Part 1 - A Roadmap To Successfully Design A Cloud Data Management Program -Six Essential Considerations

Written by Bruce McCormack and Jonathan Thornbury (20-minute read)

Cloud Data Management Program vs. Solution Architecture - both drive meaningful decision making

Over the past five years, I have collaborated often with data architecture leaders at Fortune 1000 clients. They have compelling strategies attached to their drivers for a Cloud Modernization Program unique to their desired enterprise data capabilities and forward-looking competitive advantage to be gained. Frequently I hear exiting a data center by "Date X" where they avoid "Cost Penalty Y" while "Gaining Value From Data for Data Consumers – right now!" are the platform solution needs. There is an urgency to propose and stand up a flexible collaborative enterprise cloud data architecture… but I do not hear the same urgency regarding a Cloud Data Management Program.

This situation frames the right starting point for some introductory context. If your teams are not working with the same hygienic trusted data, cannot share common business context about the data, lack a common ontology, cannot share common data lineage and traceability - and lastly – cannot enable verifiable, consistent executive beneficial results with enterprise data – your Cloud Data Modernization effort will fail. This avoidable failure frames why a sustainable Cloud Data Management Program is crucial for successful business outcomes at Fortune 1000 data-driven enterprises.

The target audience for this article includes CTOs, Managing Directors of Global Technology/Corporate Tech and Chief Data (or Digitization) Officers, plus Executive Directors of Architecture and Data Analytics. Of course Data Stewards, Business Architects and Data Scientists can gain value from it as well.

Some clarifications here – first, this article does not advocate "the only way to Cloud Data Management Platform success is buying the whole platform – Day One". Rather quite the opposite – leading CDMPs allow for a modular start – standing up just the critical CDMP services you will need on a multi-cloud platform. Microservices are one way to reduce exposure to change, ensure noisy cloud platform neighbors are held in check, and guide periodic data management change at a simpler scale. Daunting commitments (of cost, expert resources, or vendor obligation) are not being advocated in this article – but your firm should consider successful foundations with enterprise data before the CDMP journey begins.

On that topic of framing successful enterprise data foundations for a CDMP – this next item certainly has value. In late 2022 Informatica completed a survey with two hundred (200) financial services leaders and found: 1) leveraging clean, accurate, fit-for-business-use data to fuel AI and machine-learning-powered analytics is essential in gaining material benefit from their digital investments. 2) High-quality data to populate modern cloud data lakes and platforms strengthens risk management and supports the adoption of open API standards to improve business agility. Note this dual approach (data at rest and data in motion) has become a must in Fortune 1000 Enterprise Cloud Data Architecture and is certain to be a pre-requisite for any architecture review board (ARB) you will be presenting to for approval.

Ask your firm - are you trying to achieve any of these below goals in your 2023 Enterprise Data Solution? Consider the following competitive grid focusing on foundational needs of the CDMP for a moment:

