Advanced
Analytics
2021

Prolifics Agile Migration for Azure

A Business-Sensitive Migration Roadmap to Azure

By: Michael L. Gonzales, PhD

April 2021

CONTENTS

FIGURES AND TABLES

INTRODUCTION

The Prolifics Agile Migration (PAM) for Azure is a methodology to accommodate large-scale migration initiatives across the enterprise. It is based on best practices and our experience moving legacy systems onto the modern platforms.

PAM for Azure is based on five core objectives:

1. Focus on user communities. We endeavor to educate and entice their interest in Azure by using interactive workshops to demonstrate actual legacy migration use cases and stimulate discussion of the value of data science and business intelligence on Azure.

2. Examine carefully the legacy code/applications and the user communities they support.

3. Recognize and accommodate business objectives and constraints by providing multiple alternatives.

4. Ensure the success of migrating user communities and their respective code/applications through change management and technical support.

5. Acknowledge that PAM is not operating in a vacuum. We often must coordinate and collaborate with parallel initiatives in data migration and governance.

PAM for Azure is a business-sensitive approach since it explicitly acknowledges business risk, from focus on users to change management.

For brevity, outlined in this paper are the key aspects of the methodology.

Return to TOC

PAM METHODOLOGY

PAM contains four phases in order to effectively and efficiently assist business to migrate from their legacy system to a modern Azure architecture. The phases include:

1. Prioritize – here we want to ensure user communities understand the value of migrating to Azure, measure their readiness to migrate, and measure the legacy application(s) footprint that supports the community. The outcome is a prioritization of user communities ready to migrate.

2. Assess – this phase examines the code/application supporting specific user communities. The objective is to determine which migration option is applicable.

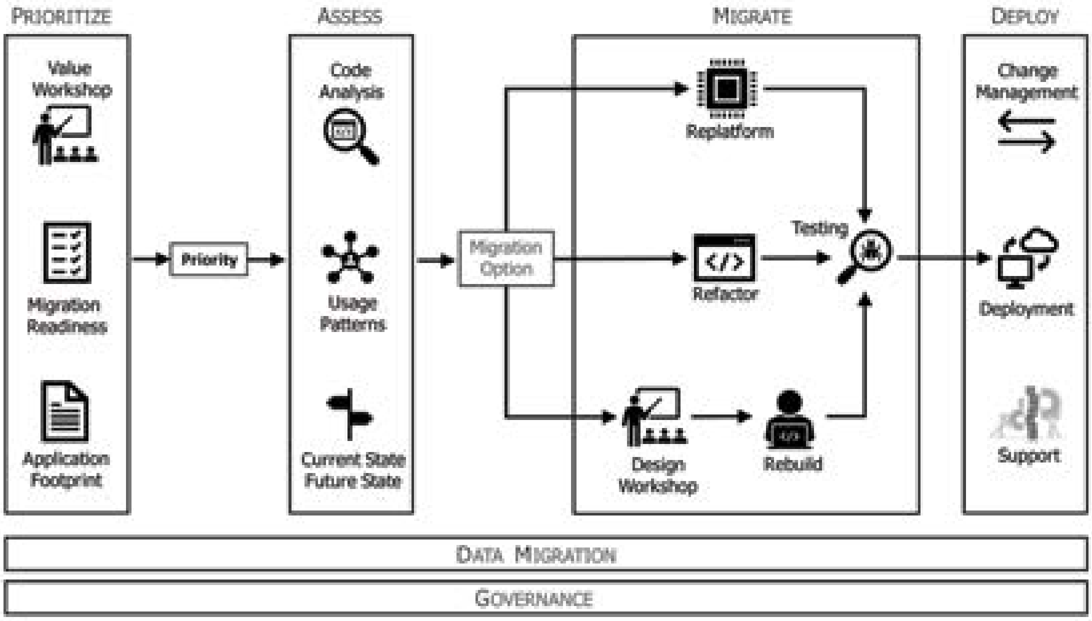
3. Migrate – the migration option selected is executed in this phase.

4. Deploy – the actual deployment of the code/application is completed in this phase as well as managing the change to business processes and providing those user communities with support.

In addition to the four phases, parallel initiatives for Data Migration and Governance are incorporated into PAM. We recognize the value and impact that these initiatives have on the success of all migrated code/applications.

The following sections provide more information about each phase and their key components. Figure 1 illustrates the PAM Methodology.

