Legend

- Data Modelling / database design
- Master Data Management
- Salesforce Data Management
- Data Governance
- Large Data Volumes
- Data Migration

External Data Sources

Non-Salesforce
Table | Table1

Salesforce Org
ObjectA | Object1

Salesforce Data Architecture

Custom Objects

Custom Objects represent domain specific data structures which do not map to Standard Objects e.g. Invoice.

Object__c
Field
Field
...

Object__c
Field
Field
...

Custom Objects support M-D relationships but can't be the master of a Standard Object.

Object Relationship Types :
Master-Detail - (Ownership) Sharing is controlled by the Master. Cascade delete. Lookup - (Associative) Lookup Filters. Delete behaviour; Clear/Restrict. Required?
Hierarchical - Same object e.g. ParentCase.

Field Data Types :
Relationship (M-D, Lookup)
Formula
Rollup Summary (RSF)
Geolocation
Literal Types (Text, Date etc.)
Abstract Types (Phone)

Big Objects

Big Objects provide BigData storage with consistent performance for a billion records or more. Data is not accessible via standard UI. Access via SOQL/Async SOQL and populate via Data Loader, API or Apex.

FieldHistoryArchive
Index
Field
...

Object__b
Index
Field
...

Standard BigObjects :
BigObjects are used by Salesforce products such as Field Audit Trail for up to 10 years of field history.

Custom BigObjects :
Index - 1 to 5 custom fields. Each has Index Position and Index Direction (ASC or DESC). Determines how records can be queried, immutable once set.
Fields - Lookups to Salesforce Objects, Literal Types.

Query Access :
SOQL - immediate (synchronous) small volume queries. Queries must filter using index fields from position 1 onward without gaps.
Async SOQL - requires additional licensing. Background massive volume queries over API. No aggregation or filtering outside of the index.

External Objects

External Objects are virtual objects which map to data stored outside of the Salesforce platform. Data is access on-demand via Web Service callouts (OData protocol) via Salesforce Connect.

Salesforce Connect
OData Adapter

Salesforce Connect
Cross Org Adapter

External Data Source — External Object Table1__x
External Data Source — External Object ObjectA__x

External Data Sources :
Connect to content & data held outside of Salesforce. Files Connect (e.g. Sharepoint, Box), Salesforce Connect (external RDBMS) or Federated Search.

Salesforce Connect :
Adapters which wrap connectivity protocols (e.g. OData 4.0).

External Object Relationships :
(a) Lookup Field - Link Child External Object to Parent Salesforce Object by 18 char Id.
(b) External Lookup - Link Parent External Object to Child Salesforce Object via External Id (Parent)==Lookup (Standard/Custom), External Column Name (External).
(c) Indirect Lookup Field - Link Child External Object to Parent Salesforce Object via External Id (Parent)==Relationship field value (Child).

Standard Objects

— Lookup Field or Indirect Lookup Field Relationship links Account to External Object child records

Sales Objects

Account

Contact

Opportunity

Quote

Lead

Order

Record Id may serve as a Globally Unique Identifier where Salesforce is the system of record. Otherwise External Id fields are used.

Service Objects

External Lookup Relationship — links parent External Object record to child Cases.

Case

Service Entitlement

Work Order

Knowledge Article

Live Agent Chat

Consent Management Objects

Individual

Authorisation Form Consent

Contact Point Consent

Communication Subscription Consent

For record merge the Individual Setting [Retain the most recently modified data privacy record?] can be used. Otherwise a manual selection is required.

External Data Providers

External data services can be accessed via Lightning Data packages and Data Integration Rules for enrichment and validation purposes.

Data Services

Examples :

Data Enrichment
—> Company Information
—> Address Completion
—> Contact Details
—> Location Geocodes

Validation
—> Bank Account details

Language Translation

Large Data Volumes

Large Data Volume considerations apply to objects with millions of records.

Query Optimiser : Determines best index for filters, or Object (Table Scan); order for remaining Objects, injects custom FK value objects for efficient join paths, optimises execution plan & updates statistics.

Database Statistics : updated nightly. The Query Optimiser maintains a table containing statistics about the distribution of data in each index.

Indexes : Standard indexes on relationship fields, RecordTypeId, Systemmodstamp, Id and Name fields. Custom indexes can be added; excl. multi-select picklist,long text, non-deterministic formula,encrypted fields. ExternalId fields are indexed.