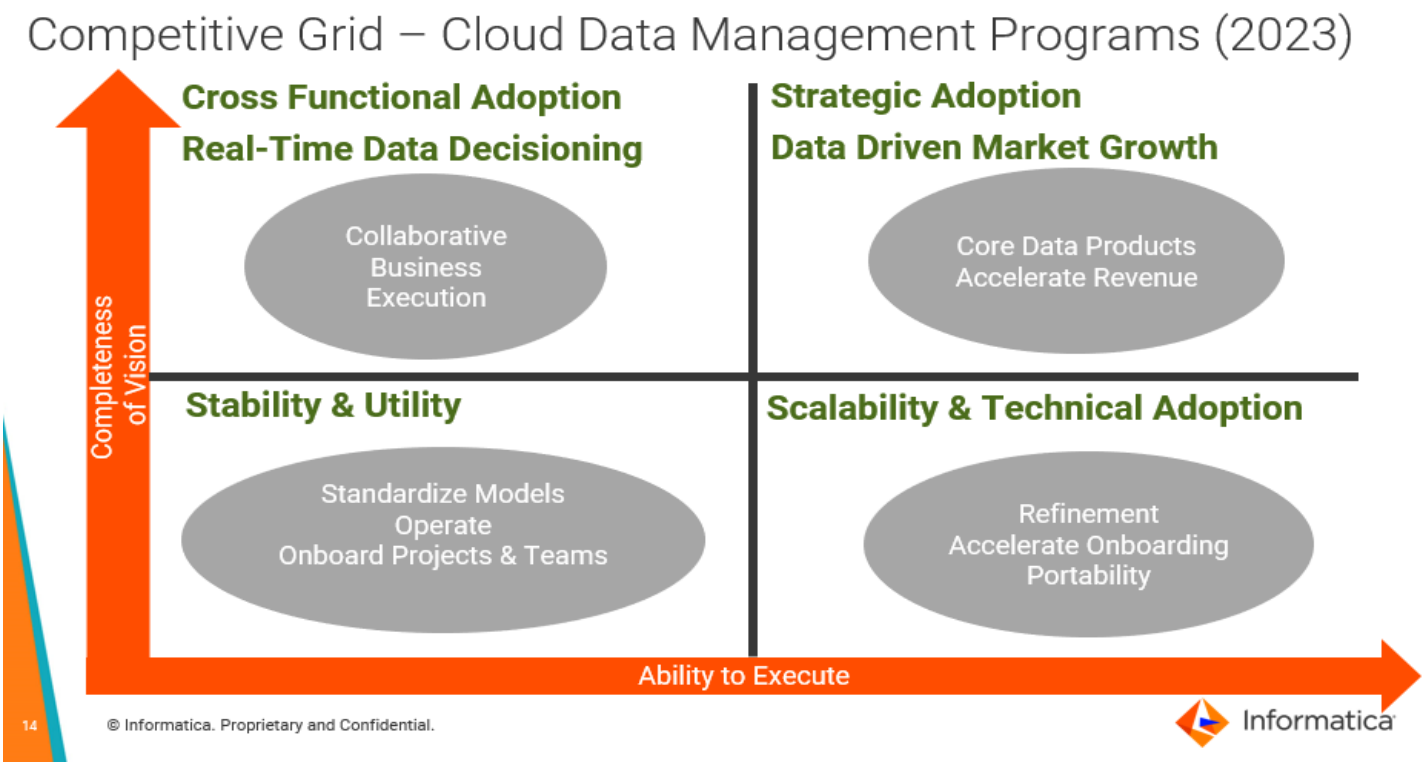

Figure X-1: Competitive Grid for CDMPs

A market defining leader and obvious example for Data As A Product is Dun & Bradstreet . The information in D&B is gathered from over 20,000 public and private sources. One result is D&Bs "gold standard" for corporate contact and executive information for firms within the public record in the US, Canada and internationally can be consumed as a data product for a subscription fee. At the outset this article highlights Data As A Product as it can be an early executive funding catalyst, and is gaining rapid acceptance as an initiative among Fortune 1000 clients. It is clear D&B is a CDMP leader - using its core data products to accelerate revenue per the above competitive grid. It made sense to share what a CDMP leader provides for Data As A Product.

The **purpose of this article** is to provide:

- An **initial toolkit** drawn from recent Fortune 1000 field experience to propose a successful Cloud Data Management Program
- Guidance toward an impactful, planned assessment of your Cloud Enterprise Data Management current state
- Identification of barriers that will surface reaching a sustainable, successful next state using a CDMP
- Real-world implemented solution examples that can accelerate next state goal and vision statements for CDMP
- Written alignment to your enterprise data goals and capabilities before defining and implementing a cloud solution architecture

This article is not advocating one practicing group or firm's CDMP learnings or framework over another. Read and use it as a repeatable toolkit and solution framework to assess, frame, prescribe and position what your firm could achieve with a CDMP - and receive feedback – before you implement.

We are confident this will be a worthwhile journey ahead – so let's get started.

