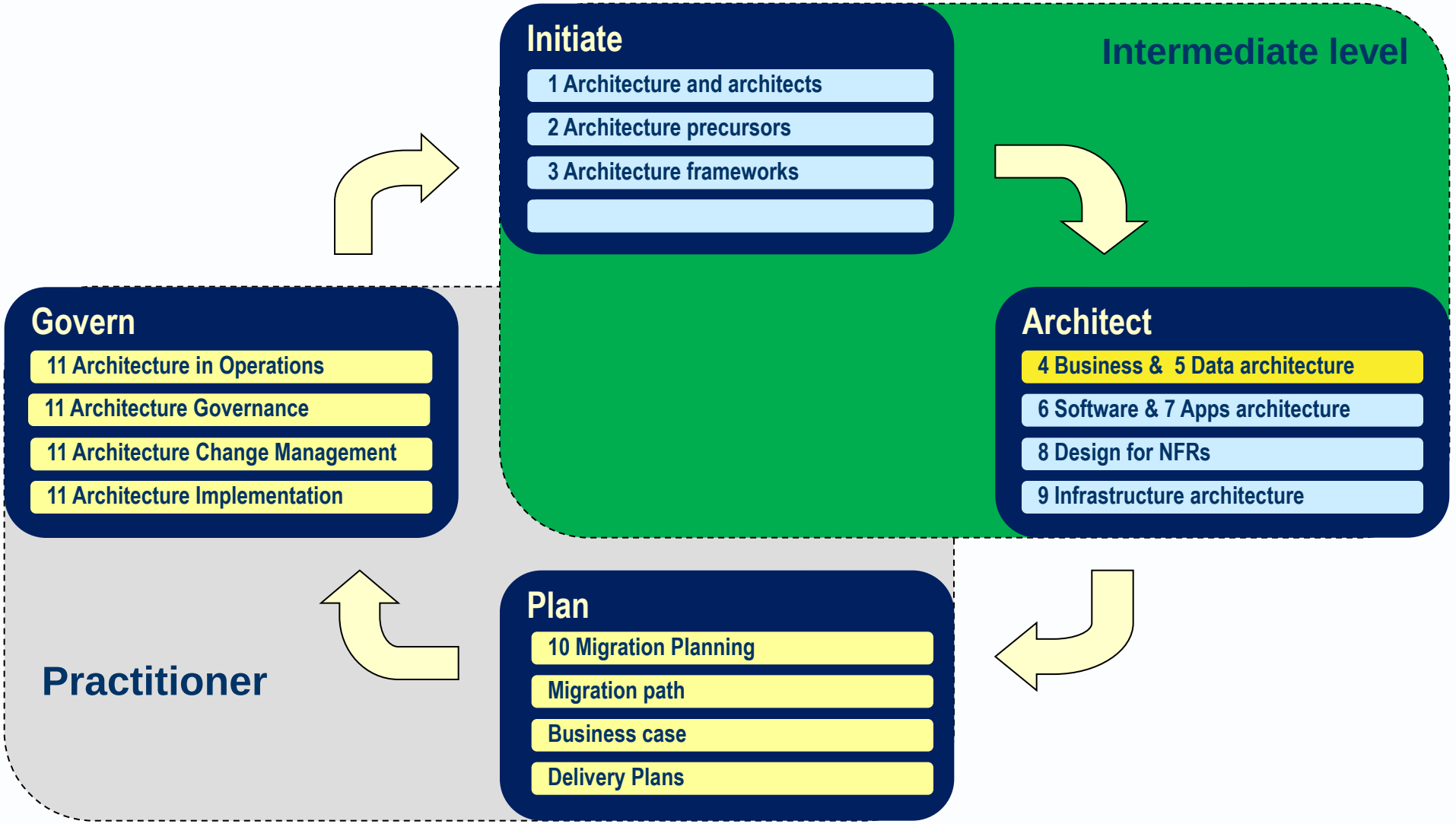
Avancier Reference Model

Data Architecture (ESA 5)

5. Data architecture

Avancier



5.1: Foundation (rarely examined)

- Fig. 5.1 Base data architecture concepts

	Data in motion	Data at rest
Data object	Data event	Data entity
Data container	Data flow	Data store

- The base elements in this domain are explained in later sections.
- This first section introduces some background concepts.

Structured v unstructured

- **Structured data** [a data object] a data store or flow that fits a pre-defined data structure.
  - It is composed of data items.
  - It usually records real word entities or events.
  - It may contain references to unstructured data.
  - All popular architecture framework focus on structured data.

- **Unstructured data** [a data object] text or images that do not fit a pre-defined data structure, as in emails, voice and video.
- It may contain recognisable structured items.

Data

- Data item
  - [a data object] an elementary unit of information, a fact.
  - An attribute of a data entity or data event.
  - A variable containing a value that is constrained by a data type.
- Meta data
  - [a description] of data.
  - It may take the form of a data structure, data type, constraint rule, derivation rule or other data quality.
- Data structure
  - [a type] a structure that arranges data items in one or more groups. It may be described as a data model or a regular expression.

Data type

- [a type] that defines the properties shared by instances of a data item or data entity. It defines the possible values for that type, the processes that can be performed on values of that type, the meaning of the data; and the way values of that type can be stored.
- **Primitive data type** [a data type] pre-defined in a programming language. e.g. character string, integer, Boolean, floating-point number (decimal).
- **User-defined data type** [a data type] defined by analyst or architects that is bespoke to the business at hand. It may be simple (e.g. credit limit, order value) or complex (e.g. address, tax reference number, order, product).

Data rule

- [a business rule] recorded in a rule repository, data dictionary or data model as a constraint or derivation rule.

- **Constraint rule** [a data rule] that limits the values of a data type.

- **Derivation rule** [a data rule] that defines how a data value is derived from one or more other values (a special kind of constraint).

Data dictionary

- [an artefact] that catalogues data types and defines their meanings.
- It may include business rules in the form of constraints on data values and derivation rules.
- It may take the form of a canonical data model.

