



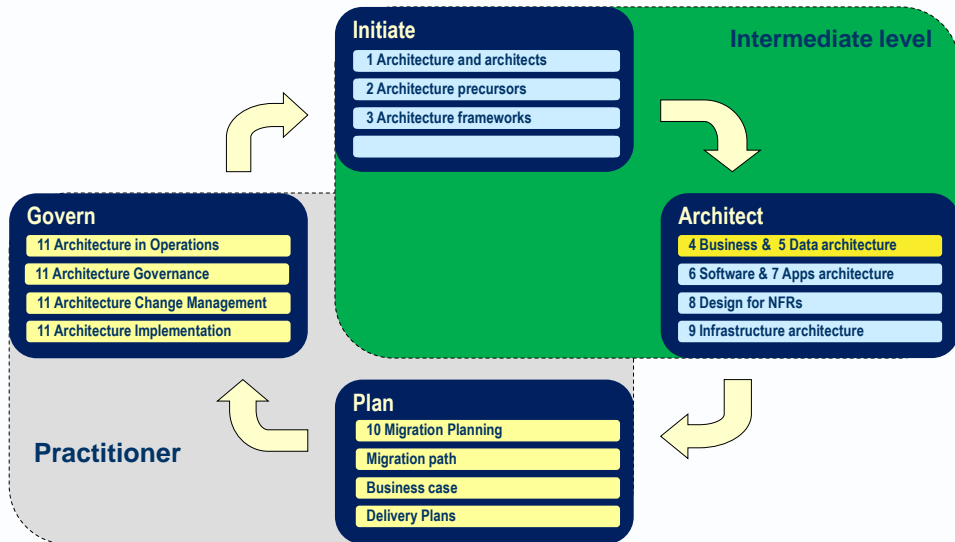
Avancier Reference Model

Data Architecture (ESA 5)

It is illegal to copy, share or show this document
(or other document published at <http://avancier.co.uk>)
without the written permission of the copyright holder

Copyright Avancier Limited 2015

5. Data architecture



Copyright Avancier Limited 2015

Data architecture in the AM core framework for EA



	<i>Passive Structure</i>	<i>Required Behaviour</i>	<i>Logical Structure</i>	<i>Physical Structure</i>
Business		Business Service Business Process	Function Role	Org Unit Actor
Data / Information	Data Entity	Data Flow	Log Data Model	Data Store
Applications		IS Service	Application Interface	Application
Infrastructure Technology		Platform Service	Platform Interface	Platform Application

Copyright Avancier Limited 2013

Foundation terms (not examined)



Entity	A structural thing that persists. A thing that is created, affected and ultimately destroyed by events.
Event	A behavioural thing that happens. May create or destroy an entity, or move an entity from one state to another in its lifecycle. May affect several entities.
Data	Can be divided into unstructured data (e.g. emails) and structured data.
Information	1: Data at the point of use by an actor in a business system. Or 2: Unstructured data as opposed to structured data.

Copyright Avancier Limited 2015

Foundation terms (not examined)



Structured data	One or more items (atomic facts) that describe one or more entities or events. Contained in data flows and data stores.
Data item	An elementary unit of information, a fact about an entity or an event. An attribute of an entity in a data model or an event in a data flow structure. A variable containing a value that is constrained by a type.
Data structure	A structure that arranges data items in one or more groups. May be defined in an XML schema or a database schema.
Data entity	The representation of an entity as a data structure that persists in a data processing system. Often corresponds to and records the state of a business process.

Copyright Avancier Limited 2015

Foundation terms (not examined)



Type	A generalisation; a form or structure common to several things; a constraint on an instance shared by all instances of that kind.
Data type	A type that defines the properties shared by instances of a data item or larger data structure. It constrains the values of the data. It defines the processes that can legitimately be performed on the data.
Data type (primitive)	A data type defined in a programming language. E.g. alphanumeric string, integer, floating-point number (decimal), and boolean.
Data type (user-defined)	A data type defined by systems analysts that is bespoke to the business at hand. E.g. customer, order, product.

