

Data product guidebook

Guide to building successful
data products

vivicta*

Introduction

Data product is a productionized, consumption-ready and reusable data asset for operational or analytical purposes.

Vivicta has delivered multiple data platform and data product projects in the Nordics.

Here we share the necessary topics to cover when working with data products.

Why data products?

Data products are critical link in the data value chain, transforming raw events into actionable insights. By improving operational efficiency, enhancing security and transparency, data products drive revenue growth, reduce costs, and reduce risk.



We have listed relevant aspects of data products that needs to be addressed to deliver data products successfully



Crystallize these data product topics to enable efficient data product delivery



For each topic we have provided a generic solution based on our expertise to guide you on the right path



Use these generic solutions and tailor the approach to fit for your organizational needs

Topics

1. Data product use cases
2. Data product features
3. Data architecture
4. Data governance
5. Data quality dimensions
6. Data quality processes
7. Data lifecycle policies
8. Optimization strategies
9. Roles and responsibilities
10. Data product development tasks
11. Data product prioritization
12. Interoperable data products
13. AI-ready data products
14. Business value of data products

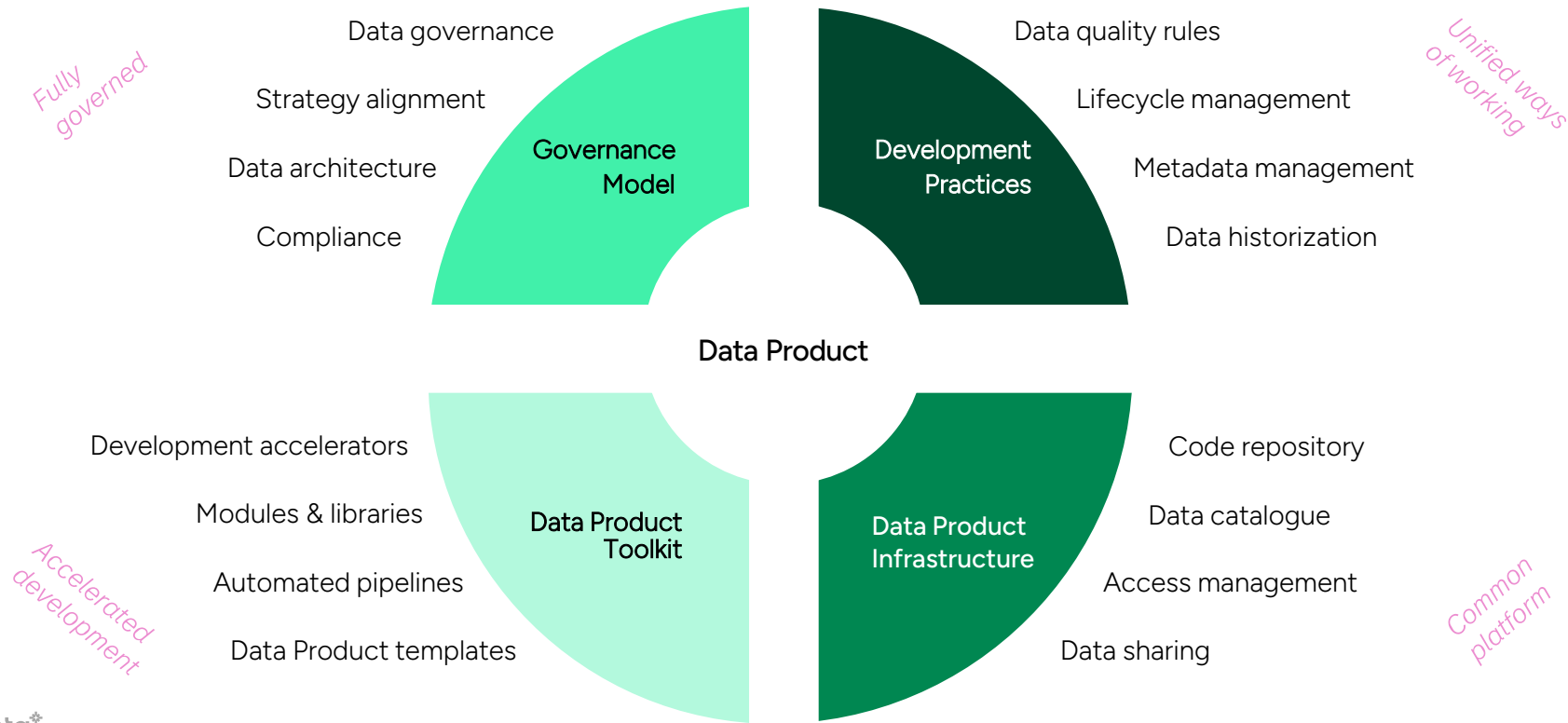
1. Define use cases for data products

Define use cases with use case templates to ensure building the right data product clarifying its purpose, aligning with stakeholders and establishing clear expected outcomes

