FIELD GUIDE · CHEATSHEET

Data Darśana

AN INTELLIGENT WAY OF SEEING DATA

In the spirit of **Sarasvatī** — goddess of knowledge, music, and flow. A living map of the data architecture behind analytics, generative AI and agents.

248	6	128	1,888
CAPABILITIES	PILLARS	ARCHETYPES	CONNECTIONS

Data Darśana

A WAY OF SEEING DATA — THE MODEL AT A GLANCE

248 capabilities · **6** pillars · **20** domains
128 archetypes · **13** families · **1,888** links

I Intent & Governance *Sankalpa · resolve / intent*

Why data exists, who is accountable, and the rules of the game.

Strategy, Value & Operating Model 10 · Governance, Ownership & Stewardship 11 · Enterprise Data Architecture & Standards 11 · Organization, Skills, Adoption & Change 11

II Supply & Platform *Aadhaara · foundation / substrate*

Getting data in and giving it a place to live, run, and flow.

Data Sources & Acquisition 10 · Integration, Ingestion & Interoperability 11 · Data Platform, Storage & Processing 12 · Real-Time, Event & Operational Intelligence 10

III Meaning & Trust *Artha · meaning*

Turning raw data into shared meaning you can trust.

Data Modeling & Semantic Architecture 12 · Metadata, Catalog, Discovery, Lineage & Provenance 12 · Data Quality, Trust & Observability 14 · Master, Reference, Identity & Relationship Data 10

IV Products & Access *Daana · giving / offering*

Packaging data as products people and agents can find and consume.

Data Products, Sharing & Marketplace 11

V AI-Native Data *Prajna · insight / gnosis*

The new dimension: context, retrieval, knowledge and memory for agents.

AI Data Engineering & Dataset Lifecycle 14 · Retrieval, Search & Context Engineering 16 · Knowledge Graphs & Machine Reasoning 11 · Agent Data Access, Tools, State & Memory 16

VI Safety & Operations *Dharma · right order / duty*

Keeping the whole system safe, reliable, affordable and improving.

Security, Privacy, Compliance & Responsible Data 18 · DataOps, Reliability, FinOps & Lifecycle Operations 14 · Evaluation, Feedback & Continuous Learning 14

THE AWAKENING PATH

STAGE 1 · 0–3 months

Trusted Foundation *Shraddha*

Establish accountability, visibility, trust, security, and repeatable delivery.

118 capabilities enter

STAGE 2 · 3–6 months

Context at Scale *Viveka*

Make data understandable, reusable, retrievable, and agent-ready.

104 capabilities enter

STAGE 3 · 6–12 months

Agentic Industrialization *Abhyasa*

Scale safe agents with knowledge graphs, memory, tools, real-time data, evaluation.

25 capabilities enter

STAGE 4 · 12+ months

Continuous Intelligence *Bodhi*

Optimize a learning data ecosystem around trusted business outcomes.

1 capabilities enter

Awakening Path

THE FOUR STAGES — CAPABILITIES BY PILLAR

248 capabilities · **6** pillars · **20** domains
128 archetypes · **13** families · **1,888** links

Stage 1 · Trusted Foundation *Shraddha · trust / faith — 0-3 months*

Establish accountability, visibility, trust, security, and repeatable delivery. · **118** capabilities

INTENT & GOVERNANCE · 34

001 Enterprise Data and AI Strategy
002 Agentic AI Data Strategy
003 Data Value Case Management
005 Data Domain Portfolio Management
006 AI-Ready Data Investment Prioritization
007 Federated Data Operating Model
008 Data Platform Product Model
010 Capability Assessment and Benchmarking
011 Data Ownership
012 Data Stewardship
013 Governance Decision Rights and RACI
014 Data and AI Governance Councils
015 Critical Data Element Management
016 Data Policy Lifecycle Management
018 Policy Exception and Waiver Management
019 Enterprise Data Issue Management
020 Responsible Data and Data Ethics Governance
022 Current-State Data Architecture
023 Target-State Data Architecture
024 Domain and Bounded-Context Architecture
025 Data Architecture Reference Patterns
026 Data Technology Standards
027 Data Naming and Design Standards
028 Enterprise Data Flow Architecture
031 Architecture Knowledge Repository
238 Data and AI Role Architecture
239 Data Platform and Enablement Teams
240 Data and AI Communities of Practice
241 Enterprise Data and AI Literacy
242 Stewardship and Ownership Enablement
243 Data Architecture Guild
244 Data and Agentic AI Change Management
245 Architecture Playbooks and Knowledge Base
248 Workforce and Capacity Planning

SUPPLY & PLATFORM · 19

033 Enterprise Data Source Inventory
034 Source Qualification and Onboarding
035 Reusable Source Connectors and Adapters
036 API-Based Data Acquisition
037 Change Data Capture
038 Batch and File Acquisition
042 Third-Party and Open Data Rights Management
043 ETL and ELT Engineering
044 Data Workflow Orchestration
045 Data API Integration
047 Schema Mapping and Transformation
050 Data Contracts
051 Interoperability Standards Management
052 Integration Error Handling and Reconciliation
054 Data Lake and Object Storage
055 Analytical Data Warehouse
056 Lakehouse Architecture
057 Operational Data Stores
062 Distributed Data Processing

MEANING & TRUST · 30

066 Conceptual Data Modeling
067 Logical Data Modeling
068 Physical Data Modeling
069 Dimensional and Analytical Modeling
071 Business Glossary Management
073 Taxonomy and Classification Management
075 Schema Registry and Evolution
077 Enterprise Identifier Strategy
078 Technical Metadata Management
079 Business Metadata Management
080 Operational Metadata Management
081 Enterprise Data and AI Catalog
083 End-to-End Data Lineage
086 Automated Metadata Harvesting
087 Metadata Quality Management
090 Data Profiling
091 Data Quality Rule Management
092 Accuracy Management
093 Completeness Management
094 Consistency and Conformity Management
095 Timeliness and Freshness Management
096 Uniqueness and Duplicate Management
098 Data Quality SLAs and SLOs
100 Data Incident Management
101 End-to-End Data Observability
104 Master Data Management Strategy
105 Golden Record Management
107 Reference Data Management
110 Survivorship and Source Precedence
111 Master Data Distribution and Synchronization