Copyright Avancier Limited 2015

Foundation terms (not examined)



Data type (user-defined)	<i>A data type defined by systems analysts that is bespoke to the business at hand. E.g. customer, order, product.</i> <i>[repeat]</i>
Constraint (rule)	A business rule that further limits the values of a data type.
Derivation rule	A business rule that defines how the value of a data item is derived from the value of one or more other items (a special kind of constraint).

Copyright Avancier Limited 2015

Foundation terms (not examined)



Meta data	Data that describes data. Includes data structures, data types, business rules, data locations and data qualities.
Data dictionary	A catalogue of data item types, which may include business rules.
File	A synonym of data store; any identified collection of data stored in the computer. May be used in a data flow. May be saved for future use. Sometimes means more specifically a “flat file”, a data store in which data is stored and accessed in sequence, starting at one end.
Data source	Any data store, actor or component from which data is received by an application.

Copyright Avancier Limited 2015

Global org seeks Enterprise Data Architect to join their team



- ▶ Drive toward strategic Architecture goals and objectives by enhancing the **coherence, quality, security and availability of the** organization's data assets through the development and deployment of Architecture roadmap.
- ▶ Be primary advocate of **standards, data modeling and data processing artefacts** and guide data architecture work across multiple projects.
- ▶ Provide strategic architectural oversight on assigned projects, including development of **data model, interface and migration formats.**

Data in storage
Data in motion
Data qualities and integration

Copyright Avancier Limited 2015

5. Data architecture



- ▶ Leading bank urgently seeking a proven EA to
- ▶ engage and lead IT projects including
 - the Enterprise Information Architecture
 - information models and flows,
 - data dictionaries, data standards
 - data quality standards and processes
- ▶ develop and maintain the logical Enterprise Information Architecture that enables seamless information interoperability of all Bank systems for efficiency and cost-effectiveness.
- ▶ eutopiaonline.com