For Custom Fields Salesforce creates an Index Table with strongly typed fields and a Database Index. Records with nulls (empty-value) are omitted by default.

Index Usage : internal statistics first checked for selectivity for the query filters.
—> Standard Indexes - filter matches less than 30% of the first million records and less than 15% of additional records.
—> Custom Indexes - filter matches less than 10% of the total records, up to a maximum of 333,333 records.

WHERE clause contains :
AND - used unless one index exceeds 20% total or 666K.
OR - used unless one index exceeds 10% total or 333K. All fields must be indexed.
LiKE - internal statistics are not used, 100K sample.

Two-Column Indexes : List views, 1 to display, 1 to sort.

Divisions : data partitioning strategy (EMEA, APAC etc.)

Skinny Tables : automatically improve the performance of read-only operations by removing the joins between underlying database tables that store standard and custom fields separately. Useful above Million record threshold. Copied to Full Copy sandboxes only. Also termed "Reporting Accelerators".

Skinny tables can be created on custom objects, and on Account, Contact, Opportunity, Lead, and Case objects. They can enhance performance for reports, list views, and SOQL. Literal Data Types (+encrypted) are supported. Skinny tables can't contain fields from other objects.

Object
Standard
Field1
Field2

Custom
Field3
Field4

Skinny
Field1
Field3
i=<100

Techniques

Mashups : quick, loosely coupled integration of the two applications. Data is never stale. Short interactions, small amounts of data.

Defer Sharing Calculation : suspend and resume sharing calculations, manage two processes: group membership calculation and sharing rule calculation until new users, rules, and other content have been loaded.

SOQL and SOSL :
—> SOQL==objects+fields are known; single object or related; number, date, checkbox fields; count the records for criteria, sort the results.
—> SOSL==objects+fields unknown; multiple objects unrelated; efficient division queries. Faster for CONTAINS queries. SOSL can tokenise multiple terms within a field (for example, multiple words separated by spaces) and build a search index.

Deleting Data : Soft delete moves data to the Recycle Bin for 15 days, the data remains resident and impacts on performance. Use Hard Delete option in the Bulk API.

Search : large data volume operations require the search index to rebuild.
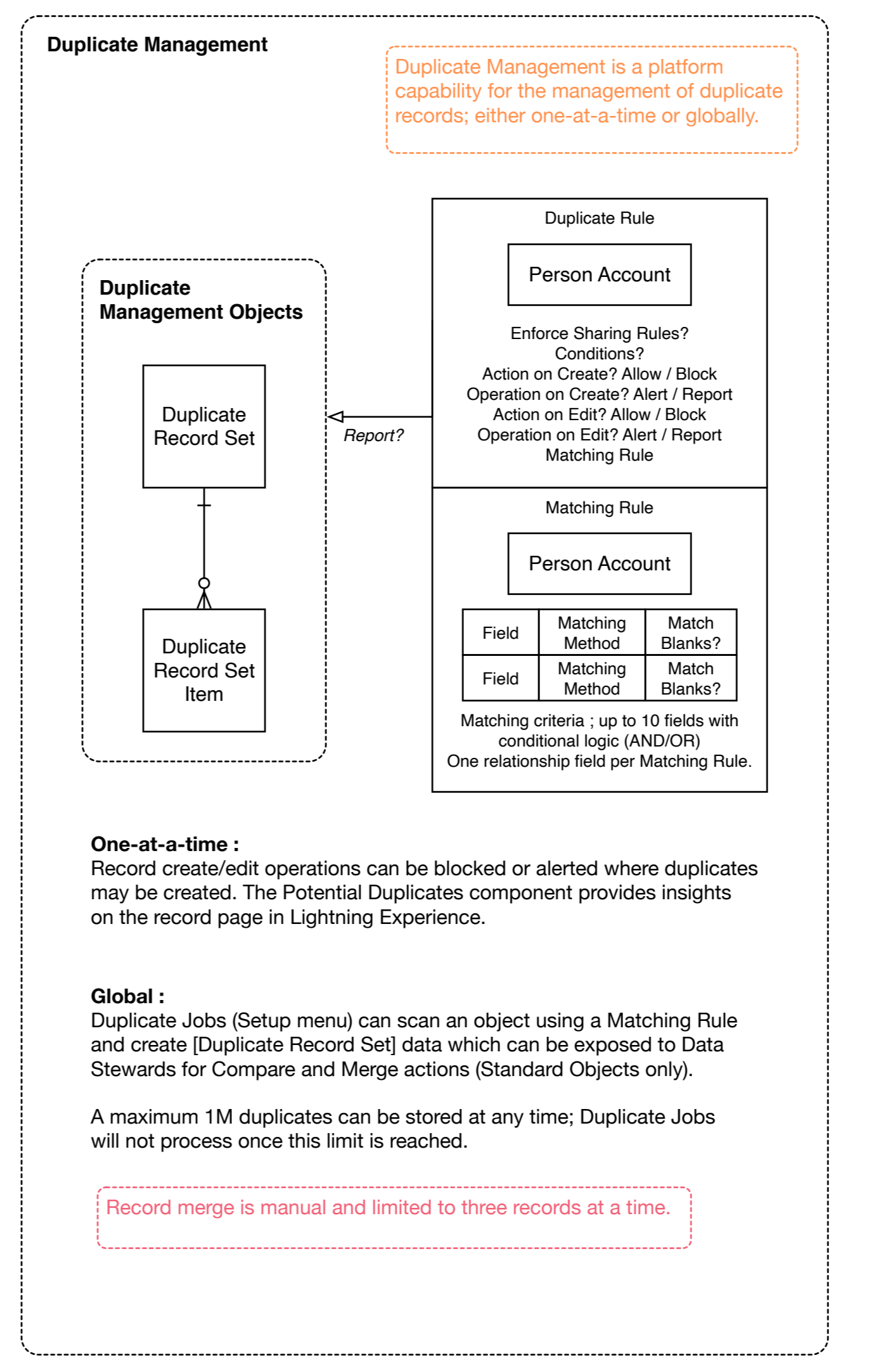
Best Practice

