

Do Diagnóstico à Ação: Maturidade em Governança de Dados no Varejo Farmacêutico via DCAM

Fabiano Arriaga de Ramanzini¹; Rebecca Manesco Paixão²

¹Coordenador de Governança de Dados · RD Saúde · São Paulo, SP · Brasil

²ESALQ/USP. Doutora em Engenharia Química. Professora orientadora associada, Piracicaba, SP, Brasil

Do Diagnóstico à Ação: Maturidade em Governança de Dados no Varejo Farmacêutico via DCAM

Resumo

Este estudo diagnosticou o nível de maturidade em governança de dados da RD Saúde (Raia Drogasil), maior rede de farmácias do Brasil, adotando a perspectiva da pesquisa-ação: o autor atuou simultaneamente como pesquisador e como agente de implementação do roteiro proposto. Utilizando o framework DCAM (*Data Management Capability Assessment Model*) como instrumento de avaliação, foi realizado um diagnóstico misto (qualitativo e quantitativo) envolvendo 62 colaboradores, 12 áreas, 432 critérios de auditoria e 25 reuniões. Os resultados evidenciaram uma assimetria estrutural: alta capacidade tecnológica (nota 3,90) contrastando com baixa maturidade em gestão e processos (notas 1,58 e 1,36, respectivamente), caracterizando o estágio “Silo/Reativo”. As evidências preliminares de implementação do roteiro proposto — Política Corporativa vigente desde fevereiro de 2026, sistema de monitoramento de qualidade de dados em produção e assunção formal do mandato sobre o Google Cloud Dataplex em abril de 2026 — validaram a capacidade preditiva do DCAM: os componentes com menores escores foram precisamente os que apresentaram as maiores barreiras na execução real. Concluiu-se que a maturidade em governança de dados não é questão de tecnologia disponível, mas de processos definidos, papéis claros e cultura analítica — fundamentos que nenhuma plataforma de nuvem substitui, mas que toda estratégia de Inteligência Artificial pressupõe.

Palavras-chave: Governança de Dados; Maturidade de Dados; DCAM; Qualidade de Dados; Varejo Farmacêutico.

INTRODUÇÃO

A atual conjuntura econômica, marcada pela transformação digital acelerada e pela intensa pressão do mercado por inovações disruptivas, posicionou a implementação de sistemas de Inteligência Artificial (IA) e *machine learning* no centro da agenda estratégica das organizações líderes. No entanto, a materialização e a confiabilidade desses projetos de alta complexidade esbarram invariavelmente em um desafio estrutural crítico: a qualidade e a governança dos dados corporativos.

Segundo Davenport e Patil (2012), o dado transcende a definição de um mero registro técnico; ele deve ser gerido como um ativo intangível fundamental que determina a eficácia de qualquer solução analítica. Uma avaliação preliminar do estado atual (“As-Is”) na organização orgânica demonstra lacunas em dimensões críticas, como Integridade, Acurácia e Conformidade Legal (LGPD). A utilização de dados não auditáveis pode induzir decisões de negócios equivocadas e riscos reputacionais severos (IBM, 2020).

Para orquestrar a resolução desses desafios, a Governança de Dados surge como obrigatória. A DAMA (2017) define Governança de Dados como o ecossistema de políticas, processos e estruturas que garantem que o ativo de informação atenda às necessidades de negócio. É uma governança que estabelece papéis claros — diferenciando *Data Owners* de *Data Stewards* — e garante a *linhagem* do dado desde a coleta até o descarte.

Neste contexto, o setor do varejo farmacêutico brasileiro, onde a informação é vital para a logística e personalização, enfrenta o paradoxo de possuir ecossistemas tecnológicos robustos coexistindo com uma crise de confiança nos dados. A ausência de um *Master Data Management* (MDM) resulta eficaz em “catálogos humanos” e crescimento do *Shadow IT*.

O presente estudo adota a perspectiva da pesquisa-ação (*action research*), conforme conceitualizada por Yin (2015), na qual o pesquisador é simultaneamente agente de diagnóstico e de mudança. Nesse delineamento, o investigador não apenas observa o fenômeno, mas participa ativamente de sua transformação, o que permite correlacionar, em tempo real, as hipóteses diagnósticas com as evidências empíricas de implementação. Tal abordagem é especialmente adequada ao contexto de gestão e governança de dados em grandes organizações, onde a distância entre o diagnóstico acadêmico e a realidade operacional frequentemente invalida modelos puramente teóricos.

O objetivo deste trabalho foi diagnosticar o nível de maturidade em governança de dados de uma grande organização do varejo farmacêutico brasileiro, mediante a aplicação do *framework* DCAM (*Data Management Capability Assessment Model*), identificar as lacunas estruturais entre a capacidade tecnológica instalada e a gestão efetiva dos dados, e correlacionar os achados do diagnóstico com as evidências preliminares de implementação do *roadmap* proposto — com ênfase na adoção do Google Cloud Dataplex como plataforma central de governança e catálogo de dados em nuvem.