Data in storage
Data in motion
Data qualities and integration

Copyright Avancier Limited

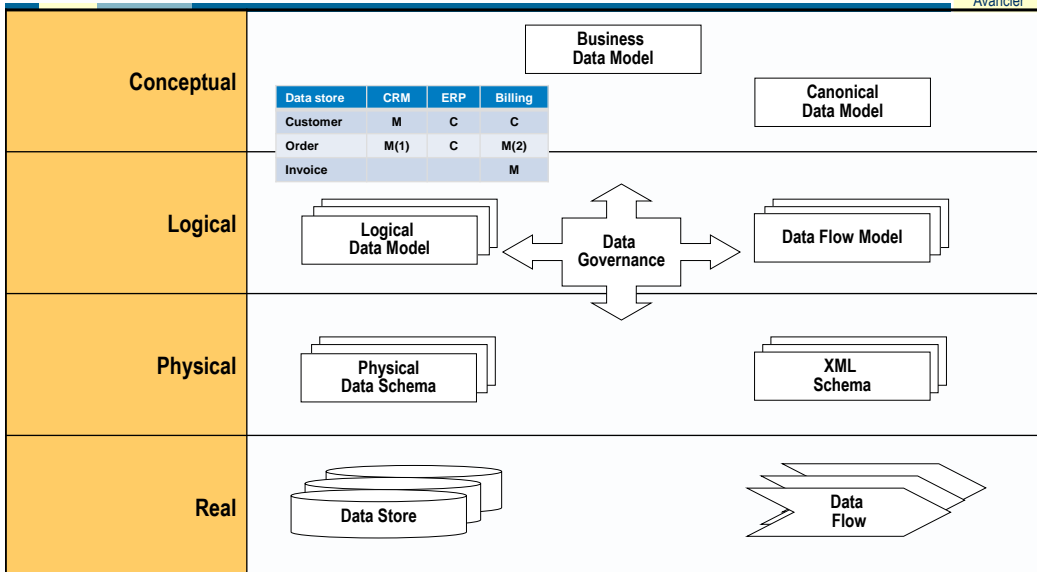
5. Data architecture



5. Data architecture
Basic concepts
Data in storage
Data in motion
Data qualities and integration

Copyright Avancier Limited 2015

Data architecture



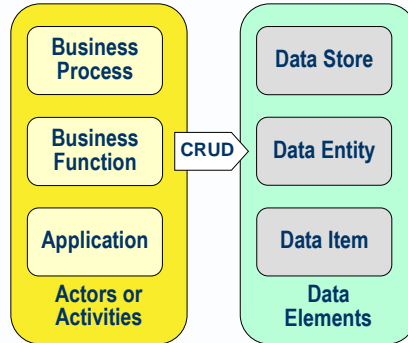
Copyright Avancier Limited 2015

Data is everywhere



- ▶ Business cannot work without data
- ▶ Data informs and enables business activities

- ▶ Various kinds of actors and activities
 - Can be mapped to
- ▶ Various levels of data element.



Process Data Entity		
	Create	Read
	Update	Create

Application Data Item		
	Create	Read
	Update	Create

Copyright Avancier Limited 2015

Mapping business data to activities

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



- ▶ A CRUD matrix shows
- ▶ Which Activities (business function, business process or event)
- ▶ Create and use which data (data store, data entity or data item)

ACTORS or ACTIVITIES	Order opening	Item addition	Order closure	Payment	Payment + 1yr
DATA ENTITIES					
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.

Copyright Avancier Limited 2015

5. Data architecture



5. Data architecture
Basic concepts
Data in storage
Data in motion
Data qualities and integration

Copyright Avancier Limited 2015

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"

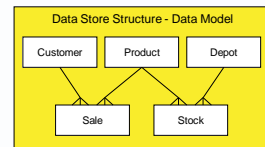


▶ Define business data in terms of



▶ **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)

Copyright Avancier Limited 2015

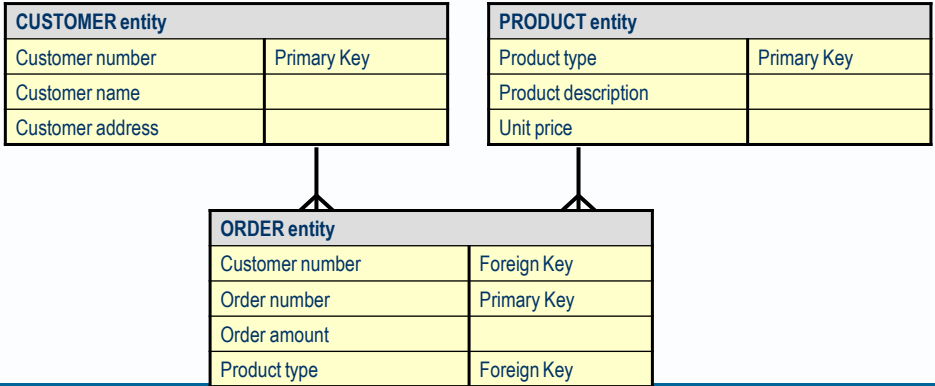
Describing data store content

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



Data model

A schema that groups data items into a data structure and defines the type of each data item.
 A structure that defines the attributes of entities and the relationships between them. It may include derivation rules for some data items.



Copyright Avancier Limited 2015

Data in storage

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

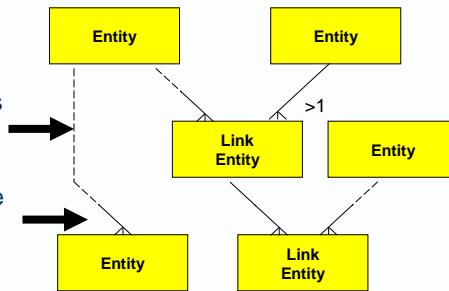


Logical data model

A definition of the data that must persist for the processes of an application to work.

