Data volume management

Archive and Backup

Siba Padhy
Program Architect, Director

spadhy@salesforce.com

Backup vs. Archiving

A **backup** is designed to facilitate information recovery, while an **archive** is designed support access to information that have lower transactional relevance.

EVENT	Backup relevance	Archiving relevance
Corruption of Records	Accidental or malicious data updates	Corporate compliance
Record deletion	Accidental or malicious data updates	Performance improvement thru data volume reduction
Data model changes	Restoration from backup	Business process changes, Migration
Metadata backup	Unintentional configuration changes	-NA-
Migration needs	All transactional data restoration	-NA-

Archiving Challenge

Responding to Regulatory Demands

- As the **Data Owner**, you are faced with escalating and more varied requirements that require you to meet the most stringent data governance standards(e.g. GDPR).
- As a **Business Owner** how to ensure you are operating within the guard rails of your industry (regulation compliancy, data archive & retention), how to guarantee data life cycle for Risk Mitigation
- As a **Business Owner** how to **delineate data** that needs transaction control, and data that no longer changes

Accelerating and Unchecked Data Growth

- As an **Administrator** & **Developer** you are seeing accelerated internal and customer utilization & engagement across the Salesforce platform functions. This is results in **order of magnitude** record **growth** across numerous **standard** and **custom** objects.
- As an **Administrator** & **Developer** you have to deal with longer timeframes for creation of full copy sandbox
- As **Business owner** you need to be able to maintain visibility to (archived) data

Business Critical Reports are Timing Out

- As **Business Owner**, at the end of **each quarter** you want **precise pulse** of the business. Salesforce records activity is growing rapidly and you are having difficulty accessing reports.
- As an **Administrator & Developer** you are seeing that too must historical data '**clogs the system**' and can cause **business issues**

Archiving

Data Life Cycle

Create – Data is created

Active – Used frequently and for transaction

Inactive – Low to Moderate Use

- Irrelevant transactional data
- Deactivated user

Archive – Data not needed for transactions and rarely/never modified

Logical Delete – Lifecycle continues until physically deleted

Physical Delete (Destroyed) – Data is physically removed and no longer available.

Archiving

LDV-Implications

Slower Query Performance
- User Experience, Reporting & Dashboard
- Search, List Views, Lookups

Data Replication Process takes longer and is more complex
- Full Copy Sandbox creation & refresh
- Non-Replicatable objects need special attention

Storage
- Inefficient Usage and increased Cost

Visibility Evaluation is more expensive
- Every record accessed requires join across 3 tables
- Sharing Rules and Role based Sharing

Risk of exceeding Governor Limits
- Too many query rows

Salesforce Operational Tasks take longer

Archiving solution

Criteria

Data Volume - Storage Capacity & Growth Rate

Audit and Compliance Needs

Data Access Requirements

Security Constraints – Data at Rest, Residency & Authorization

Data Purge Policy

Influencing factors

Archiving solution

Considerations

What objects need to be archived and what relationships do those objects have?

Do I need to see archived records in Salesforce?

When do I want to archive and which records should I archive?

- On demand vs scheduled
- Configured Set vs Inputted Values

Do I need to unarchive? When and how should I unarchive?

Do I need to auto archive/unarchive related objects?

How do I maintain relationships between live and archived records?

What is the impact of deleting data older than XX (90) days

- Trending reporting across months/years
- Customer history / Customer 360

Archiving solution

Best practices

1. Understand your data growth trends
2. Determine your success criteria
    1. Response time, Maximize current storage usage, Regular purging of data and/or providing read-only access to certain data sets, Reduce time required for routine database maintenance, backup, and disaster recovery processes
3. Establish a data retention policy
    1. Oppty-Stage open-no activities in the past year
    2. Account-Doesn't meet data quality standards-more than 6 months
4. Customize the business rules as needed
    1. Set up data quality dashboards for key objects
    - Data Quality Dashboards (https://appexchange.salesforce.com/listingDetail?listingId=a0N300000016cshEAA)
    - Percent Complete Dashboards (https://appexchange.salesforce.com/listingDetail?listingId=a0N300000016czDEAQ)
5. Test the archiving activity
    1. Run reports by object to see the number of records scheduled to be archived
6. Create user access policies
7. Ensure restoration
    1. Export data to be archived by object into .csv files including record IDs for any objects related to that object
    2. Ensure you continue to perform a weekly backup of your entire database in case of an issue where you may need to restore part of all of your archived data

Archiving Options

#	Option	Pros	Cons
1	**All on the Platform**	• On the platform simplifies access to archived data • Simple solution to implement	• Doesn't reduce storage usage • Needs to implement / refactor custom VF pages to query the data • Needs to implement code to move records to archived object
2	**Data Warehouse Architecture**	• Leverages existing investments in DW & ETL • Reduces Salesforce storage utilization	• Adds operational complexity (i. e. CRM team has to work with DW team) • Needs to implement ETL to move records to DW • Needs to implement data access from Salesforce to DW
3	**Tiered Data Architecture**	• Simplifies access to archived data • Reduces Salesforce storage utilization • Leverages existing investments in DW & ETL	• Needs to implement / refactor custom VF pages to query archived data on Salesforce • Needs to implement ETL to move records to DW • Needs to implement code to move records to archived object • Needs to implement data access from Salesforce to DW • Adds operational complexity (i.e. CRM team has to work with DW team)

Archiving solution

All On platform option

- The system of records (Salesforce)
- The system of archived records (Salesforce)

Pros	Cons
• On the platform simplifies access to archived data • Simple solution to implement	• Doesn't reduce storage usage • Needs to implement / refactor custom VF pages to query the data • Needs to implement code to move records to archived object

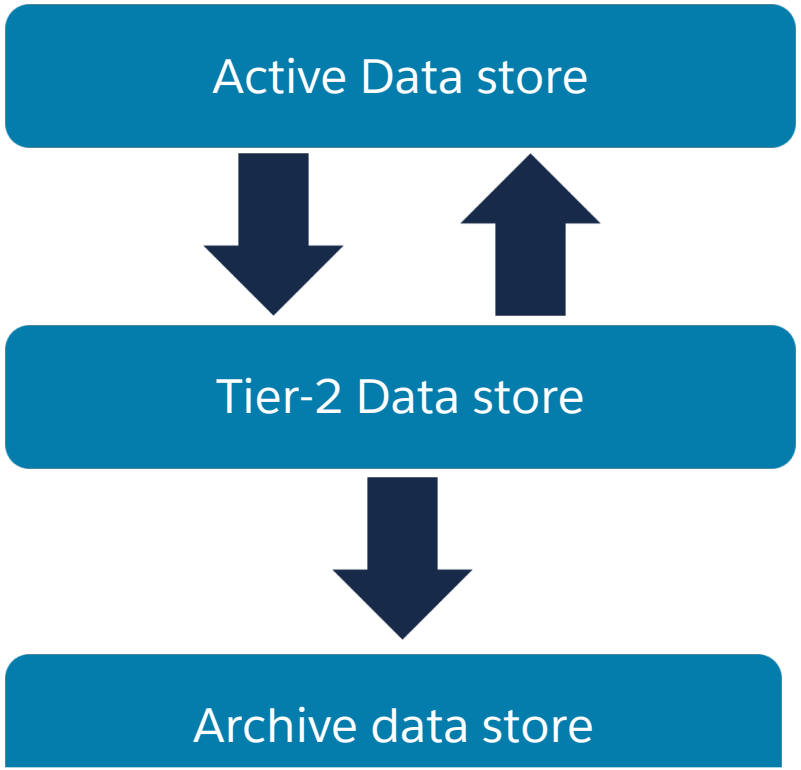
Archiving solution

Datawarehousing solution

- The system of records (Salesforce)
- Data is no longer needed by business but required to be stored for compliance reasons or analytics
- May be allows for on demand query/retrieve back into active store.

Pros	Cons
• Leverages existing investments in DW & ETL • Reduces Salesforce storage utilization	• Adds operational complexity (i.e. CRM team has to work with DW team) • Needs to implement ETL to move records to DW • Needs to implement data access from Salesforce to DW

Archiving solution



Tiered data architecture

- The system of records (Salesforce)
- Serves as an secondary store. Allows for on demand query/retrieve back into active store.
- Data is no longer needed by business but required to be stored for compliance reasons or analytics

Pros	Cons
• Simplifies access to archived data • Reduces Salesforce storage utilization • Leverages existing investments in DW & ETL	• Needs to implement / refactor custom VF pages to query archived data on Salesforce • Needs to implement ETL to move records to DW • Needs to implement code to move records to archived object • Needs to implement data access from Salesforce to DW • Adds operational complexity (i.e. CRM team has to work with DW team)