Data architecture

Avancier

Conceptual

Business Data Model

Canonical Data Model

Data store	CRM	ERP	Billing
Customer	M	C	C
Order	M(1)	C	M(2)
Invoice			M

Logical

Logical Data Model

Data Governance

Data Flow Model

Physical

Physical Data Schema

XML Schema

Real

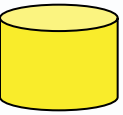
Data Store

Data Flow

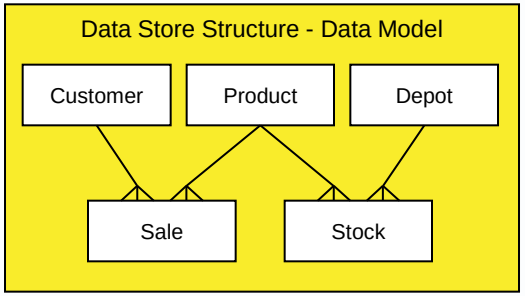
Copyright Avancier Limited 2015

Data architects

- Define business data in terms of



- **Data stores** created & used by business activities & applications
  - **Data store structures**: entities, attributes and relationships



- **Data flows** created & used by business activities & applications
  - **Data flow structures**: message and file formats

- **Data qualities** of data store/flow elements
  - Confidentiality, integrity and availability (CIA)
  - Data owners and stewards
  - Canonical data types (constraints on data item values)

5.2: Data at rest

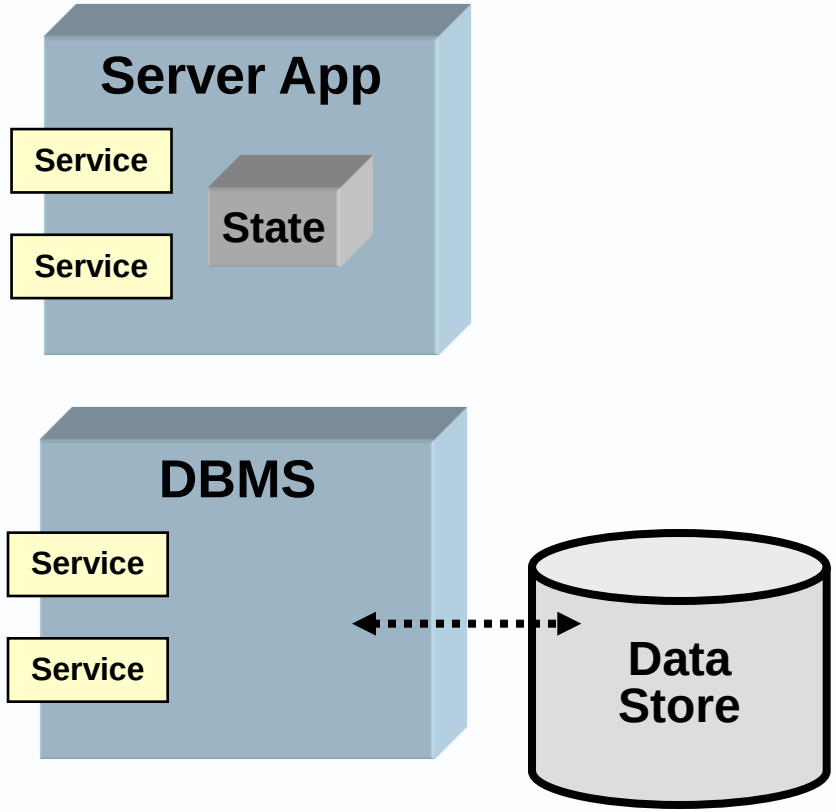
- **Data at rest** [a viewpoint] relating to data that persists.
-

Data entity

- [a data object] composed of data items that represent facts about a discrete business entity or event. It may be specified at a conceptual, logical or physical level. It may be mapped to data stores and/or data flows input to or output from IS services.
-

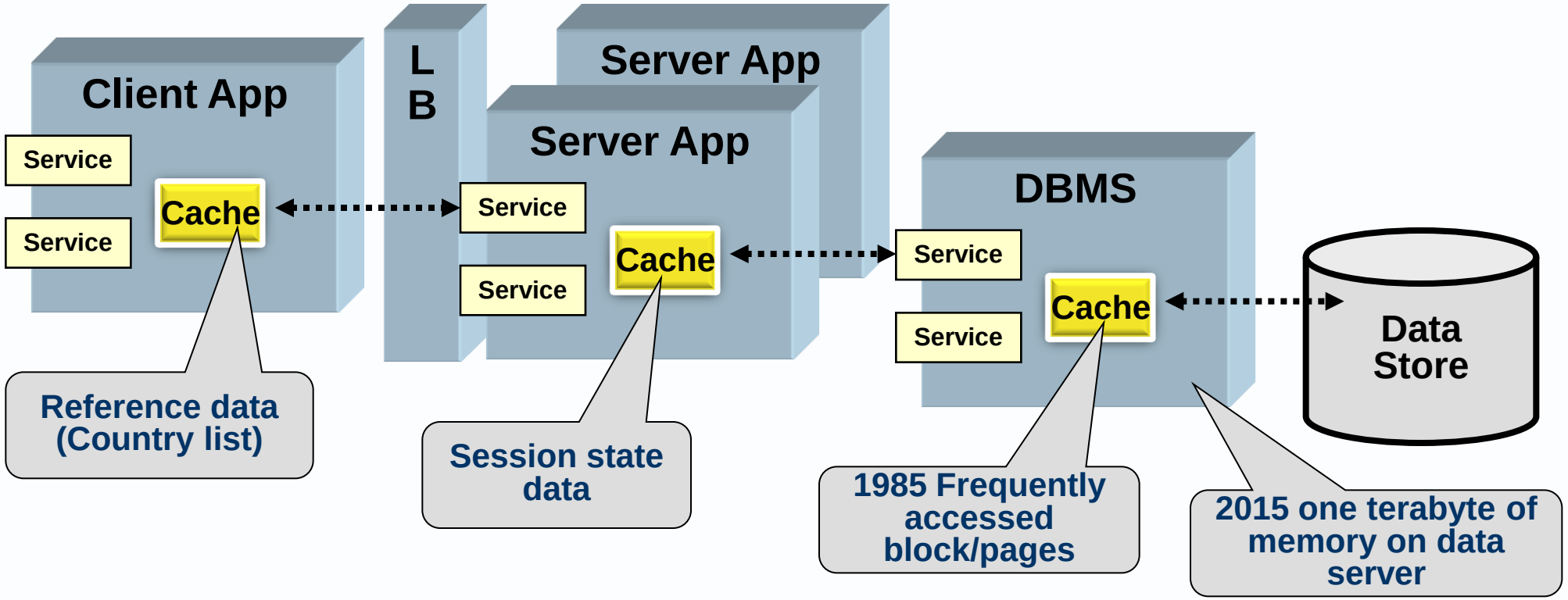
Data store

- [a platform component] that holds a persistent data structure.
- A cache, file or database from which data can be extracted by an application.
- The content of a persistent data store can be defined in a data model.