PRODUCTS & ACCESS · 6

114 Data Product Design
115 Data Product Ownership
116 Product Contracts and Interfaces
117 Data Product SLAs and SLOs
120 Internal Data Sharing
123 Data Product Lifecycle Management

AI-NATIVE DATA · 4

135 AI Dataset Discovery and Selection
141 Sensitive Data, PII and Bias Screening
143 Dataset Versioning and Reproducibility
147 Dataset Documentation and Data Cards

SAFETY & OPERATIONS · 25

192 Data Classification and Sensitivity
193 Data Identity and Access Management
194 Fine-Grained Data Authorization
195 Least Privilege and Just-in-Time Access
196 Encryption and Key Management
197 Secrets and Credential Management
198 Masking, Tokenization and Redaction
199 Privacy Engineering and Data Minimization
200 Consent, Purpose and Lawful-Use Management
201 Retention, Deletion and Right-to-Erasure
202 Data Loss Prevention
203 Data Residency and Sovereignty
204 Data Access and Action Auditing
209 Data Licensing, Copyright and IP Controls
210 Data CI/CD
211 Data Infrastructure as Code
212 Automated Data Pipeline Testing
213 Environment and Promotion Management
214 Code, Configuration and Artifact Version Control
215 Data Release and Change Management
216 Data Reliability Engineering and SLOs
217 Backup, Restore and Data Recovery
218 Business Continuity and Disaster Recovery
221 Storage Tiering, Archival and Disposal
224 Data Readiness Assessment

Stage 2 · Context at Scale *Viveka · discernment — 3-6 months*

Make data understandable, reusable, retrievable, and agent-ready. · **104** capabilities

INTENT & GOVERNANCE · 9

004 Outcome-to-Data Traceability
009 Funding, Chargeback and Showback
017 Machine-Executable Policy Management
021 Agent Data Governance
029 Workload and Data Placement
030 Hybrid and Multi-Cloud Data Architecture
032 Data Architecture Debt Management
246 Reusable Data and AI Accelerators
247 Capability Adoption and Experience Measurement

SUPPLY & PLATFORM · 21

039 Streaming and Event Acquisition
040 Document and Content Acquisition
041 Multimodal Data Acquisition
046 Event and Message Integration
048 Data Federation and Virtualization
049 Secure Data Exchange
053 Agent-to-Data Integration Patterns
058 NoSQL Data Platforms
059 Graph Data Platforms
060 Vector Storage and Indexing
061 Time-Series Data Platforms
063 Low-Latency Data Serving
064 Workload Isolation and Multi-Tenancy
065 Elastic Capacity and Performance Engineering
125 Event-Driven Data Architecture
126 Event Schema and Contract Management
127 Stream Processing
129 Operational State and Profile Serving
130 Real-Time Change Propagation
131 Event Replay and Time Travel
133 Streaming Data Quality

MEANING & TRUST · 15

070 Canonical and Event Data Modeling
072 Semantic and Metrics Layer
074 Ontology Engineering
082 Search and Data Discovery
084 Change Impact Analysis
085 Data and AI Provenance
097 Data Anomaly Detection
099 Data Trust and Health Scoring
102 Authoritative Source and Content Validation
103 Retrieval and Context Quality
106 Entity Resolution and Matching
108 Hierarchy and Relationship Management
109 Enterprise Identity Graph
112 Temporal and Historical Master Data
113 Identity and Entity Resolution Services

PRODUCTS & ACCESS · 3

118 Data and Knowledge Marketplace
119 Self-Service Data and AI Access
124 Agent-Ready Knowledge Products

AI-NATIVE DATA · 40

136 Training Corpus Curation
137 Fine-Tuning and Instruction Dataset Management
138 Evaluation and Gold Dataset Management
140 Data Labeling and Annotation Operations
142 Deduplication and Contamination Control
144 Feature Engineering and Feature Store
145 Embedding Generation and Lifecycle
146 Multimodal Preprocessing
148 Reusable AI Data Pipelines
149 Enterprise Search
150 Lexical and Keyword Search
151 Vector Similarity Search
152 Hybrid Retrieval
153 Content Parsing and Extraction
154 Chunking and Segmentation
155 Embedding Model Selection and Management
156 Retrieval Index Lifecycle Management
159 ACL-Aware and Policy-Aware Retrieval
160 Structured Data Retrieval for Agents
162 Context Assembly and Compression
163 Citation, Grounding and Evidence Packaging
164 Retrieval Evaluation and Benchmarking
165 Knowledge Graph Strategy and Use-Case Design
166 Domain Ontology Development
167 Knowledge Graph Modeling
168 Entity Linking and Disambiguation
169 Graph Ingestion and Synchronization
172 Graph Query and Traversal Services
173 Knowledge Graph Quality
174 Knowledge Provenance and Explainability
176 Agent Data Access Architecture
177 MCP Server Engineering and Governance
178 Tool and Data Service Registry
179 Agent Tool Schema and Description Design
180 Delegated Agent Authorization
181 Read/Write Separation and Least Agency
183 Conversation and Workflow State Management
184 Short-Term Working Memory
190 Human Approval and Intervention Checkpoints
191 Secure Agent Execution Sandbox

SAFETY & OPERATIONS · 16

205 Prompt Injection and Data Exfiltration Controls
206 Sensitive Context and Output Filtering
207 Runtime Policy Enforcement and Guardrails
208 Bias, Fairness and Representativeness Controls
219 Data and AI FinOps
220 Data Performance Engineering
222 Capacity and Demand Management
223 Unified Data and Agent Telemetry
225 Retrieval Benchmark Suites
226 Groundedness and Factuality Evaluation
227 Agent Task and Business Outcome Evaluation
230 Human Feedback Collection
231 Feedback Ingestion and Governance
234 Data and Agent Regression Testing
236 Agent Failure Taxonomy and Root-Cause Analysis
237 Continuous Improvement Backlog

