

---

# **sfdump Documentation**

***Release 2.6.4***

**Kevin Steptoe**

**Feb 19, 2026**

---

# Contents

<b>1</b>	<b>User Guide</b>	<b>1</b>
1.1	New Users Start Here . . . . .	1
1.2	Additional Guides . . . . .	1
1.3	Advanced Topics . . . . .	1
<b>2</b>	<b>Developer Guide</b>	<b>33</b>
2.1	Architecture Overview . . . . .	33
2.2	Modules and Data Flow . . . . .	34
2.3	Limits, Chunking and Resume Strategy . . . . .	35
2.4	Indexing Logic . . . . .	36
2.5	CLI Reference . . . . .	37
2.6	Redaction System . . . . .	37
2.7	Testing and CI . . . . .	38
<b>3</b>	<b>Admin &amp; Finance Documentation</b>	<b>38</b>
3.1	Contents . . . . .	38
<b>4</b>	<b>Reference</b>	<b>43</b>
4.1	Changelog . . . . .	43
4.2	Contributing . . . . .	43
4.3	License . . . . .	44

# User Guide

Welcome to the **sfdump User Guide**. This guide is designed for non-technical users who need to export files from Salesforce and find documents in the archive.

---

## 1.1 New Users Start Here

Follow these guides in order:

1. *Getting Started* — Install, configure, export, and view your data
2. *Finding Documents* — Search tips and browsing techniques
3. *FAQ* — Common questions answered

That's all most users need. The Getting Started guide walks you through:

- Installation (Step 1)
  - Salesforce credential setup (Step 2)
  - Running your first export with `sf dump` (Step 3)
  - Viewing your data with `sf view` (Step 4)
- 

## 1.2 Additional Guides

Once you're comfortable with the basics:

4. *Database Viewer* — Advanced viewer features and navigation
  5. *Shared Network Drive* — Set up multi-user access to the viewer
  6. *Installation* — Platform-specific installation details
- 

## 1.3 Advanced Topics

For power users and custom workflows:

7. *Exporting Files* — Custom exports, chunking, filtering
8. *Verifying and Retrying* — Manual verification commands
9. *Generating Reports* — Missing file reports for compliance
10. *Interpreting Reports* — Understanding export metrics

### 1.3.1 Getting Started

This guide walks you through installing sfdump, connecting to Salesforce, running your first export, and viewing your data. No technical experience required.

**Time required:** 15-20 minutes (plus export time, which varies by data size)

---

## Step 1: Installation

### What You Need

Before starting, make sure you have:

- **Windows 10 or 11** (macOS and Linux also supported)
- **40 GB+ free disk space** for your Salesforce export
- **Salesforce credentials** from your IT department (see Step 2)

### Installing on Windows

1. Press the **Windows key** on your keyboard
2. Type **PowerShell**
3. Click on **Windows PowerShell** (the blue icon)
4. Copy and paste this command, then press **Enter**:

```
irm https://raw.githubusercontent.com/ksteptoe/sfdump/main/bootstrap.ps1 | iex
```

The installer will:

- Download sfdump to your home folder (C:\Users\YourName\sfdump)
- Install Python if needed (no admin rights required)
- Guide you through credential setup

If you see “**Running scripts is disabled**”, use this command instead:

```
powershell -ExecutionPolicy Bypass -Command `
"irm https://raw.githubusercontent.com/ksteptoe/sfdump/main/bootstrap.ps1 | 
iex"
```

To update later: Double-click `install.bat` and choose option [1] to upgrade to the latest version.

### Installing on macOS / Linux

```
curl -LO https://github.com/ksteptoe/sfdump/archive/refs/heads/main.zip
unzip main.zip
cd sfdump-main
make bootstrap
```

---

## Step 2: Salesforce Setup

This is the most important step. You need credentials from your Salesforce administrator or IT department.

