VERISTA

A Process Model for Defining & Detecting Analytics

IMPLEMENTING DATA GOVERANCE STANDARDS TO ACCELERATE YOUR BUSINESS TOWARD GROWTH

—

EBOOK

WHAT'S INSIDE

✓ A PROCESS MODEL FOR DEFINING AND
   DETECTING GOVERNANCE-AWARE ANALYTICS

✓ IN-DEPTH EXAMPLES OF ANALYTICS USE-CASES

✓ A TECHNOLOGY-AGNOSTIC REPRESENTATION OF
   MODERN ANALYTIC CAPABILITIES

✓ REFERENCE ARCHITECTURE OF THE CAPABILITY MODEL
   CAST AGAINST SPECIFIC USE-CASES

TABLE OF CONTENTS

INTRODUCTION

In our modern digital age, analytics are on everyone's radar. As organizations go "all-in" with analytics, the effort can get very noisy. Today's complex world requires a modern approach to define and detect analytics in a manner that ensures analytics stays on everyone's radar, empowering businesses to discern the signal from the noise.

A modern approach to analytics requires a systematic and nimble process model to capture use-cases, capabilities, and patterns. The analytic process model is a governance-aware triangulation of three foundational elements:

√ Analytic use-cases defined through perpetual business engagement and outreach

√ An articulation of modern analytic capabilities that constantly evolve to meet the business demand for insights

√ Discernible patterns derived from reference architectures of use-cases and capabilities that reinforce established conventions or reveal new paths across the analytic continuum

By design, the process model minimizes the number of steps - heavy on details and visual representations of each element. Let us illustrate.

THE GOVERNANCE-AWARE PROCESS MODEL FOR DATA & ANALYTICS
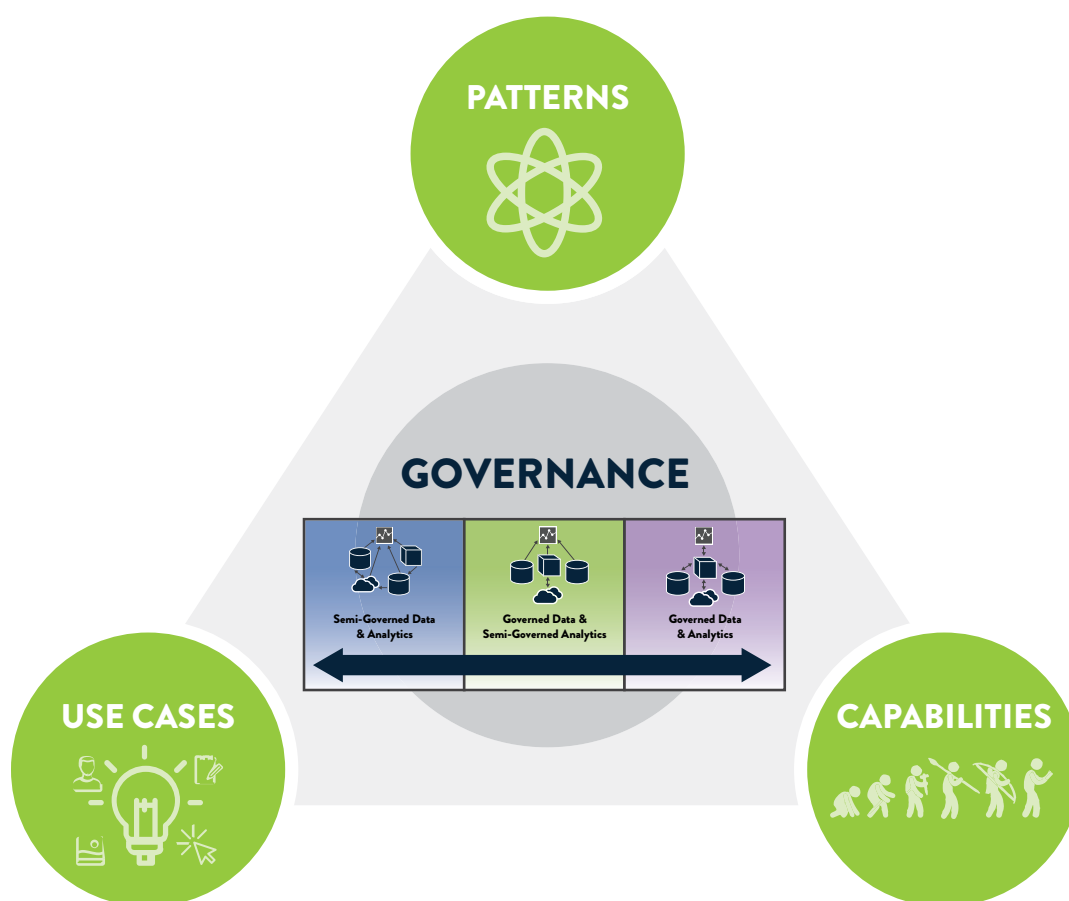
A CONTINUUM FOR GOVERNANCE-AWARE ANALYTICS

Analytics requires more than enabling technology and tools. It requires new roles and enhanced coordination between IT and the business. This coordination invokes a four-letter word: governance.