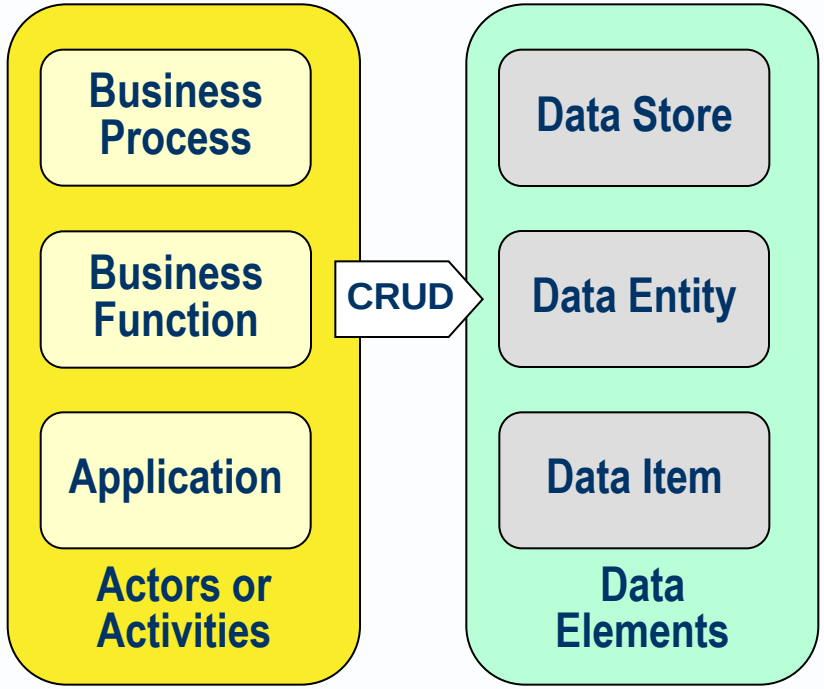
Data in storage

Cache	A local store of data that has been copied from a master data store, usually for the purpose of speeding up response or cycle time.



Data entity / business function matrix

- [an artefact] that maps data entities to the business functions that create and update them, and perhaps use them also.
- Cluster analysis can be used to cluster data that is created by the same functions, and functions that create the same data.



Function Data Entity		
	Create	Read
	Update	Create

Application Data Entity		
	Create	Read
	Update	Create

CRUD matrix

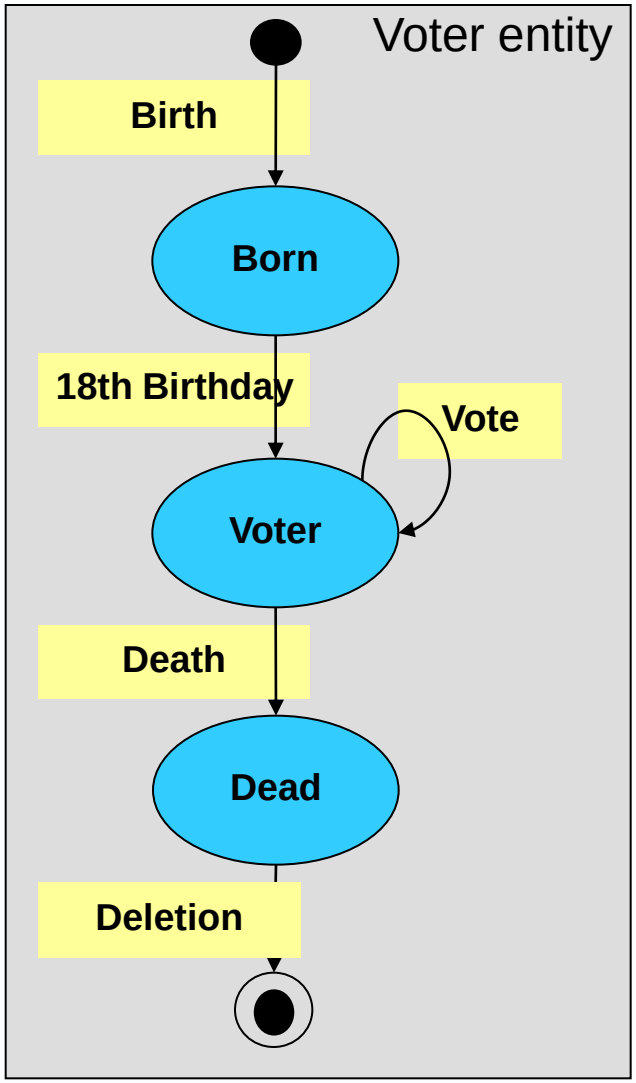
- Which Activities (business function, business process or event)
- Create and use which data (data store, data entity or data item)

ACTORS or ACTIVITIES DATA ENTITIES	Order opening	Item addition	Order closure	Payment	Payment + 1yr
Order	Create	Read	Update	Update	Delete
Order item		Create	Update		Delete
Product type		Read	Update	Update	Update
Depot stock			Update		

- You can map **data to activity** at different levels granularity.
- Better not at several levels; do it at the lowest level you can maintain.

Data entity lifecycle diagram

- [an artefact] that shows life of a data entity in terms of the states it passes through from creation to deletion, and the data events that trigger state transitions.