MATERIAIS E MÉTODOS

Delineamento da Pesquisa

O presente estudo configurou-se como uma pesquisa aplicada de natureza exploratória e descritiva, conduzida através de um estudo de caso único. A investigação ocorreu entre os meses de setembro e novembro de 2025, tendo como objeto de análise a estrutura de gestão de dados da RD Saúde (Raia Drogasil), a maior rede de farmácias do Brasil, com aproximadamente 75 mil funcionários, mais de 2.800 unidades em operação e receita bruta superior a R\$ 32 bilhões (RD Saúde, 2025). A organização opera em regime de alta complexidade transacional, com ecossistema tecnológico distribuído entre plataformas de nuvem AWS e GCP em processo de migração. O método de trabalho foi estruturado em quatro fases sequenciais: (1) Mobilização e Contextualização; (2) Avaliação (Aplicação do Diagnóstico); (3) Análise dos Resultados; e (4) Definição do Roteiro Estratégico.

Amostragem e Seleção dos Participantes

Para garantir a representatividade do diagnóstico, utilizou-se uma amostragem intencional por julgamento. Não se buscou uma seleção solicitada, mas sim a participação de

colaboradores-chave que detinham conhecimento profundo sobre a criação, consumo e gestão de dados corporativos. O universo da pesquisa compreende 62 colaboradores distribuídos em 12 áreas específicas, abrangendo tanto o espectro técnico (Engenharia de Dados, Arquitetura, Ciência de Dados) quanto o espectro de negócios (Suprimentos, Lojas, Financeiro e Operações).

A Tabela 1 apresenta um resumo do universo da pesquisa, evidenciando a amplitude e representatividade do diagnóstico realizado.

Tabela 1. Universo da pesquisa — participantes, áreas e materiais analisados

Reuniões Realizadas	Áreas Representadas	Pessoas Engajadas	Horas de Entrevistas	Materiais Analisados
25	12	62	12	80

Fonte: Resultados originais da pesquisa

Essa diversidade foi fundamental para evitar as vitórias puramente tecnológicas e capturar as "dores" reais de quem utiliza as informações para a tomada de decisão.

Instrumento de Coleta de Dados: Framework Detalhado

Para garantir a capilaridade necessária ao diagnóstico de uma organização complexa como esta, o instrumento de pesquisa não se limita a grandes temas. O framework DCAM v2.2 (*Data Management Capability Assessment Model*) foi expandido e estruturado em uma matriz de quatro níveis hierárquicos, totalizando 432 critérios de auditoria.

A lógica de construção do instrumento aumenta uma distribuição simétrica para evitar vieses de análise:

- **8 Componentes Principais** (Nível Estratégico);
- **3 Passos DCAM** por componente (Nível Tático), totalizando 24 passos;
- **3 Dimensões de Detalhamento** por passo (Nível Operacional), totalizando 72 tópicos de controle;
- **6 Perguntas de Verificação** por detalhamento (Nível de Auditoria), perfazendo um total de 432 pontos de controle.

Essa estrutura, detalhada na Tabela 2, permitiu identificar não apenas se houve falha de processo, mas *onde* ele falhou (se definição de visão, nos princípios ou na arquitetura).

Tabela 2. Estrutura Analítica do Framework de Avaliação.

(continua)

Componente:(Nível 1)	Subitem / Capacidade: (Nível 2)	Foco da Avaliação: (18 perguntas focam em:)
1. Estratégia e Negócio de Dados (54 Questões)	1.1. Definir Visão	Formalização do "Sonho Grande" e alinhamento C-Level.
	1.2. Caso de Negócio	ROI, Business Case e justificativa financeira.
	1.3. Princípios	Ética, valores e diretrizes fundamentais do dado.
2. Programa e Financiamento (54 Questões)	2.1. Definir Papéis	Nomeação formal de Proprietários e Administradores.
	2.2. Modelo Operacional	Como a governança funciona no dia a dia (RACI).
	2.3. Financiamento	Orçamento aprovado e sustentabilidade financeira do programa.
3. Arquitetura de Negócio e Dados (54 Questões)	3.1. Glossário de Negócios	Definição semântica única para termos corporativos.
	3.2. Modelagem Lógica	Desenho das entidades e relacionamentos (DER).
	3.3. Classificação	Taxonomia, ontologia e categorização de dados.
4. Arquitetura de Dados e Tecnologia (54 Questões)	4.1. Catálogo de Dados	Ferramentas de Catálogo de Dados e Discovery.
	4.2. Linhagem (Lineage)	Rastreabilidade da origem ao consumo (End-to-End).
	4.3. Arquitetura Híbrida	Integração entre Legacy (On-Premise) e Cloud.
5. Gerenciamento de Qualidade (DQ) (54 Questões)	5.1. Regras de Qualidade	Definição técnica (formato, nulls, ranges).
	5.2. Monitoramento de DQ	Painéis de Pontuação de Saúde e alertas.
	5.3. Remediação	Processos de Correção (Limpeza) e causa raiz.
6. Governança de Dados (54 Questões)	6.1. Aplicação de Políticas	Criação e publicação de normas internacionais.
	6.2. Gestão de Mudanças	Controle de impacto em alterações de estrutura.
	6.3. Disputas/Escalonamento	Fórum para resolver conflitos de propriedade.
7. Operações, Risco e Controle (54 Questões)	7.1. Monitoramento de Oleodutos	SLAs de ingestão, falhas e observabilidade.
	7.2. Backup e Retenção	Política de guarda e recuperação de desastres.
	7.3. Gerenciamento de Riscos	Identificação de vulnerabilidades e compliance.
8. Gerenciamento de Análises (54 Questões)	8.1. Entrega de Dados	Disponibilização de Conjuntos de Dados confiáveis para consumo.
	8.2. Gov. de Dashboards	Certificação de relatórios oficiais (Gold).