The G-bomb is dropped not for the sake of dropping it, but done so out of necessity. This is because analytic use-cases span a continuum of semi-governed and governed data sets and analytic use-cases. It is expected that the data sets to be fit for analytic use fall under some form of governance, i.e., never completely ungoverned. For instance, authoritative master data is at the far end of the highly governed data spectrum. At the other end of the spectrum can be found semi-governed data like machine-generated data or customer surveys or comments posted to a company-sanctioned YouTube channel which exist and have potential value, but exist beyond the reach of sanctioned stewardship. Falling off the radar would be completely ungoverned data which is completely unqualified or inappropriate for analytic use.
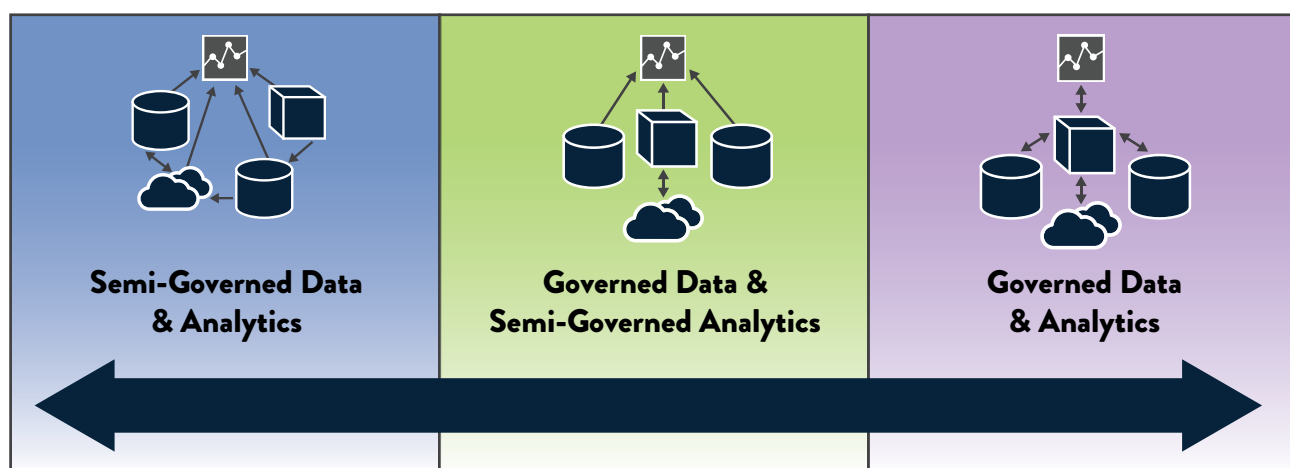
I.T.'s MISSION: FOCUS ON SELF-SERVICE ANALYTIC CAPABILITIES

Qualifying the disposition of data against a continuum of governance awareness is an act of self preservation. Specifically, data sets essential to analytics may be very familiar to business users, but unfamiliar to IT staff. It is unrealistic to expect IT to scale to support the growing data sets required by analytic use-cases, and at the same time support the governed data sets required by everyday business operations. Thus, IT's mission should not be to govern all data, but rather to establish authoritative governed data sets that can be aligned with semi-governed data, and focus on enabling technologies and platforms that allow sanctioned access and use of data by qualified, trained business users. In other words, IT should focus on building user-friendly self-service analytic capabilities, not building an army to deliver analytics.

This does not give IT a license to abandon its governance post.

IT must continue to play a supporting role in governing semi-governed data by ensuring data is fit for analytic use and consumption, properly stored and classified, not redundant to existing sources or superseded by a more authoritative source, adheres to standards and regulations, etc. In short, IT is involved enough with data to the point it is semi-governed for analytic purposes, and as needed fully governed.

DATA'S ANALYTIC UTILITY

Data alone does not qualify as analytics. Data must be stewarded, curated, prepared, integrated, enriched, visualized, and accessible for it to provide analytic utility. Each step in the analytic chain operates along the same continuum of governance awareness. If the use-case requires precise details that are reported to investors or published in a quarterly report or submitted to the FDA during Phase 3 of a clinical trial, then the data and the analytics must be fully governed. On the opposite end of the spectrum, other use-cases occupy a happy medium, such as allowing business users to combine sales with promotional spend - two governed data sets - in a self-service, semi-governed manner.

Based on this understanding of data and analytics, a continuum of governance-awareness is formed that covers 3 scales, and establishes boundaries throughout the analytic process model:
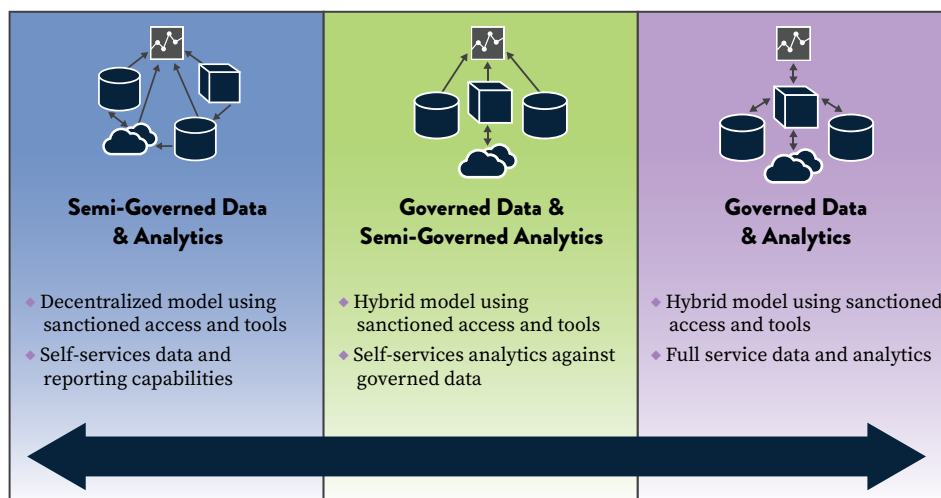
Semi-governed data & Semi-governed analytics

Governed data & Semi-governed analytics

Governed data & analytics

NOTE: Semi-governed data & governed analytics is not suitable for any analytic use-case, hence we have a continuum and not a quadrant.

THE GOVERNANCE-AWARE PROCESS MODEL FOR DATA & ANALYTICS

Semi-Governed Data & Analytics

- ◆ Decentralized model using sanctioned access and tools
- ◆ Self-services data and reporting capabilities

Governed Data & Semi-Governed Analytics

- ◆ Hybrid model using sanctioned access and tools
- ◆ Self-services analytics against governed data

Governed Data & Analytics

- ◆ Hybrid model using sanctioned access and tools
- ◆ Full service data and analytics

A critical aspect of the blueprint is to deliver self-service capabilities that allow users to govern un-governed Data & Analytics.

The dominant theme from the use-cases suggests users want data ingestion, curation & blending capabilities beyond Excel.