Define data lifecycle view – simple view

Avancier

Simplistic data entity life cycle

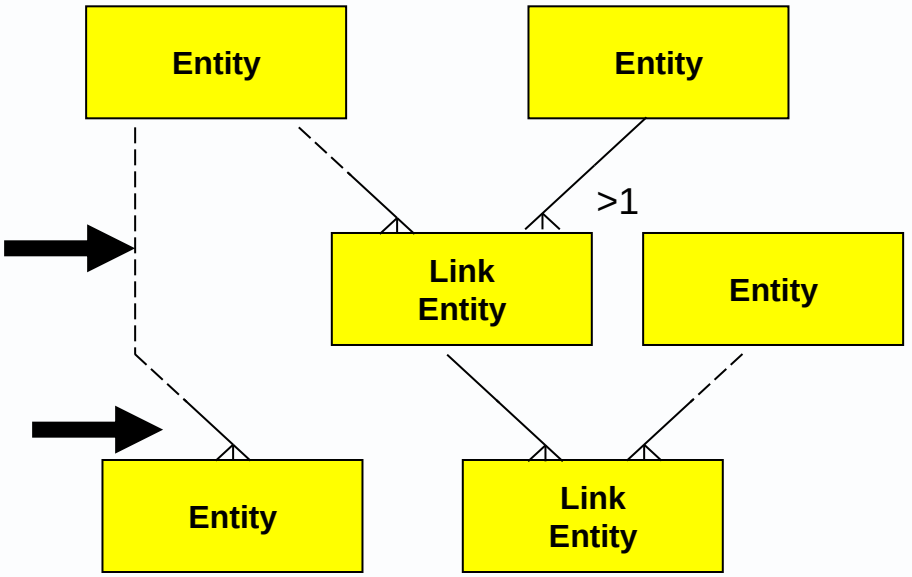
Entities	Data Stores	CREATE events	READ events	UPDATE events	DELETE events
Customer	CRM system. Call-center system. Contact-management system	Visit to Web site. Visit to facility. Account created.	Contextualized views based on credentials of viewer	Address. Discounts. Phone number. Preferences. Credit accounts	Death. Bankruptcy. Liquidation. Do-not-call.
Product	ERP system. Order-processing system.	Product purchased. Product manufactured. SCM involvement.	Periodic inventory catalogues.	Packaging change. Raw materials change.	Canceled. Replaced. No longer available.
Asset	GL tracking. Asset database.	Purchase Order. Unit Acquisition. Approval process.	Periodic report. Depreciation calculation. Verification.	Transfer. Maintenance. Accident report.	Obsolete. Sold. Destroyed. Stolen. Scrapped.
Employee	HR LOB system.	HR hire. Numerous forms. Orientation day. Benefits selection. Asset allocation. Office assignment.	Office access. Reviews. Insurance-claims. Immigration.	Immigration status. Marriage status. Level increase. Raises. Transfers	Termination. Death.

Data model

- [an artefact] that shows the content of a data store.
- A structure of inter-related data entities.

CACI data model notation

- Dashes mean the relationship is optional at *this end* of the line
- Crowsfoot allows more than one at *this end* of the line



Business data model

Avancier

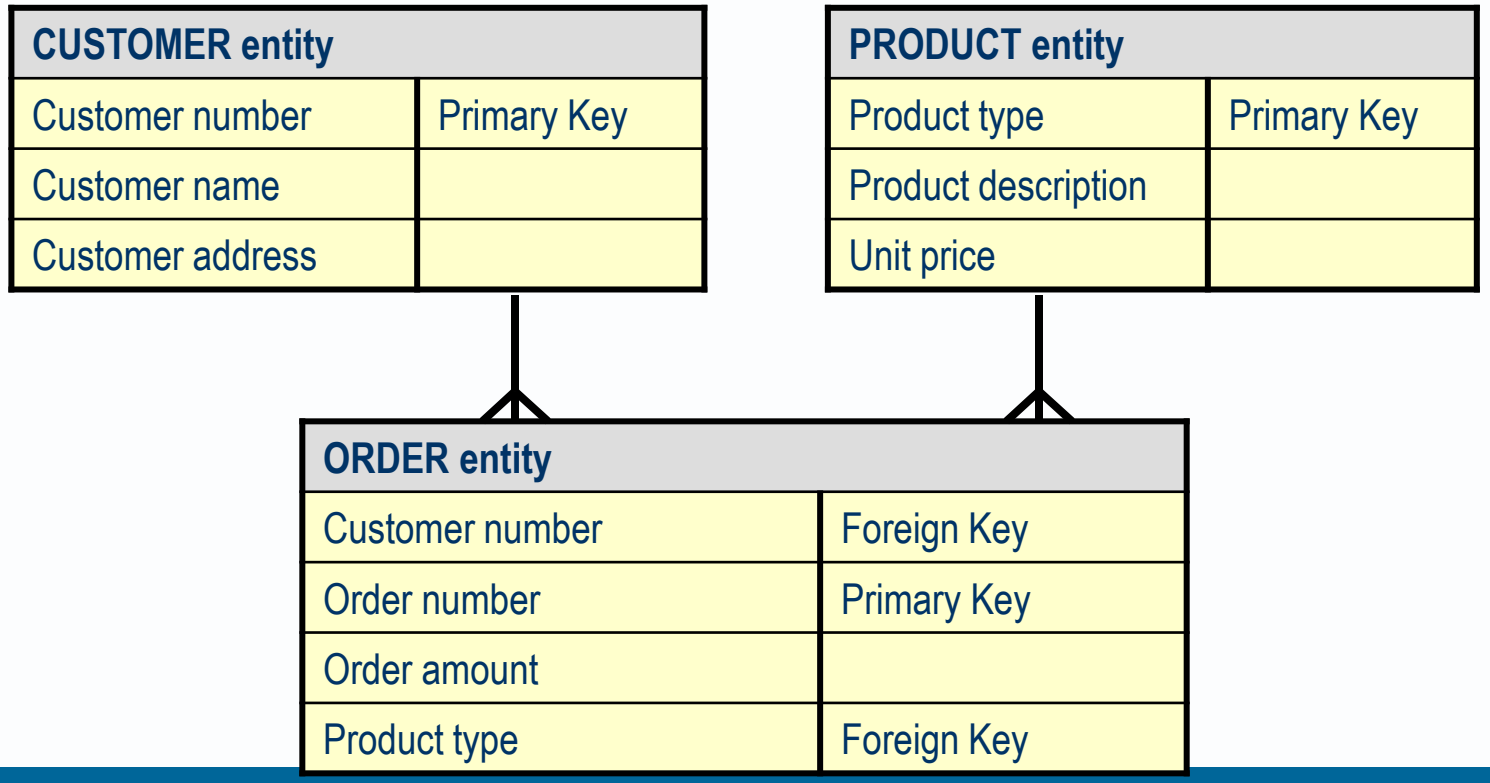
- [an artefact] that names and describes things in the business world that business people need to remember, regardless of computing. It identifies data entities that may appear in several data stores.
- It may define some business terms and concepts, but usually excludes details such as data types.

Possible business data entity groups

- Products (products, services)
- Properties (maintained resources, offices, vehicles, assets)
- Promotions (campaigns, adverts, mailings)
- Processes (transactions, events, orders, payments, applications)
- Places (areas, invoicing and delivery addresses)
- Pipes (routes, networks)
- Parties and people (customers, suppliers, organisations, employees)
- Points in time (calendar, dates, times)
- Pounds and Pennies (accounts, budgets, currencies)
- Papers (documents)

Logical data model

- [an artefact] that shows the data structure that must persist for the processes of an application to work.
- It shows relationships between instances of data types.
- It is usually drawn as a normalised data structure.