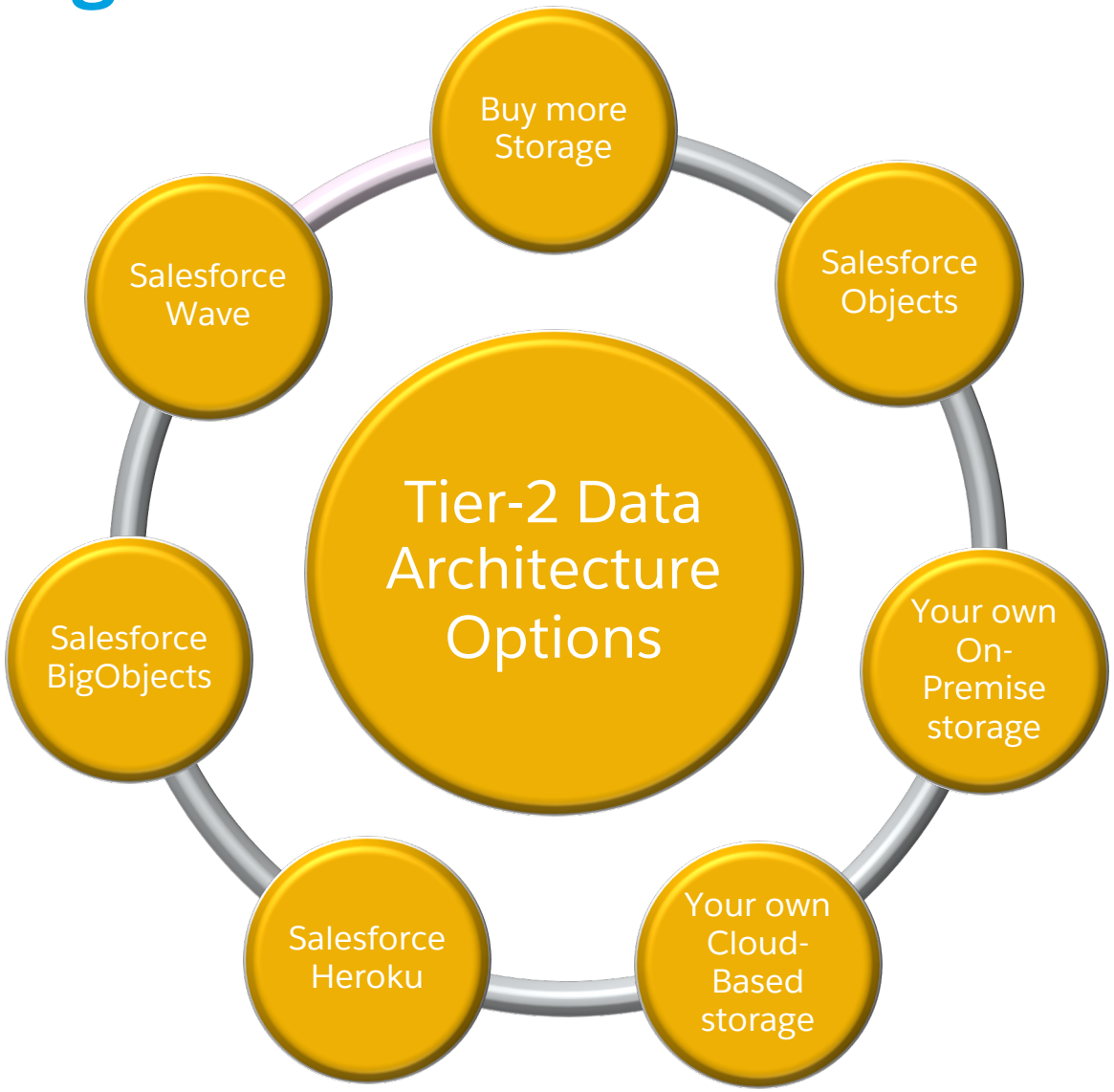
Archiving solution

Tiered data characterstics

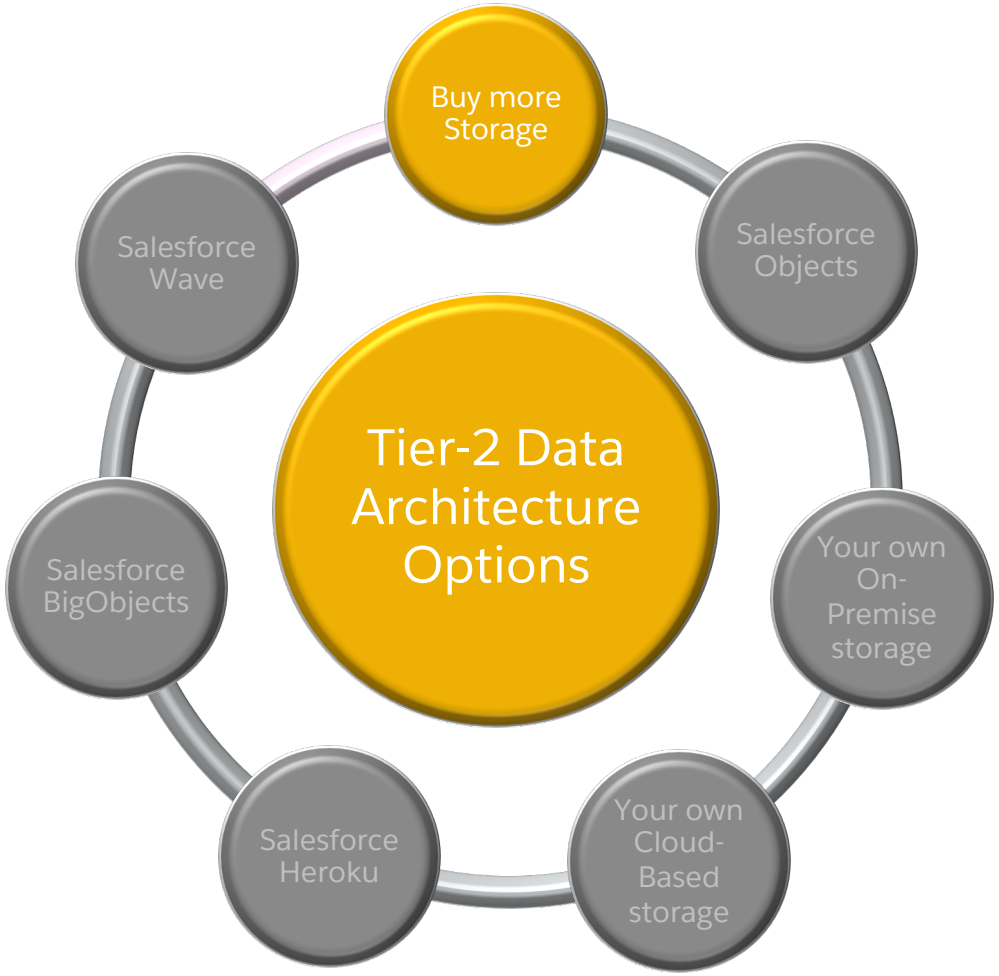
Tier	Description	Storage	Viewable	Editable	Search	Reports
Tier 1	Highly **searched and reported** on data set;	Salesforce	Yes (Standard User Interface)	Yes (if granted access)	Yes (**sidebar and advanced search**)	Yes (**Standard Reports & Dashboard)**
Tier 2	Historical data that is **not** typically searched for or reported on	Salesforce	Yes (Standard User Interface)	No	Yes [TBC by Customer] (sidebar and advanced search, excluded in embedded SOQL)	Yes [TBC by Customer] (but using ***pre-defined pages and filters*** to optimize performance)
Tier 3	Data that **requires mass batch replication** on a frequent basis	Customer Data Warehouse Repository	Yes (**mash-ups if desired)**	No	No	No

Archiving solution

Tier 2 options



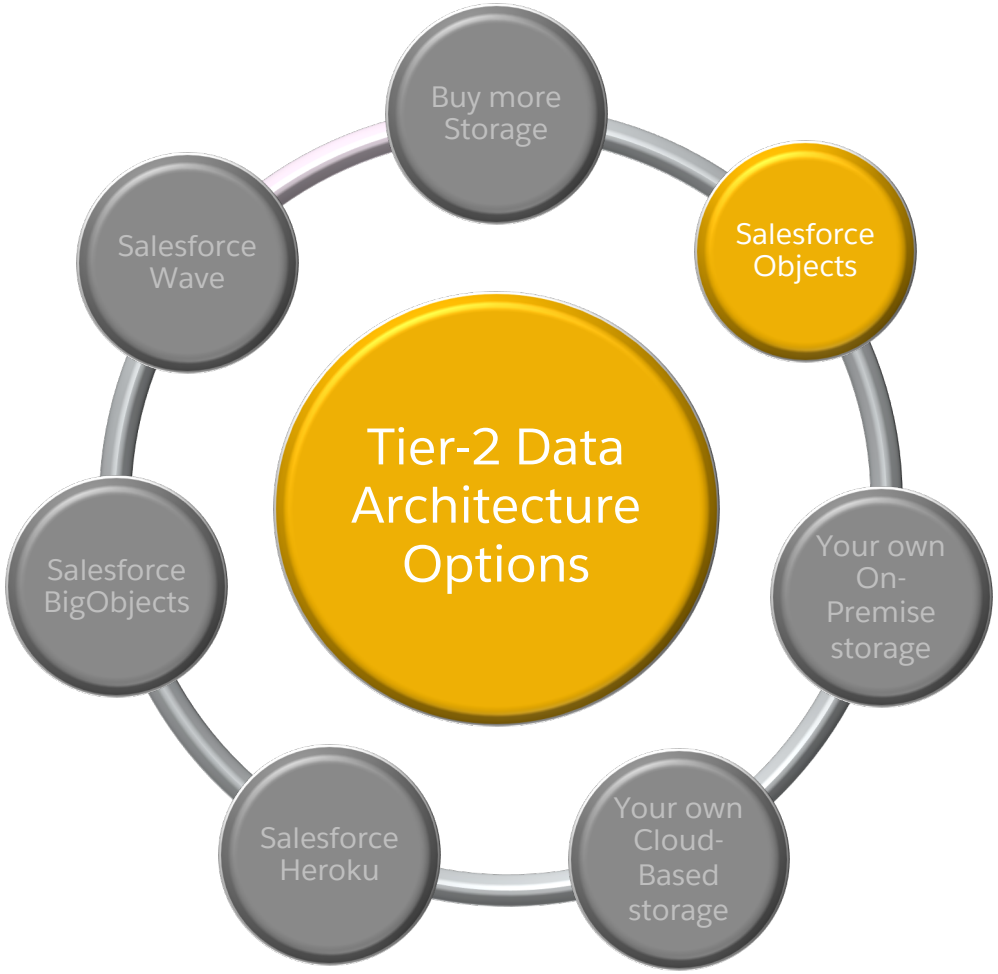
Archiving solution



Tier 2 options

Pros	Cons
• Data has full functionality of salesforce, accessed every day by users, edited very frequently, searchable, workflowable, validation rules, reportable • Highly available, DR / site switching, Security, ISO 27001, SOC reports etc.	• Can be more costly - essentially buying more storage in salesforce. • Scaling from 10 million records per table can add complexity to design or cause platform features to operate differently (e.g. reporting, sharing model, API calls).

Archiving solution



Tier 2 options

Storage Object

- Create Custom Storage Objects to store Archived Records
- Move Archival Records to Storage objects

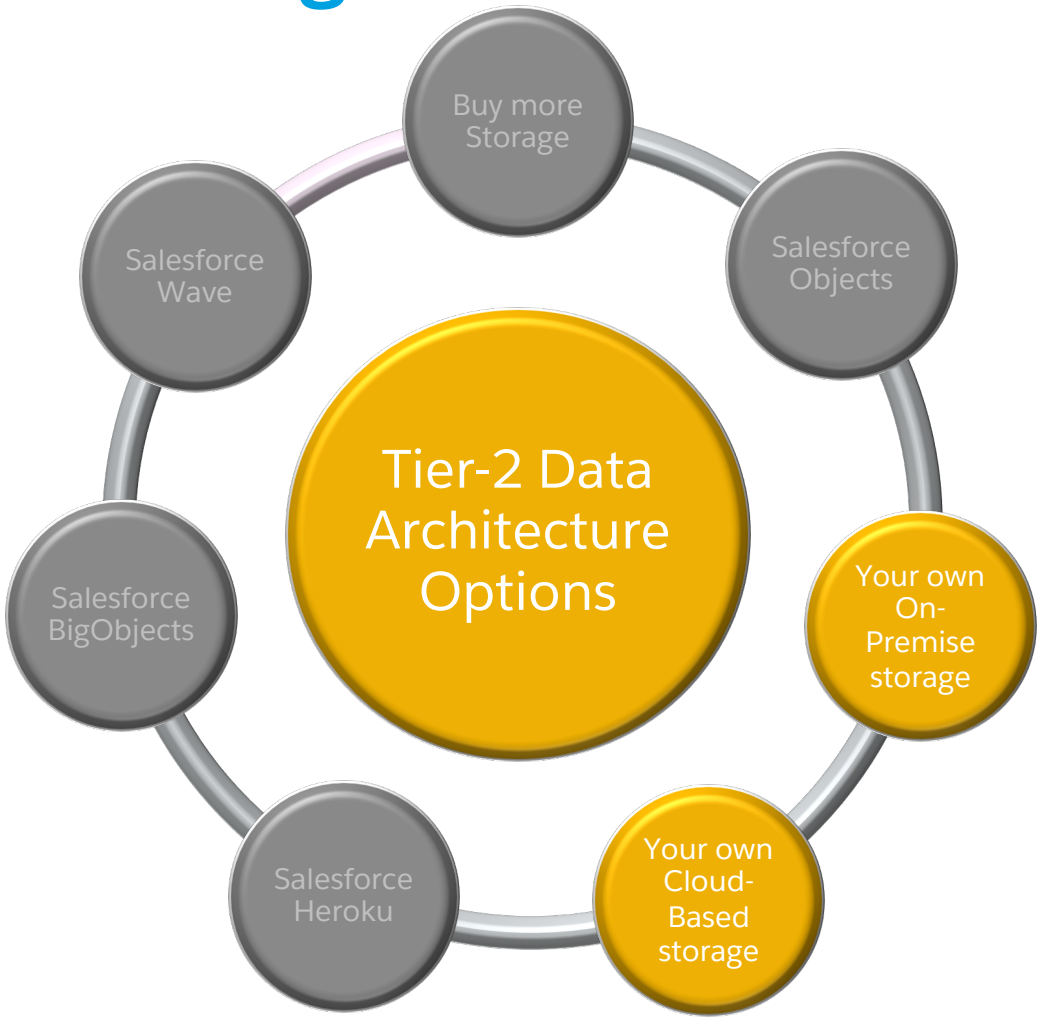
Pros	Cons
• Data in Salesforce & can be part of Salesforce Reports/Dashboards • No changes to existing Salesforce Artifacts (SOQL, Reports) • Applicable to any Salesforce Objects including System Objects	• Storage Usage and Cost issue remains • Custom Solution required to move data to Storage Object

Archive Indicator

- Archive indicator flag added to Object
- SOQL, batch jobs, Reports & Dashboards enhanced to use archive flag included in selection criteria

Pros	Cons
• Data in Salesforce & can be part of Salesforce Reports/Dashboards • Low Complexity and Highly Flexible	• Storage Usage and Cost issue remains • Cannot be used for non-customizable objects • Requires changes to Salesforce solution – Queries, Reports, etc