IT cannot always anticipate or meet the demand to govern ungoverned datasets.
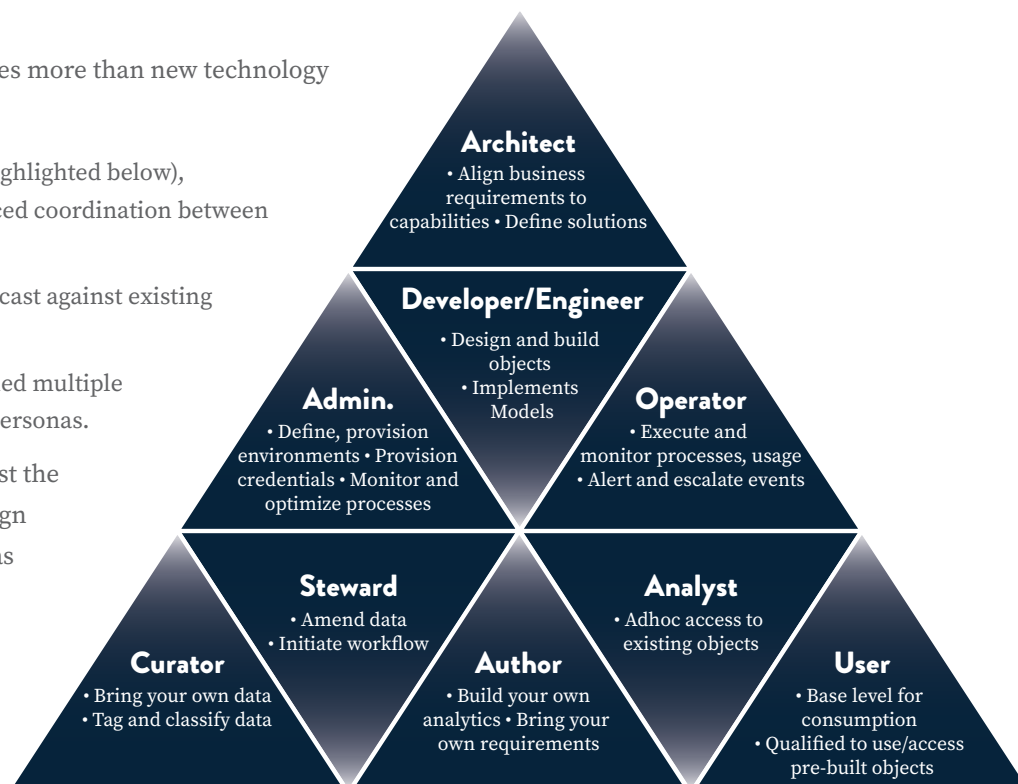
ASSIGNING DATA & ANALYTIC ROLES

One other aspect of governance-aware analytics is a new set of roles emerging to support each individual use-case. For instance, a curator role emerges to represent users that can bring-their-own data into the environment. These new roles can be cast against existing personnel. Personnel can be assigned multiple roles, as shown by the personas. This model will be cast against the personas and use-cases to align wants and needs to the personas and roles. The following represents common roles in the analytic process model. Feel free to add Data Scientist, which is represented by skill sets as opposed to a role in this particular scheme. Generally speaking, the number of roles will decrease as we move from the semi-governed data and reporting to the governed end of the continuum.

Data & Analytics Roles

Enhanced self-service requires more than new technology and tools.

√ It requires new roles (highlighted below), environments & enhanced coordination between IT and the business.

√ These new roles can be cast against existing personnel.

√ Personnel can be assigned multiple roles, as shown by the personas.

This model will be cast against the personas and use-cases to align wants & needs to the personas and roles.

Architect
• Align business requirements to capabilities • Define solutions

Developer/Engineer
• Design and build objects • Implements Models

Admin.
• Define, provision environments • Provision credentials • Monitor and optimize processes

Operator
• Execute and monitor processes, usage • Alert and escalate events

Steward
• Amend data • Initiate workflow

Analyst
• Adhoc access to existing objects

Curator
• Bring your own data • Tag and classify data

Author
• Build your own analytics • Bring your own requirements

User
• Base level for consumption • Qualified to use/access pre-built objects

TO THE POINT

Once you have determined which alignment along the data analytics continuum is right for your organization, you can assign existing personnel multiple roles to support each individual use-case.

DETECT AND DEFINE ANALYTIC USE-CASES

Governance-awareness is baked into the process model, allowing us to proceed with confidence as we begin to detect and define analytic use-cases. We begin the process by documenting obvious, "motherhood and apple pie" use-cases. Prevalent business analytic use-cases include clinical study design, KOL management, customer segmentation/deciling, attribution modeling, process optimization, fraud detection, A/B testing, etc. Also, be aware of technical IT use-cases such as log analysis and threat detection. Finally, while recently in vogue, analytics have been around for decades and some stakeholders can offer working examples day 1 that could use a fresh coat of paint or second set of eyes. It is worth documenting these analytic use-cases up front.

Having an initial starting point of possible use-cases, IT can establish credibility and engage stakeholders in a working session to further qualify the initial use-cases, which begins the real work of digging in deeper for non-obvious use-cases. Start by seeking out specific examples where stakeholders are blending data sets or crunching numbers in Excel. These are clear signs that applications and platforms like the data warehouse are not meeting the everyday analytic needs of users, and knowledge workers are spending precious cycles in non-productive data preparation tasks. For instance, users could be spending hours each week blending third party data with sales in Excel, or downloading data sets from Google Analytics 5,000 records at a time. Technical staff might also be wrestling with how to archive growing data sets in a cost-effective manner, or where to put noisy but potentially useful semi-governed data. There is a huge opportunity to deliver on these use-cases and achieve tangible benefits to the business and IT, and thus are relevant analytic use-cases in the process model.

DEFINING ROLES WITHIN YOUR ANALYTIC COMMUNITY

Of critical importance is to capture each of these scenarios by user and role so you can qualify and quantify the needs of each individual use-case. Once the needs are understood, suitable roles can be aligned to each member of the analytic community. New roles can emerge at any point so governance is constantly evaluating the needs of users. As roles are cast against users, distinct personas emerge that align users to existing and emerging analytic tendencies. Be sure to capture the personas, roles, and use-cases in a systematic, noninvasive manner using specific, succinct wording. This ensures the use-case translates into concrete business terms, such as these use-cases for a Clinical Biostatistician and a Commercial Operations Analyst at a biopharma company:

Persona	#	Use-case(s)	I want to...	So that I can...	Data	Roles
Clinical Biostatistician	CB.1	Clinical Study Design	Optimize the process to qualify study participation, length, dosing, and analysis	Reduce the costs of drug development and time to market	Electronic Health Records (EHRs) Clinical Data Management System (CDMS) Laboratory Information Management Systems (LIMS) Steward Curator Analyst	Steward Curator Analyst
	CB.2	Clinical Study mplementation & Monitoring	Clinical Study Implementation & Monitoring	Analyze the results of a clinical trial	Electronic Health Records (EHRs) Clinical Data Management System (CDMS) Patient Monitoring Devices Curator Analyst	Curator Analyst
Commercial Operations Analyst	CO.1	Patient Journey	Track key milestones (e.g., initial diagnoses, patient enrollment) for patients and therapies in a particular therapeutic area	Improve patient services and better serve the HCPs and payers in each market	Payer / HCP / Product Master Patient Demographics Diagnoses Claims Procedures Steward Curator Analyst	Steward Curator Analyst
	CO.2	Aggregate Spend	Determine the amount of payments, gifts, honoraria, travel, and other material compensation to HCOs and HCPs	Comply with federal and state reporting requirements	HCP / HCO / Product / Employee Master Meeting and Events Speaker Bureau Employee Expenses Grants and Clinical Spend	