Componente:(Nível 1)	Subitem / Capacidade: (Nível 2)	Foco da Avaliação: (18 perguntas focam em:)
	8.3. Governo de Modelos (IA)	Explicabilidade, vida e ciclo de vida do ML.

Fonte: Resultados originais da pesquisa

A relação integral dessas perguntas encontra-se detalhadamente no **Apêndice A**, organizado por bloco temático.

Nota sobre Proteção de Dados:

Por decisão metodológica de proteção de dados operacionais e competitivos da RD Saúde, as pontuações granulares atribuídas a cada um dos 432 critérios foram omitidas deste documento. O Apêndice A apresenta a estrutura integral do framework de avaliação, demonstrando o rigor metodológico sem expor informações estratégicas sensíveis. Os dados brutos encontram-se arquivados pelo autor e disponíveis para consulta acadêmica mediante solicitação formal.

Procedimentos de Análise e Evidenciação

A análise dos dados avançou uma abordagem mista (triangulação de métodos), combinando a pontuação quantitativa com a validação qualitativa documental.

Mecânica de Pontuação (Pontuação)

Para cada um dos 432 critérios, atribui-se uma nota em escala de maturidade de 6 níveis (0 a 5). A atribuição da nota baseada na metodologia de "Auditoria Baseada em Evidências", exigindo contratos comprobatórios (políticas, logs, atas) para notas iguais ou superiores a 3 ("Definido").

Tabela 3. Matriz de Critérios para Atribuição de Maturidade (Lógica de Pontuação).

(continua)

Nota Atribuída	Descrição da Nota	Critério Lógico de Auditoria (Obrigatório para a Nota)	Exemplo Prático Encontrado (Proxies)
5	Otimizado	Processo automatizado, com métricas de SLA e melhoria contínua ativa.	Monitoramento do tempo de atividade de servidores (Tech) ou consumo automático de dados por meio de rastreadores.
4	Gerenciado	Processo executado consistentemente, apoiado por ferramentas oficiais.	Uso disseminado de Data Lake (S3/Redshift) e pipelines padronizados.

Nota Atribuída	Descrição da Nota	Critério Lógico de Auditoria (Obrigatório para a Nota)	Exemplo Prático Encontrado (Proxies)
3	Definido	Processo documentado e comunicado, mas com manual de execução ou parcial.	Existência de Business Case ou Visão Estratégica aprovada pelo C-Level.
2	Repetível	A execução depende de “heróis” ou conhecimento tribal. Existem pilotos isolados.	Iniciativas pontuais de qualidade de dados em uma equipe específica, sem padrão corporativo.
1	Inicial	Inexistência de padrão. Processo reativo (“apagar incêndios”).	Criação de tabelas sem documentação, regras de negócio verbais e duplicidade de esforços.
0	Inexistente	O tema nunca foi abordado na organização.	Ausência completa de discussão sobre ética em IA ou descarte de dados.

Fonte: Resultados originais da pesquisa

Lógica de Ponderação e Agrupamento Dimensional

Para mitigar a subjetividade, o processamento analítico não se limita à média aritmética simples. Os critérios foram reclassificados em cinco macro dimensões analíticas (Estratégia, Tecnologia, Gestão, Processos e Pessoas), permitindo isolar a influência da infraestrutura tecnológica sobre a governança de processos.

A Tabela 4 apresenta a síntese dessa engenharia analítica, demonstrando como a disparidade entre as dimensões impactou o índice geral.

Tabela 4. Consolidação Analítica por Dimensão.

(continua)

Macro dimensão Analítica	Componentes DCAM Agrupados	Qtd. Critérios (Perguntas)	Média Calculada (μ)	Diagnóstico Lógico (Racional da Nota)
1. TECNOLOGIA	Comp. 4 (Arquitetura de TI) + Comp. 8 (Análise)	114	3,9	Alta Maturidade: Reflete o investimento massivo em Cloud (AWS/GCP), Data Warehouse (Redshift) e ferramentas modernas (Airflow/Databricks). As evidências técnicas sustentam a nota.

Macro dimensão Analítica	Componentes DCAM Agrupados	Qtd. Critérios (Perguntas)	Média Calculada (μ)	Diagnóstico Lógico (Racional da Nota)
2. ESTRATÉGIA	Comp. 1 (Estratégia de Dados)	54	3,75	Alta Intenção: O C-Level compreende o valor dos dados (Data Driven). A nota é sustentada pela existência de financiamento e visão, embora a execução tática falhe.
3. GESTÃO	Comp. 2 (Gov. Dados) + Comp. 3 (Arq. Dados) + Comp. 6 (MDM)	144	1,58	Vácuo de Governança: Nota penalizada pela ausência de papéis formais (Proprietários/Administradores) e falta de um Conselho de Dados deliberativo.
4. PROCESSOS	Comp. 5 (Qualidade) + Comp. 7 (Operações)	120	1,36	Baixa Maturidade: O principal ofensor. Processos manuais, reativos e sem documentação. A alta volumetria de correções (retrabalho) justifica a nota próxima a 1.
5. PESSOAS	Transversal (Cultura e Capacitação)	(Qualitativo)	1,3	Dependência Tribal: Baixo letramento em dados (Data Literacy) e alta dependência de conhecimento tácito, gerando gargalos operacionais.
MÉDIA GERAL	TODOS OS COMPONENTES	432	2,38	Estágio Silo / Reativo