Archiving solution



Tier 2 options

On-Premise/Cloud-Based Data Warehouse

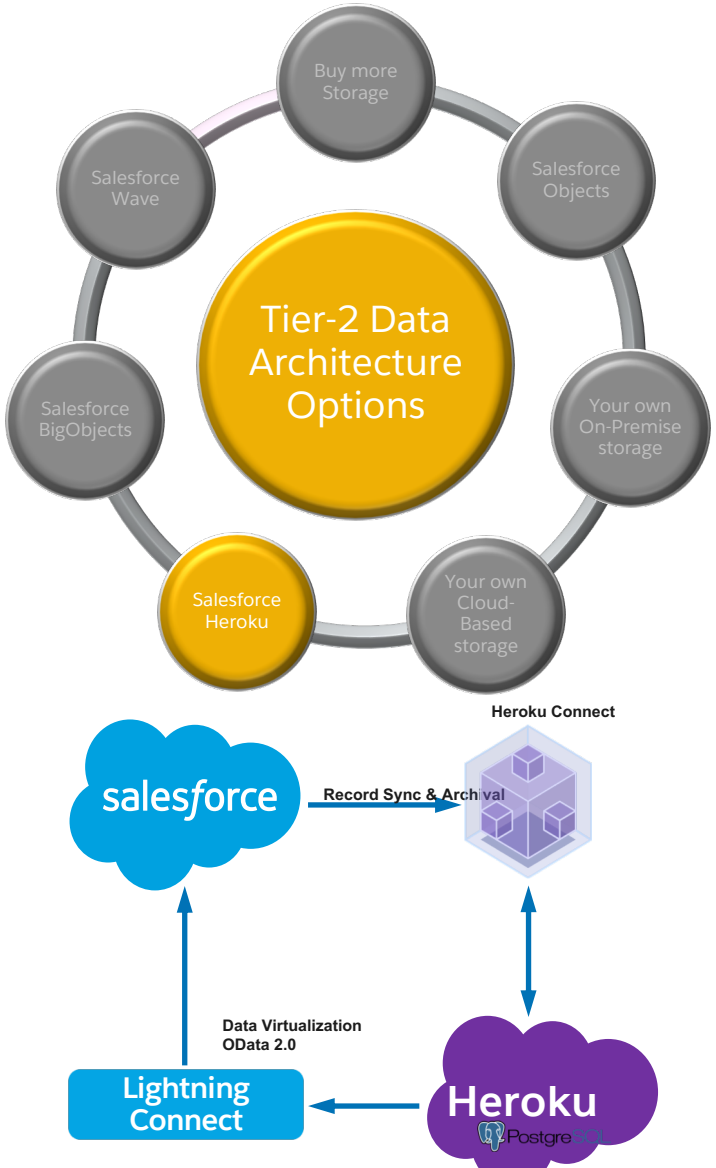
Pros	Cons
• Accessing archival data online via callout or Mash-up • Used by Internal Applications or BI tools for reporting • Can be aligned with Enterprise Archival Strategy • Archive Repository Data model can be customized	• Solution Complexity and Cost are High • Significant development, setup and operations effort • Archived Data access not tightly integrated within Salesforce

On-Premise/Cloud-Based Flat File Repository

Pros	Cons
• Cost and Solution Complexity low • Low Cost storage such as Tapes can be used for File Storage • Supports easy offline access using scripts • No change to Salesforce Architecture	• Accessing Archived data may take extended period of time • Archived Data access not tightly integrated within Salesforce

salesforce

Archiving solution



Tier 2 options

Replicate records to your Salesforce Heroku repository and delete them from Salesforce

Uses Heroku Connect to handle the data archiving & ongoing synchronizations between SFDC & Heroku Postgres

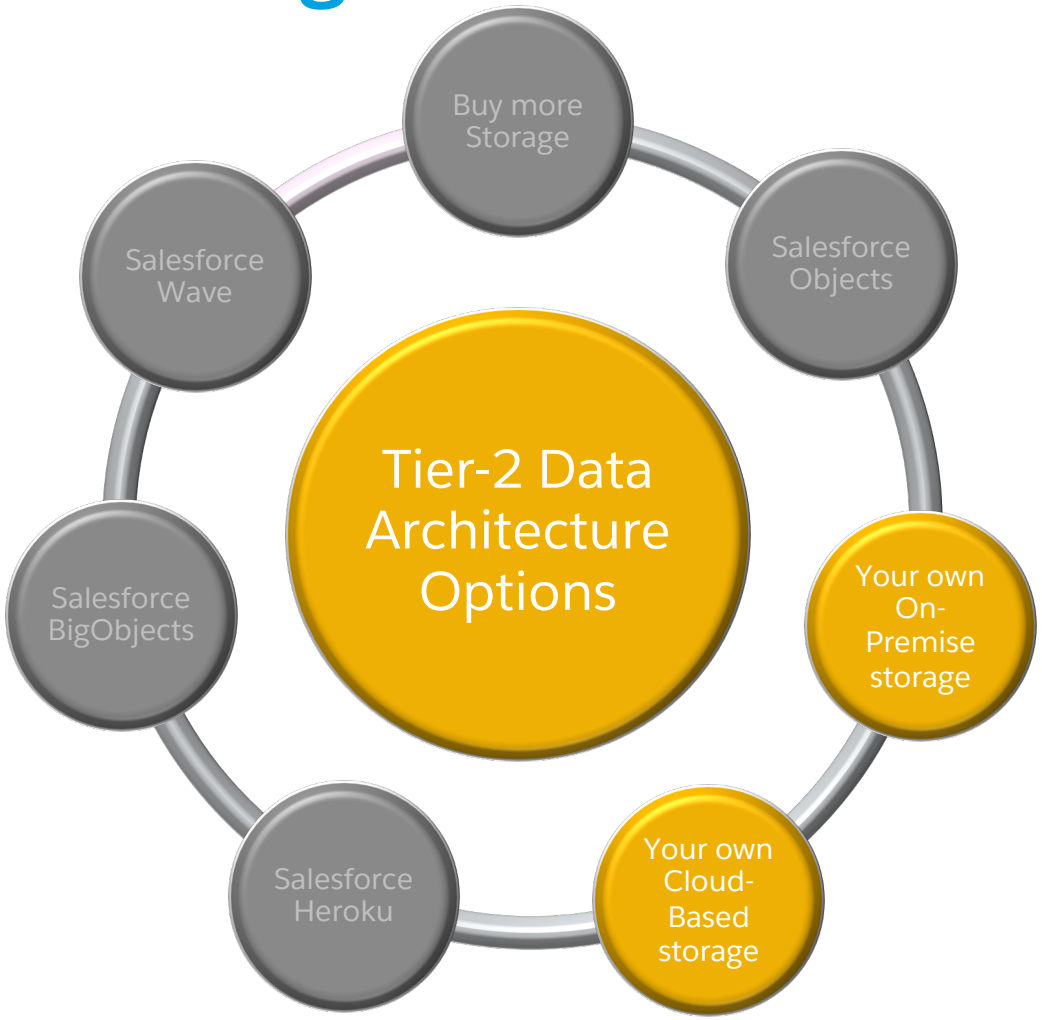
Data Replication & Purge Model

Uses Lightning Connect to virtualizes the archived Heroku data back into Salesforce for viewing & reporting

Pros	Cons
• Potentially more cost effective • Very flexible since custom built • Data can still be used every day, but would typically be edited less frequently (if at all) • Uses open and common standards • These records could be searchable inside of salesforce.com and would appear in the salesforce user interface, apex and visualforce can still be used with more functionality coming on roadmap. • Can comfortably store hundreds of millions of records.	• Customer managed security & availability, e.g. you need a DBA to manage. • The objects would not be able to be used with salesforce workflow or validation rules or reporting (on Roadmap for OData 4.0, Oct 2015 to June 2016) • When reaching a billion records or above may hit scaling issues.

salesforce

Archiving solution



Tier 2 options

On-Premise/Cloud-Based Data Warehouse

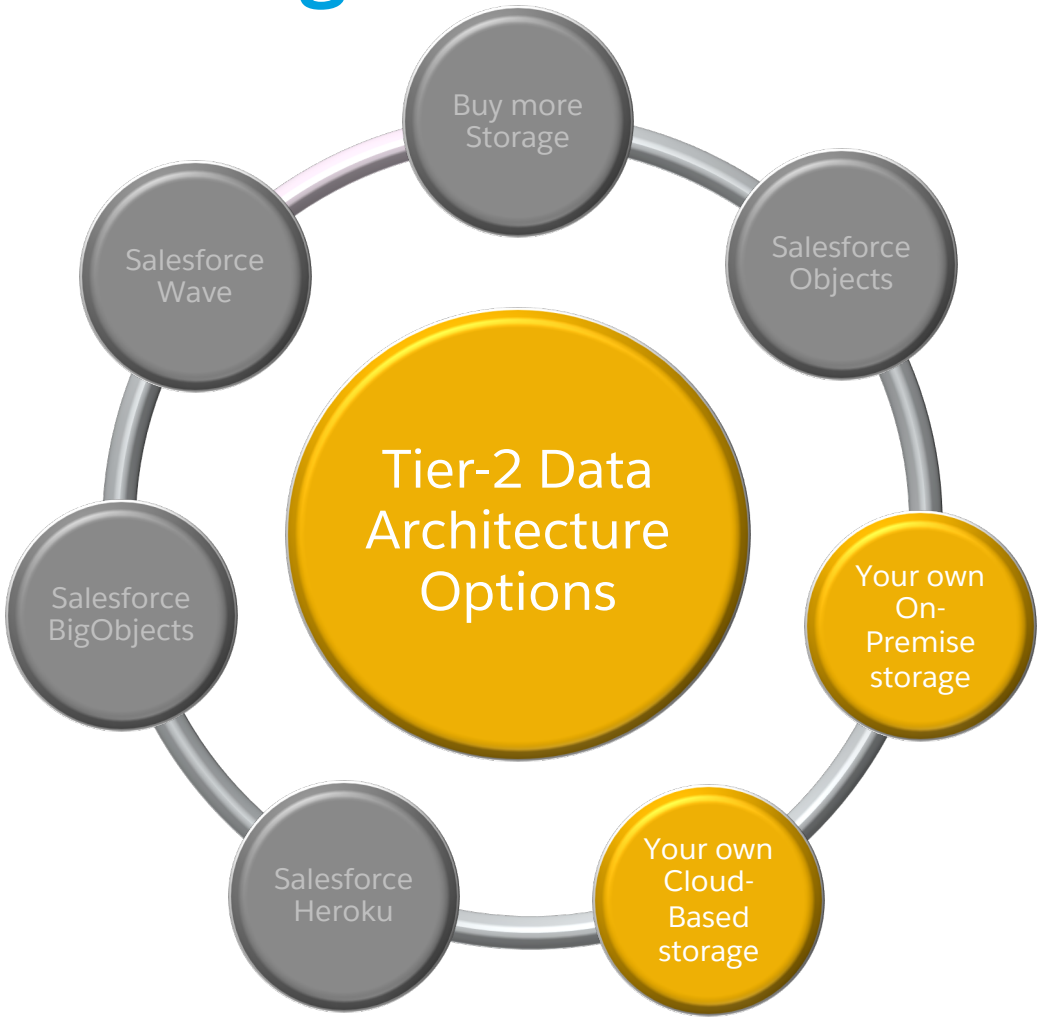
Pros	Cons
• Accessing archival data online via callout or Mash-up • Used by Internal Applications or BI tools for reporting • Can be aligned with Enterprise Archival Strategy • Archive Repository Data model can be customized	• Solution Complexity and Cost are High • Significant development, setup and operations effort • Archived Data access not tightly integrated within Salesforce

On-Premise/Cloud-Based Flat File Repository

Pros	Cons
• Cost and Solution Complexity low • Low Cost storage such as Tapes can be used for File Storage • Supports easy offline access using scripts • No change to Salesforce Architecture	• Accessing Archived data may take extended period of time • Archived Data access not tightly integrated within Salesforce