Data access path diagram

- [an artefact] that shows the route that a process takes through a data model or database. It is used to validate a data model structure and study performance issues.
-

Normalisation

- [a technique] for defining a data store structure that assists data integrity by storing each fact once. It also optimises update processes by minimising redundant data storage. The outcome of relational data analysis.
-

De-normalisation

- [a technique] that optimises input and/or output processes by structuring a data store structure to reflect the most important input or output data flow structures, at the expense of duplicating some stored data.
-

Data warehouse

- [a data store] that is optimised for the production of management information reports.
- It usually holds a non-normalised data structure.

Data server

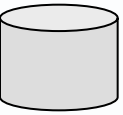
- [a platform component] that hosts a database management system and enables a data store to be accessed by applications.
- It usually enables direct access to any data entity instance in the data store, using its primary key.

Data architects

"engage and lead IT projects including
•the Enterprise Information Architecture
•information models and flows,
•data dictionaries, data standards, and data quality standards and processes"

Avancier

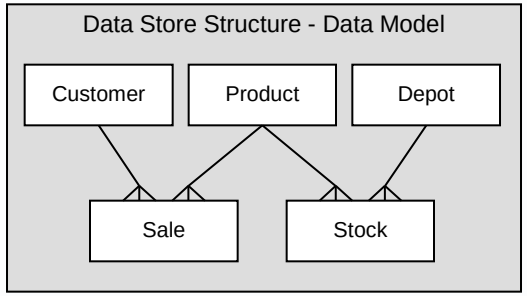
- Define business data in terms of

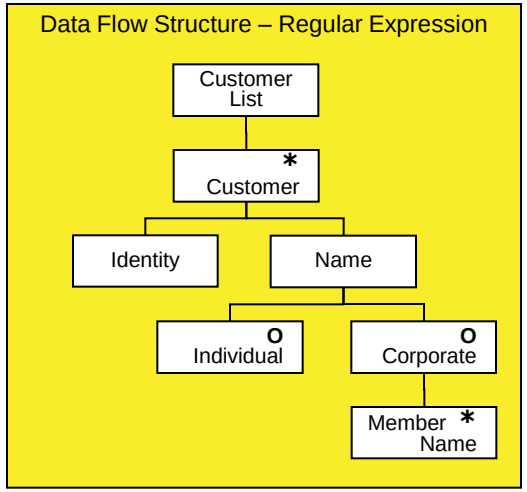


- **Data stores** created & used by business activities & applications
  - **Data store structures**: entities, attributes and relationships

- **Data flows** created & used by business activities & applications
  - **Data flow structures**: message and file formats

- **Data qualities** of data store/flow elements
  - Confidentiality, integrity and availability (CIA)
  - Data owners and stewards
  - Canonical data types (constraints on data item values)



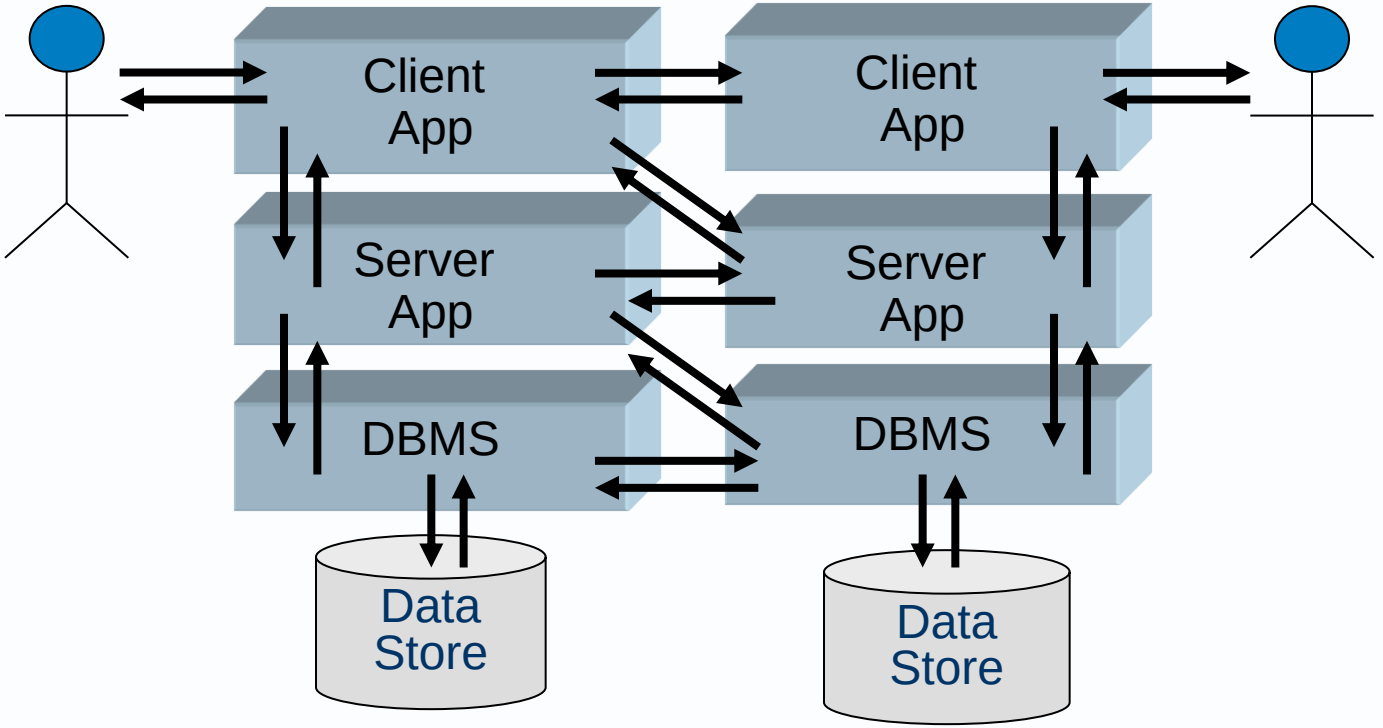


5.3: Data in motion

- **Data in motion** [a viewpoint] relating to data flows.
-
- **Data event** [a data object] representing an event in an input or output data flow.

