DATA INTELLIGENCE FRAMEWORK

BEST PRACTICES FOR DATA INTELLIGENCE

HOW CAN AIS HELP INCREASE VALUE IN DIVERSE DATA SETS

Looking to discover the relationship between your data and your mission? AIS developed a process for interactive integration of data sources that allows organizations to discover and analyze new factors that impact the line of business through visualizing data in an appropriate and trusted manner.

Throughout this white paper, we'll cover the best practices and trusted frameworks we've used to help our clients begin to glean more insight from previously disparate data and inform their business strategy with actionable insights.

What you'll learn:

- **Common Challenges** – Acquiring and ingesting data sources, keeping data secure and manageable, and more.

- **Data Intelligence Tooling** – The best of a breed and cutting-edge industry tools that will support your team on your next data intelligence project.

- **AIS' Data Intelligence Framework** – Inside AIS' repeatable framework and approach to data analytics, including technical and non-technical challenges addressed, and why it works.

- **Use Cases & Success Stories** – How to apply our data intelligence approach in realworld scenarios.

COMMON CHALLENGES

As you set out to discover hidden insights in your data, it helps to understand some of the common challenges you'll face. Below we've outlined common challenges every organization faces on their data intelligence journey. As part of creating this shared vision, the executive sponsor needs to ensure:

The organizations business strategy and goals are translated into iterative activities such as:

- **Choosing Data Sources** – What systems and sources should you ingest data from to support and inform the goal at hand?

- **Data Acquisition** – How can you acquire data from the identified data sources? This involves finding the right POC, getting approval, and the technical aspects involved with permissions and ingestion.

- **Integrating and Aligning Data for Analysis** – Review of data format, creation and refresh frequency, and other dependencies to understand how we can model the data.

- **Data Modeling** – Understanding the business case the raw data needs to be modeled for. You need to work with the right SME(s) to support the use case for modeling the data to meet project requirements.

- **Deploying and Monitoring** – When deploying the data to a production environment, the sample data in your development and Q&A environment may not align with the production data. Monitoring presents its own challenges with limitations on tooling in highly regulated environments that would otherwise be available to most commercial organizations.

Data Intelligence Tooling

DATA INTELLIGENCE TOOLING

Depending on the environment, the selection of integration tools can be limited. Below is a list of common tools used on data intelligence projects to help ensure best practices are followed, help teams stay organized, and streamline and automate.

While the list is not exhaustive, these tools will serve your team in getting to better data insights and intelligent solutions faster and with better outcomes.

- **Microsoft Word, Excel, and Visio** – Office products are used to gather and represent high-level dataflow, data mapping, and requirement gathering.

- **Microsoft Project** – Helps you with project management.

- **Azure DevOps, Jira, and Confluence** – Collaboration tools for managing work items, code repository and deployment pipeline, issue, and documentation.

- **eMASS and XACTA** – Cybersecurity risk management system.

- **Azure Data Factory** – Cloud-based data integration service.

- **HDInsight** – Azure-hosted Hadoop service used as part of the data transformation process.

- **Databox** – A Microsoft appliance for moving terabytes of data into Azure.

- **Import/Export Service** – Securely transfer data to and from Blob Storage or Azure file to and from Azure Data Center.

- **Equinix Data Center** – For customers with Express Route connectivity, data can be moved to a customer data center and Azure.

- **Azure SQL** – Cloud-based relational database.

- **Azure Analysis Services** – A tabular data model for a large database in Azure.

- **Visio, Erwin, Enterprise Architect** – A design tool for data model or application.

- **Machine Learning Service** — Azure-hosted machine learning services.

- **Databricks** — A managed spark service hosted in Azure.

- **PowerBI and PowerBI Online** — Microsoft's data visualization tool for dashboard building and its online version.

- **Azure Automation** — A Microsoft tool used to automate, configure, and install updates across hybrid environments.

- **PowerShell Desired State Configuration** — A management platform in PowerShell that enables you to manage your IT and development infrastructure with configuration as code.

- **Azure DevOps** — A platform with a set of modern dev services, including agile planning boards, pipelines, repository, test plans, artifacts, and more.

## AIS DATA INTELLIGENCE FRAMEWORK

AIS has developed a well-defined, repeatable process to address both technical and nontechnical requirements for data intelligence projects.