Fonte: Resultados originais da pesquisa

Esta segregação analítica evidencia que a média geral (2,38) seria uma "medida de tendência central enganosa" se observada isoladamente, pois esconderia a assimetria estrutural da organização: uma infraestrutura de ponta subutilizada por processos imaturos.

O FRAMEWORK DCAM

Para garantir que a análise da maturidade dos dados da organização não se baseie em percepções subjetivas ou métricas puramente tecnológicas, este estudo foi desenvolvido como Alicerce Teórico e Metodológico o DCAM (*Data Management Capability Assessment Model*). Desenvolvido pelo EDM Council (*Enterprise Data Management Council*), uma

associação global que reúne mais de 250 organizações líderes, o DCAM é reconhecido como o padrão da indústria para a estruturação de programas de Gestão e Governança de Dados.

A escolha do DCAM em detrimento de outros modelos justifica-se pela sua abordagem pragmática e externa para *capacidades*. Diferentemente dos frameworks teóricos que descrevem "o que" é governança (como o DAMA-DMBOK), o DCAM fornece os critérios de auditoria para mensurar "como" a organização executa essa governança. Ele permite traduzir conceitos abstratos em resultados tangíveis, facilitando o *benchmarking* e a comunicação com a alta direção.

A Estrutura dos 8 Componentes Nucleares

O framework estrutura a gestão de dados em 8 componentes interconectados, que cobrem desde a definição estratégica até a execução analítica avançada

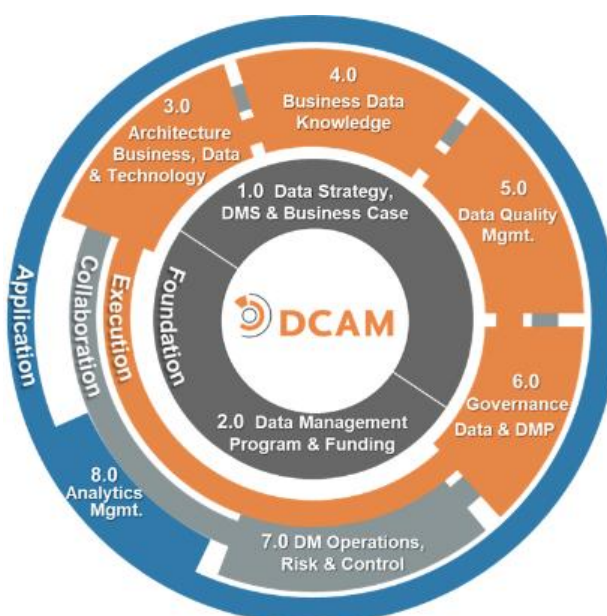


Figura 1 - O modelo DCAM v3 do EDM Council e seus oito componentes principais.

Fonte: Framework DCAM EDM Council. <https://edmcouncil.org/frameworks/dcam/>

Estratégia de Dados

O ponto de partida do modelo. Avalia se a organização possui um *Business Case* formal para dados, alinhamento com os *stakeholders* executivos e um modelo de financiamento sustentável.

Programa de Governança

Defina o "quem faz o quê". Este componente verifica a existência de um modelo operacional formalizado, incluindo a nomeação de *Data Owners* e *Data Stewards*.

Arquitetura de Dados

Avalia a estruturação lógica e física da informação, focando na existência de modelos de dados e glossários de negócios.

Tecnologia e Arquitetura de TI

O "alicerce digital". Mede a capacidade da infraestrutura tecnológica para apoiar a estratégia de dados.

Qualidade de Dados

Definir os processos de perfilamento, regras de negócio e monitoramento contínuo. Avaliar se a qualidade é gerida de forma preventiva (*by design*) ou reativa.

Governança de Dados Mestres (MDM)

Foca na gestão das "Entidades Douradas" (Clientes, Produtos, Filiais), verificando processos de *match & merge*.

Operações de Dados

Tratamento do ciclo de vida, conformidade (LGPD), retenção e descarte seguro.

Analytics e Ciência de Dados

Avalia se os modelos de IA são construídos sobre dados governados, explicáveis e éticos.

RESULTADOS E DISCUSSÃO

Os resultados do diagnóstico expõem uma assimetria estrutural na organização: a empresa possui "músculos tecnológicos", evidenciados por uma nota de **3,90** no pilar de Tecnologias, mas carece de um "sistema nervoso" de coordenação, refletido nas notas alarmantes de Gestão de Dados (**1,58**), Processos (**1,36**) e Pessoas (**1,30**). A classificação geral de maturidade atingiu **2,38**, classificando a organização no estágio "Silo/Reativo", distante do alvo estratégico de 4,00 ("Plataforma").