On the other end of the spectrum, let's also examine a few industry-agnostic use-cases for a Director of Digital Marketing:

Persona	#	Use-case(s)	I want to...	So that I can...	Data	Roles
Director of Marketing (DM)	DM.1	Cross-Channel Digital Marketing ROI	Blend marketing spend, hit-level clickstream, web analytics, social media engagement, and sales	Understand which promotions are maximizing ROI through cross-channel engagement and increased sales	GL from ERP HCP/HCO & Product from MDM Channel sales from a Commercial Data Mart Promotional spend from GL Clickstream from weblogs Web analytics from Google Analytics Social media	Steward Curator Analyst
	DM.2	KOL Management	Measure the influence that HCPs have in a particular therapeutic area or individual therapy	Identify and engage key-opinion-leaders throughout the entire therapy development lifecycle		
	DM.3	Customer Segmentation	Cluster customers into distinct groupings using obvious and nonobvious features	Identify opportunities to target key messages and promotion materials to HCPs		
	DM.4	Promotions Page A/B Testing	Simultaneously run multiple versions of a promotions web page	Determine which content, layout, and UX perform best		

VERISTA

ALIGNING ROLES ALONG THE ANALYTIC SPECTRUM

Recognizing that individual metrics do not count as a use-case, when all is said and done, most organizations might gather 50-75 true analytic use-cases in the initial sweep of the stakeholder community. This might represent 20-30 personas. Depending on the industry, the use-cases will hover across the spectrum. For a Biopharmaceutical company, the use-cases will lean towards governed data & semi-governed analytics AND governed data & analytics. Once the underlying details of a use-case have been gathered, you can align each one to the analytic spectrum of semi-governed data and analytics. Similarly, you can cast the personas against the spectrum, providing you a radar for defining and detecting analytic use-cases.
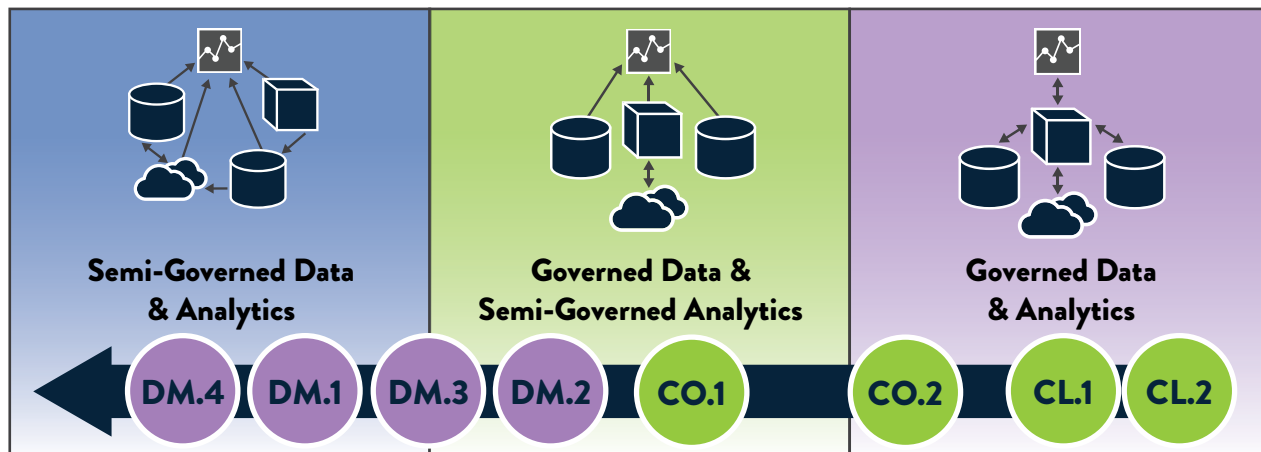
TO THE POINT

Defining use-cases at the outset is essential in anticipating knowledge workers' need and establishing a system that empowers users while creating an environment that promotes optimal productivity.
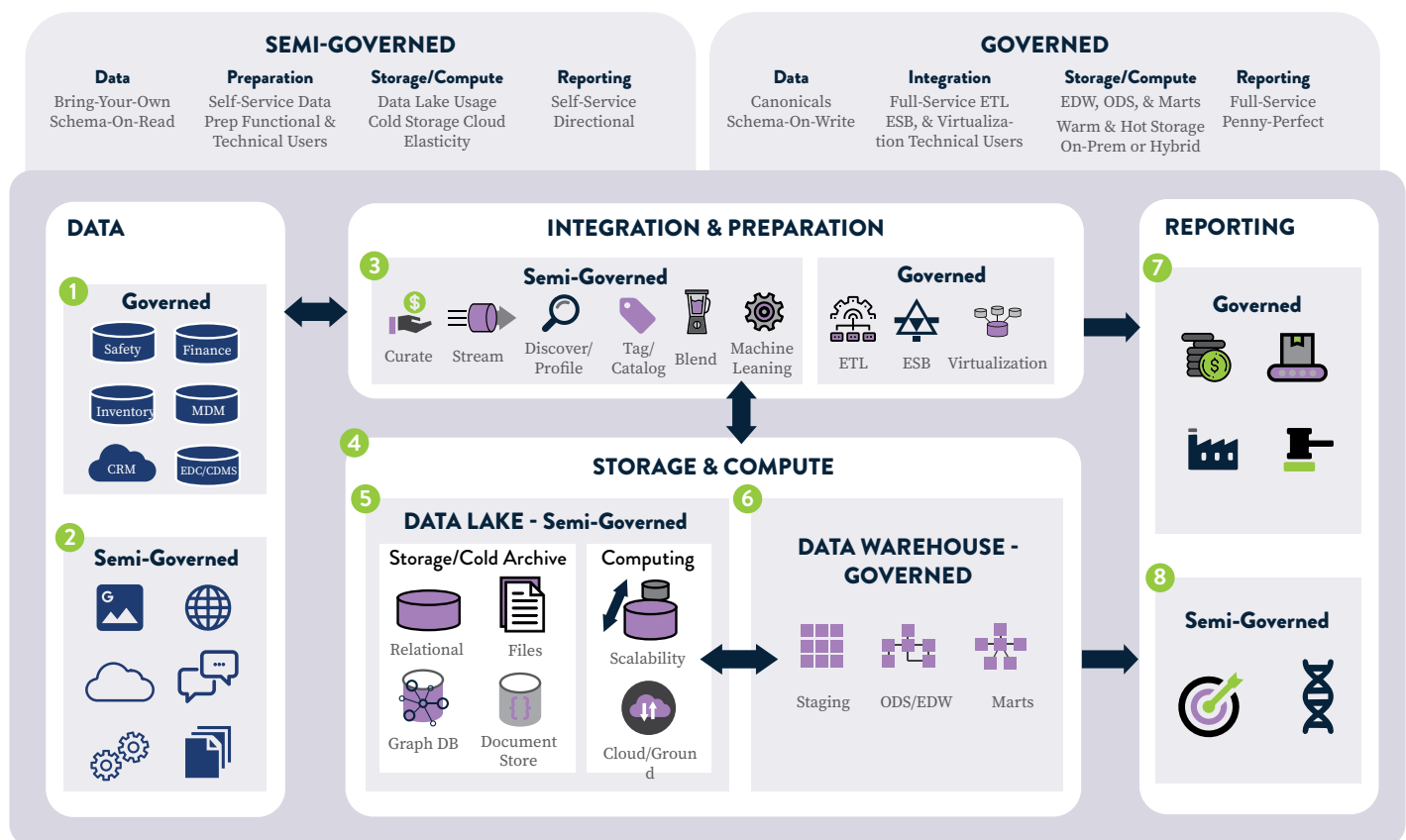
MODERN ANALYTIC CAPABILITIES