I. Define (or refine) current mission statement for enterprise data

Create a CDMP Mission Statement – outline the goals, follow proven disciplines

Without input and participation from your executive corporate sponsors – the CDMP is certain to fail or at least endure a difficult launch. Any Data Management Platform can only succeed when aligned to disciplines and goals called out in a charter or mission statement. The below bulletized list is not exhaustive but is a reliable proven place to start:

*Defining CDMP Goals – "Doing and Acting" - Being Lean**

- Be constantly focused on customers and the data consumers that service them
  - In 2023 these will be both internal and external
- Automate data-driven processes
  - Use AI/ML tools to reduce manual data wrangling and synchronization
- Seek out continuous improvement
- Exploit, demand and measure data quality - early and often
- Design a CDMP that can absorb change
- Take the total CDMP and optimize where possible
- Use rigorous measurements to eliminate waste and remove technical debt

*Following Proven CDMP Disciplines – "Monitor, Correct and Control" – Being Lean**

- Financial Management
  - Avoid isolating data driven activities, bring metrics/opportunities to business sponsors for cost savings and revenue bearing outcomes
  - Repeatedly seek out incentives or capabilities where enterprise data (see below) adds to top-level revenue and brand impact
- Resource Management
  - Early on, manage the skillset required to plan, design, integrate, maintain and operate your CDMP as a core business function – not a technology only function
  - As your CDMP grows, avoid the common trap of reliance on exorbitant or prohibitive skillsets simply to harden and operate it
  - Without resource management (RM) in check – the below management and methodology needs will impede CDMP success - make RM a priority
- Business Process Management (BPM)
  - Encourage CDMP designs with online feedback (BPEL processes, voting rights, facet/data asset rating) from key business users
  - Maintain awareness that your business will evolve to competitive pressures over time – recruit business architects who keep a keen eye on this and maintain an adaptable framework, ontology (and CDMP)
  - Carefully research and assess cloud provider solutions that realize full business processes but align to your CDMP principles
    - Used properly, this can lessen complexity, while providing time to value and operational savings - especially when multiple BUs exploit them
- Modeling Management
  - Propose and iterate through an ontology that captures your business functions and BU relationships.
  - Realize that proven mature (process and data) models for energy sector, financial sector and healthcare firms already exist
  - Define meaningful abstractions that decouple core business processes (how, why, when) from data driven applications that deliver them (what, when/frequency).
- Metadata Management
  - Adopt and expand safe proven standards such as MIMB and OWL as foundational for your CDMP exchange and ingestion of metadata
  - Treat metadata and data as equally important in your design and planning. Without consistent timely metadata you begin the trap of "manually keeping up" with data quality, lineage, traceability changes
  - If existing business metadata domains have served you well retain them. If business sponsors dispute or do not use the metadata domains – planning your cloud solution offers the opportunity to revisit this
- Integration Methodology

- Start with a pliable, flexible platform approach to your integration. Gather a roster of proven, innovative CDMP platform providers – allowing projects and applications to grow over time as necessary.
- Assess then add services to your architecture that enhance existing abilities or provide new distinct capabilities
  - Yes exploit massive scale data, application and workload pipelines on cloud…but focus on reuse of safe OOTB features (e.g. mass ingestion, automated data hygiene, self-healing DI to enforce model integrity) that avoid painful rework and dirty data
- Do not build silos…enforce periodic voting rights on the core components of your integration methodology

*Above bullets do borrow from Schmidt & Lyle's publication "Lean Integration". They are slightly re-ordered with some recent field observations inserted.

Impactful Incentives and Capabilities Driving Change – Cloud Data Management Programs