Today, there is not much of a defined, standardized process (much like there is in software development) for scheduling, cost projections, risk assessment and mitigation, and other components of a successful data intelligence project. Our innovative framework for data projects helps to better scope, roadmap, budget, and ensure project goals are met.

Whether you are looking to migrate existing data warehouses to a centralized platform or better process and transform data with cloud-native tooling and modeling for actionable insights and visualization, this framework will keep your project on time, within budget, and exceeding expectations.

The Data Intelligence project is tribute across five categories:

1. Strategic Assessment Factory

2. Data Ingestion Factory

3. Data Model Factory

4. Analysis Factory

5. Publishing/Visualization Factory

| Strategic Assessment | Ingestion | Modeling | Analytics | Publishing |
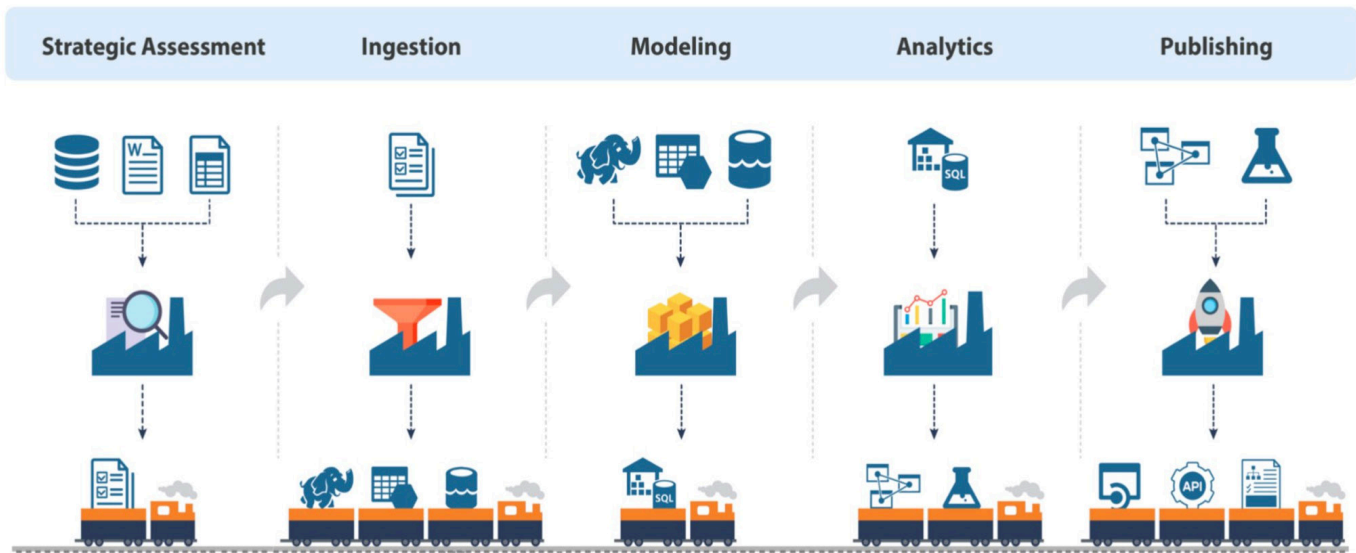
COMPONENTS OF THE FRAMEOWRK

- IScaled Agile Framework (SAFe) methodology

- DevSecOps

- Cloud-centric technical approach

WHY IT WORKS

- Allows you to scale agile.

- Emphasizes repeatable processes with standard outputs.

- Flexible enough to incorporate cloud-native technologies, such as data factory, Azure data analytics, data lake, Synopses Analytics, and more.

TECHNICAL CHALLENGES

- **Ingestion** – Gaining access to secure source systems and moving their data / connecting that data to the cloud in a secure manner, where the right end-users have access to appropriate data.

- **Modeling** – Availability of the SME's to design and re-architect existing data they have. How do we create a model to satisfy the business need?

- **Analytics** – Layering visualization specifications on top of the data model to support the presentation of data. This state also involves the optimization of large data sets for optimum performance.

- **Publishing** – Ensuring proper end users have access to the data they care about. This involves security level, governance, and visualization of data to create actionable insights for various audiences.

NON-TECHNICAL CHALLENGES

- **Data Acquisition** – Systems are gated by security and access, which needs to be overcome before the project team can begin using it. It's often hard to acquire data that follows the appropriate process/access control.