### What to Request from IT

Contact your Salesforce administrator and ask for **Connected App credentials** configured for the **Client Credentials OAuth flow**. They will provide:

Credential	What It Looks Like	Example
<b>Client ID</b> (Consumer Key)	Long alphanumeric string	3MVG9_YOUR_CONSUMER_KEY_HERE
<b>Client Secret</b> (Consumer Secret)	Long alphanumeric string	YOUR_CONSUMER_SECRET_HERE
<b>Login URL</b>	Your Salesforce instance URL	<a href="https://yourcompany.my.salesforce.com">https://yourcompany.my.salesforce.com</a>

**Note:** This uses OAuth Client Credentials flow — no username or password is required. Your IT department configures the Connected App to authenticate directly.

## Running Setup

Once you have your credentials, open a terminal and run:

```
sf setup
```

You'll be prompted to enter each credential. The setup wizard saves them to a .env file.

**What the .env file looks like:**

```
SF_AUTH_FLOW=client_credentials
SF_CLIENT_ID=YOUR_CONSUMER_KEY_HERE
SF_CLIENT_SECRET=YOUR_CONSUMER_SECRET_HERE
SF_LOGIN_URL=https://yourcompany.my.salesforce.com
SF_API_VERSION=v60.0
```

## Test Your Connection

Verify everything is configured correctly:

```
sf test
```

**Success looks like this:**

```
Testing Salesforce Connection
=====
Config: C:\Users\YourName\sfdump\.env

Connecting... OK
Instance: https://yourcompany.my.salesforce.com
Testing query... OK

Connection successful! Ready to export.
```

## Common Setup Problems

Error	Solution
SF_CLIENT_ID not set	Run <code>sf setup</code> to enter your credentials
SF_LOGIN_URL not set	Add your Salesforce instance URL to <code>.env</code>
Invalid client credentials	Double-check Client ID and Client Secret with IT
Token request failed (400)	Verify the Connected App is configured for Client Credentials flow
Token request failed (401)	Client ID or Secret is incorrect
Connection failed	Check your network connection and verify SF_LOGIN_URL is correct

Still stuck? Ask your IT department to verify:

1. The Connected App is configured for **Client Credentials** OAuth flow
2. The app has the required API permissions (e.g., `api`, `refresh_token`)
3. A “run as” user is configured with appropriate Salesforce data access

## Step 3: Export Your Data (sf dump)

Now you’re ready to download your Salesforce files and data.

### Running the Export

```
sf dump
```

This command automatically:

1. Authenticates to Salesforce using your credentials
2. Downloads all Attachments and ContentVersions (files)
3. Exports business data (Accounts, Contacts, Opportunities, etc.)
4. Builds searchable indexes for fast lookups
5. Creates a SQLite database for offline browsing
6. Verifies all downloads and retries any failures

### What to Expect

The export runs through several stages. You’ll see progress updates:

```
=====
sfdump - Salesforce Data Export
=====

Step 1/6: Authenticating...
Connected to: https://yourcompany.my.salesforce.com
```

(continues on next page)

(continued from previous page)

```

Step 2/6: Querying file metadata...
  Found 12,479 files to download

Step 3/6: Downloading files...
  [#####
  #####] 12,456 / 12,479 (99.8%)

Step 4/6: Exporting object data...
  Account: 2,341 records
  Contact: 8,127 records
  Opportunity: 1,892 records
  ... (more objects)

Step 5/6: Building indexes...
  Creating search indexes...

Step 6/6: Creating database...
  Database ready: ./exports/export-2026-01-26/meta/sfdata.db

```

### How long does it take?

This depends on your data size:

- Small org (< 1,000 files): A few minutes
- Medium org (1,000-10,000 files): 15-60 minutes
- Large org (10,000+ files): Several hours

You can leave it running and check back later.