Business incentive or core capability	Surfacing questions	Examples of applicable impact, metrics and existing solutions
Composable Applications	Do repeating apps impose too much operational complexity? Can innovations scale as they come to market (e.g. AI trained models)?	Packaged Business Capabilities (PBCs) become reusable Key Principles to standup a suite of cloud-only apps in shorter time with less end-user expertise required.
Hyperautomation, Digitize Customer Experience (CX)	Are you losing clients to peers who have adopted hyperautomation? What first steps can counter this?	Order to Cash (O2C) fee-based apps take rapidly changing inputs (email, API requests) to accurately create, update delete orders and can reconcile backend data at rest systems.
Provide purpose-driven timely data for critical decisions	Is current data latent for timely (or even useful) decision making? Will this worsen on cloud?	A volatile complex financial product lost 50% valuation in one hour – client was not alerted to act.
Prevent fraud from data or digital processes	Does our fraud prevention strategy need to be revisited for cloud? What recent events cause concern here? Can cloud first AI be trained to assess and avoid fraud?	An outlier order (due to expense) from an approved pharma vendor is auto-approved for a healthcare account – with no context-based alert. The account is drained to $0 and locked out.
Data As A Product	How do our sector peers exploit Data As a Product (revenue stream, brand expansion, market credibility)? Is there a pre-built framework to reuse or refactor?	Dun and Bradstreet has built an entire data product business model starting with credit reporting on firms in all sectors. A description of DQ in their Data Product process is here.

The narrative and bullets from above are drawn from the fact that C-level sponsors are not paying for a crusade, nor will they budget for "flying car" scenarios servicing enterprise data. Be deliberate, outcome oriented and focused outlining the CDMP goals to executive sponsors. Realize there are completely new (but viable) ways to monetize data and bolster your competitive brand position with data that did not exist five years ago. Feel free to shape the above goals to your culture… but be clinical in validating your CDMP mission and proposal are *aligned to business goals*.

If your firm is struggling to connect these business capabilities (or other compelling ones) to their CDMP goals – several trusted proven cloud modernization providers exist ready to provide a business value assessment.

A business ontology – research, assess and propose

At a Fortune 1000 firm it is likely that Business SMEs, Product Owners and Business Architect are available who have created an ontology for your firm, or at past firms in your market space. Whether working in financial services, healthcare, public service or another sector there are very mature ontologies for financial services and healthcare you can examine.

An ontology provides a common set of entities, properties, relationships, verbs and terms that represent core concepts in a domain and how they interact. Using your firm's tribal knowledge, critical analysis, and relevant new facts, items which are not codified in the original ontology can be refined or derived. Collaborative repeated assessment of your ontology allows your firm to placate long-standing concerns and bolster standards that improve external and internal business processes.

An ontology for an insurance P&C (Property & Casualty) is listed below:

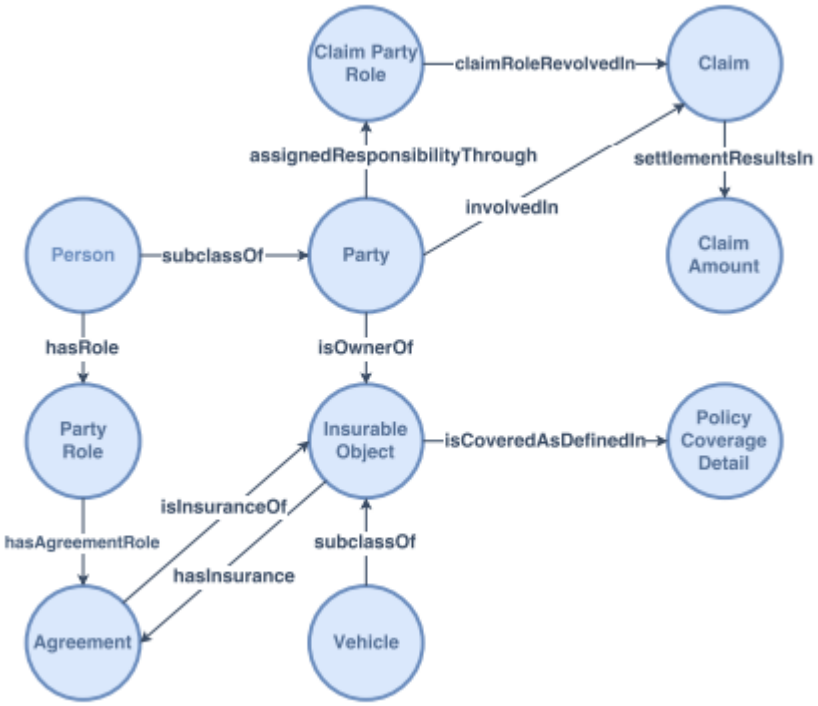

Figure X-2 Representing a P&C ontology including Claims, Parties, Policies and Insurable Objects