- **Governance** – How do you manage access to the data at each level of the process (during ingestion, modeling, etc.)? When data is changed or merged, how do you continually define and enforce the level of access, management, and ownership of this data, from stakeholders to end-users as new and more data becomes available?

- **ATO Authorization** – Starting the process early on and using automation to speed up the ability to create a cloud platform to store and manage your data. Hiring a partner who has been through the process and has documentation to support ATO access can ensure timelines are met.

INTRODUCING THE FACTORIES & SAFE

AIS' data intelligence framework relies on "Factories". Factories are cross-functional agile teams applying a disciplined process to generate repeatable products. Multi-disciplinary teams using an agile development approach to modernize the data ingestion, data modeling, analysis, and visualization. Each factory has discovery, design, build, and launch activity.

Arrivals / Departure

Departing Station	Cargo	Departure time	Arriving Station	Arrival Time
Strategic Assessment	ICD, PoC	6/1/2018	Ingestion	6/2/2018
Ingestion	Stored Raw Data	6/15/2018	Modeling	6/15/2018
Modeling	Data mart	7/1/2018	Analytics	7/15/2018
Analytics	Insight, Prediction	8/1/2018	Publishing	8/1/2018
Publishing	API, Dashboard	8/15/2018	Environment	8/15/2018
Analytics	Transformation Logic	9/1/2018	Modeling	9/1/2018
Publishing	Defects, Logs	9/15/2018	Strategic Assessment	9/15/2018

SCALED AGILE FRAMEWORK

- **Lean-Agile leaders** orchestrate factory functions to deliver value.
- **Agile release train (ART)** carries artifacts, products, and people between factories.
- **Express ARTs** enable agility by providing JIT delivery.
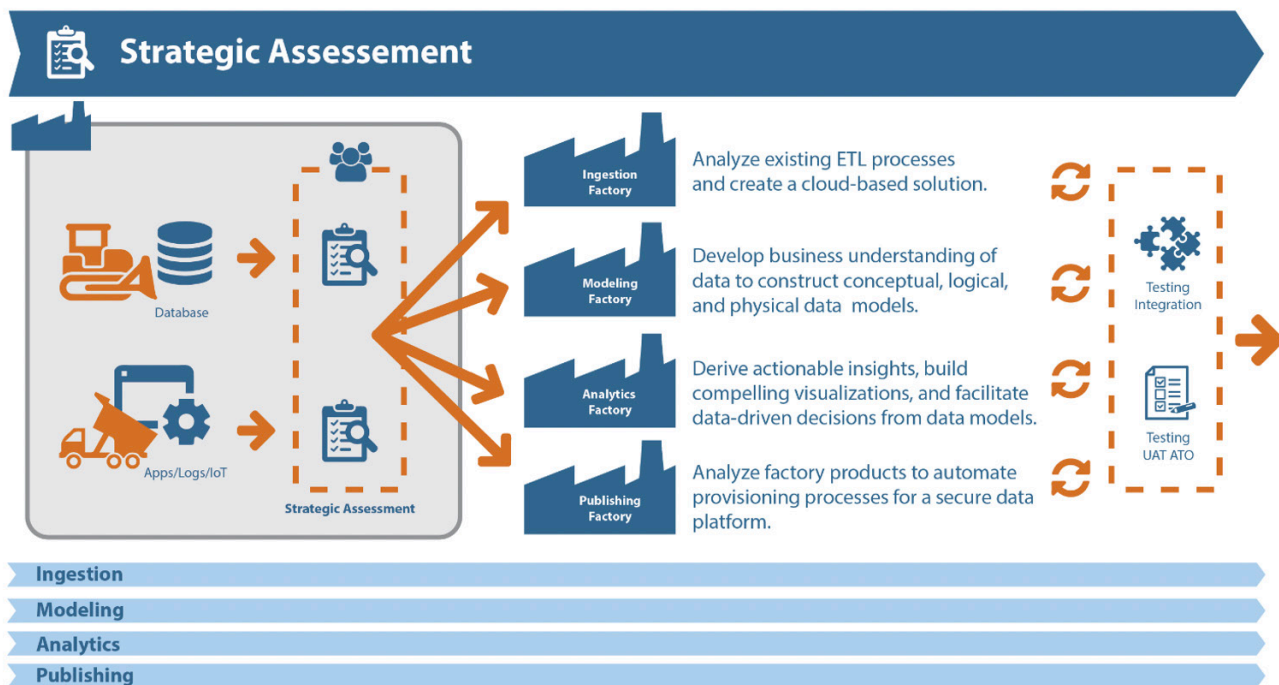- **Multiple ARTs** are scheduled and running simultaneously.