A figura abaixo representa a maturidade das macros dimensões analíticas identificados ao longo das análises dos processos operacionais

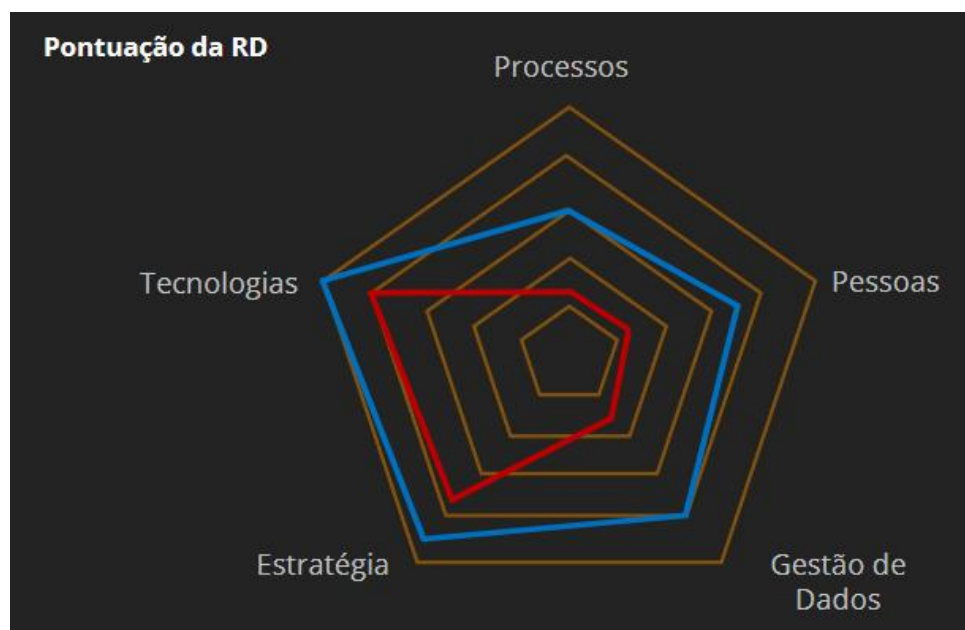


Figura 2: Gráfico de Radar - Nível de Maturidade Consolidado (Cenário As-Is).

Fonte: Resultados originais da pesquisa

Análise de Aderência: Discrepância entre Potencial e Realizado (Gap Analysis)

Para aprofundar o diagnóstico além das médias aritméticas, foi feita uma análise de aderência absoluta, confrontando a pontuação obtida (*Score Realizado*) com a pontuação máxima possível (*Score Ideal*) para cada componente do framework. Esta visualização permite medir a magnitude do “Passivo de Governança” existente na organização.

Considerando que o framework DCAM está avaliando a modernidade em uma escala de eficiência máxima, os gráficos a seguir evidenciam o abismo operacional entre as dimensões de infraestrutura e as dimensões de controle.

Tabela 5: Análise de Gaps - Comparativo entre Pontuação Máxima Possível e Atingida por Componente.

(continua)

Componente DCAM	Nota Alcançada	Nota Máxima
1. Estratégia e Negócio de Dados	200	270
1.1. Definir Visão	72	90
1.2. Caso de Negócio	66	90
1.3. Princípios	62	90
2. Programa e Financiamento	72	270
2.1. Definir Papéis	24	90
2.2. Modelo Operacional	24	90
2.3. Financiamento	24	90
3. Arquitetura de Negócio e Dados	88	270

Componente DCAM	Nota Alcançada	Nota Máxima
3.1. Glossário de Negócios	24	90
3.2. Modelagem Lógica	35	90
3.3. Classificação	29	90
4. Arquitetura de Dados e Tecnologia	208	270
4.1. Catálogo de Dados	65	90
4.2. Linhagem de Dados (Lineage)	68	90
4.3. Arquitetura Híbrida	75	90
5. Gerenciamento de Qualidade de Dados	76	270
5.1. Regras de Qualidade	22	90
5.2. Monitoramento de DQ	27	90
5.3. Remediação (Correção)	27	90
6. Governança de Dados	88	270
6.1. Aplicação de Políticas	29	90
6.2. Gestão de Mudanças	35	90
6.3. Disputas / Escalonamento	24	90
7. Operações de Dados, Risco e Controle	72	270
7.1. Monitoramento de Pipelines	25	90
7.2. Backup e Retenção	21	90
7.3. Gerenciamento de Riscos	26	90
8. Gerenciamento de Análises	137	270
8.1. Entrega de Dados	46	90
8.2. Governança de Dashboard	49	90
8.3. Governança de Modelos (IA/ML)	42	90
Total Geral	941	2160

Fonte: Resultados originais da pesquisa

Aderência Tecnológica (Alta)

Nos componentes de Tecnologia e Arquitetura de TI, a barra de "Realizado" preenche quase a totalidade da capacidade instalada. Isso confirma que a compra de ferramentas (Capex) foi realizada com sucesso.

Déficit de Processos (Crítico)

Nos componentes de Qualidade (DQ) e Governança (DG), observa-se a maior discrepância absoluta. Uma vasta área "em branco" entre o realizado e o máximo representa os processos que deveriam existir para sustentar a tecnologia, mas que hoje são inexistentes ou informais.