Use Case: Data Product	
Description This data product combines and standardizes data from operational data sources for monthly reporting and predictive analytics. It delivers quality-validated, historized and comparable metrics aligned with regulatory frameworks.	
Objectives • Integrate data and apply standardized calculation logic • Craft features for ML model • Report monthly figures	Business value • Reduces manual reporting effort • Increases transparency and trust • Supports strategic decision-making • Enables accurate predictions
Data sources • On-premises database • Cloud file storage • External APIs • Manual files	Users • Reporting team • Data science team • Citizen developers
Risks • Exposing sensitive data • Compute costs • Vendor risks	SLA • Availability: 99% for aggregated reporting tables • Monitored 08.00 – 16.00 during weekdays

2. Define data product dimensions

Integrate organizational context, ways of working and technologies into the data product approach to fully capture organizational needs and deliver data products efficiently



3. Choose data architecture

Select a data architecture, such as the Medallion architecture, that provides a structured and repeatable framework for building secure, flexible and well-governed data products

Bronze



Raw, unvalidated and immutable data

Decoupling data from source systems and making it available for data processing by ingesting raw data from the data source and storing it in its original format.

Silver



Filtered, validated and historized data

Validating data, schema, and keys by performing data quality checks, lifecycle management, historization, security labelling, and policy validations.

Gold



Use case optimized data

Joining, transforming and modelling data to publish actionable data product optimized for reporting, analytics or AI use cases.

4. Align between centralized and decentralized governance

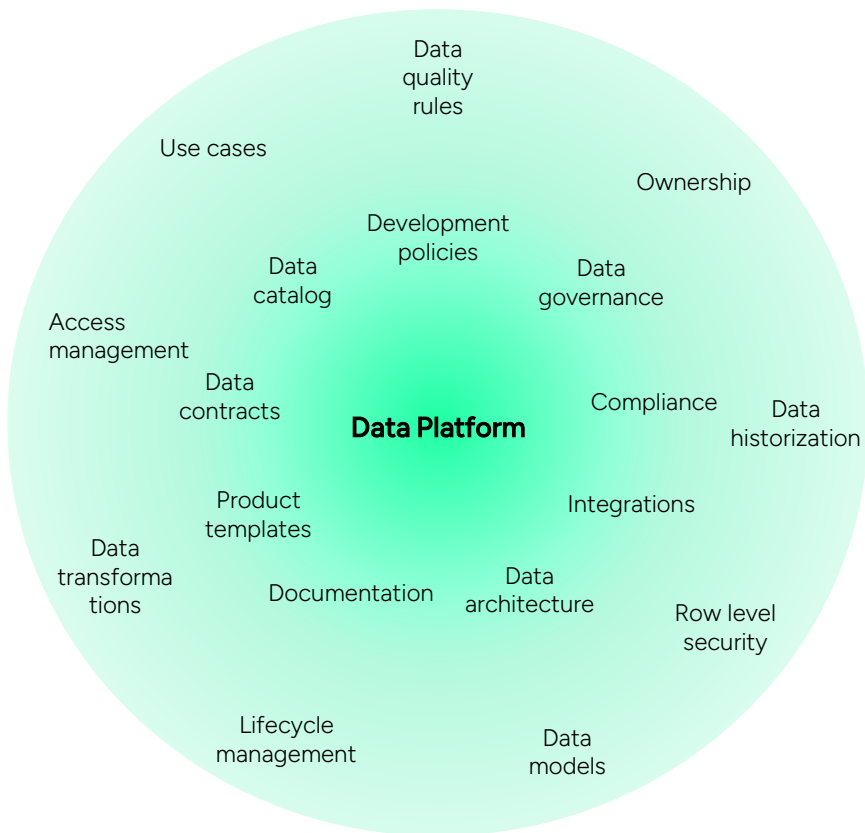
Find balance between decentralized and centralized data governance to deliver unified and business optimized data products