Shortly after the process model detects and defines analytic use-cases, the required capabilities will quickly surface. Modern analytic capabilities include:

- √ Self-service data curation, prep, stewardship and BI
- √ Schema-on-read (i.e., unstructured) and schema-on-write (structured).
- √ Separation of storage from compute
- √ Cloud and hybrid deployments
- √ Flexible integration styles (e.g., streaming, ETL, ESB, virtualization)

- √ Tiered storage and memory, i.e., in-memory, active storage, active archive, cold storage
- √ Auto-scaling
- √ Advanced visualizations
- √ Machine learning

ORGANIZING ANALYTIC CAPABILITIES

This laundry list of capabilities must be organized in a manner that corresponds to the continuum of governance-awareness. Namely the analytic use-cases might be delivered through the capabilities of existing technologies and platforms. Given the tech debt that companies accumulate over decades, the required capabilities will be beyond the horsepower of existing platforms and applications. Thus, it is expected to evaluate options to deliver the required capabilities to meet the growing demand of analytics. A technology-agnostic representation of modern analytic capabilities would look like:

MODERN ANALYTIC CAPABILITIES

KEY ASPECTS OF ANALYTIC CAPABILITIES

This is a baseline architecture intended to be a starting point, not a destination. Without dwelling too long, there are several key aspects that we will highlight:

1. **Governed data,** primarily structured and historically on-prem, but steadily transitioning to cloud or hybrid deployments.

2. The nature of analytics will produce an increasing demand for **semi-governed data**, from anywhere, at any time, in any shape.

3. **User self-service** is critical to meeting the growing business demand for data & analytics. Combined with governed data integration, semi-governed data integration is a focal point of the reference architecture.

4. The **storage and compute layer** brings together data for analysis, be it in the lake or the EDW.

5. A **Data Lake** provides flexible storage and scalable computing to manage governed & ungoverned data in a cost-effective manner.

6. The **EDW** is and will be essential for governed reporting. Represented are the 3 layers of the EDW.

7. **Governed reporting** requires the highest level of confidence in the data. Governed data from the EDW layer or the operational data sources themselves will serve the reporting needs of these users.

8. Not all business users need penny-perfect reporting, in which case **semi-governed reporting** will suffice. These users will benefit the most from self-service and flexible means of storage and compute.

It seems like not a lot of bang for your buck, but by design, the process has few steps and moves fast (without using the A-word, aka agile). The focus is on the details of the implementation, and not slideware. **Those concerned about architectural coherence and standards should be assured that governance is the organizing element of the process model. It pertains to all aspects of architectural governance - IT, business, and of course data.**
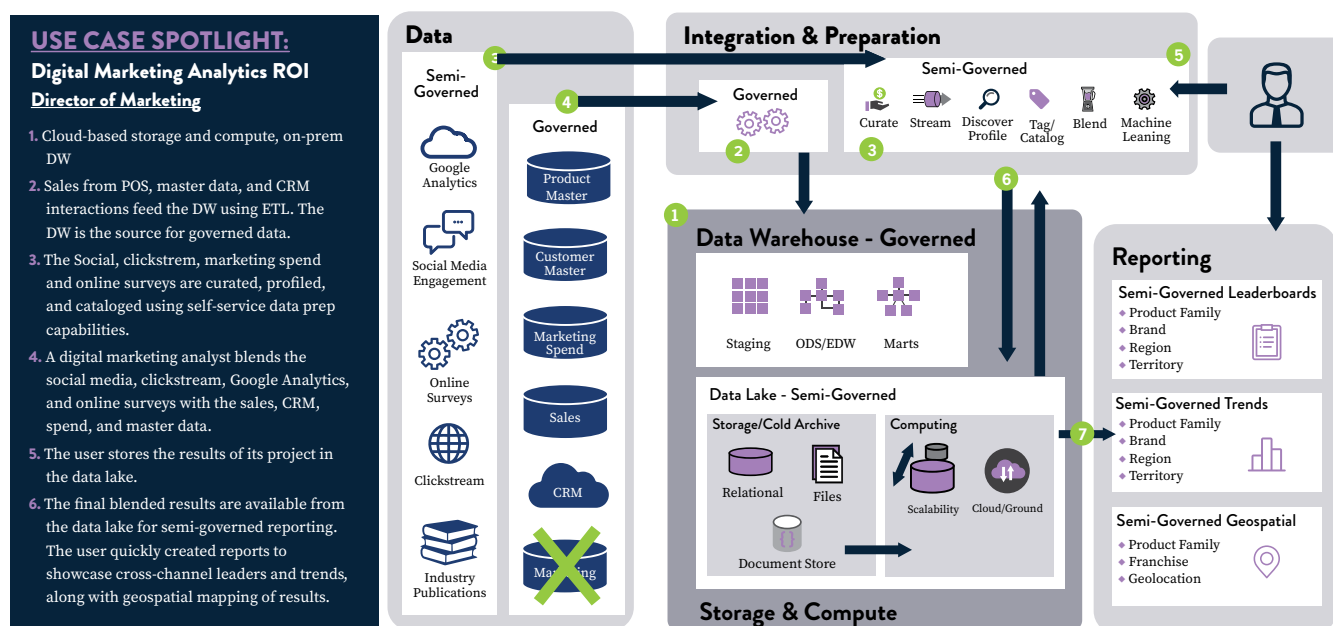
TO THE POINT

As the demand for analytics continues to expand it has become paramount to embrace a process model not shaped by technology but by governance focused on business development and expansion.