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



Copyright Avancier Limited 2015

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

Business data model

Avancier

Business data model	A conceptual or domain model; can be a vehicle for documenting business semantics (the meaning of business terms).
---------------------	--

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)

Copyright Avancier Limited 2015

5. Data architecture

Avancier

5. Data architecture
Basic concepts
Data in storage
Data in motion
Data qualities and integration

Copyright Avancier Limited 2015

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)

Copyright Avancier Limited 2015

Data architects care about data qualities

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

Data quality (CIA)	A characteristic of a data item, data structure or data store. Notably: Confidentiality, Integrity and Availability.
---------------------------	---

▶ "Drive toward strategic Architecture goals and objectives by enhancing the **coherence, quality, security and availability of the organization's data assets** through the development and deployment of Architecture roadmap."

Copyright Avancier Limited 2015

Why?

Data quality	The 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: <ul style="list-style-type: none"> •Consistent: the same in all data stores. •Conformant: conforms to business rules. •Correct: represents what it is meant to. •Controlled: not changed without authorization (in store or flow.)
Availability	The data (or systems) are available when needed.

Copyright Avancier Limited 2015

Reference model entries

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

Data integrity	<p>Data integrity (1): A data Item has the same value in every part of a distributed system. A fact (e.g. customer name) has the same value in all locations that data item is stored.</p> <p>Data integrity (2): A data Item obeys relevant business rules, sometimes in relation to another data Item. The value of a data item is consistent with all invariant business rules e.g. an order must be for a known customer.</p> <p>Data integrity (3): A data Item accurately represents a fact about an entity or event. The value of a data item in a data processing system is consistent with a fact in the real world.</p>
Data flow (or message) integrity	The requirement that a data flow has the same data content when it reaches its destination as it did when it left its source.

Copyright Avancier Limited 2015

engage and lead IT projects including

- the Enterprise Information Architecture
- information models and flows,
- data dictionaries, data standards, and data quality standards and processes

Threats to data qualities

Data quality	Security and data architects consider
Confidentiality	<p>Deliberate theft. Identity theft is a common goal of criminal attacks against systems.</p> <p>Accidental revelation through loose identity management (including loose roles and authorities).</p>
Integrity	<p>Unauthorised creates, updates, deletes of data in data stores</p> <p>Tampering with data being transported in data flows.</p> <div style="border: 1px solid black; padding: 2px; display: inline-block; margin: 5px;"> Duplication of data storage Duplication of data entry Low quality data entry </div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; display: inline-block; margin: 5px; background-color: #ffffcc;"> Data architects especially interested </div>
Availability	<p>Attacks that disable access to systems.</p> <p>Denial-of-service attacks (can cost as much)</p> <p>Inadequate design for reliability and disaster recovery</p>

Copyright Avancier Limited 2015

engage and lead IT projects including

- the Enterprise Information Architecture
- information models and flows,
- data dictionaries, data standards, and data quality standards and processes

Data dissemination view

Data dissemination view A view showing the dispersal (and perhaps duplication) of data between data stores or locations.
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

Copyright Avancier Limited 2015

Improving data quality



- Data integrity solutions can involve
- One-off data quality improvement exercises
 - Data warehouse
 - Master data management

Data warehouse	A kind of database system designed to hold a non-normalised data structure that is optimised for the production of management information reports.
Master data management	The systems and processes that enable an enterprise to maintain and/or find one “master” version of any data item or data structure, typically customer or product data.

See
App Integration

Supported by a range of approaches and technologies, including middleware technologies that hide the reality of multiple disparate data sources from data consumers.

Copyright Avancier Limited 2015

Scoring data qualities (Tom Peltier)

engage and lead IT projects including

- the Enterprise Information Architecture
- information models and flows,
- data dictionaries, data standards, and **data quality standards** and processes



▶ 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.

Copyright Avancier Limited 2015

Data security

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

Data security

1: Confidentiality alone.

Or 2: a combination of Confidentiality, Integrity and Availability.

Tom Peltier suggests rating the security level of a data item, data structure or data store as equal to the highest of the individual ratings (one, two or three) awarded for Confidentiality, Integrity and Availability.