Centralization:

- ✧ Standardized data products to be centrally joined with other domain's data products
- ✧ Development policies and templates applied across data products for scalability and reusability
- ✧ Common data catalogue for data discovery

Decentralization:

- ✧ Derived data products optimized for business value
- ✧ Decision making of data sharing and usage
- ✧ Business related data quality validation



5. Define data quality dimensions

Define data quality dimensions based on data product's objectives and choose the quality aspects that directly affect its consumers



Accuracy

The degree to which the data correctly represents the entity or attribute



Completeness

The percentage of missing data from a given data set



Consistency

The absence of difference in data irrespective of the data source



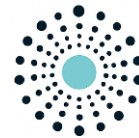
Validity

The degree of value attributes that are aligned with the specific domain



Integrity

The validity of relationships across various data entities



Uniqueness

The degree of duplicate or overlapping data

6. Design data quality processes

Design data quality processes to ensure that data products are validated, compliant, and readily consumable across the data platform

Data profiling

Data profiling assesses data structure, content, and quality by examining patterns, distributions and anomalies. It helps detect issues early, and guide cleansing, transformation, and governance decisions.

Data quality validation

Data quality validation defines structured rules describing acceptable values and constraints. A rule engine retrieves these rules, evaluates datasets, flags violations and produces governed validation results.

Data quality reporting

Data quality reporting summarizes the results of data validation to clearly show whether data meets expected standards. Its purpose is to make data issues visible, track quality over time, and understand where problems occur.

Data cleansing

Data cleansing applies corrections to improve data quality by fixing errors, standardizing formats and isolating invalid records. It uses insights from profiling and validation to produce accurate, consistent, and reliable data products.

7. Configure data lifecycle management policies

Configure data lifecycle management policies ensure data products stay relevant over time and meet regulatory and compliance requirements

Lifecycle management policies

Policy	Description	Priority
Data protection	Data deletion is prevented	100
Event-based data deletion	Data deletion is triggered by specific business or legal event	200
Data retention	Retain data for number of days before automatic deletion	300

Lifecycle management process



Configure data lifecycle management policies for data product



Add the lifecycle management policies into data product metadata



Perform data lifecycle management by reading and applying policies as a part of the regular data product runs

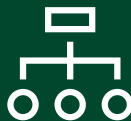
8. Design optimization for data product workloads

Use optimization strategies to ensure speed, cost-efficiency and scalability of data products



Autoscaling

Autoscaling automatically adjusts compute capacity by dynamically expanding or shrinking worker nodes based on real-time metrics, optimizing both performance and cost.



Parallelization

Parallelization distributes work across multiple tasks or data partitions so independent operations run simultaneously, accelerating processing, reducing bottlenecks enabling faster execution.



Workload isolation

Workload isolation separates different compute activities into dedicated environments so they cannot impact one another, ensuring each workload maintains consistent performance.

9. Define roles & responsibilities

Assemble a data product team with the right data skills to design, build and maintain data products



Data Platform Architect

Data Platform Architect shapes the overall design of data platform, making sure storage, processing, integration, and governance layers work together in a scalable and secure way.



Data Engineer

Data Engineer focuses on designing and building the data pipelines and notebooks that ingest, transform, and publish data products into the data platform.



DataOps Engineer

DataOps Engineer automates data pipelines, implements CI/CD practices and automated recovery actions, and optimizes workflows while ensuring seamless collaboration.



Data Steward

Data Steward ensures data is well-defined, trustworthy, and used responsibly across the organization by managing metadata, quality standards, and policies.