Want to learn more about our approach for Data Intelligence? We'd love to start the conversation.

CONTACT US

The Five Factory Steps

STEP ONE: STRATEGIC ASSESMENT

Purpose

In this step we assess existing applications and data streams to document basic assumptions and complexities to understand what's required to move the system(s) to the cloud. The assessment is based on a detailed business strategy and stakeholder interviews at the start of the project to understand the existing systems and data sources. AIS works with the customer to evaluate the applications and identify which source feeds the system for the data intelligence project. The team provides a backlog report consisting of system migration and data movement activities decomposed into epics and features for agile delivery as well as a high-level design document. The assessment process also documents the connectivity to the cloud, overall Azure subscription design, security posture, and governance plane.

The strategic assessment step clarifies project initiatives and provides alignment among the team by developing an actionable and well-documented plan.

- **Crystalizes the Vision** – All of the project stakeholders are aware of the plan, timeline, and dependencies.

- **Evangelizes the Mission** – The why behind the project is defined and communicated as part of the Strategic Assessment step.

- **Defines the Strategic Plan** – The roadmap for getting to the desired end state is determined and committed.

Tools

- Microsoft Word, Excel, Visio

- Microsoft Project

- Azure DevOps, Jira, Confluence

Challenges

Risks are identified during the strategic assessment step. To mitigate risk, AIS creates a proof of concept to clarifying the requirements and identify external dependencies, such as access to the source data from the cloud.

- **RMF Implementation** – Risks often vary in size and quantity, and the team must ensure risks are monitored, categorized, and assessed appropriately.

- **Team Collaboration** – Identifying POCs, establishing communications, and engendering cooperation.

- **Cloud Connection** – Establishing a secure cloud access point (CAP) is a long process, sometimes delaying initial project starts.