salesforce

Archiving solution



Tier 2 options

On-Premise/Cloud-Based Data Warehouse

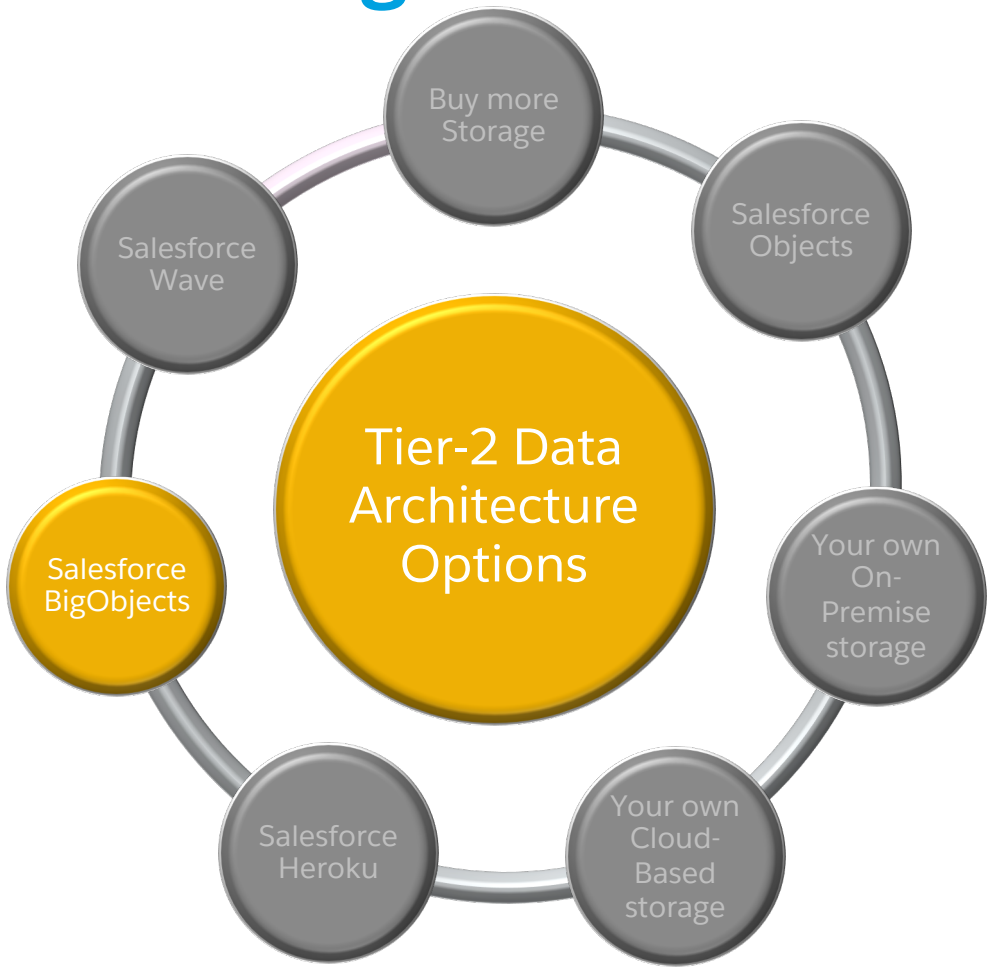
Pros	Cons
• Accessing archival data online via callout or Mash-up • Used by Internal Applications or BI tools for reporting • Can be aligned with Enterprise Archival Strategy • Archive Repository Data model can be customized	• Solution Complexity and Cost are High • Significant development, setup and operations effort • Archived Data access not tightly integrated within Salesforce

On-Premise/Cloud-Based Flat File Repository

Pros	Cons
• Cost and Solution Complexity low • Low Cost storage such as Tapes can be used for File Storage • Supports easy offline access using scripts • No change to Salesforce Architecture	• Accessing Archived data may take extended period of time • Archived Data access not tightly integrated within Salesforce

salesforce

Archiving solution



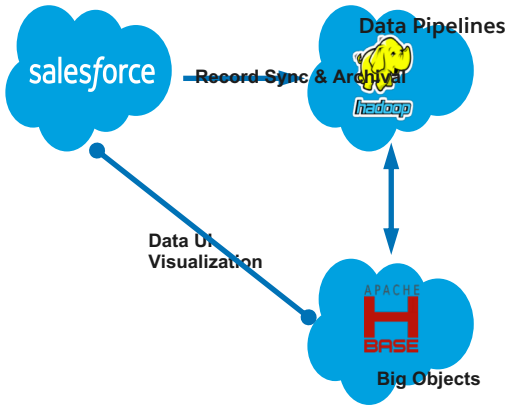
Tier 2 options

Replicate records to BigObjects repository and delete them from Salesforce

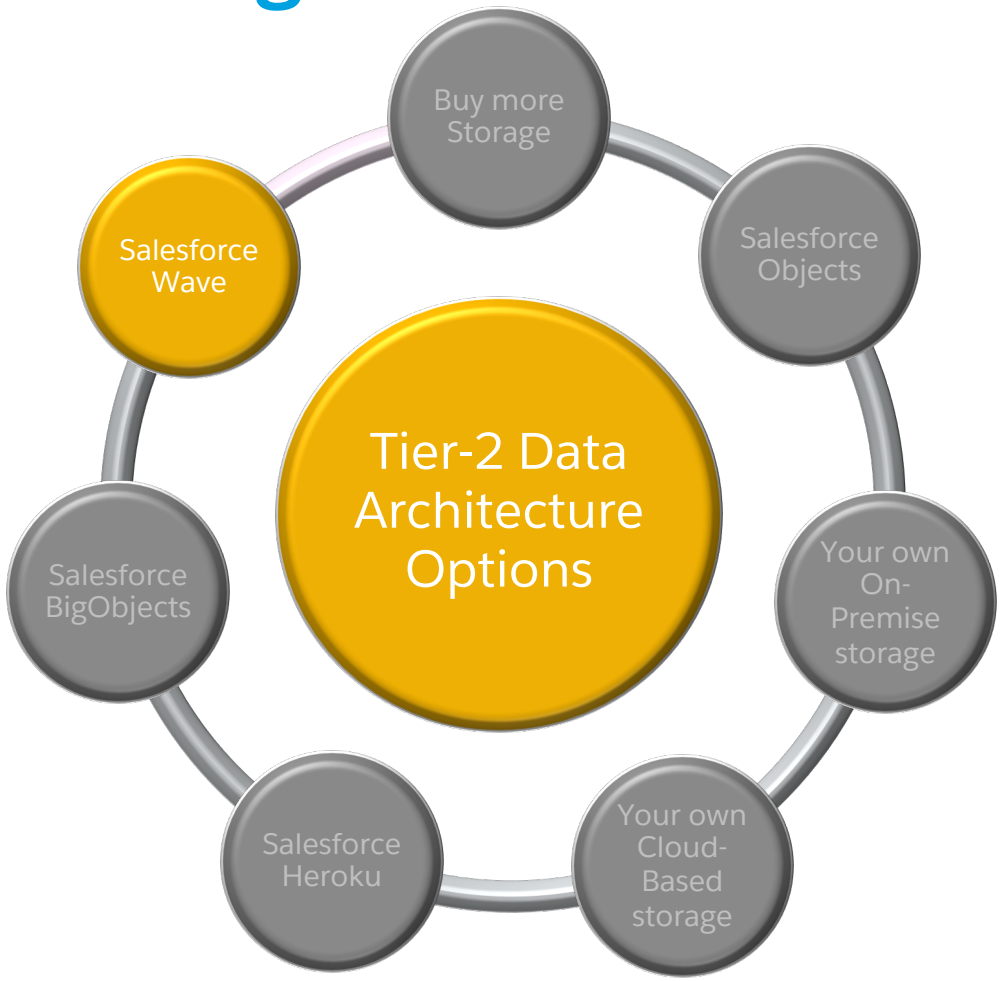
Uses Apex & Data Pipeline to handle the data archiving & ongoing synchronizations to BigObjects

Data Replication & Purge Model

Pros	Cons
• Ability to surface records directly w/Aloha UI • Declarative tooling to define archive and retention policies • API access to archived data • Access management via object-level permissions • Data retention for compliance • Data storage for platform scale and increased storage	• Records stored are immutable (not editable). • Apex triggers, workflow, validation rules reporting, sharing model, etc cannot be used natively. • In pilot and currently not GA.



Archiving solution



Tier 2 options

Replicate records to Wave repository and delete them from Salesforce

Uses Apex, Data Pipeline (or DataFlow) to handle the data archiving & ongoing synchronizations to Wave DataSet

Data Replication & Purge Model

Pros	Cons
• Potentially more cost effective • API access to archived data • Access management • Data retention for compliance • Data storage for platform scale and increased storage	• Maximum of 250 million rows of data stored for all registered datasets combined.

salesforce

Custom Big Objects

Pieces of the puzzle

High-Volume data storage on Platform

- Geared for billions of records
  Geared for Immutable data – perfect for Archiving

Familiar, object based model

- Simple data types – string, number, date, JSON
  Exposed in SOAP, REST, Bulk and Metadata API
  Synchronous / Asynchronous query patterns

High throughput Ingress / Egress

- Second level is eighteen point Salesforce Sans

Separate

- Second level is eighteen point Salesforce Sans

BigObject

APACHE PHOENIX

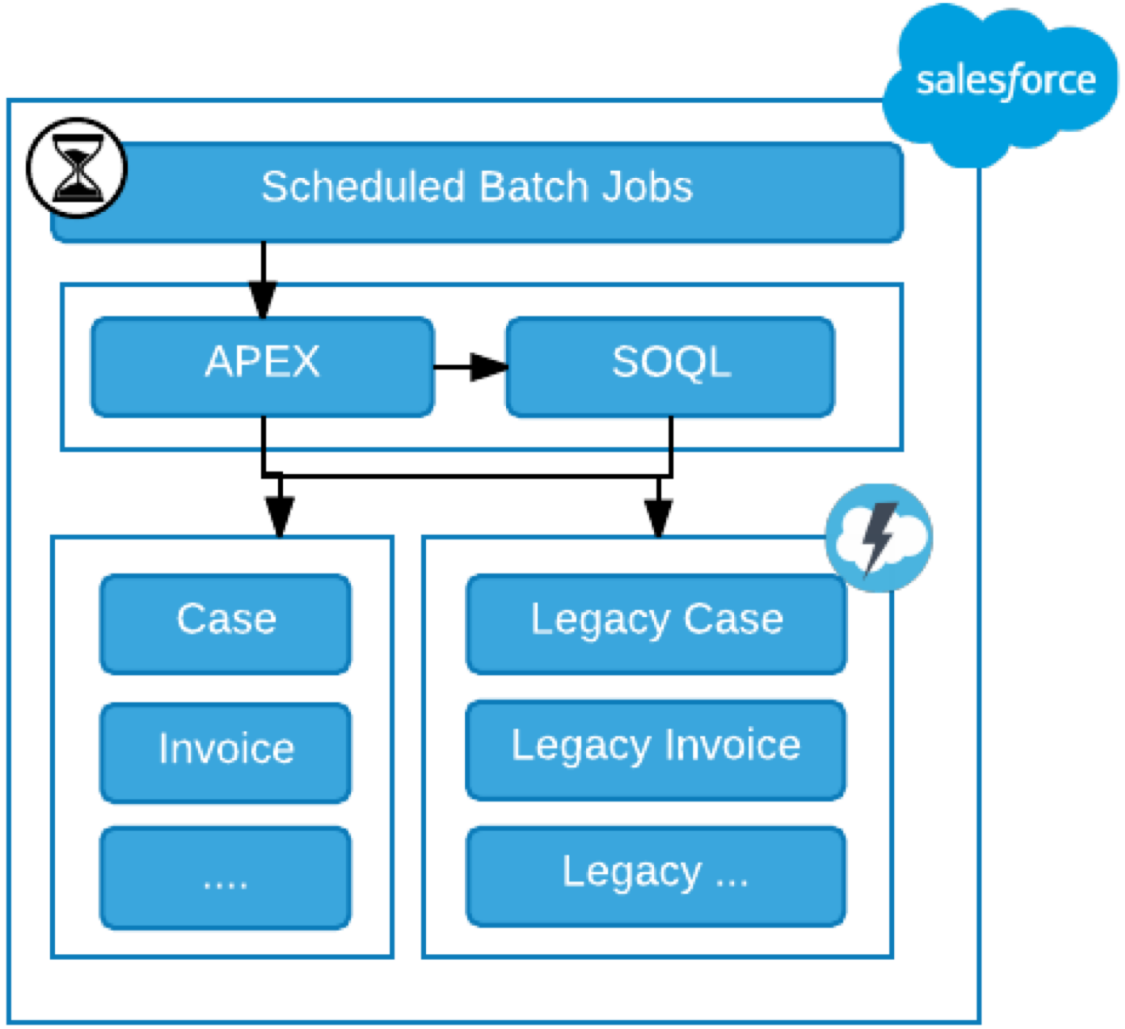
APACHE HBASE

salesforce

Big Objects

Scheduled Batch Jobs (supporting transactions) Periodically archive records to Salesforce Big Objects

Access through familiar tech (Apex, SOQL)



salesforce

Big Objects

Considerations

Consistency: Only Archive Stable Objects

Row Level Security **

It is possible to “re-hydrate” records,
but standard ETL issues will apply

Preservation of Chatter

Features

Traditional Salesforce UI
(Lightning Experience)
(Page Layouts)
(List Views) *
(Lookups / Relationships)
(Reports) *

Accessible for End Users / Search **

* - Not available if High Volume Enabled
(search > 100K records/Hour)

Big Objects

Considerations

Consistency: Only Archive Stable Objects

Row Level Security **

It is possible to “re-hydrate” records,
but standard ETL issues will apply

Preservation of Chatter

Big Objects subqueries can only run against indexed fields (defined at object definition)

Big Objects do not support changes to indices or fields past object definition

Features

Traditional Salesforce UI
(Lightning Experience) *
(Page Layouts)
(List Views) *
(Lookups / Relationships)
(Reports) *

Accessible for End Users / Search **

* - Not available if High Volume Enabled
(search > 100K records/Hour)

Comparison Chart

Two Options

	Archive via Heroku Postgres	Archive via Big Objects
Administration Efforts	Dev assistance to define tables	Dev assistance to define tables
Encryption Support	Manual	Manual
Sharing	Object and field level Permissions	Object and field level Permissions
Administration Efforts	Medium Low Effort Dev assistance to define tables	Low Effort Dev assistance to define tables
Archive Coding	Medium Effort Postgres stored procedure + Cron / Heroku Scheduler	Medium Effort Batch Apex
Compliance	Differing levels depending on Private Spaces, Shield, etc	Core data covered by MSA
Schema Flexibility	More Options	Schema fixed at Object Definition

Backup

Background

The purpose of this document is as follows:

- To propose a backup and restore approach Salesforce Production Data
- To present a long term solution based on 3rd party vendors

Agenda

1. Data Recovery – Overview
2. Operational Backup and Restore Context
3. Native Backup and Restore Options
4. Metadata Backup and Restore Options
5. AppExchange Partners

Data Recovery

Overview

Data Recovery is usually required in two circumstances

1) In case of a full Salesforce **Disaster Recovery** scenario:

Salesforce will provide recovery services, in the highly unlikely event of either an impacting Database or Data Centre failure (http://trust.salesforce.com)

There may however be some additional validation or data correction required in a highly integrated enterprise estate.