10. Create data product development tasks

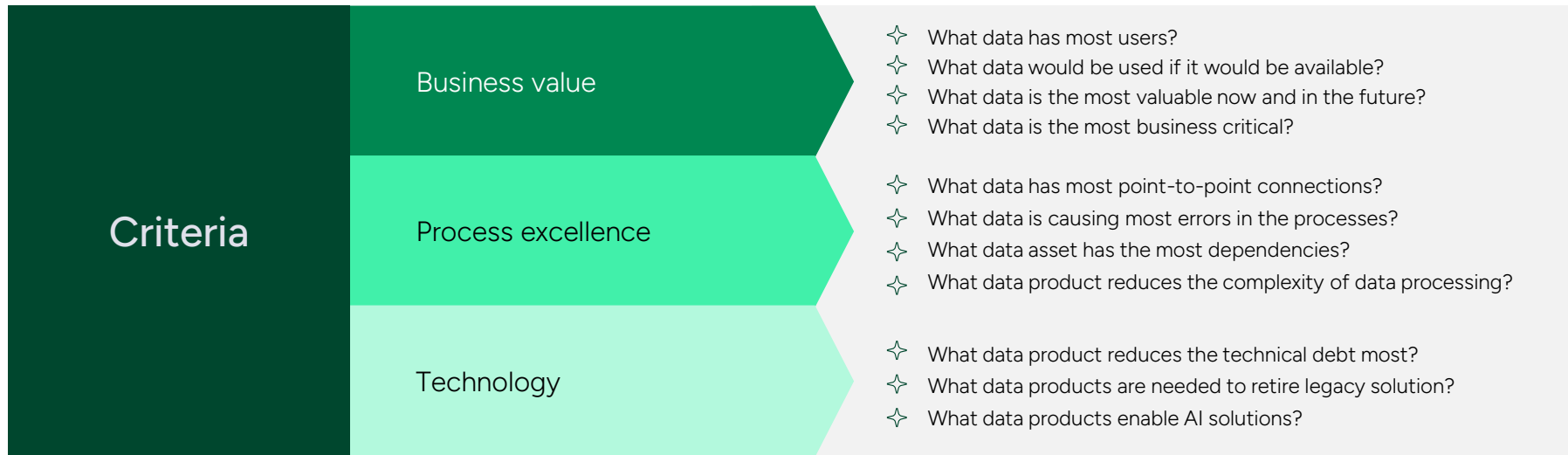
Document development tasks to keep the work organized, transparent and business-aligned enabling teams to deliver high-quality data products efficiently

Development Task	Description	Responsible
Infrastructure setup	Configure data platform infrastructure including environments, networks, policies, and security.	Data Platform Architect
Connect to data sources	Establish secure and reliable connections between data sources and the platform.	Data Engineer
Review source data	Validate source data integrity and ensure readiness for ingestion.	Data Engineer
Implement data quality rules	Define and enforce rules to maintain data accuracy and consistency.	Data Steward
Define metadata	Create metadata for tables including names, columns, and data types for governance.	Data Steward
Bronze: Ingest source data	Develop ingestion pipelines to load raw data into the platform efficiently.	Data Engineer
Silver: Validate and manage data	Apply data quality checks, lifecycle policies, and historization for curated datasets.	Data Steward
Gold: Transform and model data	Build joins, transformations, and models to create business-ready data products.	Data Engineer
Develop orchestration pipeline	Implement pipelines to orchestrate data processing following architectural standards.	Data Engineer
Develop CI/CD pipeline	Automate deployment processes for data pipelines and platform components.	DataOps Engineer
Promote to test environment	Execute integration and performance tests in a controlled environment.	DataOps Engineer
Promote to production	Deploy validated pipelines and data products into the production environment.	DataOps Engineer
Optimize pipeline execution	Enhance job performance and resource utilization for efficient compute.	Data Engineer
Add documentation	Generate automated documentation for data products and pipeline status.	Data Steward
Set up monitoring & alerts	Configure monitoring tools and alerts for pipeline health and SLA compliance.	DataOps Engineer
Implement DataOps	Establish DataOps practices to ensure continuous delivery and operational excellence.	DataOps Engineer
Final data product review	Conduct stakeholder review and obtain approval for the data product release.	Data Platform Architect

Development progress

11. Define criteria for data product prioritization

Combine business and technology viewpoints to prioritize data product development



12. Design interoperable data products

Design data product processes to support organization-wide use cases



Data catalog

A data catalog provides a unified place to discover data products. It standardizes metadata, enforces governance and lineage ensuring data products interoperate.



Combined DataOps

Combined DataOps processes unify data products by providing consistent environment, standardizing workflows, and coordinating changes across products.