Note above that the time was taken to answer these required concerns:

- What key entities must be identified (e.g. Party, Claim, Policy) to begin addressing this solution space
- What are critical interactions and dependencies for entities (e.g Which is a subclass – are these handled differently?)
- What interactions require logic-based criteria
    - When is a party involved/not involved in a claim
    - If the claim exceeds $1000, is different process logic invoked

Having an ontology allows the cross-BU team (here at a P&C firm) to start useful conversation about their enterprise data **goals and core concerns for a CDMP in a substantial targeted manner**. Getting this ontology socialized allows stakeholders who are not SQL tuning, Spark tuning, or data definition experts to access and understand what is being proposed and assessed for this CDMP in a collaborative manner. Experienced practitioners can then partner with your firm and start to assess what capabilities are best suited to a private vs. public cloud for a CDMP, what expected data bottlenecks surface for complex third-party use cases with the ontology, do non-traditional (NoSQL, key-pair database)

data stores provide quick wins here, when do elastic compute and storage help with volume, bottleneck and seasonal barriers, what data should belong or be prohibited (sensitivity, validity, periodicity, relevance) in a VPC for this ontology?

Hopefully it is clear an ontology can break down the barriers to gathering business processes, what is applicable in a cloud solution, and a phased CDMP together in a common conversation. Or conversely you may have a multi-national P&C firm that has to segment this into two ontologies to start (perhaps EU and North America) due to differing regulations and business models.

By design this article puts your firm in the right more comfortable position to ask:

- What business capabilities must be fulfilled with enterprise data in a CDMP?
- What architecture capabilities support those business capabilities for the CDMP?

Without knowing the business behaviors and outcomes you are reaching with data….throwing public cloud commodity services at your CDMP vision may yield unpredictable outcomes that block progress and reduce stakeholder credibility.

Desired business capabilities with enterprise data

Below is not an exhaustive list but should provide foundational context for charting vision and goals to start your CDMP:

- Name enterprise data consumers (role, individual person [e.g. data product owner] or integrated service (e.g. a downstream marketing AI model trainer) who have defined goals – such as
    - Route data UX toward business track leads, steer 50% more activity to templates and wizards
    - Reduce complex data finishing for business data consumers, or prior to model training - enable self-service
    - Reduce data-driven ticket requests
        - Can data products and data services be consumed at an internal site or “no tech” provider site
- What are business objectives, constraints and goals
    - Examples: exit data center within twelve months, reduce cost of enterprise data operation by 20% over two years, turn in-house marketing analytics into for-profit data product within three years
    - Gain executive budget awareness (or better still guidance) around their multi-year CDMP vision
- What business processes or key projects benefit from improved data quality, observability, lineage, SLA outcomes
    - Be specific and name these – looking for quick wins as well as foundational CDMP starting points
- Which of your firm’s executive(s) have articulated cloud capability accelerators, impactful goals, market changes related to enterprise cloud data in an article
    - Most important – did your executive propose a cloud framework or methodology for the entire firm
    - Have these executives changed entire service models or what is available to clients based on the CDMP vision
- Avoid how and where at this point – focus on why, what and “what is urgent” for these goals

Desired architectural capabilities related to enterprise data

The most important requirements before exploring the below topics - you have a business sponsor and business architect in the room and a documented result was captured from the business capabilities for data gap analysis listed we examined prior:

- Is there an overarching mission (or vision) statement you can share capturing enterprise data goals for 2023
- What documented key goals or mandates from business sponsors can be drawn from that mission statement
- Are there a set of business models (or an ontology) that align to that mission statement
- What adopted industry standard data architectures, vocabularies or OOTB “common data models” could provide long-term gains (and quick wins) per the above goals & mission statement
- Focus on goals and outcomes - using the above where can you reduce technical debt over time

    - Has above items changed (or disappeared) since prior corporate architecture review
- Avoid how at this point – focus on why, what within the goals and priorities

With your business goals, ontology, desired enterprise data business capabilities and architectural capabilities starting to take shape – this the right place to prepare a deeper assessment of CDMP readiness with your team.