Esta representação gráfica isola a causa-raiz da baixa maturidade: a organização possui o "motor" (Tecnologia) pronto para desempenho máximo, mas opera com "freio de mão puxado" devido à ausência de ritos de gestão e papéis definidos.

Uma análise de eficiência técnica revelou um passivo arquitetônico significativo. Identificou-se a existência de **1.690 DAGs** (*Directed Acyclic Graphs*) ativos na ferramenta Airflow, operando com uma redundância estrutural onde existem, em média, dois processos para cada tabela. Esse modelo impede a escalabilidade e inflaciona os custos de infraestrutura de nuvem.

A figura abaixo representa a volumetria dos processos identificados ao longo das análises dos processos operacionais.

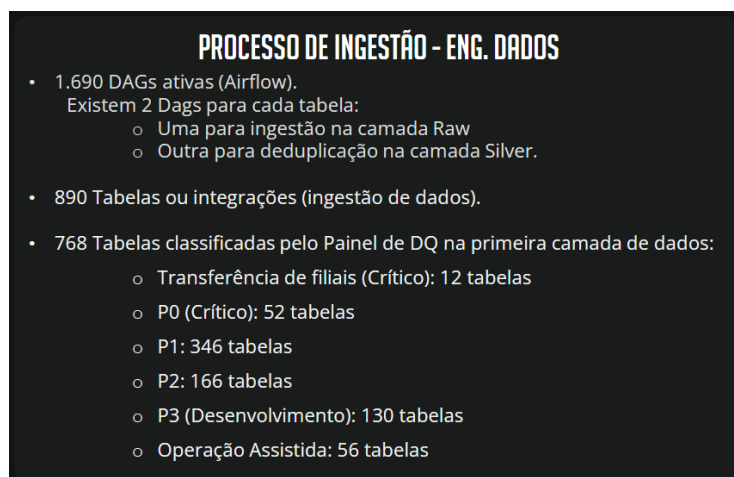


Figura 4: Volumetria de processos e identificação dos processos Operacionais.

Fonte: Resultados originais da pesquisa

No que diz respeito à Qualidade de Dados, foram mapeadas 768 tabelas monitoradas, sendo 52 com criticidade máxima (P0). Apesar desse monitoramento, a percepção dos usuários finais permanece de baixa qualidade, impulsionada por um Catálogo de Dados apenas técnico.

Análise prévia dos Resultados por Componente

Uma análise granular dos componentes revela uma discrepância acentuada entre a capacidade tecnológica e a de gestão.

A Disparidade Tecnológica (Componente 4 vs. Componente 2)

O resultado mais alarmante é o desequilíbrio estrutural. O Componente 4 (Tecnologia) obteve **3,90**, evidenciando investimentos massivos em *Big Data* e *Redshift*. Em contrapartida, a Componente 2 (Governança) obteve **1,58**. Isso indica que faltam “pilotos habilitados” para as “ferraris” tecnológicas. As evidências mostram que não existe um Conselho de Dados deliberativo e os papéis dos *Proprietários de Dados* são informativos.

A Fragilidade nos Processos de Controle (Componente 5 e 7)

O Componente 5 (Qualidade de Dados) pontuou apenas **1,36**. A qualidade é gerida de forma reativa (*combate a incêndios*), sem regras preventivas na ingestão, ou que explica a duplicidade de DAGs para limpeza. Da mesma forma, o Componente 7 (Operações) pontual **1,30**, refletindo a ausência de gestão do ciclo de vida, expondo a organização a riscos jurídicos e custos de armazenamento de dados obsoletos (*Dark Data*).

O Paradoxo da Estratégia (Componente 1)

A Componente 1 (Estratégia) obteve uma nota razoável de **3,75**. Isso sugere que a alta direção tem visão clara do valor dos dados, mas existe uma *lacuna* de execução tática entre o desejo do C-Level e a realidade operacional.

Validação Preliminar do Diagnóstico: Evidências de Implementação

O diagnóstico DCAM atribuiu nota 1,58 ao pilar de Gestão de Dados, identificando como causa-raiz a ausência de papéis formais, catálogo de dados abandonado e ausência de processos preventivos. A varredura técnica realizada em abril de 2026 — após a nomeação formal do coordenador de governança como proprietário do Google Cloud Dataplex — confirmou objetivamente: (a) catálogo EDC com mais de 14.000 ativos sem nenhum aspecto de domínio aplicado; (b) cinco zonas do *Data Lake* com classificação incorreta (todas configuradas como brutas, mesmo as camadas *Silver*, *Gold* e *Trusted*, que deveriam ser refinadas); (c) nenhum *dataset* com metadados de *Data Owner* preenchidos. A correspondência entre o diagnóstico teórico e o achado técnico empírico valida a capacidade preditiva do *framework* DCAM neste contexto.

O Componente 2 (Governança de Dados) obteve nota 1,58, evidenciando ausência de papéis formais e normas publicadas. Como primeira iniciativa de resposta, foi elaborada e publicada a Política Corporativa de Governança e Gestão de Dados (16 páginas), vigente desde 6 de fevereiro de 2026, com escopo abrangendo toda a Vice-Presidência de Operações. O documento estabelece critérios técnicos de bloqueio de produção, define papéis de *Data Owner* e *Data Steward* por domínio e institucionaliza o poder de veto da Governança sobre ingestões fora de conformidade — exatamente as lacunas apontadas pelo diagnóstico.