STEP TWO: INGESTION FACTORY
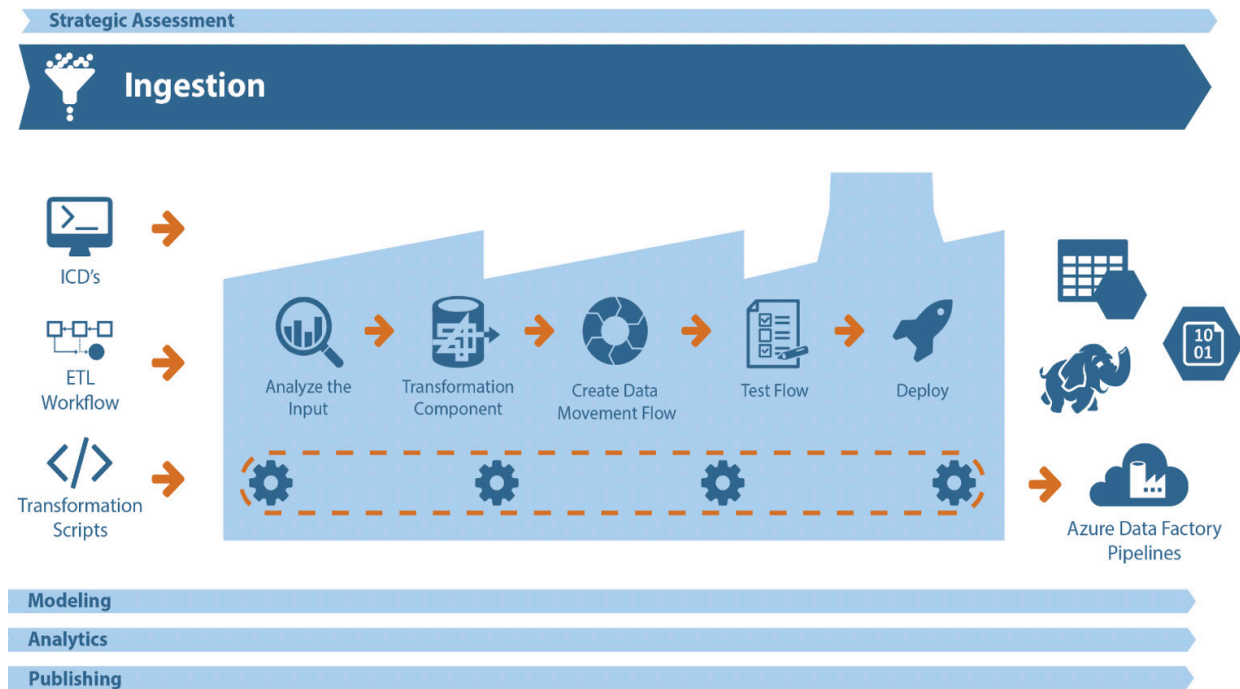
Purpose

The ingestion process is developed based on the outcome of the strategic assessment plan. The ingestion team performs the analysis of the input, such as data mapping, existing ETL script, and transformation rule. The team decomposes the epics and features created in the strategic assessment step into a detailed list of user stories with acceptance criteria and provides updated design documentation for building and implementing the system.

The design, documentation, and implementation include:

- **Platform Build** – Building a unified data platform capable of handling an infinite amount of all data types.

- **Data Connectivity** – Establishing connectivity between data sources and the data platform to transfer data at the required frequency and volume.

The output of this Ingestion Factories may include transformation Spark Script, Data factory pipeline, and storage for ingested data.

Tools

- Azure Data Factory

- HDInsight

- Databox

- Import/Export Service

- Equinix Data Center

Challenges

When it comes to accessing and moving data from enterprise systems, expect delays and challenges.

- **Large Data Size** – The initial data size can be too large for an extract, transform, and load job – It may be difficult to expediently move terabytes of data.

- **Accessing the Source Data** – The location of the source system may not be reachable for security reasons.

- **Satisfying Security Concerns** – The business owners need to verify their data is moved in a secure way.

- **Extracting Data from Legacy Systems** – Legacy systems may have an outdated protocol, outdated data formats, data may not be accessible, etc.