Stage 3 · Agentic Industrialization *Abhyasa · disciplined practice — 6-12 months*

Scale safe agents with knowledge graphs, memory, tools, real-time data, evaluation. · **25** capabilities

SUPPLY & PLATFORM · 3

128 Complex Event Processing
132 Digital Twin and Context State
134 Agent Event Triggers and Subscriptions

MEANING & TRUST · 3

076 Semantic Constraint Validation
088 Active Metadata and Policy Automation
089 Data and Context Usage Analytics

PRODUCTS & ACCESS · 1

121 External Sharing and Data Collaboration

AI-NATIVE DATA · 13

139 Synthetic Data Generation and Validation
157 Query Understanding and Decomposition
158 Reranking and Relevance Optimization
161 Graph-Based Retrieval and GraphRAG
170 Relationship and Claim Extraction
171 Rules, Inference and Constraint Reasoning
175 Knowledge Graph and RAG Integration
182 Agent Transaction and Commit Controls
185 Long-Term Agent Memory
186 Episodic and Semantic Memory
187 Memory Summarization and Compaction
188 Cross-Agent Context Sharing
189 Agent-to-Agent Data Exchange

SAFETY & OPERATIONS · 5

228 Tool Execution Quality Evaluation
229 Data-to-Outcome Correlation
232 Active Learning and Data Improvement
233 Data, Retrieval and Behavior Drift Detection
235 Controlled Experiments and A/B Testing

Stage 4 · Continuous Intelligence *Bodhi · awakening — 12+ months*

Optimize a learning data ecosystem around trusted business outcomes. · **1** capabilities

PRODUCTS & ACCESS · 1

122 Data Monetization and Commercialization

I · Intent & Governance

SANKALPA · RESOLVE / INTENT

43 capabilities

Data Darśana

Why data exists, who is accountable, and the rules of the game.

D1 Strategy, Value & Operating Model 10

- 001 Enterprise Data and AI Strategy
- 002 Agentic AI Data Strategy
- 003 Data Value Case Management
- 004 Outcome-to-Data Traceability
- 005 Data Domain Portfolio Management
- 006 AI-Ready Data Investment Prioritization
- 007 Federated Data Operating Model
- 008 Data Platform Product Model
- 009 Funding, Chargeback and Showback
- 010 Capability Assessment and Benchmarking

D2 Governance, Ownership & Stewardship 11

- 011 Data Ownership
- 012 Data Stewardship
- 013 Governance Decision Rights and RACI
- 014 Data and AI Governance Councils
- 015 Critical Data Element Management
- 016 Data Policy Lifecycle Management
- 017 Machine-Executable Policy Management
- 018 Policy Exception and Waiver Management
- 019 Enterprise Data Issue Management
- 020 Responsible Data and Data Ethics Governance
- 021 Agent Data Governance

D3 Enterprise Data Architecture & Standards 11

- 022 Current-State Data Architecture
- 023 Target-State Data Architecture
- 024 Domain and Bounded-Context Architecture
- 025 Data Architecture Reference Patterns
- 026 Data Technology Standards
- 027 Data Naming and Design Standards
- 028 Enterprise Data Flow Architecture
- 029 Workload and Data Placement
- 030 Hybrid and Multi-Cloud Data Architecture
- 031 Architecture Knowledge Repository
- 032 Data Architecture Debt Management

D20 Organization, Skills, Adoption & Change 11

- 238 Data and AI Role Architecture
- 239 Data Platform and Enablement Teams
- 240 Data and AI Communities of Practice
- 241 Enterprise Data and AI Literacy
- 242 Stewardship and Ownership Enablement
- 243 Data Architecture Guild
- 244 Data and Agentic AI Change Management
- 245 Architecture Playbooks and Knowledge Base
- 246 Reusable Data and AI Accelerators
- 247 Capability Adoption and Experience Measurement
- 248 Workforce and Capacity Planning

II · Supply & Platform

AADHAARA · FOUNDATION / SUBSTRATE

43 capabilities

Data Darśana

Getting data in and giving it a place to live, run, and flow.

D4 Data Sources & Acquisition 10

- 033 Enterprise Data Source Inventory
- 034 Source Qualification and Onboarding
- 035 Reusable Source Connectors and Adapters
- 036 API-Based Data Acquisition
- 037 Change Data Capture
- 038 Batch and File Acquisition
- 039 Streaming and Event Acquisition
- 040 Document and Content Acquisition
- 041 Multimodal Data Acquisition
- 042 Third-Party and Open Data Rights Management

D5 Integration, Ingestion & Interoperability 11

- 043 ETL and ELT Engineering
- 044 Data Workflow Orchestration
- 045 Data API Integration
- 046 Event and Message Integration
- 047 Schema Mapping and Transformation
- 048 Data Federation and Virtualization
- 049 Secure Data Exchange
- 050 Data Contracts
- 051 Interoperability Standards Management
- 052 Integration Error Handling and Reconciliation
- 053 Agent-to-Data Integration Patterns

D6 Data Platform, Storage & Processing 12

- 054 Data Lake and Object Storage
- 055 Analytical Data Warehouse
- 056 Lakehouse Architecture
- 057 Operational Data Stores
- 058 NoSQL Data Platforms
- 059 Graph Data Platforms
- 060 Vector Storage and Indexing
- 061 Time-Series Data Platforms
- 062 Distributed Data Processing
- 063 Low-Latency Data Serving
- 064 Workload Isolation and Multi-Tenancy
- 065 Elastic Capacity and Performance Engineering

D12 Real-Time, Event & Operational Intelligence 10

- 125 Event-Driven Data Architecture
- 126 Event Schema and Contract Management
- 127 Stream Processing
- 128 Complex Event Processing
- 129 Operational State and Profile Serving
- 130 Real-Time Change Propagation
- 131 Event Replay and Time Travel
- 132 Digital Twin and Context State
- 133 Streaming Data Quality
- 134 Agent Event Triggers and Subscriptions

III · Meaning & Trust

ARTHA · MEANING

48 capabilities

Data Darśana

Turning raw data into shared meaning you can trust.