Figure 1 - PAM Methodology



PRIORITIZE

The objective of this phase is to prioritize those user communities ready to migrate. Each will be included in the Scrum Product Backlog to begin the migration sprints

A model is established to score the overall priority of a user community to migrate. This model is based on a 5-point Likert scale. Information used by the model scoring is gathered from the structured interviews as well as the application footprint. Each are described in the following sections.

.

Return to TOC

VALUE WORKSHOP

The workshop is specifically designed to articulate the value of moving to a Azure. A few benefits include[2]:

- Speed: how quickly teams can create, test, and deploy new applications in data science and business intelligence across the organization.
- Enhanced Flexibility: for business and data science teams, flexibility comes in the form of being able to scale computing power and storage up or down to better meet changing computing requirements.
- Integrated Delivery Pipeline: a significant amount of processing in legacy systems is often a manual coordination and execution of numerous ETL and data wrangling jobs. Much of this effort will be readily automated in the unified delivery pipeline.

Using actual use cases, the workshop will draw clear comparisons between the legacy environment and the cloud-based platform for data science, machine learning, and business intelligence. Moreover, we expect to provide a couple of actual legacy applications migrated to the target architecture in order to demonstrate the differences in the platforms.

This two-hour workshop will be individually scheduled and conducted with all interested user communities.

MIGRATION READINESS

This assessment is designed to establish an unbiased measurement of the ability and willingness of specific legacy user communities to migrate. But that is only part of the value. The overall information gathered across user communities allows the PAM team to[1] ...

- Establish performance metrics to measure and monitor the initiative
- Create a roadmap for migration improvement
- Ensure both business and the PAM team are collaborating

The information gathered will come in a form of structured interviews. These are planned and scripted interviews to ensure consistent information is gathered. An interview instrument is used that includes short answer, single answer, and open response.

APPLICATION FOOTPRINT

The footprint of an application provides critical information when deciding how easy or complex it may be to migrate. An instrument is designed to consistently inventory the legacy applications and their impact on the user communities they support. It covers information such as technology licenses, the data used, user communities supported, and application maturity.

ASSESS

The combination of Information from code analysis and usage patterns provides critical insight into the code. The objective of this phase is to identify the best migration option.

Return to TOC

CODE ANALYSIS

Only through careful examination of the legacy code/applications can we assess the level of effort it will take to migrate to the new platform. This examination is completed by using different techniques, including:

- Metrics of the utilization of legacy processes, their execution, the users and departments executing those processes, etc.
- Semantic analysis is executed against the legacy code in order to conduct link analysis, cluster analysis, visualization, and information retrieval.

USAGE PATTERNS

Often legacy systems exhibit clear usage patterns that are relevant to the migration effort. Some patterns include:

- ETL or data wrangling
- Ad hoc queries
- Standardized reporting and analysis
- Advanced Analytics

Isolating the primary usage of the code/application provides guidance in the complexity of the migration as well as the future state in the target environment.

CURRENT AND FUTURE STATE

A final step in the assessment phase is to create a high-level current and future state of the specific legacy code/applications being examined. This is done to define what components of the Azure environment will be leveraged as well as give guidance to any business process impact.

MIGRATE

Many legacy systems remain productive because they are stress-hardened, built over years of testing, validation, and performance tuning. Being stress-hardened, however, also makes them brittle and inflexible, unable to adapt beyond their current environment. And while these brittle applications likely require complete replacement to migrate, we do not assume that the entire legacy system will require the same. It is important to remember that complete replacement often leads to the following[4]:

- Significant management overhead
- Errors due to damaging implicit business rules
- Scope creep

In reality, there are often a wide range of applications on legacy systems, each distinguished by:

- Use Case: for example, ETL, ad hoc queries, standardized reporting, and advanced analytics
- Application Maturity: New, Growing, Mature, Legacy, and Sunset
- Application Footprint: impact on business and technical scope

These differences are evidence that alternatives to complete re-coding must be offered. To that end, PAM applies best-practice by providing three distinct methods to better accommodate business in the migration of their legacy applications.[3,5]

Return to TOC

As shown in Figure 1, PAM includes the following migration paths for business to consider:

1. Replatform – redeploy to a new runtime platform, making minimal changes to the legacy code. This migration option can be leveraged in multiple ways, for example:
    a. Used as a POC for business to quickly evaluate the new ecosystem
    b. As a phased migration to the target platform to be followed by either Refactor or Rebuild as time, budget, and requirements dictate.
2. Refactor – for example, translate the legacy code/applications to Python and improve nonfunctional elements, reduce technical debt, but not alter general functionality. Similar to the Replatform option, Refactor can be a phased approach to migration, for example:
    a. Establish an MVP.
    b. Consider the Rebuild option once time, budget, and requirements dictate.
3. Rebuild – fundamentally rewriting the code/application in order to optimize, modernize, and innovate. There is one extra component when considering the Rebuild option to migration, that being a Design Workshop. This workshop is specifically conducted to help business explore the modernization, optimization, and innovation landscape possible on the Azure platform.