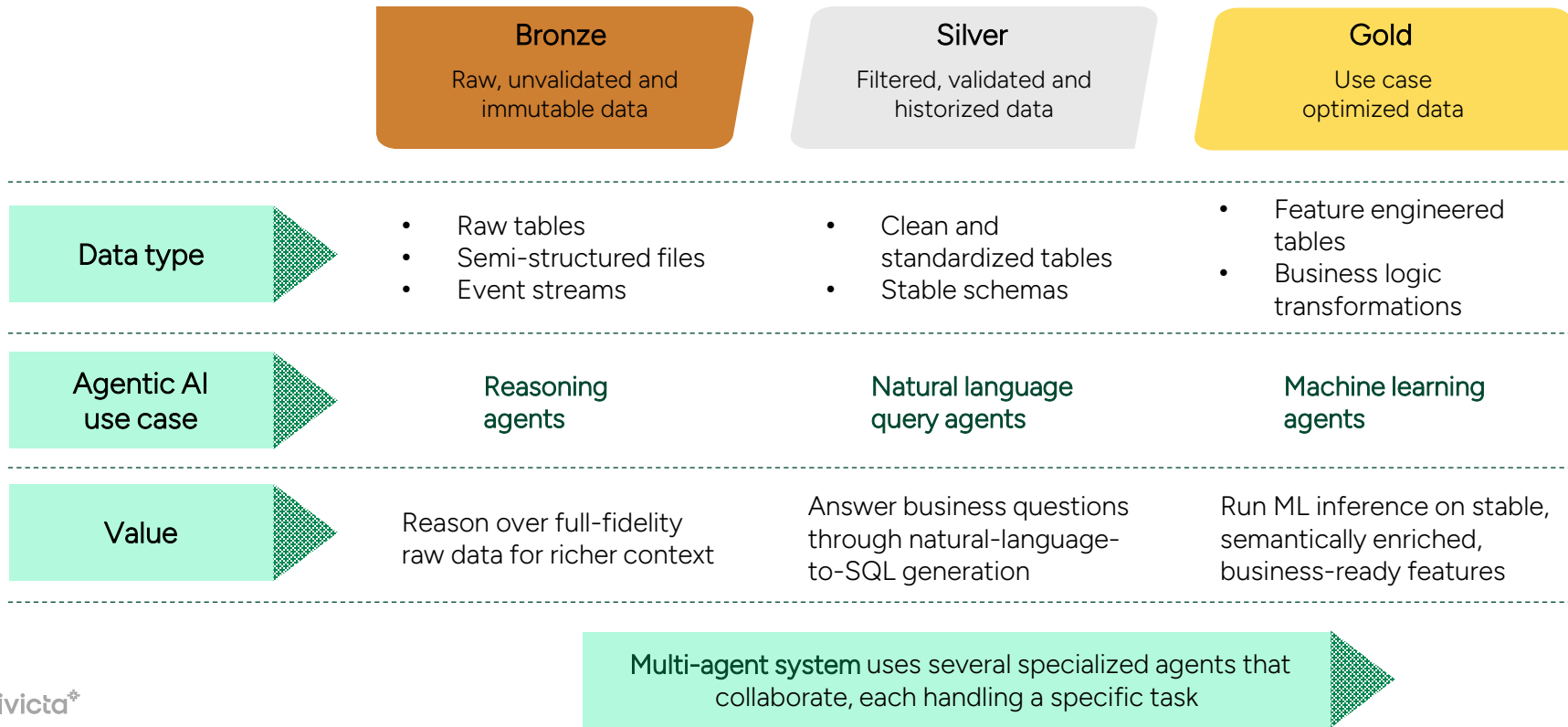


Product templates

Data product templates ensure cross-product compatibility by enforcing consistent standards. They define shared structures that make data products joinable.

13. Align data products with AI use cases

Expose the appropriate layers of the medallion architecture to AI agents based on each use case, while preserving robust data governance and security



14. Define and measure business value of data products

Define and measure the business value of data products to ensure alignment with organizational goals and drive business value realization.



Trust

Data products increase trust by treating data like any other product, built with clear ownership, guaranteed quality, and reliable service levels. Defined SLAs, transparent lineage, consistent interfaces, and proactive monitoring transform data into trusted data product.

Measure: Service level indicators



Adoption

Data product adoption reveals the impact data products have to the organization. With reusable data products, teams can launch new services rapidly, build AI solutions and innovate at scale turning data products into a catalyst for continuous business growth.

Measure: Data product usage, time-to-insight, number of new services launched and legacy services retired



FinOps

FinOps practices link data product spending to business outcomes. Value emerges through optimized operations and new services. By tracking efficiency and impact, FinOps quantifies ROI and shows how data product generates business value.

Measure: Revenue from new services, cost reduction from retired services and reduced operational risk.

vivicta*