Confidentiality	Integrity	Availability	Security Level
Impact of unauthorized use or disclosure	Impact of data inaccuracy, incompleteness or unauthorized modification	Impact of unavailable information	DERIVED DATA
High	High	High	High if any high to left
Moderate	Moderate	Moderate	Moderate if any moderate to left
Low	Low	Low	Else low

Copyright Avancier Limited 2015

Why score security levels?

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

▶ **For EAs to impose enterprise-wide standards that constrain SAs!**

Security Level	Data store encryption	Inter-App Data flow encryption	Remote disaster recovery
High	Yes	Yes	Yes
Moderate		Yes	Yes
Low			

Copyright Avancier Limited 2015

Unstructured data



Knowledge and/or content management

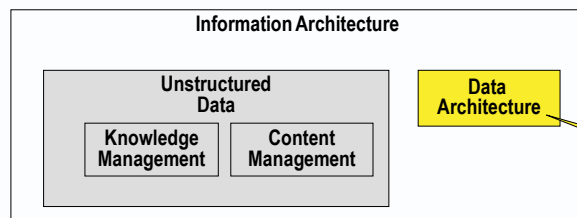
The organisation, systems and processes for producing, storing, editing, sharing and searching unstructured data.

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.

Copyright Avancier Limited 2015

Unstructured data is out of scope for us here



EA focus is on structured data

- ▶ 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.

Copyright Avancier Limited 2015

Brands can be identified by various elements



- ▶ 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

Copyright Avancier Limited 2015

5. Data architecture – end of pass 1



- ▶ SHOW RELEVANT MOCK EXAM QUESTIONS

Copyright Avancier Limited 2015

5. Data architecture



5. Data architecture
Basic concepts
Data in storage
Data in motion
Data qualities and integration

Copyright Avancier Limited 2015

Define data lifecycle view – simple view



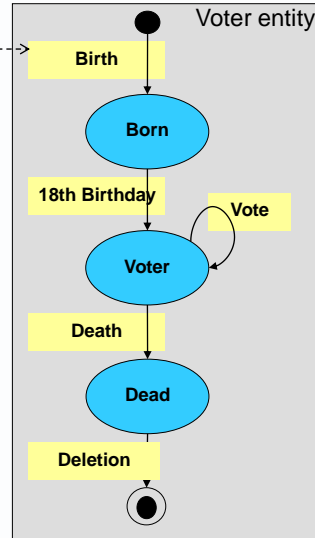
		Simplistic data entity life cycle			
Entities	Data Stores	CREATE events	READ events	UPDATE events	DELETE events
Customer	CRM system. Call-center system. Contact-management system	Visit to Web site. Visit to facility. Account created.	Contextualized views based on credentials of viewer	Address. Discounts. 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.

Copyright Avancier Limited 2015

Foundation terms



Data event	The representation of an event as a data structure in a data flow input to one or more data processing systems.
Data lifecycle	The life of a data entity expressed in terms of the states it passes through from creation to deletion, and data events that cause state transitions.

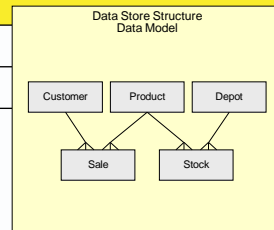


Copyright Avancier Limited 2015

5. Data architecture



5. Data architecture
Basic concepts
Data in storage
Data in motion
Data qualities and integration



Copyright Avancier Limited 2015

Data in storage



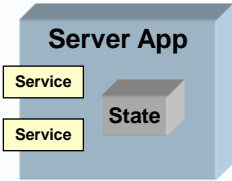
Data in storage	Those aspects of data architecture relating to data that persists in a location.
Data store	<p>A data structure that is held in persistent memory.</p> <p>Any file or database from which data can be extracted by an application.</p> <p>You can define the state of any data store (be it a database, cache or component) in a data model (though it may contain only a flat list of attributes).</p>

Copyright Avancier Limited 2015

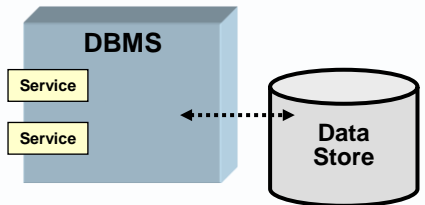
Data in storage



State The data structure maintained inside the memory of a process or component.