D7 Data Modeling & Semantic Architecture 12

- 066 Conceptual Data Modeling
- 067 Logical Data Modeling
- 068 Physical Data Modeling
- 069 Dimensional and Analytical Modeling
- 070 Canonical and Event Data Modeling
- 071 Business Glossary Management
- 072 Semantic and Metrics Layer
- 073 Taxonomy and Classification Management
- 074 Ontology Engineering
- 075 Schema Registry and Evolution
- 076 Semantic Constraint Validation
- 077 Enterprise Identifier Strategy

D8 Metadata, Catalog, Discovery, Lineage & Provenance 12

- 078 Technical Metadata Management
- 079 Business Metadata Management
- 080 Operational Metadata Management
- 081 Enterprise Data and AI Catalog
- 082 Search and Data Discovery
- 083 End-to-End Data Lineage
- 084 Change Impact Analysis
- 085 Data and AI Provenance
- 086 Automated Metadata Harvesting
- 087 Metadata Quality Management
- 088 Active Metadata and Policy Automation
- 089 Data and Context Usage Analytics

D9 Data Quality, Trust & Observability 14

- 090 Data Profiling
- 091 Data Quality Rule Management
- 092 Accuracy Management
- 093 Completeness Management
- 094 Consistency and Conformity Management
- 095 Timeliness and Freshness Management
- 096 Uniqueness and Duplicate Management
- 097 Data Anomaly Detection
- 098 Data Quality SLAs and SLOs
- 099 Data Trust and Health Scoring
- 100 Data Incident Management
- 101 End-to-End Data Observability
- 102 Authoritative Source and Content Validation
- 103 Retrieval and Context Quality

D10 Master, Reference, Identity & Relationship Data 10

- 104 Master Data Management Strategy
- 105 Golden Record Management
- 106 Entity Resolution and Matching
- 107 Reference Data Management
- 108 Hierarchy and Relationship Management
- 109 Enterprise Identity Graph
- 110 Survivorship and Source Precedence
- 111 Master Data Distribution and Synchronization
- 112 Temporal and Historical Master Data
- 113 Identity and Entity Resolution Services

IV · Products & Access

DAANA · GIVING / OFFERING

11 capabilities

Data Darśana

Packaging data as products people and agents can find and consume.

D11 Data Products, Sharing & Marketplace 11

- 114 Data Product Design
- 115 Data Product Ownership
- 116 Product Contracts and Interfaces
- 117 Data Product SLAs and SLOs
- 118 Data and Knowledge Marketplace
- 119 Self-Service Data and AI Access
- 120 Internal Data Sharing
- 121 External Sharing and Data Collaboration
- 122 Data Monetization and Commercialization
- 123 Data Product Lifecycle Management
- 124 Agent-Ready Knowledge Products

V · AI-Native Data

PRAJNA · INSIGHT / GNOSIS

57 capabilities

Data Darśana

The new dimension: context, retrieval, knowledge and memory for agents.

D13 AI Data Engineering & Dataset Lifecycle 14

- 135 AI Dataset Discovery and Selection
- 136 Training Corpus Curation
- 137 Fine-Tuning and Instruction Dataset Management
- 138 Evaluation and Gold Dataset Management
- 139 Synthetic Data Generation and Validation
- 140 Data Labeling and Annotation Operations
- 141 Sensitive Data, PII and Bias Screening
- 142 Deduplication and Contamination Control
- 143 Dataset Versioning and Reproducibility
- 144 Feature Engineering and Feature Store
- 145 Embedding Generation and Lifecycle
- 146 Multimodal Preprocessing
- 147 Dataset Documentation and Data Cards
- 148 Reusable AI Data Pipelines

D14 Retrieval, Search & Context Engineering 16

- 149 Enterprise Search
- 150 Lexical and Keyword Search
- 151 Vector Similarity Search
- 152 Hybrid Retrieval
- 153 Content Parsing and Extraction
- 154 Chunking and Segmentation
- 155 Embedding Model Selection and Management
- 156 Retrieval Index Lifecycle Management
- 157 Query Understanding and Decomposition
- 158 Reranking and Relevance Optimization
- 159 ACL-Aware and Policy-Aware Retrieval
- 160 Structured Data Retrieval for Agents
- 161 Graph-Based Retrieval and GraphRAG
- 162 Context Assembly and Compression
- 163 Citation, Grounding and Evidence Packaging
- 164 Retrieval Evaluation and Benchmarking

D15 Knowledge Graphs & Machine Reasoning 11

- 165 Knowledge Graph Strategy and Use-Case Design
- 166 Domain Ontology Development
- 167 Knowledge Graph Modeling
- 168 Entity Linking and Disambiguation
- 169 Graph Ingestion and Synchronization
- 170 Relationship and Claim Extraction
- 171 Rules, Inference and Constraint Reasoning
- 172 Graph Query and Traversal Services
- 173 Knowledge Graph Quality
- 174 Knowledge Provenance and Explainability
- 175 Knowledge Graph and RAG Integration

D16 Agent Data Access, Tools, State & Memory 16

- 176 Agent Data Access Architecture
- 177 MCP Server Engineering and Governance
- 178 Tool and Data Service Registry
- 179 Agent Tool Schema and Description Design
- 180 Delegated Agent Authorization
- 181 Read/Write Separation and Least Agency
- 182 Agent Transaction and Commit Controls
- 183 Conversation and Workflow State Management
- 184 Short-Term Working Memory
- 185 Long-Term Agent Memory
- 186 Episodic and Semantic Memory
- 187 Memory Summarization and Compaction
- 188 Cross-Agent Context Sharing
- 189 Agent-to-Agent Data Exchange
- 190 Human Approval and Intervention Checkpoints
- 191 Secure Agent Execution Sandbox

VI · Safety & Operations

DHARMA · RIGHT ORDER / DUTY

46 capabilities

Data Darśana

Keeping the whole system safe, reliable, affordable and improving.