STEP THREE: MODELING
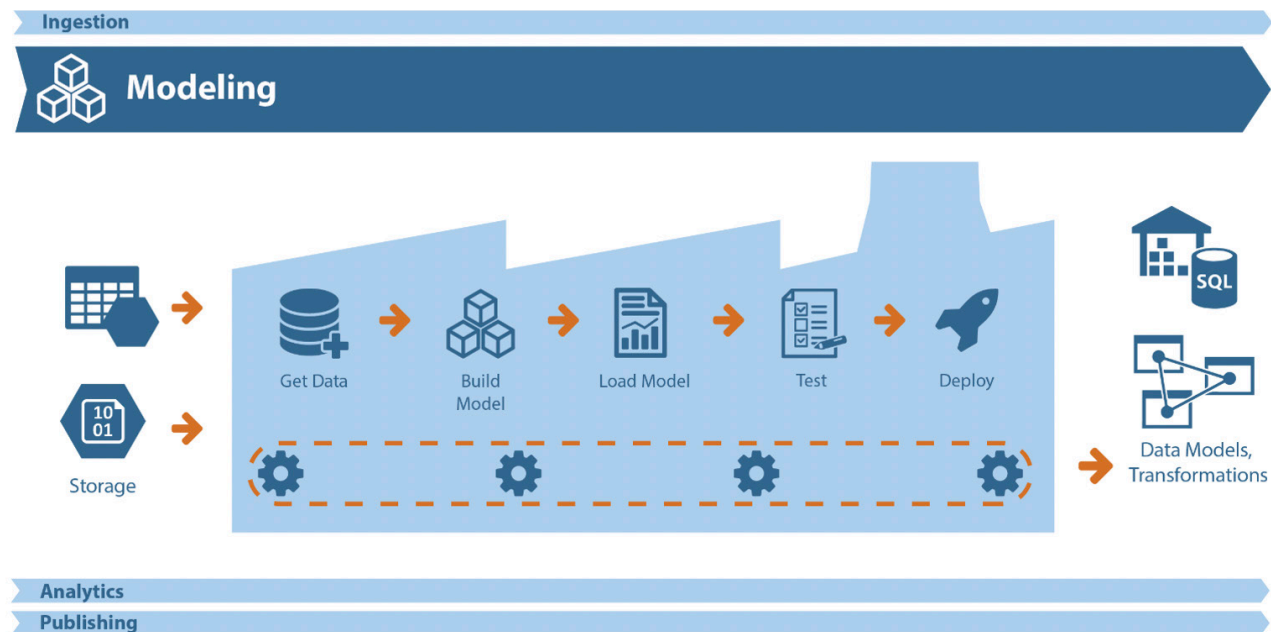
Purpose

The team reviews the requirements for a report, dashboard documentation, and wireframes. Guided by this information, the team will engage SMEs to help create epics, features, user stories, and acceptance criteria based upon current operational needs. A model is built and data is loaded, tested, and deployed to either Azure SQL, Azure Data warehouse or Azure Analysis Services.

The outcomes of this step are:

- **Actionable Insights** – Report and dashboard users can locate and reveal the truths within the data sources.

- **Better Data Models** – In this step, we develop easily understood models by integrating and correlating the data.

Tools

- Azure SQL

- Azure Analysis Services

- Visio, Erwin, Enterprise Architect

Challenges

- **Data Governance** – Ensure proper process and security documentation is determined for managing user access and tagging when aggregating data of varying security levels from multiple data sources.

- **Coordinating with SME's** – Aligning availability within the anticipated project timeframe can sometimes pose challenges and delays.

- **Maintaining Data Lineage and Provenance** – The acquisition, transformation, and maintenance of data needs to be traceable.

STEP FOUR: ANALYTICS
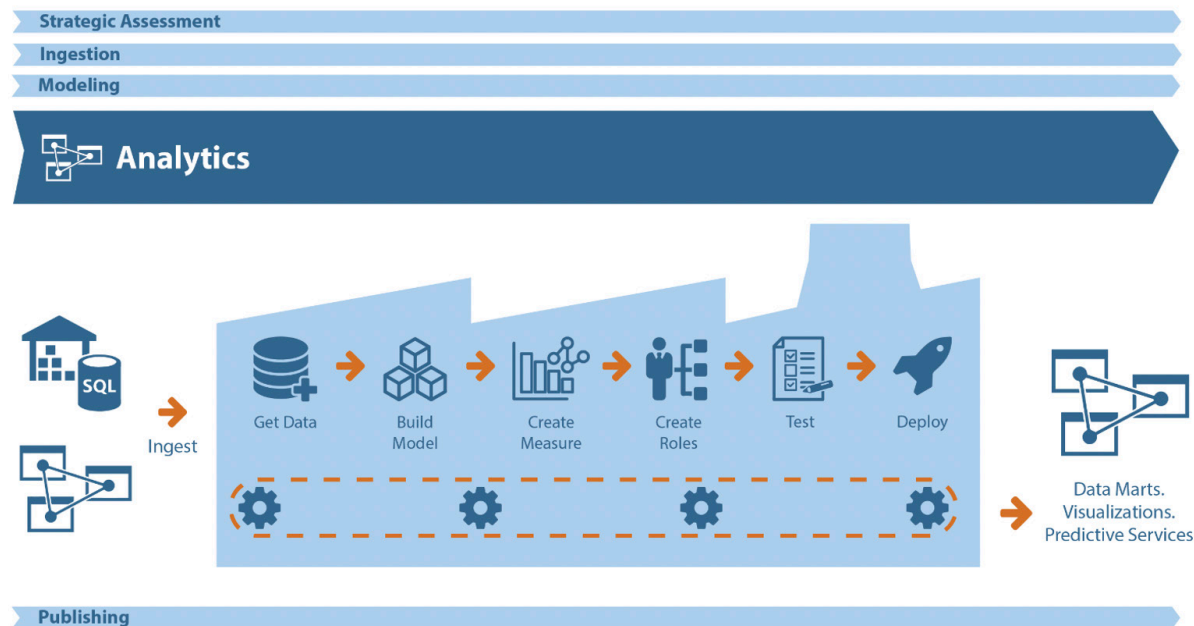
Purpose

The models produced in step three are then used to build semantic models that can be consumed by client applications. These semantic models provide a more accessible and unified view of data retrieved from multiple disparate source systems. This process is comprised of six stages: requirements elicitation, model construction, securitization, user acceptance testing, and deployment. These stages will be executed within an Agile SCRUM framework with well-defined user stories and acceptance criteria.