II. Clinically assess – your Enterprise Data challenges that exist today

Before starting this clinical assessment – take a moment. Consider where you firm is within the Enterprise Data Gap (Creators v. Consumers) challenge depicted below. This problem absolutely exists from Informatica's the past five years of hearing vision statements and executive consultation with stakeholders at North America's fifty largest banks. Simply put:

- Project Track Leads measure value in timely project delivery and project sponsor satisfaction
- Data Scientists often measure value in
- Business Architects often measure success in terms of frameworks and standards that efficiently, safely deliver key capabilities and projects for executive business sponsors

An executive authority will need to prioritize and delineate enterprise data goals, success measurements and capabilities to keep your CDMP on-track – while applying the innovative talents of the competing roles above longer term.

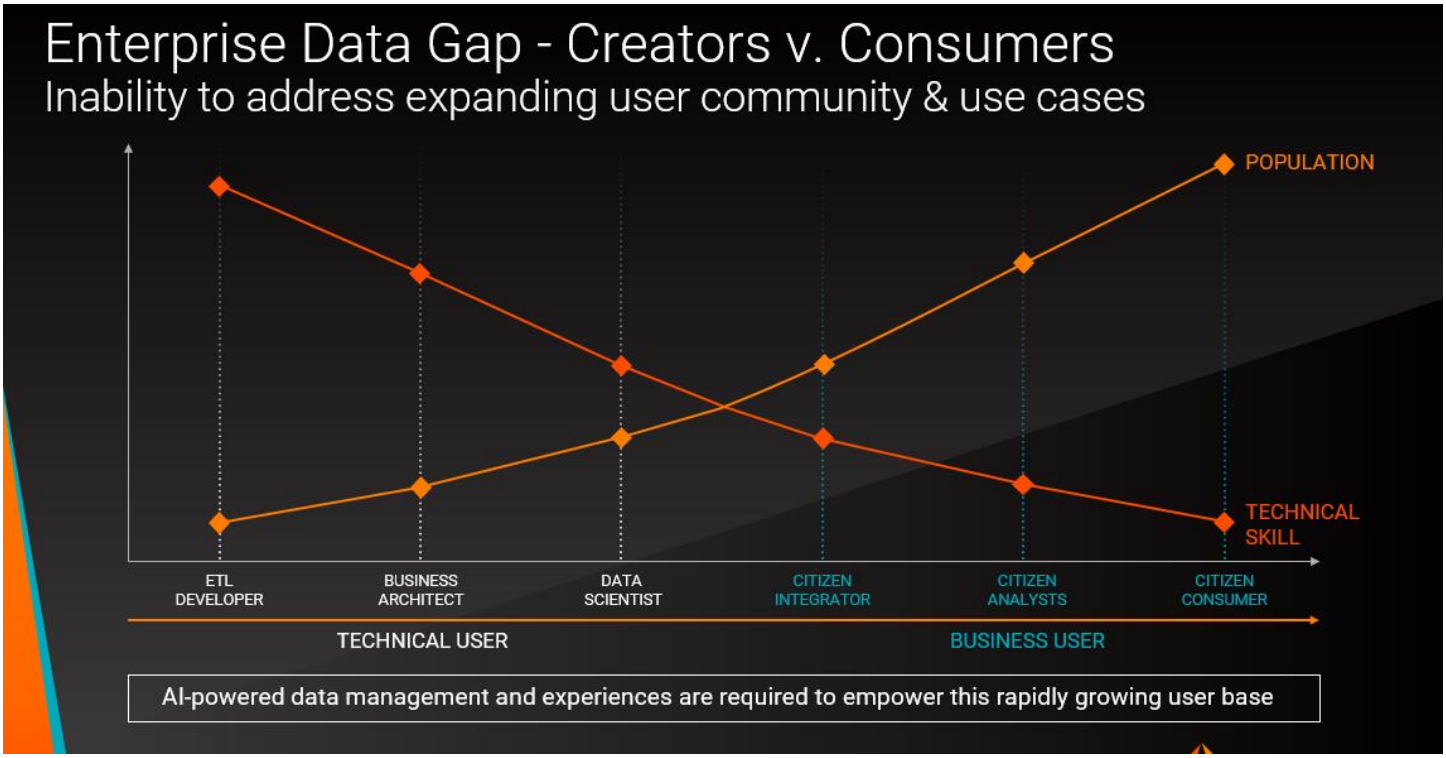

Figure X-3 Visualizing the Enterprise Data Creators vs. Consumers Gap

As an Enterprise Data Leader it is essential that you are considering the above in your assessment exercise tied to "what we are trying to solve for as an organization – the future state of a CDMP". If you do not consider the gaps in user experiences and user outcomes depicted above the future state CDMP will surely fall short in solving them.

At this point your firm should be starting to shape a common CDMP strategy, the candidate logical architecture for the Cloud Data Management Plan has directional agreement from voting rights participants, a list of trusted GSIs and vendors are participating who will maintain and operate the CDMP, and all parties agree to the goals and capabilities of it.

Avoid the four bottlenecks

Working left to right these blockers are prevalent and constant across all of the Fortune 1000 clients I have engaged the past five years seeking to create a Cloud Data Management Program. It is important to avoid block diagrams that generalize this complex problem set…and be very specific with clinical analysis of what needs to be solved, and in what order.