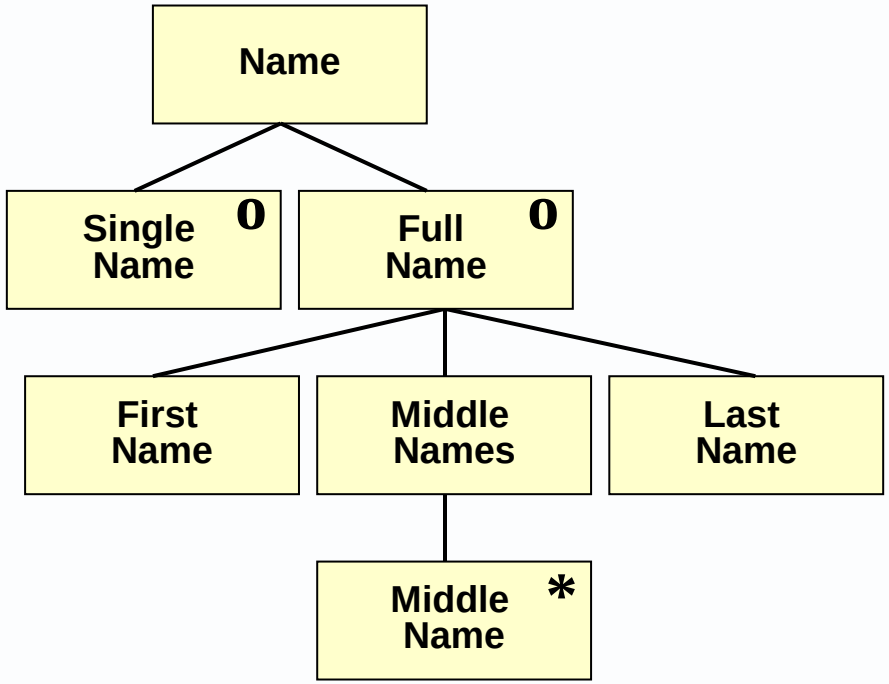
Data flow catalogue

- [an artefact] that lists the data structures transported from senders to receivers within a given application portfolio or system family.
- A data flow can take the form of a message, file, report, form, display format or other data stream.



Regular expression

- [a model] that describes the structure of a data flow
- (as a logical data model describes a structure of a data store).
- It is drawn using the universal grammar for defining the structure of a data flow or message.
- It is a hierarchical structure of in which every element is part of a sequence, or an option of a selection or an occurrence of an iteration.



Data format

- **Data format** [a standard] for the organisation of a data structure, such as
  - Comma Separated Values (CSV),
  - JSON or
  - Extensible Mark Up Language (XML).

- **Data format standard** [a standard] for the content of a data structure, such as
  - EDIFACT
  - a domain-specific XML Schema Definition (XSD).

Order, Invoice, Payment ,etc.

Flat text file
Fixed position fields
Fixed length fields?

Canonical data model

- [a standard] that provides the “one true definition” of data types used by an enterprise.
- It is an important tool in SOA and application integration.
- It defines what data can appear in messages between applications, and in the signatures of automated services.
- It may be defined at a physical level using a data format standard such as XML.

Data architects

"engage and lead IT projects including
- •the Enterprise Information Architecture
- •information models and flows,
- •data dictionaries, data standards, and data quality standards and processes"

- Define business data in terms of
- **Data stores** created & used by business activities & applications
  - **Data store structures**: entities, attributes and relationships
- **Data flows** created & used by business activities & applications
  - **Data flow structures**: message and file formats
- **Data qualities** of data store/flow elements
  - Confidentiality, integrity and availability (CIA)
  - Data owners and stewards
  - Canonical data types (constraints on data item values)

5.4: Data qualities and integration

- **Data quality** [a property] of a data item, data structure or data store. Notably, Confidentiality, Integrity and Availability (CIA).

Data quality	Aims are to ensure that
Confidentiality	Enterprise data is protected Private data remains private, accessible only to authorized readers.
Integrity	Business decisions (especially if safety-critical) are right, because a data item value is: •Consistent •Conformant to rules. •Correct •Controlled
Availability	The data (or systems) are available when needed.

Data integrity

Avancier

- [a property] that may embrace any or all of four qualities:

- **Consistent:** a data item (e.g. customer name) has the same value in every part of a distributed system, in all locations that data item is stored.
- **Conformant:** a data item obeys relevant business rules, sometimes in relation to another data item. E.g. an order must be for a known customer.
- **Correct:** a data item accurately represents a fact about an entity or event. The value of a data item is consistent with a fact in the real world.
- **Controlled:** a data flow has the same data content when it reaches its destination as it did when it left its source. OR data in a data store is not changed without authorisation.

Scoring data qualities (Tom Peltier)

Avancier

- Score each data item/group/store H/M/L thus

Confidentiality	Integrity	Availability
Impact of unauthorized use or disclosure	Impact of data inaccuracy, incompleteness or unauthorized modification	Impact of unavailable information
Severely impairs business operations, make a segment of the company unable to function or cause high monetary loss.	Causes failures of operations, revenue loss, wrong decisions to be made, loss in productivity or loss of customer confidence or market share.	Impairs business operations, affects customer service or makes it impossible to process revenues.
Does not severely affect operations or does not result in high monetary loss.	Makes it impossible to make some decisions, but the problem is not difficult to detect and correct, and does not severely impact business operations.	Causes productivity loss, but does not interrupt customer service or revenue generation.
Does not affect operations or result in significant monetary loss.	Does not disable business operations, since alternative validations of the information make it possible to continue	Does not severely impact business operations.

Threats to data qualities

Data quality	Security and data architects consider
Confidentiality	Deliberate theft. Identity theft is a common goal of criminal attacks against systems. Accidental revelation through loose identity management (including loose roles and authorities).
Integrity	Unauthorised creates, updates, deletes of data in data stores Tampering with data being transported in data flows. Duplication of data storage Duplication of data entry Low quality data entry
Availability	Attacks that disable access to systems. Denial-of-service attacks (can cost as much) Inadequate design for reliability and disaster recovery

Data architects especially interested

Data disintegrity

- [a property] an issue that may be redressed by one-off data quality improvement exercises, and by a variety of application integration patterns.