The outcomes of this step are:

- **Intelligent Insights** – Predict the events of tomorrow using the truths of today.

- **Applied ML to Data** – Build robust analytical products through deduction, induction, and machine learning techniques.

Tools

- Machine Learning Service

- DataBricks

- PowerBI

Challenges

- **Data Preparation** – The majority of effort is spent preparing the data to be consumed by machine learning services.

- **Machine Learning Model Selection** – Choosing the appropriate ML model for your mission is critical in producing the desired outcomes.

- **Satisfying End User Expectations** – Ensuring the insights gained from the analytical products produced in this stage meets user expectations must be validated early on to ensure project success.

- **Maturing the ML Model** – Creating a method for input and feedback to continually improve the machine learning model will increase the value of the analytical product and user satisfaction.

STEP FIVE: PUBLISHING
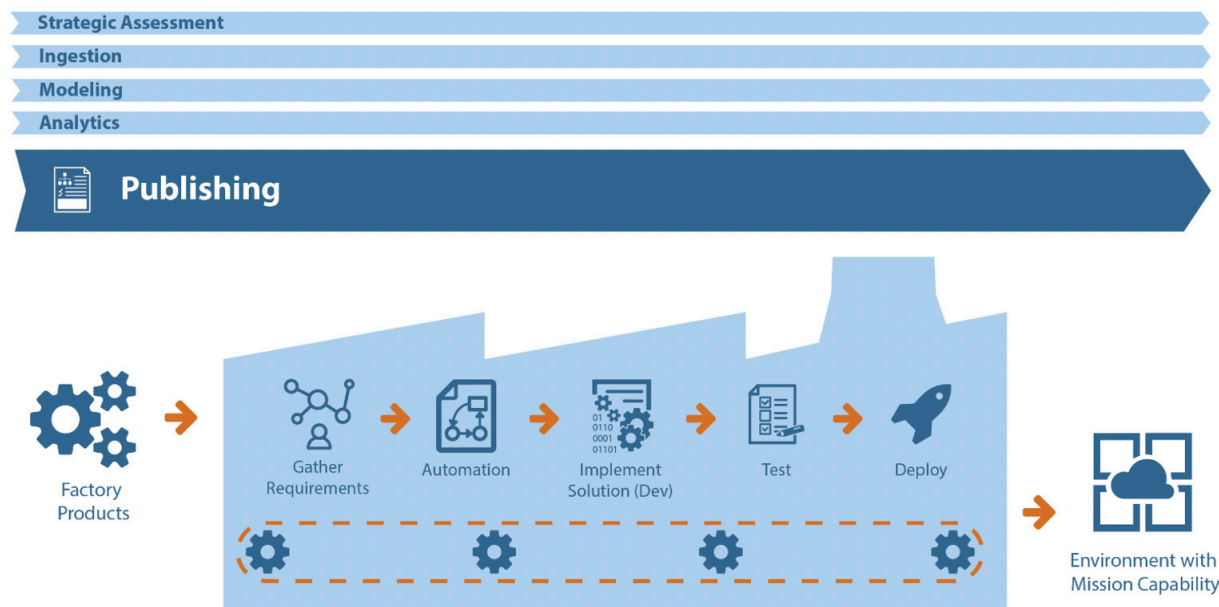
Purpose

The DevOps team works with the system architect and lead developer(s) to create a final deployment package and implement an automated deployment script. The DevOps team creates deployment pipelines that promote the script and code to the Pre-Production and Production environments.

The outcomes of this step are:

- **Data at Your Fingertips** – Every decision-maker and analyst can always access all available intelligence necessary based on role.

- **New Report Notifications** – When a new report is published, the right team is notified.

Tools

- Azure Automation
- PowerShell Desired State Configuration
- Azure DevOps
- PowerBI Online

Challenges

- **User Access** – Integrating Enterprise Identity Access Management.
- **Operational Training** – Transition to the sustainment team to operate in a new paradigm.

Getting Started

YOUR NEXT DATA INTELLIGENCE PROJECT

Whether you're knee-deep in requirements gathering for a data intelligence project or just starting the discussions around the art of the possible, AIS' data intelligence framework can help your team make data visualization a reality.

Contact AIS to learn more about our framework and how we can partner to get your enterprise to the cloud.

CONTACT US