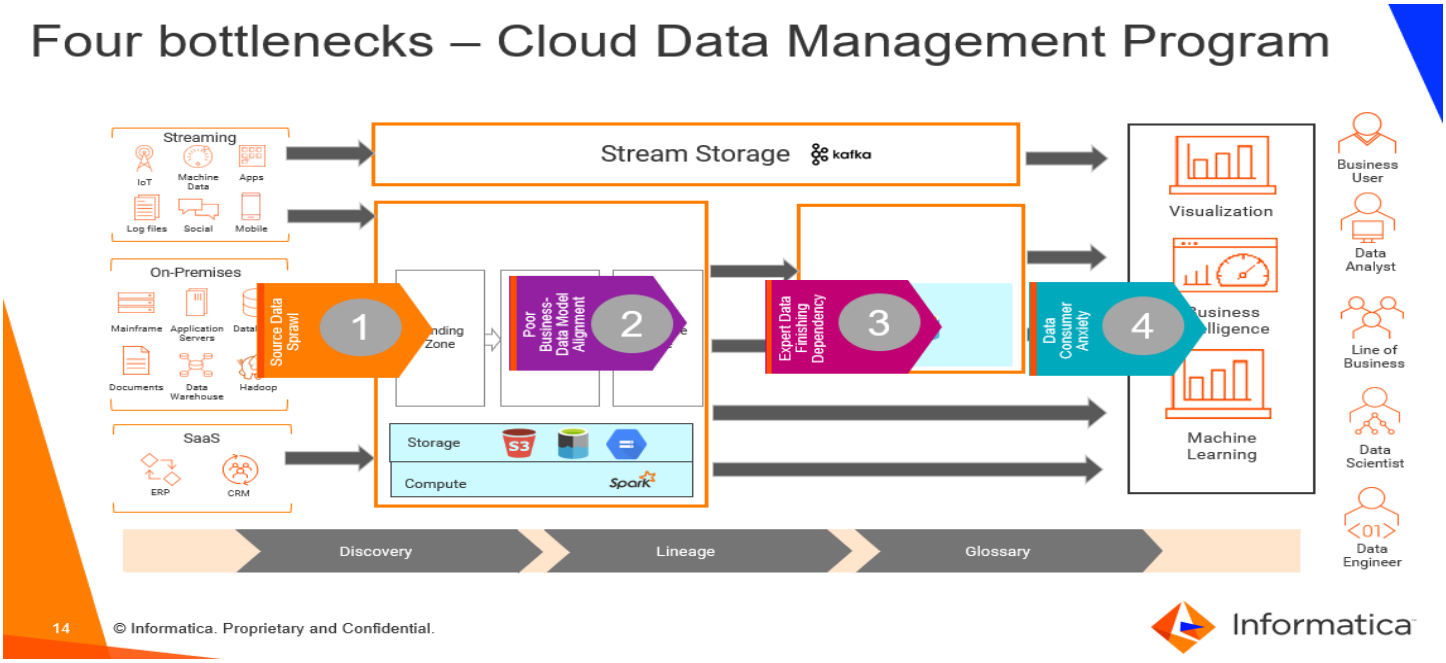

Figure M-1 The Four Bottlenecks

1.Source Data Sprawl – data discovery is not in place so basic "fit for use" data hygiene cannot be surfaced at enterprise scale for raw staged data (a.k.a bronze stage data). Data controls are not in place with acceptance thresholds (e.g. <90% correct/complete is unacceptable, above 90% is marginally acceptable, above 95% is acceptable). Data stewardship is not in place where template-driven scorecards are curated, driven by automated data controls that are reused and published for standard scorecards as well as custom scorecards as required. Business data owners lack a metadata driven mechanism to assign high or low trust to sources of record that data stewards can approve or interact with online.

2.Poor Business-Data Model Alignment – this can be as simple as not having a contemporary business data dictionary in place (with voting rights and stewards to maintain the dictionary). Without AI-driven data suggestion (based on a search request or input model) the business data architect relies upon tribal knowledge or a DDL-driven "best guess" to find sources that comprise the data aligned to business processes, decision needs and desired outcomes. Ontologies are not agreed upon and updated so there is not a common vocabulary of how upstream data applies to business processes or entities. Active metadata does not classify upstream assets in context with the ontology or applicable domains (e.g. insurance claims per prior example). Data quality, traceability and data literacy are incomplete – so business architects cannot attest transformed data meets business logic or make corrections "pointing to the best trusted source".

3.Expert Data Finishing Dependency – the data output either does not match the business consumer's content, context or form expectation. The CDMP lacks intelligent data enrichment or Data-As-A Service (DaaS) to fill the gaps surfaced in data that is not quite correct or complete but is traceable to entity of interest (e.g. customer, household, patient) Symptoms of this situation include – data consumers "send data back" to have content, context or form fixed by a data steward (or worse a data engineer), what I hear most often is data is not ready for its *last mile* to the consumer. Manual