Irrespective of the migration option selected, a thorough testing must be conducted on the target platform prior to deployment.

DEPLOY

PAM recognizes there are many aspects of deployment that go beyond the scope of this paper, including development, testing, and production platforms, containers and microservices, storage options, data propagation, etc. Moreover, the level of architectural planning will require input, if not dictated by, other department resources outside the PAM team.

For the purpose of this paper, we briefly describe three high-level components of deployment specific to migrated user communities.

DEPLOYMENT

A deployment strategy and related architecture are required. The PAM team will collaborate with other client resources to ensure a plan and process are established to facilitate the deployment of newly migrated legacy code/applications. A high-level Azure reference architecture is provided in order to begin the needed planning. Refer to the Azure Reference Architecture section of this paper.

CHANGE MANAGEMENT

With each migration of legacy code/applications there will be changes to the business process, or at the very least, the user interaction with the migrated code/application.

SUPPORT

In order to properly support the new user communities, PAM will identify, define, and create the support materials needed for the specific target architecture components employed by the user communities. This requires a parallel initiative that provides the prerequisite materials needed by the client's technical support organization.

Return to TOC

DATA MIGRATION

Large migration efforts are often associated with enterprise-wide initiatives for data migration executed by a different department. In these cases, PAM will adhere to, and align with the constraints and cadence of the organizations' data migration program. However, there are migration efforts that will invoke a data migration effort focused entirely on the migration itself.

The data strategy for migration of legacy applications can represent a range of alternatives, including:

- an enterprise initiative beyond the scope of the target legacy platform
- only the scope of data leveraged by the legacy platform
- a limited movement of specific data into the cloud-based architecture supporting a hybrid data platform

Table 1 outlines a few key elements of data migration planning.

The planning and execution of a data migration initiative goes beyond the scope of this paper. Prolifics represents years of expertise in Information Management and the migration of data between ecosystems. Request more information about our Information Management practice.

GOVERNANCE

The migration effort affords us the opportunity to establish critical governance processes. If your organization has an active governance program, we recommend embracing the process to ensure the target platform is established with continuity. If, on the other hand, there is incomplete or no governance program in place, we recommend the organization consider establishing that oversite while the migration initiative is executed. Table 1 outlines key elements for incorporating governance to the migration effort.

Table 1 - Governance and Data Migration Planning

	Assessment & Development Planning	Deployment Configuration & Testing	Migration
Inputs	Migration Assessment & Planning	Final Deployment Plan	Execute Migration & Unit Testing
	Validate Governance Use Cases	Execute Deployment & Connectivity	Capture and Record Lineage
	Define and Agree to Project Governance	Deployment Testing & User Acceptance Testing	Integration Testing
Deliverables	Roadmap & Timeline	Establish Development, Test, and Prod	Test Outcome Reports
	Governance Use Case Documentation	Formal Testing Criteria	End-to-End Lineage Report
	Formal Migration Plan		
	Project Governance Artifacts		