D17 Security, Privacy, Compliance & Responsible Data 18

- 192 Data Classification and Sensitivity
- 193 Data Identity and Access Management
- 194 Fine-Grained Data Authorization
- 195 Least Privilege and Just-in-Time Access
- 196 Encryption and Key Management
- 197 Secrets and Credential Management
- 198 Masking, Tokenization and Redaction
- 199 Privacy Engineering and Data Minimization
- 200 Consent, Purpose and Lawful-Use Management
- 201 Retention, Deletion and Right-to-Erasure
- 202 Data Loss Prevention
- 203 Data Residency and Sovereignty
- 204 Data Access and Action Auditing
- 205 Prompt Injection and Data Exfiltration Controls
- 206 Sensitive Context and Output Filtering
- 207 Runtime Policy Enforcement and Guardrails
- 208 Bias, Fairness and Representativeness Controls
- 209 Data Licensing, Copyright and IP Controls

D18 DataOps, Reliability, FinOps & Lifecycle Operations 14

- 210 Data CI/CD
- 211 Data Infrastructure as Code
- 212 Automated Data Pipeline Testing
- 213 Environment and Promotion Management
- 214 Code, Configuration and Artifact Version Control
- 215 Data Release and Change Management
- 216 Data Reliability Engineering and SLOs
- 217 Backup, Restore and Data Recovery
- 218 Business Continuity and Disaster Recovery
- 219 Data and AI FinOps
- 220 Data Performance Engineering
- 221 Storage Tiering, Archival and Disposal
- 222 Capacity and Demand Management
- 223 Unified Data and Agent Telemetry

D19 Evaluation, Feedback & Continuous Learning 14

- 224 Data Readiness Assessment
- 225 Retrieval Benchmark Suites
- 226 Groundedness and Factuality Evaluation
- 227 Agent Task and Business Outcome Evaluation
- 228 Tool Execution Quality Evaluation
- 229 Data-to-Outcome Correlation
- 230 Human Feedback Collection
- 231 Feedback Ingestion and Governance
- 232 Active Learning and Data Improvement
- 233 Data, Retrieval and Behavior Drift Detection
- 234 Data and Agent Regression Testing
- 235 Controlled Experiments and A/B Testing
- 236 Agent Failure Taxonomy and Root-Cause Analysis
- 237 Continuous Improvement Backlog

Architecture Eras

THE SAME 128 ARCHETYPES, BY THE WORK THEY SERVE AND THE GENERATION THEY BELONG TO

248 capabilities · **6** pillars · **20** domains
128 archetypes · **13** families · **1,888** links

Workload	**Proven** Long-settled patterns that remain the right answer for governed, well-understood work.	**Mainstream** The current default — widely deployed, well-tooled, low argument.	**Frontier** Actively forming. High leverage, thinner track record, faster change.
Transactional Running the business: operational writes, consistency, service data.	11-01 Normalised transactional core 11-02 OLTP/OLAP separation with read offload 11-07 Core-system / mainframe offload 12-04 MDM centralised transaction hub	02-09 Operational data store 03-08 API-led integration 03-09 Event sourcing & CQRS 11-03 Database-per-service 11-04 Transactional outbox & saga 11-05 Immutable ledger / append-only system of record 11-06 NoSQL operational store	02-10 HTAP / unified operational-analytical platform 03-10 Active-data / reverse-ETL architecture
Integration Moving and reconciling data between systems.	03-01 Batch ETL 11-07 Core-system / mainframe offload 12-01 MDM registry style 13-06 ESB / SOA hub-and-spoke integration 13-07 EDI / B2B partner exchange 13-08 Point-to-point & file-based batch integration	02-05 Logical data warehouse 02-06 Data virtualisation / federation 02-09 Operational data store 03-02 ELT 03-05 Event-driven architecture 03-07 CDC-based architecture 03-08 API-led integration 04-03 Data-as-a-Service 04-04 API-accessible data products 11-04 Transactional outbox & saga 12-03 MDM coexistence style 13-03 Data Vault 2.0	02-04 Data fabric 03-10 Active-data / reverse-ETL architecture
Real-time Continuous, event-shaped data with freshness measured in seconds.	10-03 Fraud and risk intelligence	03-03 Lambda architecture 03-04 Kappa architecture 03-05 Event-driven architecture 03-07 CDC-based architecture 03-09 Event sourcing & CQRS 10-02 IoT and digital twin 10-05 Real-time personalisation 10-06 Supply-chain intelligence 10-12 Industrial / manufacturing data 11-04 Transactional outbox & saga	02-10 HTAP / unified operational-analytical platform 03-06 Streaming-first architecture 05-09 Real-time / streaming RAG 06-05 Event-driven autonomous agents 08-08 Continuous controls monitoring
Analytics Exploration, modelling and decision support over history.	02-01 Enterprise data warehouse 04-01 Central curated data marts 04-10 Self-service analytics architecture 10-01 Customer 360 10-03 Fraud and risk intelligence 10-10 Healthcare / life-sciences data 10-11 Financial-services data 11-02 OLTP/OLAP separation with read offload 12-02 MDM consolidation style 13-01 Inmon corporate information factory 13-02 Kimball dimensional bus 13-04 Star & snowflake schema with slowly-changing dimensions	02-02 Data lake 02-03 Lakehouse 02-05 Logical data warehouse 02-06 Data virtualisation / federation 03-02 ELT 03-03 Lambda architecture 04-07 Feature-store architecture 09-03 Privacy-preserving analytics 09-04 Data clean-room architecture 10-02 IoT and digital twin 10-06 Supply-chain intelligence 10-07 Geospatial intelligence 10-09 Scientific and research data 10-12 Industrial / manufacturing data 12-06 Entity resolution & identity graph 13-03 Data Vault 2.0	02-10 HTAP / unified operational-analytical platform 10-08 Document intelligence
Reporting & BI Governed, repeatable answers with agreed definitions.	02-01 Enterprise data warehouse 04-01 Central curated data marts 04-10 Self-service analytics architecture 10-04 Regulatory reporting 13-01 Inmon corporate information factory 13-02 Kimball dimensional bus 13-04 Star & snowflake schema with slowly-changing dimensions 13-05 OLAP cube & multidimensional semantic layer	04-02 Domain data products 04-03 Data-as-a-Service 04-04 API-accessible data products 05-07 SQL / structured-data agent	04-05 Semantic products 04-06 Metrics-as-a-Service 04-08 Knowledge-product architecture 04-09 Marketplace-based consumption

