider possibility of scheduling at end of APSPOC if time allows
- Avoid scope creep!

Managing the Scope of an APS POC

Deliver the Final Scoping Document

- Schedule Time To Discuss All Aspects Of The Scope
- Ensure All Stakeholders Are Present
- Make Sure Key Milestones Are Identified And Dates Established
- Ground Rules Established For:
 - Obtaining Artifacts
 - Problem Resolution
 - Data Strategy
 - Data Expansion
- *Obtain Customer Signature(s) To Ensure All Parts Are Understood*

Module Summary

In this module, you learned...

- The purpose, importance and structure of the Scoping Document and of the customer-provided artifacts
- To validate sample data availability and requirements
- How to set up and access the APS SME Sharepoint site
- Validating/Collating/Cataloging DDL and Queries