Database A persistent data structure that can be accessed by applications. Usually accessed via a DBMS that enables direct (rather than serial) access to any part of the data structure.

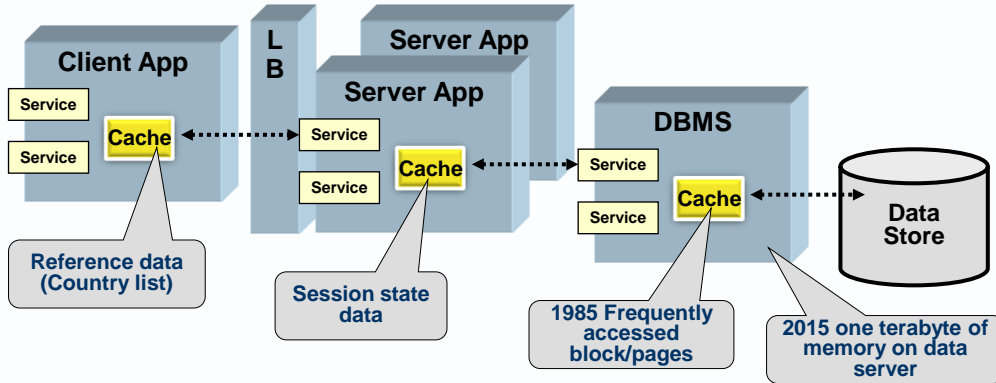


Copyright Avancier Limited 2015

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.



Copyright Avancier Limited 2015

5. Data architecture



5. Data architecture
Basic concepts
Data in storage
Data in motion
Data qualities and integration

Copyright Avancier Limited 2015

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)

Data Store Structure - Data Model

Data Flow Structure - Regular Expression

Copyright Avancier Limited 2015

Data in motion

Avancier

Data flow A data structure that is transported from sender to receiver. It is carried from data source to destination in a message, file, report or other data transport vehicle.

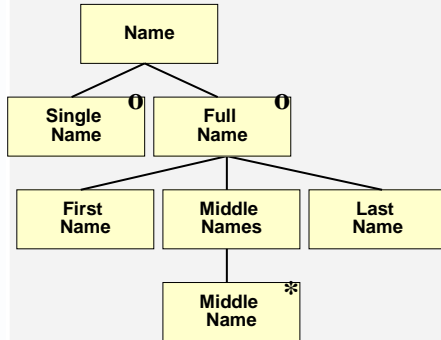
Serial files
 Message queues
 Transaction I/O messages
 UI display
 Any data record(s) sent from sender to receiver

Copyright Avancier Limited 2015

Data in motion



Regular expression



A hierarchical structure of elements arranged so that every element is part of a sequence, or is an option of a selection or is an occurrence of an iteration.

You can define every data flow structure as a regular expression (after Kleene's theorem). Though many messages contain no more than a list of data items.

(You can also define a process structure as a regular expression, under a logical control flow in which loops and alternative paths are governed by conditions.)

Data in formats and standards



Data format

A format or language for presenting data flow structures.

E.g. Comma Separated Values (CSV), JSON, Extensible Mark Up Language (XML).

Data format standard

A standard for the content of data flow structures. E.g. EDIFACT, and domain-specific XML Schema.

Flat text file
Fixed position fields
Fixed length fields?

Order
Invoice
Payment
etc.