Workload	**Proven** Long-settled patterns that remain the right answer for governed, well-understood work.	**Mainstream** The current default — widely deployed, well-tooled, low argument.	**Frontier** Actively forming. High leverage, thinner track record, faster change.
Master & Reference The shared nouns — parties, products, places, code sets.	10–01 Customer 360 12–01 MDM registry style 12–02 MDM consolidation style 12–04 MDM centralised transaction hub 12–05 Reference-data management & code-set distribution	05–06 Knowledge-graph-first AI 12–03 MDM coexistence style 12–06 Entity resolution & identity graph	06–10 Knowledge-graph-coordinated agents
Platform & Topology Where compute and storage physically live.	07–01 On-premises 07–09 Centralised global platform	02–02 Data lake 02–03 Lakehouse 02–07 Polyglot data platform 02–08 Cloud data platform 07–02 Cloud-native 07–03 Hybrid cloud 07–04 Multi-cloud 07–07 Regionally isolated deployment 07–08 Air-gapped / disconnected environment 07–11 Multi-tenant platform 07–12 Dedicated / client-isolated platform 11–03 Database-per-service 11–06 NoSQL operational store	02–04 Data fabric 06–12 Edge-agent architecture 07–05 Edge and cloud 07–06 Sovereign cloud 07–10 Regional platform federation
Governance & Trust Control, policy, privacy and assurance.	01–01 Centralised enterprise data organisation 08–01 Centralised governance 09–07 Tokenised-data architecture 10–04 Regulatory reporting 10–10 Healthcare / life-sciences data 10–11 Financial-services data 12–05 Reference-data management & code-set distribution	01–02 Decentralised / domain-owned model 01–03 Federated data operating model 01–05 Hub-and-spoke 01–06 Platform-led shared-service model 01–07 Embedded data teams 04–02 Domain data products 08–02 Federated governance 08–05 Domain stewardship 08–07 Risk-tiered governance 09–01 Zero-trust data architecture 09–03 Privacy-preserving analytics 09–04 Data clean-room architecture 09–08 Sovereign-data architecture 09–09 Attribute-based access architecture 11–05 Immutable ledger / append-only system of record	01–04 Data mesh 01–08 Hybrid central-platform / domain-product model 02–04 Data fabric 04–09 Marketplace-based consumption 08–03 Computational governance 08–04 Policy-as-code 08–06 Data-product governance 08–08 Continuous controls monitoring 08–09 Zero-trust data governance 08–10 Purpose- & consent-based governance 09–02 Confidential-computing architecture 09–05 Federated-learning architecture 09–06 Differential-privacy architecture 09–10 Identity-propagating agent architecture
AI & Agentic Retrieval, context and autonomous consumers of data.	*nothing here yet*	04–07 Feature-store architecture 05–01 Basic RAG 05–02 Advanced modular RAG 05–03 Hybrid-search RAG 05–06 Knowledge-graph-first AI 05–07 SQL / structured-data agent 05–12 Domain-specific retrieval pods 05–13 Enterprise shared retrieval platform 10–05 Real-time personalisation	04–05 Semantic products 04–08 Knowledge-product architecture 05–04 Federated RAG 05–05 GraphRAG 05–08 Multimodal RAG 05–09 Real-time / streaming RAG 05–10 Agentic RAG 05–11 Corrective & adaptive RAG 06–01 Centralised agent context platform 06–02 Domain-aligned agent knowledge pods 06–03 Tool-mediated agent architecture 06–04 Memory-centric agent architecture 06–05 Event-driven autonomous agents 06–06 Workflow-orchestrated agents 06–07 Human-in-the-loop agents 06–08 Multi-agent shared-memory architecture 06–09 Multi-agent blackboard architecture 06–10 Knowledge-graph-coordinated agents 06–11 Federated multi-agent architecture 06–12 Edge-agent architecture 09–05 Federated-learning architecture 09–10 Identity-propagating agent architecture 10–08 Document intelligence

Archetype Catalogue

HOW YOU ASSEMBLE IT — 13 FAMILIES, 128 ARCHETYPES

248 capabilities · **6** pillars · **20** domains
128 archetypes · **13** families · **1,888** links

Data Operating-Model Archetypes — How the organisation and ownership of data are structured. · 8

ARCH–01–01 **Centralised enterprise data organisation** — One central team owns data platforms and delivery for the enterprise.
ARCH–01–02 **Decentralised / domain-owned model** — Business domains own and run their own data independently.
ARCH–01–03 **Federated data operating model** — Domains own data under shared, centrally-set standards and policies.
ARCH–01–04 **Data mesh** — Domain-oriented ownership of data-as-a-product on self-serve platform with federated governance.
ARCH–01–05 **Hub-and-spoke** — A central hub sets standards and shared services; spokes deliver locally.
ARCH–01–06 **Platform-led shared-service model** — A platform team provides paved-road services consumed by many teams.
ARCH–01–07 **Embedded data teams** — Data specialists embedded directly inside business/product teams.
ARCH–01–08 **Hybrid central-platform / domain-product model** — Central platform plus domain-owned data products on top.

Core Data-Platform Archetypes — The shape of the core storage and processing platform. · 10

ARCH–02–01 **Enterprise data warehouse** — Central, schema-on-write store for governed analytics.
ARCH–02–02 **Data lake** — Low-cost store for raw multi-format data at scale.
ARCH–02–03 **Lakehouse** — Lake storage with warehouse-grade tables, ACID and governance.
ARCH–02–04 **Data fabric** — Metadata-driven integration layer spanning distributed data.
ARCH–02–05 **Logical data warehouse** — Virtualised warehouse over multiple physical stores.
ARCH–02–06 **Data virtualisation / federation** — Query data in place across sources without moving it.
ARCH–02–07 **Polyglot data platform** — Multiple purpose-built stores (SQL, NoSQL, graph, vector) combined.
ARCH–02–08 **Cloud data platform** — Managed cloud-native storage and processing services.
ARCH–02–09 **Operational data store** — Integrated store for current operational data and lookups.
ARCH–02–10 **HTAP / unified operational-analytical platform** — Single platform serving transactions and analytics together.