VERISTA

ANALYTIC REFERENCE ARCHITECTURE AND PATTERNS

Modern analytic capabilities take shape through the implementation of each individual analytic use-case. In the process model, a reference architecture instantiates the appropriate baseline capabilities in a manner that tightens the coordination between business and IT in a governance-aware fashion.

Using the digital marketing use-cases previously mentioned, the puzzle pieces fall into place. For instance, the Cross-channel Marketing ROI use-case requires self-service data prep capabilities that allow users to curate and blends clickstream and social media with governed data-like sales, CRM interactions, marketing spend, and master data. Cast against the capability model, the reference architecture for this use-case takes the following form:

## DIGITAL MARKETING ROI REFERENCE ARCHITECTURE

USE CASE SPOTLIGHT:
Digital Marketing Analytics ROI
Director of Marketing

1. Cloud-based storage and compute, on-prem DW
2. Sales from POS, master data, and CRM interactions feed the DW using ETL. The DW is the source for governed data.
3. The Social, clickstrem, marketing spend and online surveys are curated, profiled, and cataloged using self-service data prep capabilities.
4. A digital marketing analyst blends the social media, clickstream, Google Analytics, and online surveys with the sales, CRM, spend, and master data.
5. The user stores the results of its project in the data lake.
6. The final blended results are available from the data lake for semi-governed reporting. The user quickly created reports to showcase cross-channel leaders and trends, along with geospatial mapping of results.

KEY CHARACTERISTICS OF THE USE CASE

- Solid understanding of use case and data
- Governed & semi-governed data with semi-governed analytics
- Self-service data blends semi-governed data with governed data from DW
- Self-serve reporting, with a mix of dashboards and advanced data visualizations
- Data lake for storage and compute
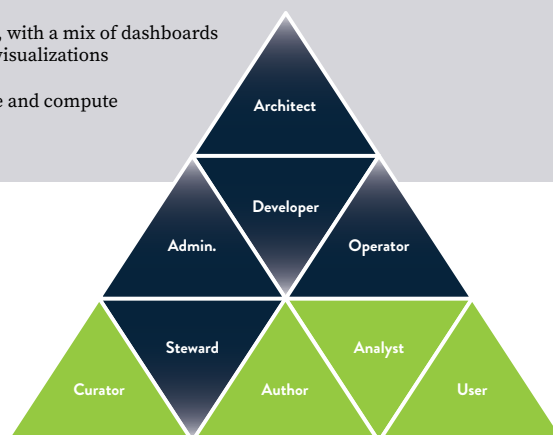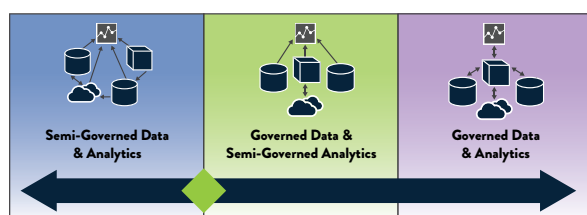
USE-CASES

The use-case for KOL management follows a similar pattern to marketing ROI. Specifically, the KOL management use-case for Digital Marketing requires a k-means cluster machine learning algorithm to segment KOL influencers from KOL followers, while adding one data source (online publications) and removing one data source (i.e., no marketing spend).

### NET PROMOTER SCORE REFERENCE ARCHITECTURE
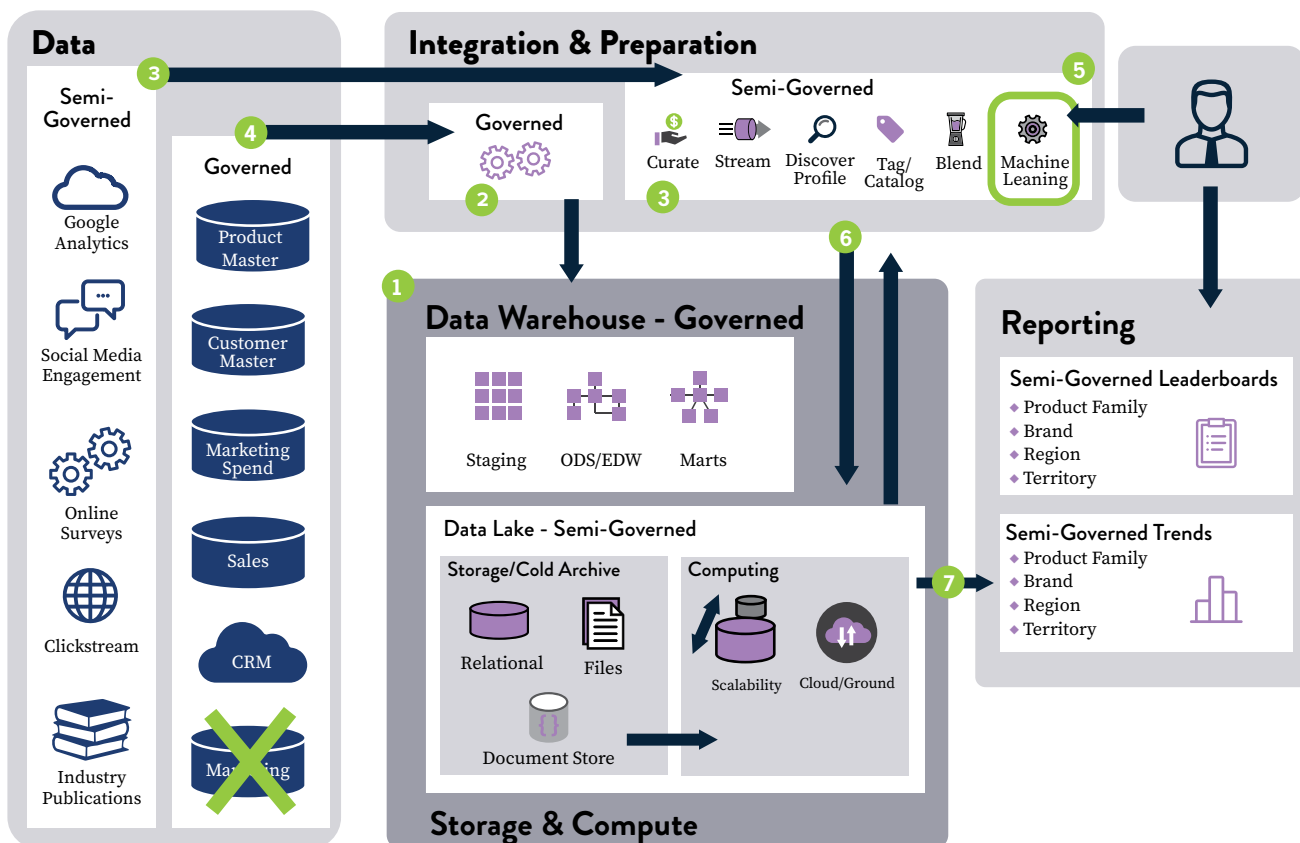
Based on these two use-cases, a clear pattern for "shovel-ready" self-service analytics is established, i.e., the users understand their data and have clear direction on how the data needs to come together. Rather than mobilizing IT resources to build out the use-case, the business users just need sanctioned IT tools and access to their governed and semi-governed data to deliver insights in a self-service manner. The final piece to the puzzle is the self-service reporting. In this pattern, there is no need for heavyweight, enterprise-scale BI with a semantic layer, etc. Users just need a license to a self-service BI platform, and connections into the data lake layer to visualize their data. Eventually, these processes can be scheduled and orchestrated into a production environment. Before long DevOps will harden to support analytics with the right mix of technical oversight, operational coordination, and timing.