O Componente 5 (Qualidade de Dados) registrou nota 1,36 — o segundo pior resultado —, caracterizado pelo modelo reativo de gestão e ausência de regras preventivas na ingestão. Em resposta, foi implementado um sistema de monitoramento de Qualidade de Dados em tempo real, integrando n8n (orquestração), *DBT* (transformação) e Airflow (agendamento), em produção desde o primeiro trimestre de 2026. O sistema monitora ingestões críticas com alertas automáticos que identificam a localização exata da falha — evoluindo do modelo reativo para o preventivo, respondendo diretamente ao diagnóstico.

O *gap* mais crítico identificado pelo diagnóstico foi a inexistência de um catálogo funcional: o sistema EDC estava operacionalmente abandonado, com *discovery* dependendo de conhecimento informal. Em 2 de abril de 2026, a Coordenação de Governança de Dados assumiu formalmente o mandato de proprietário do Google Cloud Dataplex. A varredura técnica realizada em 8 de abril de 2026 confirmou que: a linhagem automática está funcional via BigQuery/Dataflow; o Gemini/LLM está integrado ao *Knowledge Catalog* para busca em linguagem natural; dois *Aspect Types* de metadados foram criados; e a camada *Raw* do *Data Lake* foi vinculada ao catálogo. O Dataplex Universal Catalog representa a materialização do “Catálogo Funcional” proposto como ação prioritária de curto prazo nas considerações finais do diagnóstico.

CONSIDERAÇÕES FINAIS

O diagnóstico realizado confirmou que a organização opera em estado de assimetria estrutural: uma infraestrutura tecnológica de alto desempenho — avaliada em 3,90 — convive com pilares de gestão, processos e pessoas que não ultrapassam as notas 1,58, 1,36 e 1,30, respectivamente. O *score* geral de 2,38 posicionou a empresa no estágio “Silo/Reativo”, caracterizado pela dependência de conhecimento informal, ausência de papéis formais de governança e gestão reativa da qualidade de dados.

A análise de *gap* revelou que o principal obstáculo ao uso escalável de Inteligência Artificial não reside na escassez de tecnologia — que está disponível e bem avaliada —, mas na ausência de processos, normas e estruturas humanas que transformem dados técnicos em ativos confiáveis de negócio. Esse paradoxo manifestou-se concretamente na existência de 1.690 DAGs com redundância estrutural e em um catálogo EDC abandonado com mais de 14.000 ativos sem metadados aplicados.

As evidências preliminares de implementação coletadas nos primeiros meses de execução do *roadmap* proposto validaram a acurácia do diagnóstico. A Política Corporativa de Governança publicada em fevereiro de 2026, o sistema de monitoramento de qualidade de dados em tempo real implantado no primeiro trimestre e a assunção formal do mandato de proprietário do Google Cloud Dataplex em abril de 2026 responderam diretamente às três lacunas de maior criticidade apontadas pelo *framework* DCAM: ausência de normas formais (Componente 2, nota 1,58), processos reativos de qualidade (Componente 5, nota 1,36) e catálogo de dados inoperante (Componente 4, item 4.3).

A trajetória observada sugere que o *framework* DCAM demonstrou capacidade preditiva no contexto estudado: os componentes com menores *scores* foram precisamente aqueles que, na execução real, apresentaram as maiores barreiras operacionais e exigiram as primeiras intervenções. Isso reforça a utilidade do modelo não apenas como instrumento

de diagnóstico pontual, mas como bússola para priorização de esforços em programas de governança de dados em grandes organizações.

Como limitações do estudo, destaca-se que o diagnóstico capturou um momento específico da organização — caracterizado pela transição entre plataformas (AWS para GCP) e pela reconstrução de uma área de governança previamente desativada —, o que pode ter amplificado as notas negativas em processos e pessoas. Estudos futuros poderiam reaplicar o *framework* DCAM ao final das ondas de migração para mensurar a evolução do *score* geral e aferir se as intervenções implementadas foram suficientes para atingir o estágio “Proativo” (*score* 4,0), meta estabelecida para o quarto trimestre de 2026.

Conclui-se que a governança de dados em grandes organizações do varejo farmacêutico não é uma questão de disponibilidade tecnológica, mas de maturidade organizacional. O dado, como ativo estratégico, só gera valor quando sustentado por processos definidos, papéis claros e uma cultura analítica consolidada — elementos que nenhuma plataforma de nuvem substitui, mas que toda plataforma de governança pressupõe.

REFERÊNCIAS BIBLIOGRÁFICAS:

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Brasília, DF: Diário Oficial da União, 2018.

DAMA INTERNATIONAL. DAMA-DMBOK: Corpo de Conhecimento em Gerenciamento de Dados. 2ª ed. Nova Jersey: Technics Publications, 2017.

DAVENPORT, Thomas H.; PATIL, DJ. Cientista de Dados: A Profissão Mais Sexy do Século XXI. Harvard Business Review, v. 90, n. 10, p. 70-76, out. 2012.