Integration & Processing Archetypes — How data is moved, transformed and kept fresh. · 10

ARCH–03–01 **Batch ETL** — Scheduled extract-transform-load into a target store.
ARCH–03–02 **ELT** — Load raw then transform inside the target platform.
ARCH–03–03 **Lambda architecture** — Parallel batch and speed layers merged for serving.
ARCH–03–04 **Kappa architecture** — Single streaming pipeline reprocessed from the log.
ARCH–03–05 **Event-driven architecture** — Components react to published events asynchronously.
ARCH–03–06 **Streaming-first architecture** — Streams are the primary integration substrate.
ARCH–03–07 **CDC-based architecture** — Change data capture replicates source changes continuously.
ARCH–03–08 **API-led integration** — Reusable, governed APIs mediate data exchange.
ARCH–03–09 **Event sourcing & CQRS** — State as an event log with separated read/write models.
ARCH–03–10 **Active-data / reverse-ETL architecture** — Warehouse data pushed back into operational systems.

Data-Product & Consumption Archetypes — How data is packaged and consumed by people and systems. · 10

ARCH–04–01 **Central curated data marts** — Governed, subject-oriented marts for reporting.
ARCH–04–02 **Domain data products** — Domain-owned, contracted, discoverable data products.
ARCH–04–03 **Data-as-a-Service** — Data delivered as a governed, subscribable service.
ARCH–04–04 **API-accessible data products** — Data products exposed through managed APIs.
ARCH–04–05 **Semantic products** — Products defined by a governed semantic/metrics layer.
ARCH–04–06 **Metrics-as-a-Service** — Consistent metric definitions served centrally.
ARCH–04–07 **Feature-store architecture** — Curated ML features served for training and inference.
ARCH–04–08 **Knowledge-product architecture** — Curated knowledge assets for retrieval and reasoning.
ARCH–04–09 **Marketplace-based consumption** — Internal/external marketplace for discovering and requesting data.
ARCH–04–10 **Self-service analytics architecture** — Governed self-service exploration for business users.

AI & Retrieval Archetypes — How context is retrieved and grounded for generative AI. · 13

ARCH–05–01 **Basic RAG** — Single vector search over chunks feeding the model.
ARCH–05–02 **Advanced modular RAG** — Composable retrieval with pre/post-processing stages.
ARCH–05–03 **Hybrid-search RAG** — Combines lexical and vector retrieval with re-ranking.
ARCH–05–04 **Federated RAG** — Retrieval spanning multiple isolated corpora/sources.
ARCH–05–05 **GraphRAG** — Retrieval guided by a knowledge graph structure.
ARCH–05–06 **Knowledge-graph-first AI** — Reasoning primarily over a curated knowledge graph.
ARCH–05–07 **SQL / structured-data agent** — Agent queries structured data via generated SQL.
ARCH–05–08 **Multimodal RAG** — Retrieval across text, image, audio and other modalities.

ARCH–05–09 **Real-time / streaming RAG** — Retrieval over continuously updated live data.

ARCH–05–10 **Agentic RAG** — Agents plan, retrieve and iterate over multiple steps.

ARCH–05–11 **Corrective & adaptive RAG** — Retrieval that self-checks and adapts on low confidence.

ARCH–05–12 **Domain-specific retrieval pods** — Purpose-built retrieval services per domain.

ARCH–05–13 **Enterprise shared retrieval platform** — One governed retrieval platform for all use cases.

Agentic-Data Archetypes — How agents access data, tools, state and memory. · 12

ARCH–06–01 **Centralised agent context platform** — One platform serving context to all agents.

ARCH–06–02 **Domain-aligned agent knowledge pods** — Per-domain knowledge and context services.

ARCH–06–03 **Tool-mediated agent architecture** — Agents act only through governed tools/APIs.

ARCH–06–04 **Memory-centric agent architecture** — Shared long/short-term memory drives agent behaviour.

ARCH–06–05 **Event-driven autonomous agents** — Agents triggered and coordinated by events.

ARCH–06–06 **Workflow-orchestrated agents** — Agents run inside orchestrated, auditable workflows.

ARCH–06–07 **Human-in-the-loop agents** — Human approval gates on agent actions.

ARCH–06–08 **Multi-agent shared-memory architecture** — Multiple agents collaborate over shared memory.

ARCH–06–09 **Multi-agent blackboard architecture** — Agents post/read on a shared blackboard workspace.

ARCH–06–10 **Knowledge-graph-coordinated agents** — A shared graph coordinates multi-agent context.

ARCH–06–11 **Federated multi-agent architecture** — Autonomous agents cooperate across boundaries.

ARCH–06–12 **Edge-agent architecture** — Agents run at the edge with local data/context.

Deployment & Topology Archetypes — Where the platform physically runs and how it is isolated. · 12

ARCH–07–01 **On-premises** — Runs in owned data centres.

ARCH–07–02 **Cloud-native** — Built on managed cloud services.

ARCH–07–03 **Hybrid cloud** — Workloads split across on-prem and cloud.

ARCH–07–04 **Multi-cloud** — Spread across multiple cloud providers.

ARCH–07–05 **Edge and cloud** — Processing split between edge and central cloud.

ARCH–07–06 **Sovereign cloud** — Cloud constrained to jurisdictional/sovereignty rules.

ARCH–07–07 **Regionally isolated deployment** — Isolated stacks per region for residency.

ARCH–07–08 **Air-gapped / disconnected environment** — No external network connectivity.

ARCH–07–09 **Centralised global platform** — Single global platform instance.