Note: Salesforce Business Continuity and DR Plans are not covered in this document.

2) In case of a **Customer Requested Recovery** scenario (for example, customer introduced data corruption):

Salesforce does not currently offer an on demand service to restore data from backup

Customer does not have access to the Salesforce created database backups due to the nature of being a multi-tenant database

We recommend Customers to ensure a data backup and restore process is in place if Customer Requested Recovery is a requirement.

Salesforce Disaster Recovery

Salesforce’s Disaster Recovery Plan

100% Data Centre replication (Network, Storage and Servers)

Near real time data replication

Target recovery objectives:

- Restoration of the Salesforce service within 12 hours after SFDC's declaration of a disaster
- Maximum customer data loss of 4 hours;excluding
  - Disaster or multiple disasters causing the compromise of both data centres at the same time
  - Development and testbed environments, such as the Sandbox service

Continuously working to improve Disaster Recovery processes

Salesforce's Business Continuity Plan

Salesforce has developed a global Business Continuity and Disaster Recovery Program for its Force.com (core) platform and its online services running on the Force.com platform; hired Certified Business Continuity Planners (CBCP) and has retained the services of leading consultants to assist in the on- going development of Business Continuity and Disaster Recovery plans and procedures. This program is overseen by senior management for each of the key functional areas within Salesforce, and is supported by executive leadership at the highest level.

Salesforce has a Crisis Management Team (CMT) comprised of select executives from key departments globally. The CMT is mobilized when a crisis or significant event occurs, and is responsible for evaluating the situation and responding accordingly. Depending on the severity and nature of an incident the CMT Leader may request engagement from various support teams to assist with mitigation of the incident. The CMT meets periodically for training, education, and review of the documented CMT Action Guide, or as required due to a crisis or significant event. CMT members have specified roles and responsibilities and are expected to be available at all times (24/7/365). The CMT conducts table-top exercises, at minimum, once per annum.

Salesforce maintains a Mirror Site that is 100% staged warm site with block-level data replication. The secondary data center is replicated at 100% of capacity (host, network, and storage) of the Production data center.

As a part of developing a viable Disaster Recovery plan and program for the production environment and platforms, Salesforce schedules Disaster Recovery exercises which are conducted at least annually. Overall Salesforce has conducted more than 40 Disaster Recovery exercises in the production environment.

Additionally, disaster communication processes are exercised using the mass notification system during each exercise, which includes call-outs with response requests to Salesforce Crisis Management Team and the production Disaster Recovery teams. Salesforce will test its disaster recovery plan at minimum on an annual basis and will continue to enhance and develop processes and its technology related to disaster recovery to further reduce RPOs and RTOs.

Salesforce has developed additional procedures, processes and plans, including a Pandemic plan

Customer Recovery Plan

Requested Recovery may fall under other scenarios

Integration errors

An integration interface has malfunctioned and corrupted data

An error in an integrated system passes invalid data over an interface

Data Migration / Mass Update Errors

Errors during data migration activities

During a mass update of existing records using Data loader or similar tool

Human error

Accidental deletion of records

Malicious user actions

Rogue system administrator or other user in Salesforce

Configuration Deployment

Removing field from object definition

Changing field data type

Note: this is not intended as an exhaustive list.

The following can also irreversibly modify data

Salesforce UI

Changing field data type

Using Import Wizard

Mass Transfer Records

Mass Delete Records

Mass Reassign Account Teams

Mass Reassign Opportunity Teams

Mass Update Addresses

Salesforce API

Performing an update

Performing a mass delete or physical delete

Operational Backup Restore

No Rollback of an Org possible

IMPORTANT

There is no equivalent to a database rollback that can be managed by the organization's system administrator.

A full data backup of a Salesforce org must be processed as a data migration in order to restore its data. We cannot simply recreate an org from a database dump.

This can be achieved manually by a data export, or by using 3rd party vendor solutions that automate the restore process and make it transparent.

Operational Backup and Restore Context

1. Salesforce recommends keeping a local copy of data or use a partner to avoid the fee based back up and restore.
2. Salesforce publishes a multipart series on back up and restore as well as other best practices for managing data, which should be consulted as part of your implementation.
3. Business records can be soft deleted and go into a recycle bin or hard deleted to bypass the recycle bin as noted in the best practices on deleting data.
4. You cannot delete standard objects.
5. When you delete a custom object the object and the data are soft deleted and remain in the recycle bin for a short period of time (15 days).
6. Rules for master detail relationship apply to deleted records per the attached troubleshooting guide.
7. Based on the system of record, multiple solutions may be required (e.g., email, content, Salesforce records).

Restoring data is a complex task

→It is not possible to “simply” restore Salesforce with a set of files containing data to restore the system to a point in time

→Manual intervention or specific tools are required

→Full automation of the data restore process is **not** possible

→A full large restore might take several days. Restore should target specific goals.

→A full Metadata restore process is **not** possible. A bugfixing approach must be considered instead, targeting specific objects or classes to troubleshoot any issues.

Native Backup and Restore Options

Native Backup and Restore Options

The following options are available to export and restore data from Salesforces through simple configuration:

1. Data Export Service: Manual or scheduled exports of your data via the UI.
2. Data Loader: Manual on-demand exports of your data via the API and the non-programmatic means in importing data.
3. Report Export: Manual on-demand exports of your data via reports

See the multipart series on back up and restore for build options using the Salesforce APIs. AppExchange partners should be considered as an alternative to building a solution.

Metadata Backup and Restore Options

Metadata Backup and Restore Options

Back up and restore options for Metadata may be achieved through one or more of the following approaches:

1. Change Sets: Copy metadata from your production org to a sandbox or developer org.
2. Sandbox Refresh: By refreshing a related sandbox, your configuration metadata is copied over automatically.
3. Force.com Migration Tool: Java/Ant-based command-line utility for moving metadata between a local directory and a Salesforce org. See also: https://developer.salesforce.com/page/Migration_Tool_Guide
4. Force.com IDE : Client application for creating, modifying, and deploying Force.com applications.

An external repository (e.g., Git) should be used in concert with the Force.com migration tool and should also be backed up.

Backup and Restore

AppExchange

AppExchange: Backup and Restore Options

1. The Salesforce AppExchange offers several backup solutions that customers could choose based on their requirements, cost/benefit, etc..
2. Selection is based on defined criteria and follow a structured decision process.
3. Salesforce out of the box capabilities offer a reuse before buy option, whereas AppExchange partners offer a buy before build (using APIs) alternative.
4. Odeseva, Backupify, CopyStorm and dataloader.io are often considered as potential solutions, but there are many alternatives that could meet the client's specific requirements.
5. Based on the system of record, multiple solutions may be required (e.g., email, content, Salesforce records).

Decision Criteria

Which approach should I adopt

Decision Points

1. What Backup Interval is needed ? Daily, Weekly, Monthly …
2. Which Objects should be backed up ? Account, Contact, Cases, Attachments …
3. What Recovery Interval is required ? Immediate ? Can wait xxx days ?
4. Data Residence : Cloud or On Premise ?

What does Salesforce offer if I need to recover data ?

- Scheduled Weekly Exports (all or specific objects)
- Salesforce Data Recovery Service – costs ca. 10000 USD for the data to be extracted for a specific point in time. Requires that the customer re-import the missing data records themselves (e.g. Usually this is a new project just to identify which records to re-import and then the effort to load the data)

Salesforce DR is a Last Resort

While Salesforce maintains backups of data for disaster recovery scenarios and can recover data, it's important to back up your own Salesforce data locally in order to be able to restore it.

Doing so will help you avoid recovery fees and allow you to quickly recover your data whenever needed. The Salesforce Data Recovery Service is a paid for service and should be considered an option of last resort.

- Salesforce can provide a customer their Data in the form of csv files
- This does not include Metadata
- A customer is then responsible to restore the required elements to their Org
- Process takes minimum 15 business days to provide csv files, cost is upward of US$10K
- Files will be extracted from a Salesforce database backup
- Recommended as the option of Last Resort only

Preparing for Data Recovery scenarios

Select a 3rd party application as the long term solution

Create a well defined and documented recovery plan

- Across the entire enterprise estate
- Variants per Customer Recovery Scenario
- Roles, Responsibilities, Approvals, etc
- Know the parameters that you are working within, for example:
  - Business expects to be operational within 1 business day
  - Business agrees to xyz data loss

Test the recovery plan

- Use sandbox environment(s), possibly trial org to restore into
- Test on a regular basis

Review and update plan and tooling

- For each major (Spring, Summer, Winter) Salesforce release
- For each major Customer Release
- To ensure new objects, features, functionalities are included

AppExchange

Copy Storm – On Premise Solution

Recommended solutions

Daily Data Backup

Use **CopyStorm**'s command line capabilities to schedule a daily batch process Mon-Fri on the Windows server, so that there's always a Full Backup with the latest data available.

This will leverage CopyStorm's incremental approach, running in minutes.

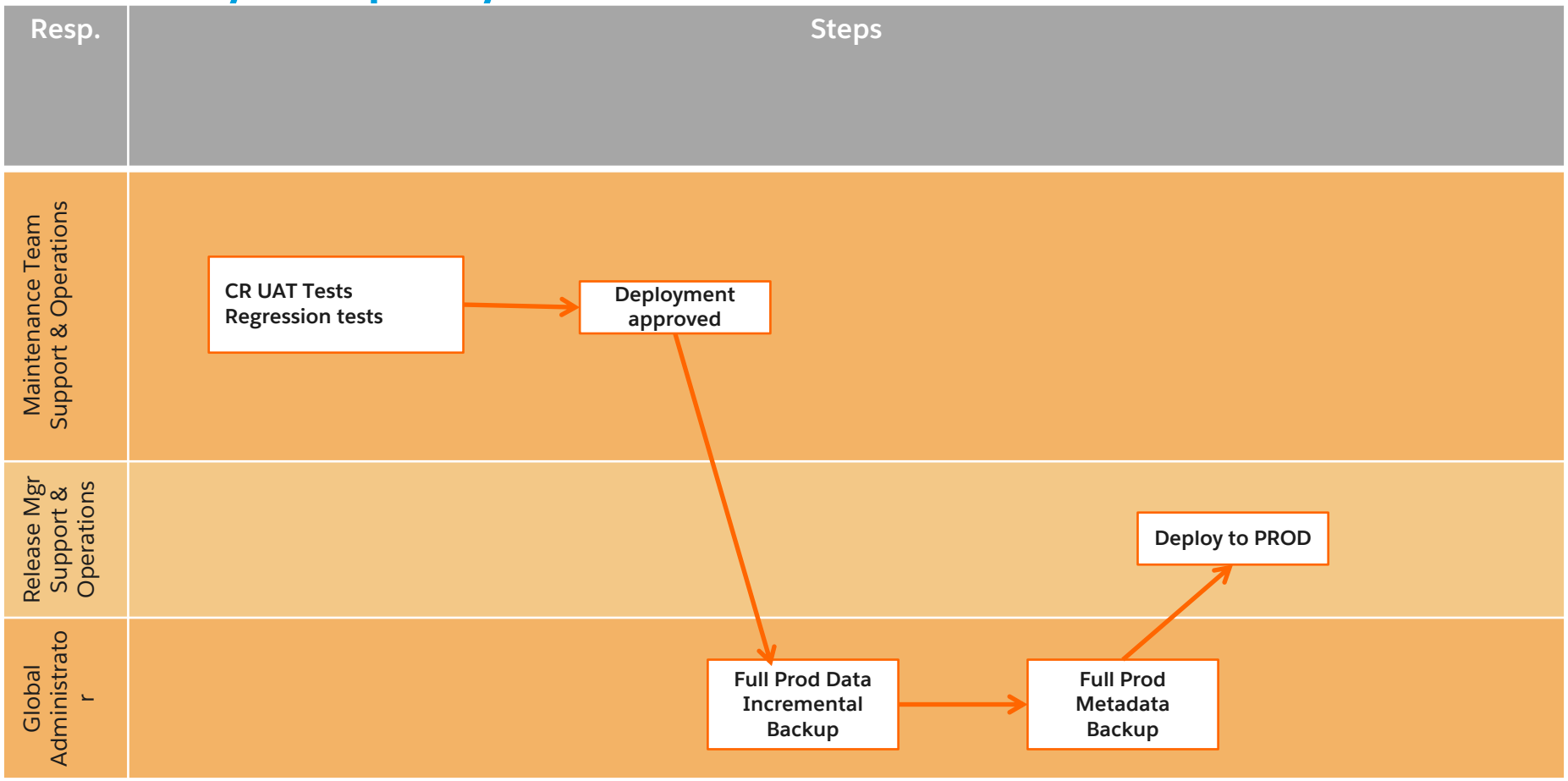
Release Full Metadata Backup

This requires a manual step of metadata discovery first, in order to generate the full package. Alternatively, 3rd party tools such as **Copado** could automate this task and even tell the differences with previous backups.