## SHOVEL-READY SELF-SERVICE ANALYTICS PATTERN
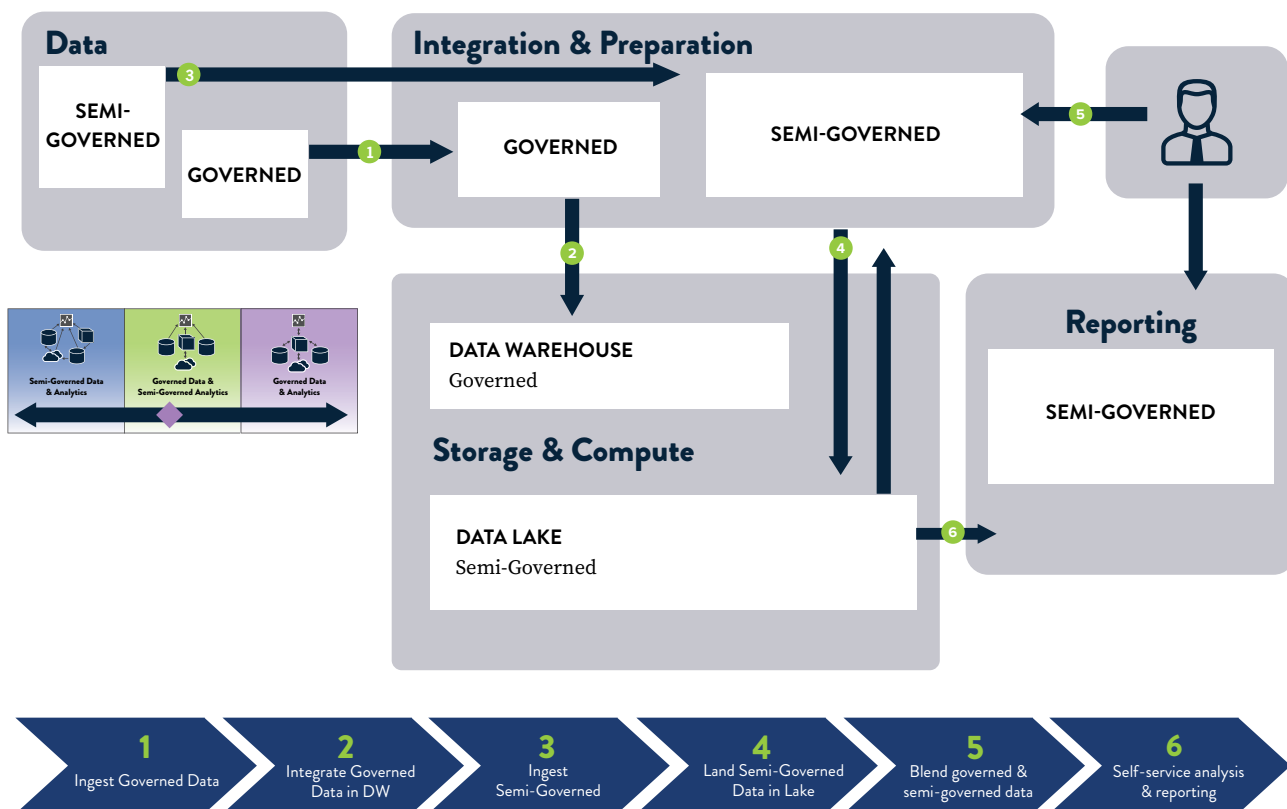
Over time new patterns will emerge. For instance, a use-case for self-service data prep and stewardship would look very different. Let's say that you are a Commercial Operations Analyst supporting a specialty pharmacy product in an open channel. Data feeds can sometimes come from smaller pharmacy outlets, and thus the formats can be varied and nonstandard. Your job is to blend the data together so it is prepped and ready – i.e., stewarded – to support governed HCPlevel channel sales reporting. This is a pedantic and mundane example of a common use-case, but it mustn't fall under the analytics radar given the model includes self-service data prep capabilities. This pattern can also extend to a Commercial Operations Analyst blending new files that contain primary investigator expenses from a clinical trial that are required for aggregate spend. This one pattern covers a wide span of the analytics continuum, depending on the nature of the reporting (i.e., self-service adhoc reporting versus full-service).