Reporting : Limit relationships, fields and records.
Partition data to match its likely use. Reduce records per object.
Reduce joins by restricting number of relationships per report.
De-normalise data when practical.
Limit the number of fields in a report, SOQL or list view to the essential.
User filters which emphasise the use of Standard then Custom indexes.

Searching : Keep searches single-object, specific and avoid wildcards.

SOQL and SOSL :
Avoid querying of formula fields especially deterministic formulae.
Tune the SOQL query, reducing query scope, and using selective filters.
Single-column indexes for picklists and foreign key fields exclude rows in which the index column is equal to null. Use a Two-Column index.

Deleting Data : Use Bulk API hard delete option, delete child records first.

General :
Avoid having any user own more than 10,000 records.
Avoid any parent record having more than 10,000 child records.

Duplicate Management

Duplicate Management is a platform capability for the management of duplicate records; either one-at-a-time or globally.

Duplicate Management Objects

Duplicate Record Set

Duplicate Record Set Item

Report?

Duplicate Rule
Person Account

Enforce Sharing Rules?
Conditions?
Action on Create? Allow / Block
Operation on Create? Alert / Report
Action on Edit? Allow / Block
Operation on Edit? Alert / Report
Matching Rule

Matching Rule
Person Account

| Field | Matching Method | Match Blanks? |
| Field | Matching Method | Match Blanks? |

Matching criteria ; up to 10 fields with conditional logic (AND/OR) One relationship field per Matching Rule.

One-at-a-time :
Record create/edit operations can be blocked or alerted where duplicates may be created. The Potential Duplicates component provides insights on the record page in Lightning Experience.

Global :
Duplicate Jobs (Setup menu) can scan an object using a Matching Rule and create [Duplicate Record Set] data which can be exposed to Data Stewards for Compare and Merge actions (Standard Objects only).

A maximum 1M duplicates can be stored at any time; Duplicate Jobs will not process once this limit is reached.

Record merge is manual and limited to three records at a time.

Salesforce Data Management

User License Types

User License types control data access & storage.