ARCH–07–10 **Regional platform federation** — Federated regional platforms under shared standards.

ARCH–07–11 **Multi-tenant platform** — Shared platform serving many isolated tenants.

ARCH–07–12 **Dedicated / client-isolated platform** — Dedicated stack per tenant or client.

Governance Archetypes — How policy and control over data are exercised. · 10

ARCH–08–01 **Centralised governance** — A central body sets and enforces all policy.

ARCH–08–02 **Federated governance** — Shared global policy, local domain enforcement.

ARCH–08–03 **Computational governance** — Policy enforced automatically in the platform.

ARCH–08–04 **Policy-as-code** — Governance rules expressed and versioned as code.

ARCH–08–05 **Domain stewardship** — Accountable stewards per domain.

ARCH–08–06 **Data-product governance** — Governance organised around data products/contracts.

ARCH–08–07 **Risk-tiered governance** — Controls scaled to data risk tier.

ARCH–08–08 **Continuous controls monitoring** — Controls verified continuously, not periodically.

ARCH–08–09 **Zero-trust data governance** — Never-trust, always-verify access decisions.

ARCH–08–10 **Purpose- & consent-based governance** — Access bound to declared purpose and consent.

Security & Privacy Archetypes — How data is protected, isolated and privacy-preserved. · 10

ARCH–09–01 **Zero-trust data architecture** — Continuous verification for every data access.

ARCH–09–02 **Confidential-computing architecture** — Data processed in hardware-encrypted enclaves.

ARCH–09–03 **Privacy-preserving analytics** — Analytics without exposing raw personal data.

ARCH–09–04 **Data clean-room architecture** — Joint analysis without sharing underlying data.

ARCH–09–05 **Federated-learning architecture** — Models trained across data that never centralises.

ARCH–09–06 **Differential-privacy architecture** — Provable privacy via calibrated noise.

ARCH–09–07 **Tokenised-data architecture** — Sensitive values replaced by reversible tokens.

ARCH–09–08 **Sovereign-data architecture** — Data kept within jurisdictional control.

ARCH–09–09 **Attribute-based access architecture** — Access decided by attributes/policy (ABAC).

ARCH–09–10 **Identity-propagating agent architecture** — End-user identity carried through agent actions.

Industry & Workload Archetypes — Recurring solution shapes for specific workloads. · 12

ARCH–10–01 **Customer 360** — Unified, resolved view of the customer.

ARCH–10–02 **IoT and digital twin** — Live sensor data mirrored as a digital twin.

ARCH–10–03 **Fraud and risk intelligence** — Real-time detection of fraud and risk signals.

ARCH–10–04 **Regulatory reporting** — Traceable, controlled reporting to regulators.

ARCH–10–05 **Real-time personalisation** — Live tailoring of experiences per user.

ARCH–10–06 **Supply-chain intelligence** — End-to-end visibility and optimisation of supply chains.

ARCH–10–07 **Geospatial intelligence** — Location-aware analytics and reasoning.

ARCH–10–08 **Document intelligence** — Extraction and reasoning over documents.

ARCH–10–09 **Scientific and research data** — Managing large-scale experimental/research data.

ARCH–10–10 **Healthcare / life-sciences data** — Regulated clinical and life-sciences data.

ARCH–10–11 **Financial-services data** — Regulated financial data and risk workloads.

ARCH–10–12 **Industrial / manufacturing data** — Operational technology and plant-floor data.

Transactional & Operational Archetypes — How the systems that run the business hold and move their own data. · 7

ARCH–11–01 **Normalised transactional core** — A 3NF, ACID system of record optimised for correct concurrent writes.

ARCH–11–02 **OLTP/OLAP separation with read offload** — Analytical reads moved off the transactional engine via replicas or extracts.

ARCH–11–03 **Database-per-service** — Each service privately owns its store; no shared database coupling.

ARCH–11–04 **Transactional outbox & saga** — Reliable messaging and long-running consistency without distributed transactions.

ARCH–11–05 **Immutable ledger / append-only system of record** — Facts are appended, never updated, giving a replayable audit history.

ARCH–11–06 **NoSQL operational store** — Document, key-value or wide-column stores for scale and flexible shape.

ARCH–11–07 **Core-system / mainframe offload** — Freeing a constrained core system by mirroring its data outward.

Master, Reference & Semantic Archetypes — How an organisation agrees on its shared nouns and their meaning. · 6

ARCH–12–01 **MDM registry style** — A thin index of cross-references; sources keep their own data.

ARCH–12–02 **MDM consolidation style** — Master data is consolidated downstream for reporting, not written back.

ARCH–12–03 **MDM coexistence style** — A curated master is synchronised bidirectionally with the sources.

ARCH–12–04 **MDM centralised transaction hub** — The hub is the system of entry and authors master data directly.

ARCH–12–05 **Reference-data management & code-set distribution** — Controlled code sets versioned and distributed to consumers.

ARCH–12–06 **Entity resolution & identity graph** — Probabilistic matching that resolves records into linked identities.

Classical Warehouse & Integration Archetypes — The modelling and integration methodologies the discipline was built on. · 8

ARCH–13–01 **Inmon corporate information factory** — A normalised enterprise warehouse feeding dependent departmental marts.

ARCH–13–02 **Kimball dimensional bus** — Conformed dimensions as the integration contract across business processes.

ARCH–13–03 **Data Vault 2.0** — Hubs, links and satellites separating keys, relationships and history.

ARCH–13–04 **Star & snowflake schema with slowly-changing dimensions** — Dimensional structures and the disciplines for tracking history.

ARCH–13–05 **OLAP cube & multidimensional semantic layer** — Pre-aggregated hierarchies for fast, governed slice-and-dice.

ARCH–13–06 **ESB / SOA hub-and-spoke integration** — A central bus mediating, routing and transforming between systems.

ARCH–13–07 **EDI / B2B partner exchange** — Standards-based document exchange with external trading partners.

ARCH–13–08 **Point-to-point & file-based batch integration** — Direct bilateral feeds and file drops — the shape most estates start in.