SELF-SERVICE PREP & STEWARDSHIP PATTERN
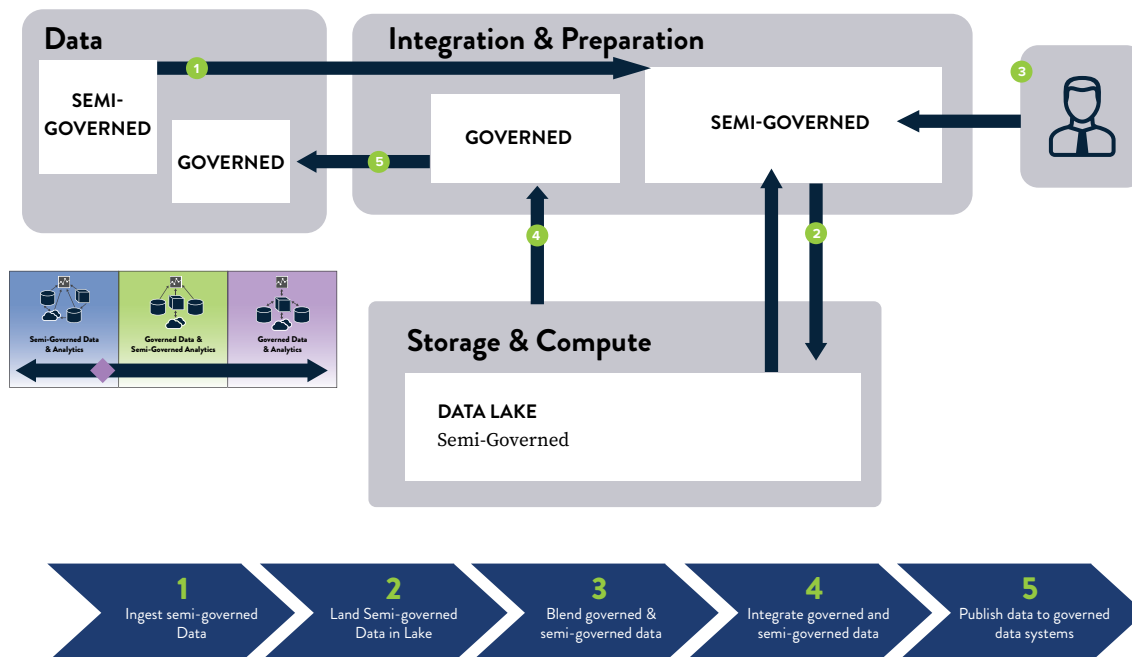
Other use-cases are more pioneering and fit a different analytic pattern. For instance, capturing and analyzing vitals from a medical device worn by a patient undergoing therapy involves a high-degree of engineering, streaming and machine learning capabilities. Given the considerations of patient consent & privacy (i.e., HIPAA, GDPR) and medical safety (i.e., Data Integrity per CFR Part 11), this pioneering use-case will require strict controls and possibly open the door for a Health Information Exchange. These use-cases and the supporting capabilities need to incubate before they transition from semi-governed to governed analytics. As the analytics incubate, the pattern will look something like this.

INCUBATOR ANALYTICS PATTERN
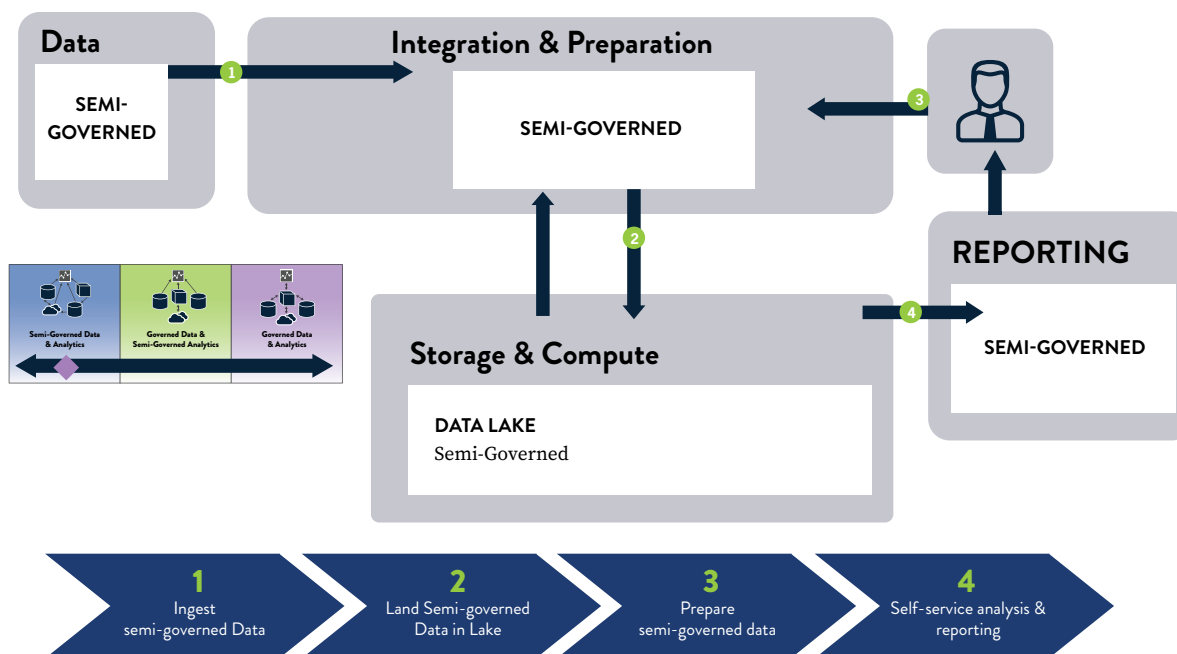
CLOSING THOUGHTS

The devil is truly in the details, and some details have been deliberately omitted from the reference architecture, such as which technologies deliver the capabilities. While important, the technologies will vary according to enterprise standard platforms, and the tech is constantly changing. More enduring and of equal or greater value is an underlying process model to define and detect governance-aware analytics, underpinned by ongoing outreach between IT and business. And with few steps in the process model, it gives ample opportunity to focus on the details and continual iteration of steps and revisiting implemented use-cases, to understand changing and new requirements, and new capabilities.

ABOUT VERISTA

Verista is a leading business and technology consultancy firm that provides systems, compliance, validation and quality solutions to life science companies enabling them to improve health and improve lives. We help clients solve their most critical and complex challenges across the GxP lifecycle, from preclinical and clinical to commercialization, manufacturing and distribution – bringing together decades of knowledge, the most advanced engagement platforms and transformative technologies. This allows clients to benefit from the ease, efficiency and trust that results from working with one partner who excels across specialties. Verista's clients trust the company's 500+ experts to deliver consistent, safe, and high-quality results across the product development lifecycle.

Visit www.verista.com