processes are relied upon with internal semi-automated frameworks that have become culturally safe and trusted but are *highly vulnerable to change* and *people powered when adapting predictable periodic* changes.

4.Data Consumer Anxiety – the data consumer points at catalog tool surfacing gaps between the data content/form he gets vs. what is desired or much worse the data consumer manually goes upstream to a raw staging area and "wrangles the data himself". The most devastating impact to your Cloud Data Management Program occurs when data consumers lose trust in your data product….and seek that trusted data elsewhere. Often Fortune 1000 clients have invested millions of dollars (and multiple years of expert effort) to insert tools massaging "the last mile of data"– but still arrive at this conundrum. That is the one behavior this article seeks to coach you away from.

To overcome the intersection of (1) and (2), start planning for tools, processes and a UX that will collect data for entities, collect data depicting the relationship between entities, and collect the necessary data (and metadata) that keeps an accurate data hierarchy up to date and intact. We will examine a very basic enterprise data hierarchy for overcoming these four bottlenecks later in the article.

Knowing the four bottlenecks – and keeping them mindful in the context of your organization's CDMP current state means your team is ready to start your enterprise assessment process.

Rely on the Mission Statement, Enterprise Goals and Capabilities and Ontology gathered above to drive your assessment as "agreed to principles, already approved by E-Staff". This will enable a time-boxed, scope-bound assessment driven by defined needs and less prone to cultural resistance – though ***some resistance is inevitable in an assessment.***