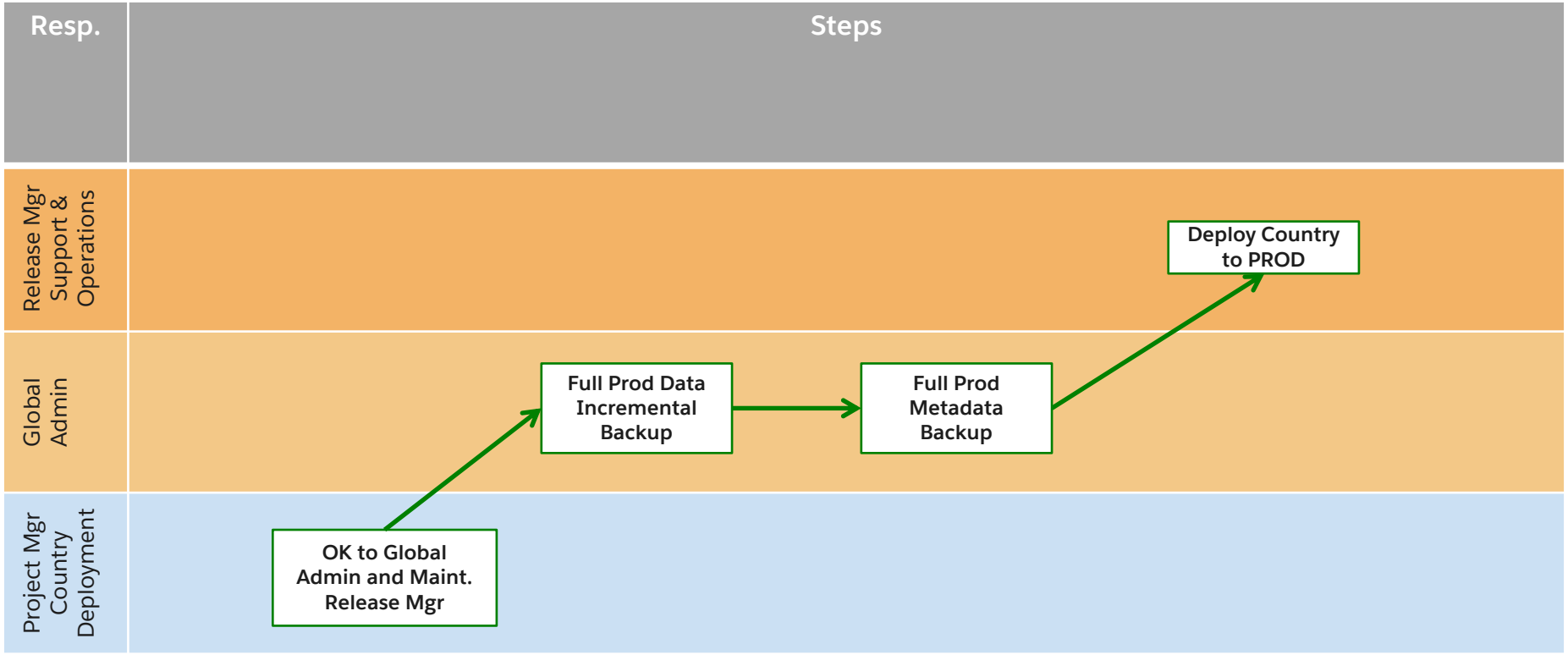
Customer Example

Backup Scenario

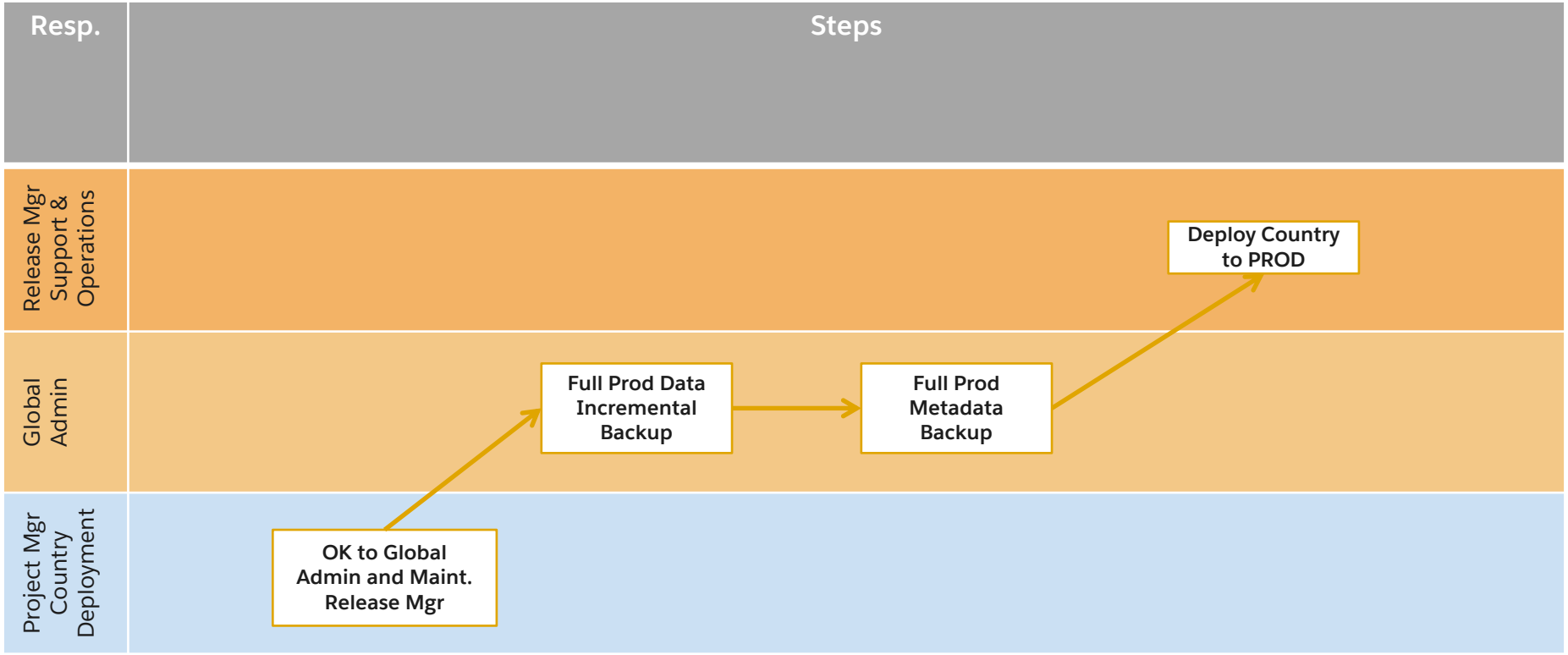
CR Bi-Weekly Deployment



Country Deployment



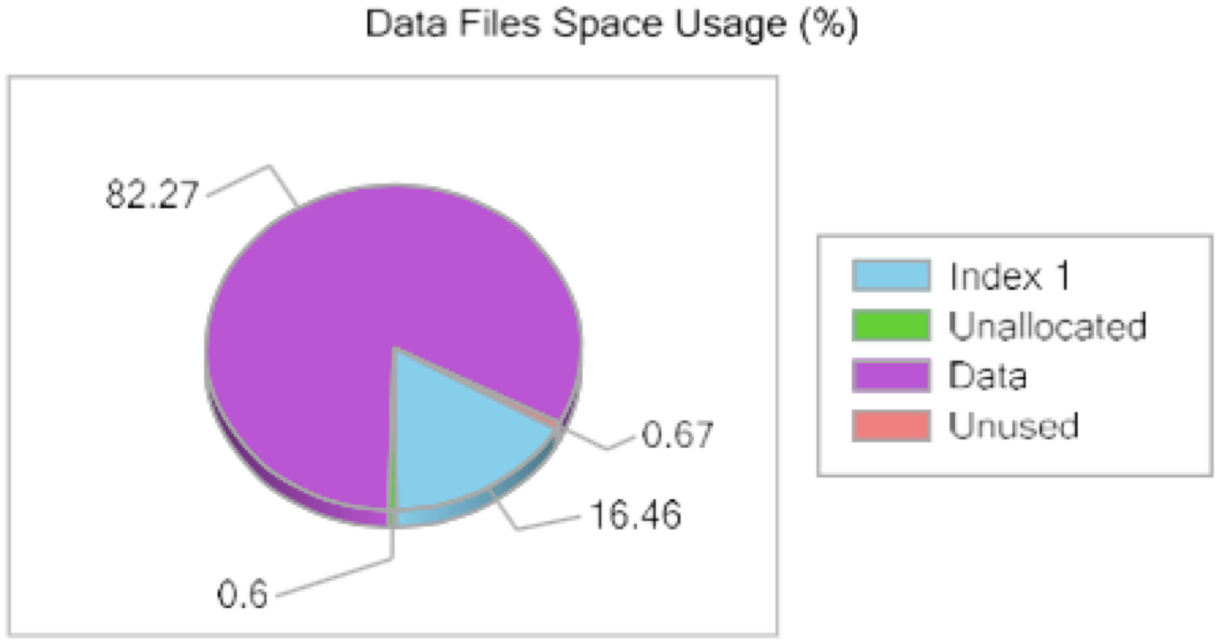
New Release


Resp.
Steps
Release Mgr Support & Operations
Global Admin
Project Mgr Country Deployment
Deploy Country to PROD
Full Prod Data Incremental Backup
Full Prod Metadata Backup
OK to Global Admin and Maint. Release Mgr

CopyStorm Prod Full Backup data storage

A Full Prod Backup takes 2.93 Gb of data (incl. chatter and history data).