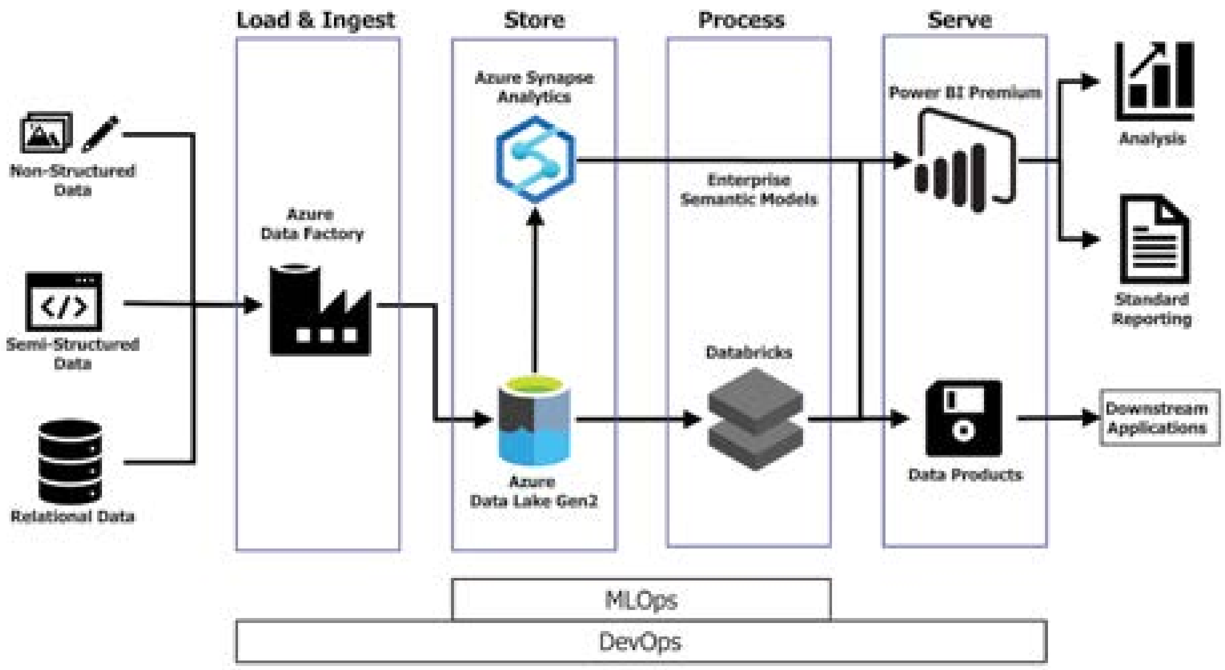
The planning and execution of a governance program goes beyond the scope of this paper. Prolifics represents years of expertise in this area. Request more information about our Governance practice.

Azure REFERENCE ARCHITECTURE

As with other elements of PAM, the Azure reference architecture will be dictated by the use cases migrated. While Azure offers a large array of services and components, Figure 2 illustrates an architecture to support standardized reporting, business intelligence, and advanced analytics. Each component is outlined in this section.

Return to TOC

Figure 2 - Reference Architecture[1]



Each of the major components is outlined below.

- Data Factory – represents the primary ETL technology, focused on the following:
  - Sourcing Data
  - Transforming Data
  - Loading Data into Target Data Stores
- Azure Synapse Analytics (ASA) – ASA offers two key advantages:
  - It serves as a single, Use Interface between the data warehouse of Synapse and the data storage in Data Lake Gen2.
  - Supports the ability to combine big data and traditional relational data into enterprise models for ready consumption by downstream technologies, processes, and applications.
  - Provides data wrangling and data science capability via Azure Synapse Spark that contain interactive notebooks similar to Databricks. (Broome, 2020)
- Databricks – A landscape of open-source (for example, Python and PySpark) capability is provided with this technology, including:
  - Extensive data wrangling
  - Extended data science capability
- Power BI Premium (PBI) – this represents the primary reporting and analysis layer for user communities. It provides:
  - A broad range of reporting and analysis
  - A superset of the capabilities compared to Azure Analysis Services[2]

[1] https://docs.microsoft.com/en-us/azure/architecture/
[2] Azure Analysis Services versus Synapse Analytics, stackoverflow.

Return to TOC

- Azure Machine Learning (MLOps) – as an extensive of DoveOps, MLOps represents a means to build, train, and deploy advanced algorithms in a repeatable pipeline. Key elements of MLOps include:
    - Create and retrain predictive and exploratory models
    - Establish scoring files and dependencies
    - Monitor the scoring output for data drift and model accuracy
- Azure DevOps – essentially affords the opportunity for Continuous Improvement/Continuous Deployment. Key components include:
    - Azure Pipelines
    - Azure Boards
    - Azure Repos

There are, of course, many other Azure services/components available based the organization's use cases. And as the use cases differ so does the reference architecture.

Return to TOC

APPENDIX A – REFERENCES

1 - Gonzales, M., Comprehensive Analytic Maturity Assessment, Prolifics, 2020.

2 - Knerl, L., Top 5 Benefits of Microsoft Azure, 2020.

3 - Moore, S., 7 Options to Modernize Legacy Systems, Gartner, 2020.

4 - Shinde, K., The New Mantra for Legacy Migration, bitwise, 2020.

5 - Wood, W., Information Technology Systems Modernization, Carnegie Mellon University, 2015.

Return to TOC

Prolifics®

Vision to Value. Faster.

twitter.com/prolifics

youtube.com/prolificstv

facebook.com/prolificstech

linkedin.com/company/prolifics

phone: 818-582-4952

email: solutions@prolifics.com

web: prolifics.com

Prolifics Advanced Analytics