EDM COUNCIL. DCAM: O Modelo de Avaliação da Capacidade de Gerenciamento de Dados. Versão 2.2. Nova York: Enterprise Data Management Council, 2020. Disponível em: <https://edmcouncil.org>. Acesso em: 10 de janeiro de 2026.

GARTNER. Glossário do Gartner: Dados Escuros. 2024. Disponível em: <https://www.gartner.com/en/information-technology/glossary/dark-data>. Acesso em: 15 de janeiro de 2026.

IBM. O custo da baixa qualidade dos dados. IBM Big Data & Analytics Hub, 2020. Disponível em: <https://www.ibm.com/analytics/data-management>. Acesso em: 12 dez. 2025.

LADLEY, John. Governança de Dados: Como Projetar, Implementar e Sustentar um Programa de Governança de Dados Eficaz. 2ª ed. Waltham: Morgan Kaufmann, 2019.

RD SAÚDE. Relatório Anual e de Sustentabilidade 2024. São Paulo: RD RI, 2025. Disponível em: <https://ri.rd.com.br>. Acesso em: 20 nov. 2025.

YIN, Robert K. Estudo de Caso: Planejamento e Métodos. 5. ed. Porto Alegre: Bookman, 2015.

APÊNDICE A — TABELA DE AVALIAÇÃO DETALHADA (CRITÉRIOS DE AUDITORIA)

Nota: Este apêndice lista os critérios do instrumento de avaliação aplicado na RD Saúde (Raia Drogasil) entre setembro e novembro de 2025. As pontuações atribuídas a cada critério foram omitidas para preservar a confidencialidade operacional e competitiva da organização pesquisada.

Nota Técnica: O instrumento de coleta de dados integral é composto por **432 critérios de verificação** (perguntas de auditoria). Devido à extensão do questionário e às limitações de paginação deste documento, apresenta-se abaixo a estrutura analítica dos **72 Tópicos de Controle** avaliados. O detalhamento granular das perguntas e das evidências brutas pode ser encontrado arquivado pelo autor e disponível para consulta acadêmica mediante solicitação.

Tabela A1. Mapeamento dos Tópicos de Controle por Componente

(continua)

Componente DCAM (Nível 1)	Capacidade (Nível 2)	Tópicos de Controle Avaliados (Nível 3)
1. ESTRATÉGIA	1.1. Visão de Dados	Declaração de Visão; Princípios de Dados; Visão de Arquitetura.
	1.2. Estudo de Caso Empresarial	Quantificação de Problemas; Projeção de Benefícios; Patrocínio.
	1.3. Alinhamento	Priorização Executiva; Roteiro de Investimento; Plano de Comunicação.
2. PROGRAMA	2.1. Papéis e Resp.	Proprietários de dados; Administradores de Dados; Custodiantes Técnicos.
	2.2. Modelo Operacional	Comitê Executivo; Conselho de Governança; Processo de Escalada.

Componente DCAM (Nível 1)	Capacidade (Nível 2)	Tópicos de Controle Avaliados (Nível 3)
	2.3. Sustentabilidade	Orçamento (Financiamento); Gestão de Fornecedores; Treinamento.
3. ARQUITETURA	3.1. Glossário	Termos de Negócio; Dicionário de Dados; Taxonomia.
	3.2. Modelagem	Modelo Conceitual; Modelo Lógico; Modelo Físico.
	3.3. Padrões	Nomenclatura; Domínios de Dados; Interoperabilidade.
4. TECNOLOGIA	4.1. Plataformas	Data Lake/Armazém; Ferramentas de ETL; Bancos de Dados.
	4.2. Integração	API/Microserviços; Ingestão Batch/Real-time; Orquestração.
	4.3. Ferramental	Catálogo de Dados; Linhagem Técnica; Ferramentas de DQ.
5. QUALIDADE (DQ)	5.1. Perfilamento	Análise de Perfil; Identificação de Regras; KPIs de DQ.
	5.2. Monitoramento	Painéis de Saúde; Alertas de Incidente; Relatórios de Erros.
	5.3. Remediação	Limpeza de Dados; Causa Raiz; Plano de Ação.
6. GOVERNANÇA	6.1. Políticas	Política Geral de Dados; Política de Segurança; Política de Acesso.
	6.2. Gestão de Mudança	Controle de versão; Impacto de Mudança; Comunicação de Alterações.
	6.3. Ética e Uso	Classificação de Sigilo; LGPD/Privacidade; Uso Ético de IA.
7. OPERAÇÕES	7.1. Ciclo de Vida	Retenção e Expurgo; Arquivamento (armazenagem frigorífica); Descarte Seguro.
	7.2. Continuidade	Backup e Restauração; Recuperação de Desastres; Alta Disponibilidade.
	7.3. Monitoramento	SLAs de Processamento; Registros de Auditoria; Desempenho.
8. ANÁLISE	8.1. Engenharia	Pipelines de Dados; Curadoria (Prata/Ouro); Disponibilização.

Componente DCAM (Nível 1)	Capacidade (Nível 2)	Tópicos de Controle Avaliados (Nível 3)
	8.2. Visualização	Catálogo de Relatórios; Certificação de Dashboards; Glossário de KPIs.
	8.3. Análises Avançadas	Governança de Modelos (MLOps); Ética em IA; Monitoramento de Drift.

Fonte: Resultados originais da pesquisa