Data integrity solutions can involve

- One-off data quality improvement exercises
- Data warehouse
- Master data management

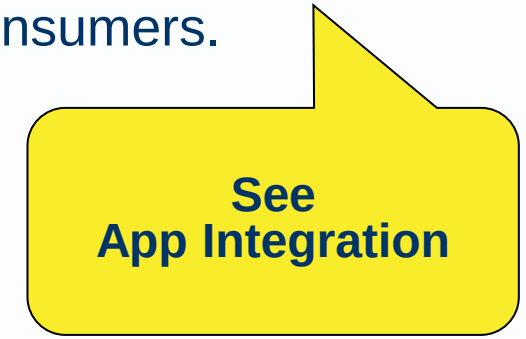
Data dissemination view

- [an artefact] that maps data entities to the applications, data stores or locations that hold them.
- This view shows duplication of data between data stores.
- It is useful in analysis of change impacts, data mastering and security vulnerabilities.

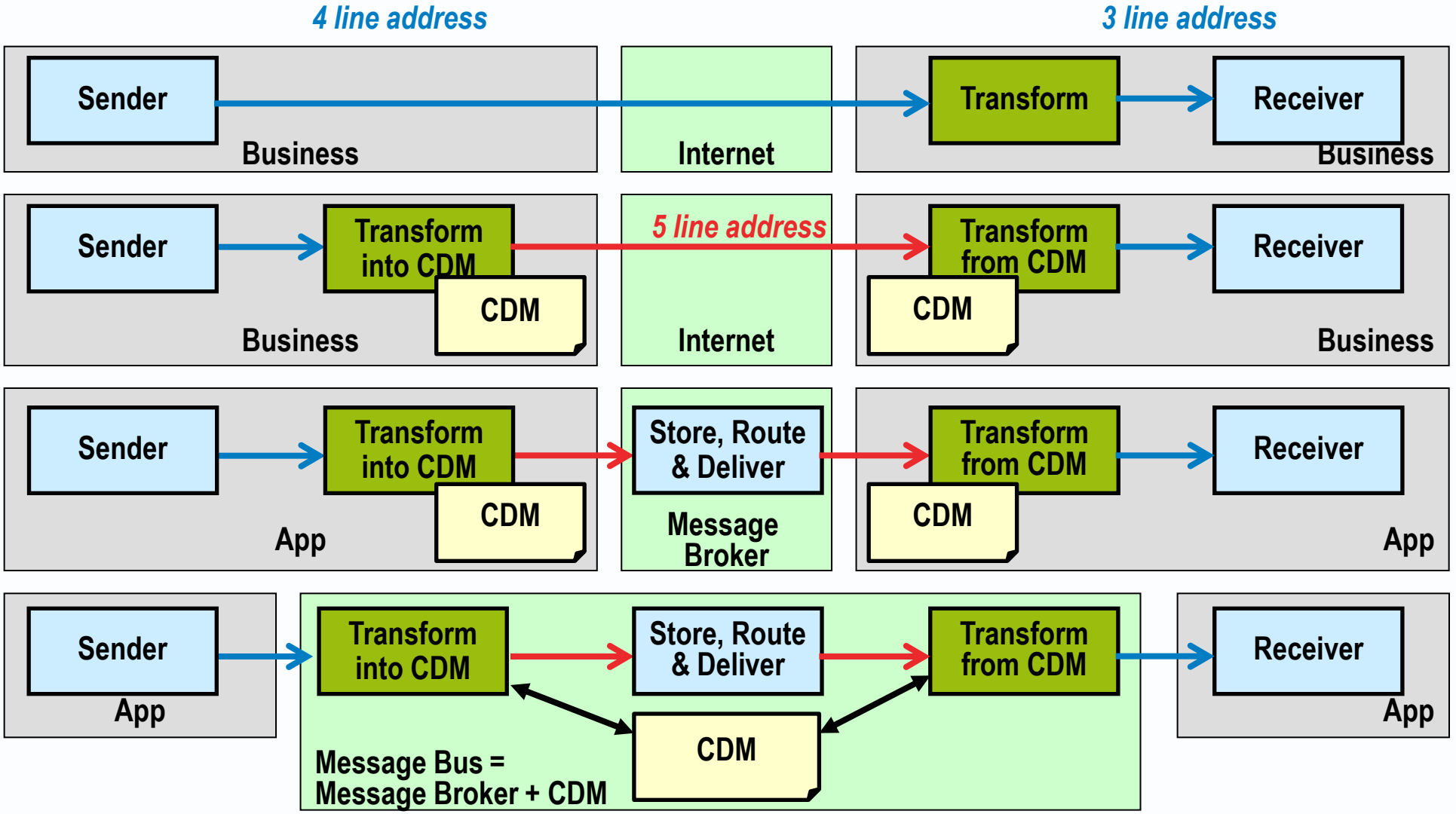
Data stores \ Common Entities	*Customer*	*Product*	*Asset*	*Employee*
CRM system.	**Master**			Copy
Call-center system.	Copy			
Contact-management system	Copy			
ERP system.		**Master**		
Order-processing system		Copy		
GL tracking			Copy	
Asset database			**Master**	
Timesheet				Copy
Expense Claim				Copy
Contract DB				Copy
Company Directory				**Master**