It also includes 200 Mb of SQL Server indexes



CopyStorm Prod Full Backup Top Tables

Table Name	# Records	Reserved (KB)	Data (KB)	Indexes (KB)	Unused (KB)
dbo.LoginHistory	1,023,557	537,880	456,712	80,896	272
dbo.AccountHistory	401,396	153,104	82,440	70,192	472
dbo.ContactShare	194,361	77,616	35,800	41,552	264
dbo.PricebookEntry	191,472	100,792	51,512	49,080	200
dbo.AccountShare	126,404	55,728	28,648	26,736	344
dbo.Event	100,354	177,760	137,400	37,560	2,800
dbo.Account	81,679	230,720	190,080	38,744	1,896
dbo.ContactHistory	59,096	19,752	9,472	9,984	296
dbo.Visit_Report__Share	54,515	20,656	9,088	11,248	320
dbo.ContactFeed	53,485	124,352	110,424	13,520	408
dbo.Contact	52,413	114,392	88,760	24,520	1,112
dbo.Visit_Report__Feed	26,804	22,784	15,512	6,760	512
dbo.Visit_Report__c	26,499	46,856	38,152	8,008	696
dbo.Visit_Attendee__Feed	24,073	12,792	6,184	6,160	448
dbo.Visit_Attendee__c	23,988	13,816	7,296	6,056	464
dbo.FieldPermissions	13,084	4,248	2,768	1,336	144
dbo.CustomObjectUserLicenseMetrics	8,462	2,648	1,808	704	136

BACKUP CONSIDERATIONS

Standard Backup offer Salesforce

Standard Backup: Weekly Export through Salesforce UI

- Creates scheduled zip files with the selected objects data and send an email with a link to download the zip file after creation
- Data only – does not include metadata
- 1 CSV file for each object that can be used with dataloader or similar tool to in/upsert into another (or same) org, or to load into a local database.
- Weekly export files should be stored in a dedicated Source Control Repository

Sandbox Copy

Full Sandbox (Data & MetaData) or Config only (Meta data)

- Only from production to sandbox, erases the destination.
- Record IDs from production are preserved
- Can copy from 0 to 180 days of Object/Case/Opportunity history, in 30-day increments.
- Exact copy for full sandbox , except email addresses (sandbox name suffixed)
- Full sandbox refresh interval is every 29 days

Main technical considerations

Data restoration approach
- Restore data from backup csv files
- Restore from on-premises or cloud database
- Fix forward
- Simple record restore from Recycle Bin
- Data refresh from Master Data records via Interface

Full or partial data restore
- One/Some/AllObjects
- One/Some/AllRecords

What special considerations are needed to handle
- Encrypted Data
- Integrated Objects/Fields vs Non-Integrated Objects/Field

Integrated Systems
- Is there any data held in any integrated systems that hold/link to Salesforce record IDs?
- AppExchange apps with local storage
- Data Warehouse or ERP

Salesforce Restore Complexity

Restoring Salesforce differs in several ways to a conventional relational database restore:

- Cannot set SFDC ID during insertion of data. If data record is lost, there is no way to insert a record with the same Salesforce ID as the record that was deleted. Salesforce generates its own primary keys and those in backup files cannot be used during restores (for inserts).
- This in turn means that foreign key references need to be adjusted to match the new generated primary keys. Need to re-map relationships using new record IDs
- Replacing or updating prior record IDs with new record IDs will need to be done externally or, alternatively, consider creating a custom External Id field to store prior record IDs for use in relating records to one another upon re-import.
- To preserve old/original create dates, request Salesforce Premier Support via a case to enable Create Audit Fields before restoring
- The interaction with Salesforce for restoring data takes place at the functional/logical layer rather than at the data layer. Salesforce automations and logic (e.g. workflows) on the affected objects will be triggered and may need to be disabled for the restore

Salesforce Restore Limitations

Various API limits to consider when restoring

- Ex: Bulk API supports up to 50M record insert/updates per day

Some data can’t be restored

- Ex : Opportunity history, AutoNumber
- @Mentions in Chatter Feeds

Metadata Considerations

- Not all configurations are exposed through Metadata API
- Some point and click settings cannot be backed up. These should be documented step by step for manual recreation
- Metadata API doesn’t allow some changes to supported metadata types

Salesforce Restore Metadata considerations 1 of 2

Apex Classes and Apex Triggers

- By default, changes to Apex code that have Apex jobs pending or in progress can't be deployed. To deploy these changes, do one of the following.
- Cancel Apex jobs before deploying changes to Apex code. Reschedule the jobs after the deployment.
- Enable deployments with Apex jobs in the Salesforce user interface in the Deployment Settings page.

Approval Processes

- To use approval processes on Salesforce Knowledge articles with the Metadata API, the article type must be deployed. For article version (_kav) in approval processes, the supported action types are: Knowledge Action, Email Alert, Field Update, and Outbound Message.
- If the approval process references any post templates that contain custom fields, then you need to resave those post templates in the originating organization before adding them to the change set. From Setup, click Create | Workflow & Approvals | Post Templates. For each post template, click Edit and then Save.
- The metadata doesn't include the order of active approval processes. You may need to reorder the approval processes in the destination organization after deployment.
- If you change the Unique Name of an approval process that was previously included in a change set and deployed in another organization, and you resend the approval process via a change set, a new approval process will be created upon deployment in the other organization. The previously deployed approval process will not be modified.

Custom Fields

- Starting in API version 30.0, when deploying a new custom field, the default values for the editable and readable fields in profile field permissions are false. To override the default values, include field permissions for the new field in your profiles.

Custom Objects

- Using API version 29.0, you can't change the sharingModel of an object using the Metadata API. You must manually make this change to the target organization through the user interface.
- Starting with API version 30.0, you can change the sharingModel of an object for internal users using the Metadata API and the user interface.

Salesforce Restore Metadata considerations 2 of 2

Connected App

- You cannot set the consumerKey in the Metadata API. It is included in a retrieve operation for informational purposes. If you try to move the connected app to another organization, you must remove the consumerKey from the .zip file before the deployment to an organization. A new key will be generated in the destination organization.
- Mobile settings of connected apps are not supported in change sets and must be manually migrated.

Page Layout

- A deployment containing page layout assignments replaces all existing page layout assignments in the destination organization with those specified in the .zip file. Existing page layouts in the organization disappear if they're not included in the .zip file. Always include all page layouts for all required record types in the .zip file.

Profiles

- If a package includes a profile with a name that doesn't exist in the target organization, a new profile is created with that name. If the deployed profile doesn't specify any permissions or settings, the resulting profile consists of all the permissions and settings in the Standard Profile.

Sharing

- Simultaneously updating the sharingModel field for an object and adding a new sharing rule isn't supported in the Metadata API, regardless of which object you're updating. For example, you can add a sharing rule when the organization-wide default is public, and subsequently update the sharingModel. This would result in a single sharing recalculation.
- You might encounter an error if you're deploying a change set with a custom object that has a parent-child relationship without the master/detail field in the same change set. To resolve this error, include the master/detail custom field in the change set, even if you haven't changed the organization-wide default.

Workflow

- Test mode for flow triggers isn't supported in the Metadata API. If you want a flow trigger to run the latest flow version when an administrator causes the workflow rule to fire, enable test mode via the user interface after deployment.

How Attachments Should Be Handled

The backup of attachments will be handled differently to other data due to the following reasons:

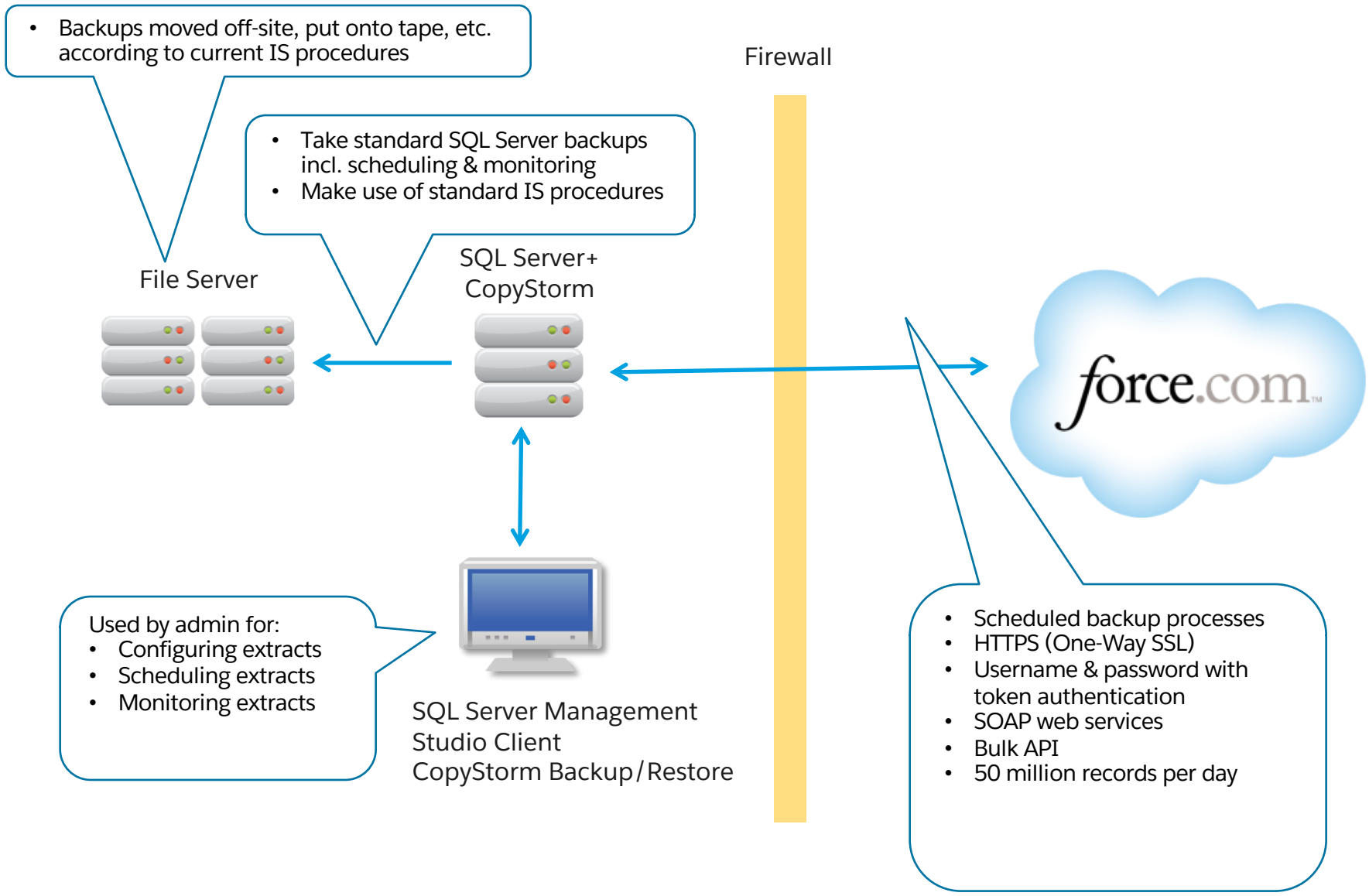
- The Bulk API cannot handle attachments (Base64 binary) due to size implications and every attachment becomes a SOAP call
  - This has a bigger impact on API call limits
- Extracting attachments can take considerable time (as the number of Salesforce attachments increase over several years time).

The following approach should always be used for attachments:

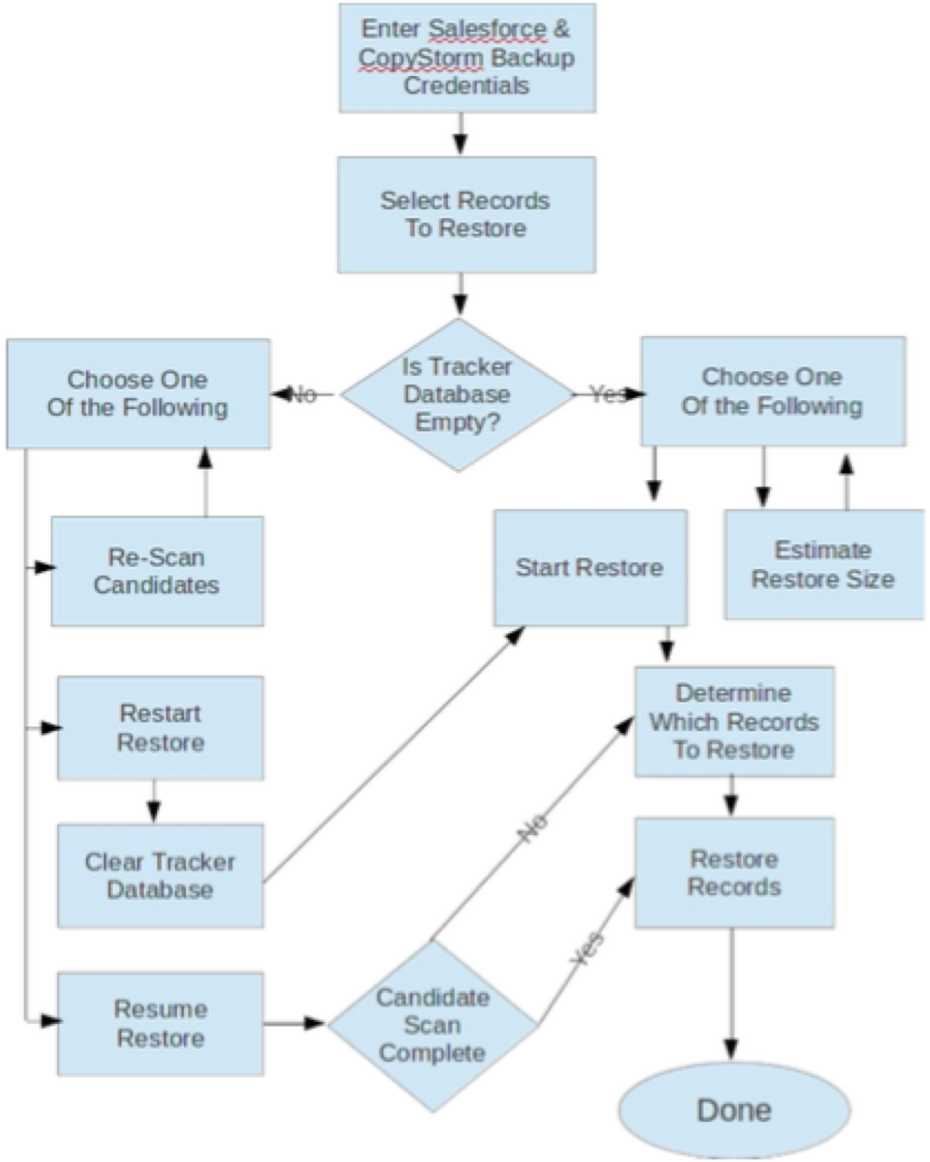
- Only if a new file has been created or an existing file has been updated, will the file be backed up
  - Therefore, for the full weekly back-up, attachments should be INCREMENTALLY backed up
- Maximum file size will be set at 15MB