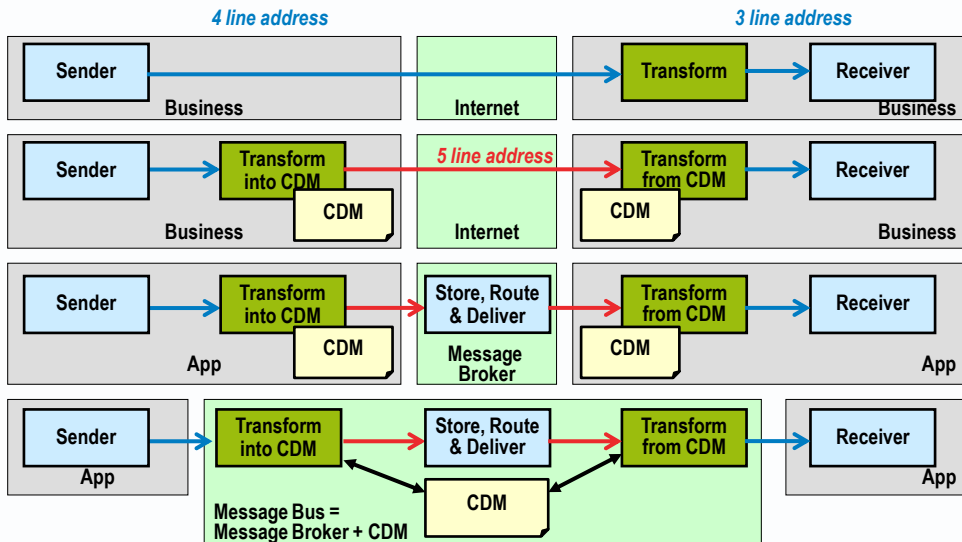
5. Data architecture



5. Data architecture
Basic concepts
Data in storage
Data in motion
Data qualities and integration

Copyright Avancier Limited 2015

Application integration using a Canonical Data Model



Copyright Avancier Limited 2015

Data in motion



Canonical data model

The “one true definition” of data items used in data flows between processes.

Used to define the data items sent to and returned from automated business services (whether or not they are provided by a broker application).

A logical data model or other form of specification that aggregates and de-duplicates the definitions of data types used in various inter-application data flows.

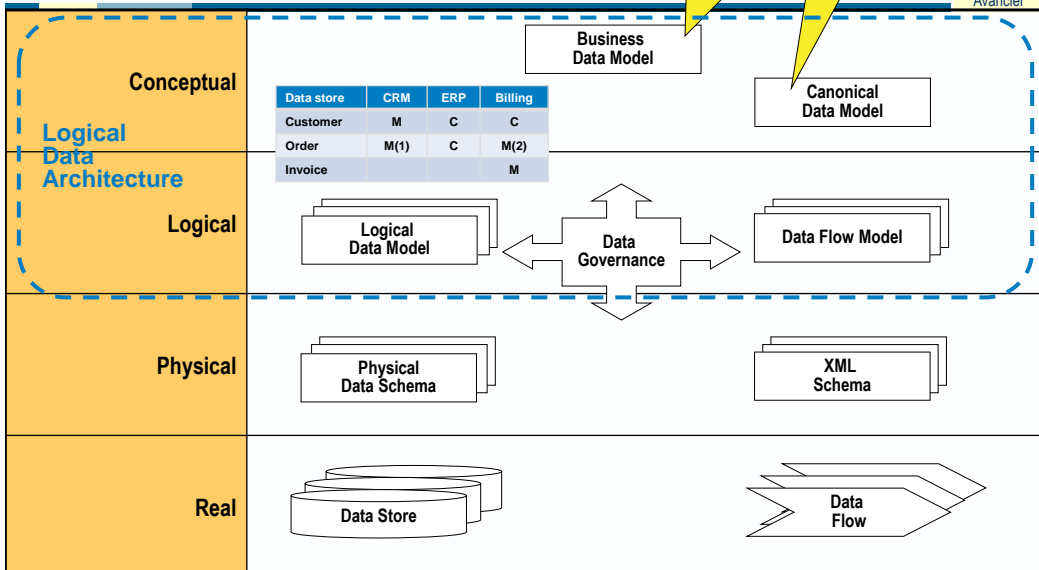
The data flows may be defined at a more physical level using XML schema and other data format standards.

Copyright Avancier Limited 2015

Data architecture

Informal - no data types

Formal data types



Copyright Avancier Limited 2015