Master data management

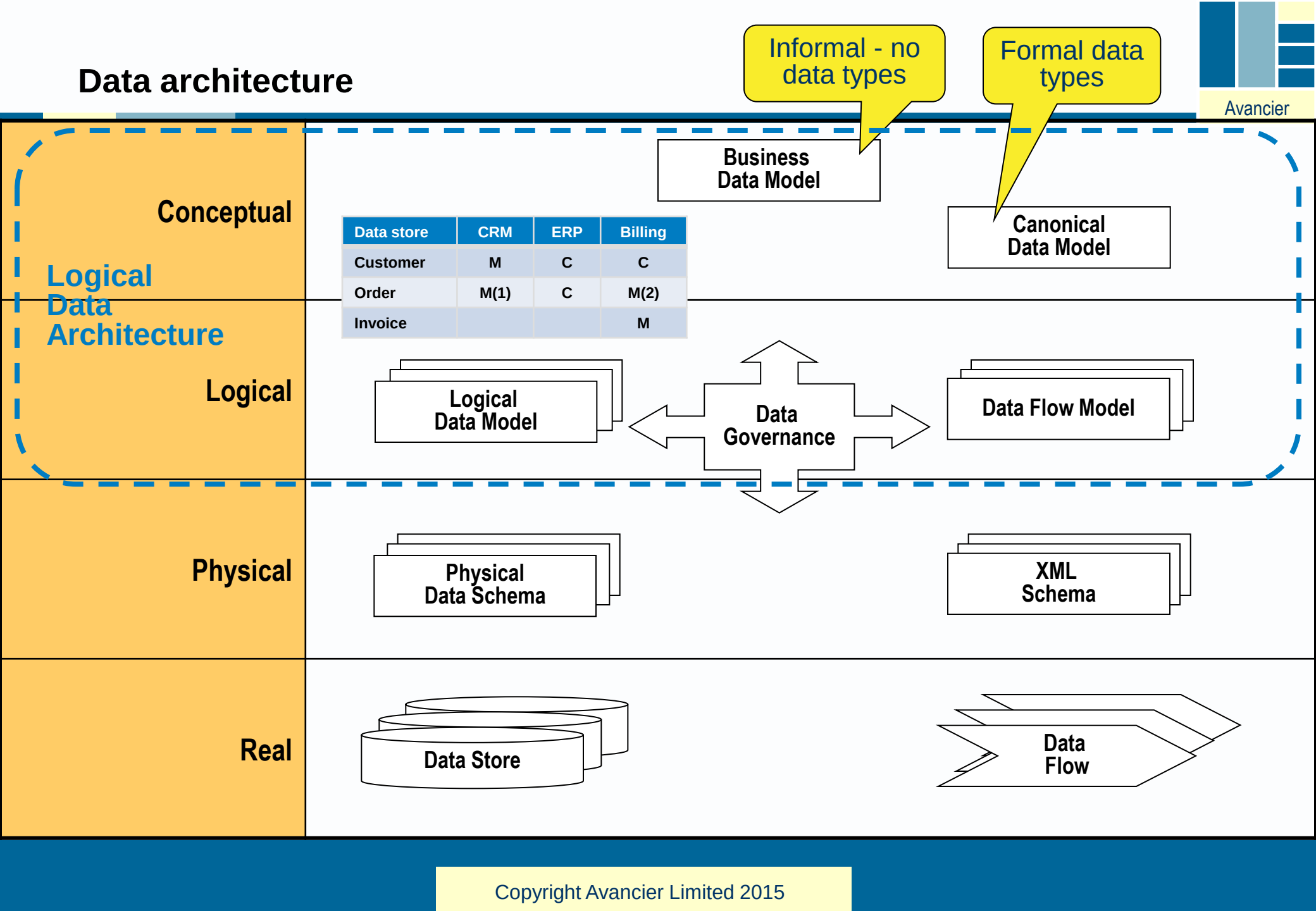
- [a technique] that enables an enterprise to maintain and/or find one “master” version of a data item or data structure, such as a customer or product data record.
- It is supported by a range of application integration patterns and technologies, including some that hide the reality of disparate data sources from data consumers.



Application integration using a Canonical Data Model



Data architecture

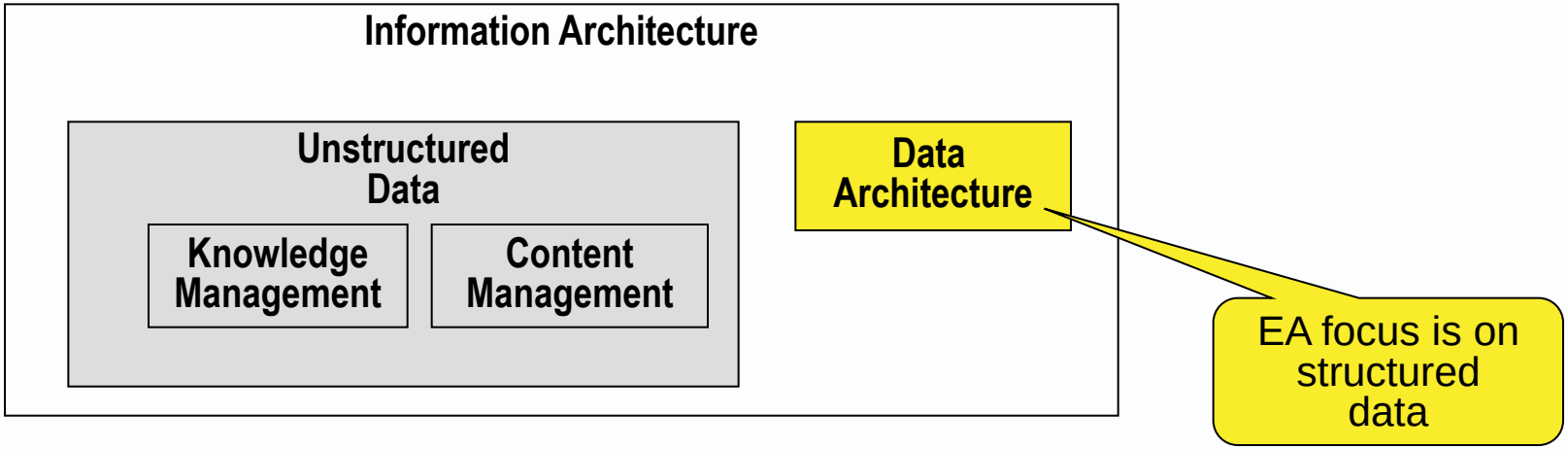


Unstructured data

Avancier

Knowledge and/or content management	The organisation, systems and processes for producing, storing, editing, sharing and searching unstructured data. Roles can include creator, editor, publisher, administrator (managing access permissions etc.) and consumer, viewer or guest. Knowledge and/or content management is regarded in this reference model as a matter for systems analysts and applications architects to address, rather than data architects.

Unstructured data is out of scope for us here



- Twitter say they have four fundamental data types and query patterns:
  - tweets,
  - timelines,
  - social graphs
  - search indices.
- For each, Twitter implemented custom data stores because existing solutions were insufficient.

Brands can be identified by various elements

Avancier

- When the levers of control are strongly centralized, content management systems are capable of delivering an exceptionally clear and unified brand message.

- All following may be trademarked as “brands”

  - **name:** word(s) used to identify a company, product, service, or concept
  - **logo:** a visual trademark
  - **tagline or catchphrase**: e.g. “Never-knowingly undersold"
  - **graphics:** e.g. the "dynamic ribbon" is a trademarked part of Coca-Cola's brand
  - **shapes:** e.g. the Coca-Cola bottle and Volkswagen Beetle shapes
  - **colors:** e.g. Owens-Corning is the only fiberglass insulation that can be pink.
  - **sounds:** e.g. a unique tune or chord: e.g. Windows
  - **scents:** e.g. the rose-jasmine-musk scent of Chanel No. 5 is trademarked
  - **tastes:** e.g. a trademarked recipe of herbs and spices for fried chicken
  - **movements:** e.g. the upward motion of Lamborghini car doors