CURRENT AVAILABLE TOOLSET

CopyStorm Backup Solution Overview

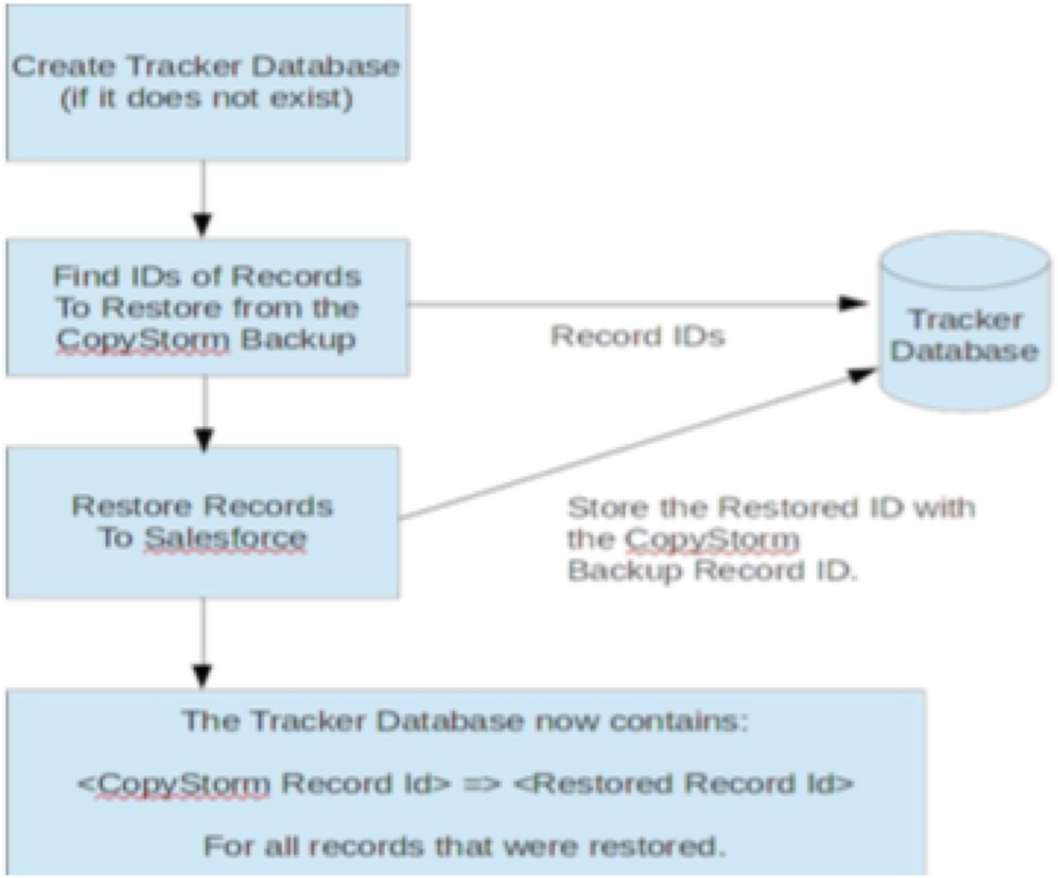


Running a CopyStorm Restore



How CopyStorm handles a Restore

CopyStorm generates a Tracker DB in order to keep track of the IDs of backed up records and map them to new IDs generated by Salesforce when executing a Restore.



How CopyStorm handles restored IDs

The usage of the tables is as follows:

When a table is referenced in the restore set, an entry is made in RestoreTable.

When a record is identified as a candidate for a restore, an entry is made in RestoreTableRecord.

When a record is a reference to a another table that is being restored, an entry is made in RestoreTableRecordReference.

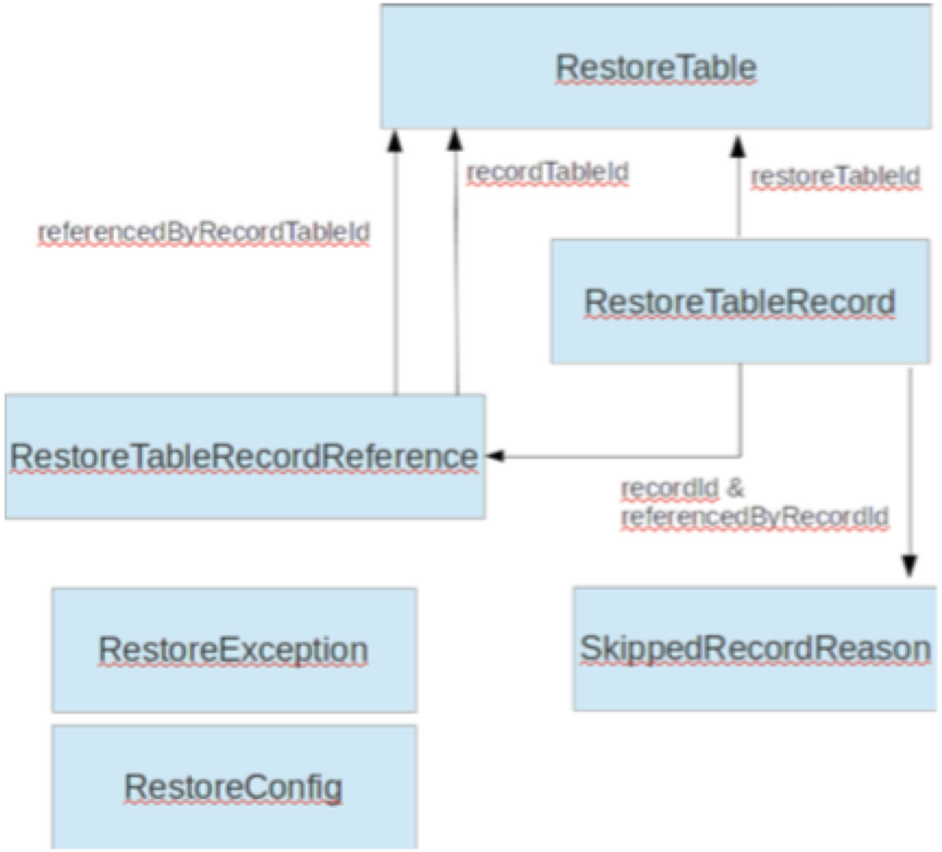
During a restore:

The restore status of a record is recorded in RestoreTableRecord.

The overall restore status of each table is recorded in RestoreTable.

Any exceptions are recorded in RestoreException.

Any skipped records are record in SkippedReason.



CopyStorm

CopyStorm

Bringing your Salesforce Data Home

★★★★★ (18)

Paid - *Details below*

OVERVIEW DETAILS REVIEWS 18 PROVIDER

App by Capstorm

Copy your Salesforce table structure and data
- to SQL/Server, Oracle, MySQL, PostgreSQL, H2
- update data & structure as often as every 5 minutes
- create an accurate, usable backup of your Salesforce
- runs on Windows, Linux, and Mac!

RELEASED

9/26/2011

thank you

App Exchange Vendors - Summary

Data Recovery options include custom or 3rd Party

A Custom data extract could be created using an ESB or data integration component

Create a custom export capability to extract data more frequently than weekly

Could utilize web-services & scripting or a data integration component

SOAP API provides near real-time access of Salesforce data from a 3rd party tool (e.g.Mulesoft)

Move data from Salesforce Database to on premise or cloud DB (e.g. CopyStorm, Azure, AWS)

The AppExchange includes several vendors with Data Backup capabilities

Backups would be undertaken through a qualified AppExchange product

Would be able to back up data to Cloud or On-premise data stores (db or CSV options typically available)

Functionality does vary but could include scheduling, and full or incremental backups

All offer backup, however data restoration is generally limited due to the complexities

A sample third party vendor is listed within the Appendix.

https://help.salesforce.com/apex/HTViewHelpDoc?id=import_which_data_import_tool.htm

Backupify - AppExchange App

AppExchange Application

Daily, automated backup of all Salesforce objects, files, chatter, and metadata

Self-service portal allows admins to monitor, manage, search and download data easily

Export data from Backupify at any time

Automated daily backup of all Salesforce objects and metadata

Easy and seamless setup by any Salesforce administrator

Self-service portal allows admins to monitor, manage and download data easily

Export data from Backupify to store a local copy

Backs up production and sandbox editions

Search function makes it easy to find objects to restore

Each daily backup is stored in the archive and data can be retrieved from any date

Administrative control of API consumption

Weekly Backup Status Summary Email

OwnBackup - AppExchange App

AppExchange Application.

OwnBackup provides easy to setup, secure, automated daily backups of all your SFDC Org data + metadata. Backups are available for preview, download, compare & restore via their site. No quotas; optional unlimited snapshot retention.

Daily backup of all your Salesforce data -including custom objects, attachments & chatter, etc

Metadata backup (reports, workflows, apex code,etc.)

Data is stored encrypted via AES 256-bit on Amazon's EBS.

Easy previewing, downloading, comparing & restoring of snapshots.

Automated comparing of sObjects across snapshots enables efficient identification of changed records.

Sophisticated data restore tools facilitate efficient recovery from a data loss event.

One-click “export all” of all data from any particular backup/snapshot.

Force backup now for manually executing a backup at any time.

Includes sandbox backup.

Informatica Cloud Data Replication

Informatica Cloud Data Replication provides a simple, intuitive way to replicate cloud application data, such as Salesforce, regardless of data volumes, or how often the application data has changed – all through an intuitive web-based interface. Most organizations with large deployments of cloud applications have persistent compliance and operational reporting issues because core data no longer exists on-premise, but instead resides in various cloud applications. Replicating relevant cloud application information and manipulating the data so that it's in the right order and format greatly speeds up your core reporting, compliance, and business intelligence processes.

Features

Replicate all Salesforce data, including any custom objects or fields, into your data warehouse for reporting and BI

Rapidly archive your Salesforce data to comply with existing regulations and internal corporate governance policies with data replication

Support for a variety of databases including SQL Server, MySQL, IBM DB2, and Oracle

Automatically detect schema changes in Salesforce and reflect those changes in your target database schema

Save time by making use of changed data capture functionality to ensure that only Salesforce data that has changed since the last replication job gets extracted into the data warehouse

Benefit from Informatica Cloud's trusted performance, and replicate hundreds of millions of rows with ease

Mulesoft CloudHub

Connect cloud to cloud

CloudHub integrates SaaS applications reliably and securely. CloudHub offers global availability and 99.99% uptime.

Compliance with the highest security standards ensures integrations are protected wherever they run.

Connect cloud to ground

SaaS may be on the rise, but legacy and on-premises applications aren't going away. With CloudHub, you can securely connect SaaS applications to on-premises systems with a single platform designed for hybrid deployments.

Write an integration once and deploy it anywhere.

A long-term solution

CloudHub is a part of the Anypoint Platform, the only complete solution that delivers SOA, SaaS Integration, and API management. From simple data sync to complex SaaS integrations, CloudHub offers capabilities to meet any current or future business need.

Evaluation Matrix App Exchange Vendors

Vendor	Mulesoft	Own Backup	Informatika	Copy Storm	Odaseva
Cloud	Y	Y	Y	N	Y
On Premise	N	N	Connector	Y	N
Cost	$$	$$ - per user	$$$	$$ per Production Org (1500 Euro / Year)	$ per user
Support	Good	Good	Good	Good	Good
Data Encryption	Y	Y	?	Y	Y
SFDC Experience Rating	Good	Fair	Fair	Good	Good
WebSite				https://www.capstorm.com/copystorm	odaseva.com