Internal User Licenses (non-exhaustive) :

Sales Cloud - Sales Objects, Custom Objects (CRUD).
Service Cloud - Sales Cloud plus extended Service Objects (CRUD).
Salesforce Platform - Account+Contact (RU), Custom Objects (RU)
Lightning Platform App Subscription Light - 10 Custom Objects (RU)
Lightning Platform App Subscription Enterprise - Account+Contact (RU), 10 Custom Objects (RU)
Employee Community - 10 Custom Objects (RU)

External User Licenses (non-exhaustive) :

Partner Portal - Below plus Opportunity C
Customer Community Plus - Below plus Account C
Service Cloud Portal - Account (RU), Contacts (RU), Assets (CRU), Cases (CRU)
External Identity - Custom Object (CRUD)

C = Create
R = Read
U = Write
D = Delete

Data Classification

Data classification enables recording of the data owner, field usage, data sensitivity level, and compliance categorisation of fields in any standard or custom object. Data classification can be used to control access, reporting and to manage data compliance.

Data Classification Metadata Fields :

Data Compliance Category - Picklist
PII, GDPR, HIPAA etc.

Data Sensitivity - Picklist
Public, Internal, Confidential

.. metadata values can be uploaded per Object-Field
E.g. Contact,BirthDate,Confidential

Data Owner - User or Public Group responsible for setting sensitivity etc.
Field Usage - Active, DeprecateCandidate,Hidden

Data Governance

Data governance is the process of managing the availability, usability, integrity and security of enterprise data. A governance framework must define standards, controls and processes that control data usage and data protection compliance.

Data Protection GDPR :
Get Buy-in and build your team - leadership buy-in, core cross-functional team to work on compliance effort. Appoint Data Protection officer.
Assess your organisation - Data inventory for personal data and register of processing activities. Establish controls and processes - create a roadmap of operational and technical changes.
—> Privacy notices - Website, cookies, tags etc.
—> Usage limitations - Administrative and technological controls to limit use of data.
—> Security - Administrative, physical and technological security measures.
—> Data Subject rights - process to manage consent preferences and complaints.
—> Vendor management - contracts with 3rd parties.
—> Incident response - detect and respond to security breaches.
—> Training - employee and vendor training for awareness of policy.
—> Assessments - Data protection impact assessments for high risk data processing activity.
Document Compliance - the protection of personal data is a fundamental right.

Enterprise Data Governance Framework :
Roles & Responsibilities -
—> Data Steward - Defines teams, roles and responsibilities for quality improvement and day-to-day maintenance. Quality Control.
—> Data Custodian. Responsible for the technical environment.
Governance Plan - Data Definitions, Data Quality Standards, Permissions, Quality Control.
Governance Framework Models -
—> Centralised - execution of rules, standards and policies by a central governance body.
—> Decentralised - individual responsibility for data quality maintenance.
—> Hybrid - combines both.

Data Import/Export

The Salesforce platform provides a rich set of APIs and tools for the import and export of data in small synchronous batches or via large volume asynchronous data loading operations.

The Bulk API is intended for large volume Data Import, Export and Delete operations.

Loading Data from the API :
—> More than 100K use Bulk API.
—> Clean data first as errors in batches forces single row processing for the batch.
—> Use read/write sharing for initial load.
—> If possible for initial loads, populate roles before sharing rules.
(i) Load users into roles.
(ii) Load record data with owners, triggering role hierarchy calculations.
(iii) Configure public groups & queues, let computations propagate.
(iv) Add sharing rules one at a time, let computations finish for each.
—> If possible, add users & data then create & assign groups & queues.
—> Minimise parent record-locking conflicts by grouping records by parent.
—> Defer Computations (disable ApexTriggers and process automations).
—> Defer sharing calculations.
—> Tune the batch size (HTTP keepalives, GZIP compression).

Export Data via the Bulk API Query :
—> Use the PK Chunking request header (Sforce-Enable-PKChunking) to enable automatic primary key (PK) chunking for a bulk query job. PK chunking splits bulk queries on very large objects into chunks based on the record IDs, or primary keys, of the queried records.

Delete Data via the Bulk API :
—> For large volume deletes use the Hard Delete option to skip the Recycle Bin (soft-delete) processing.

Cross Org Data Synchronisation :
—> Salesforce to Salesforce - Native platform capability. Object Publication/Subscription model.
—> Organisation Sync (retired) - A second Salesforce Org with a synchronised sub-set of data.
—> Salesforce Connect Cross Org Adapter - External Object access (Virtualised) to data held in a second Salesforce org (Publication org). Consumes REST API calls.