### What Success Looks Like

When complete, you'll see a summary:

```

=====
Export Summary
=====

Location:      ./exports/export-2026-01-26

Files
  Expected:    12,479
  Downloaded:  12,456
  Missing:     23
  Complete:    99.8%

NEARLY COMPLETE - 23 files could not be retrieved
(These may have been deleted from Salesforce)

Objects:  38
Database: ./exports/export-2026-01-26/meta/sfdata.db

To browse your data:
  sf view

```

**Understanding the summary:**

Metric	Meaning
<b>Downloaded</b>	Files successfully saved to your computer
<b>Missing</b>	Files that no longer exist in Salesforce (normal)
<b>Complete %</b>	99%+ is excellent; 100% is rare due to normal deletions

**Note:** A 99%+ completion rate is normal and expected. Some files in Salesforce metadata may have been deleted but their records remain. This is not an error.

**If Something Goes Wrong**

**Export interrupted?** Just run `sf dump` again. It automatically resumes where it left off and retries any failed downloads.

**Completion below 95%?** This might indicate a connection issue. Run `sf dump` again — it will retry failed downloads.

**Custom Export Location**

By default, exports are saved to `./exports/export-YYYY-MM-DD/`. To specify a different location:

```
sf dump -d /path/to/my-export
```

---

**Step 4: View Your Data (`sf view`)**

Once the export completes, you can browse your data in a web interface.

**Launch the Viewer**

```
sf view
```

This opens your web browser to the sfdump viewer (usually at `http://localhost:8501`).

**Keep the terminal window open** — closing it stops the viewer.

**Finding Documents**

The viewer opens in Explorer mode — a full-width search page:

1. Type a customer name, invoice number, or keyword in the search box
2. Results show matching documents across the entire export
3. Click any result to preview it

**Browsing Records**

Click the **DB Viewer** button (top-right) to switch to the record browser:

1. Choose an object type (Account, Opportunity, Invoice, etc.)
2. Search for the record you want
3. View details, related records, and attached documents

- 
4. Click **Back to Explorer** in the sidebar to return to search

## Viewing Files

- Click any file name to preview it
- PDFs and images display inline
- Click **Download** to save a copy to your computer
- Click **Open Folder** to see the file in Windows Explorer

## Closing the Viewer

To stop the viewer, go back to your terminal and press **Ctrl+C**.

---

## Quick Reference

Command	What It Does
<code>sf setup</code>	Configure Salesforce credentials
<code>sf test</code>	Verify connection works
<code>sf dump</code>	Export everything from Salesforce
<code>sf view</code>	Browse exported data in web viewer
<code>sf status</code>	List available exports

---

## Next Steps

- *Finding Documents* — Advanced search tips and techniques
  - *FAQ* — Common questions answered
- 

## Need Help?

**Connection issues:** Run `sf test` and check the error message against the troubleshooting table above.

**Credential problems:** Contact your IT department to verify your Connected App access.

**Export questions:** See the *FAQ* for common scenarios.

### 1.3.2 Finding Documents

**For:** Finance teams, accountants, contract managers, and anyone who needs to find documents quickly.

This guide assumes **zero technical background**. If you can use a web browser, you can find documents.

## Opening the Viewer

Run this command:

```
sf view
```

Your web browser opens automatically with the document viewer. The search page appears immediately — you can start searching right away.

**Tip:** Keep the terminal window open — closing it stops the viewer.

## Finding Documents by Invoice Number (PIN/SIN)

This is the most common task for finance users.

1. Type the invoice number in the **Search** box (e.g., “PIN010063” or “SIN001234”)
2. Results appear automatically as you type

**Results show:**

Column	Meaning
record_name	Invoice/record number (e.g., “PIN010063”)
file_name	Document name (e.g., “Invoice_2024.pdf”)
account_name	Customer name
opp_name	Deal/project name

## Using Wildcards

The search box supports wildcards for powerful searching:

Pattern	Meaning	Example
*	Any characters	PIN01006* finds PIN010060, PIN010061, etc.
?	Single character	PIN01006? finds PIN010060 to PIN010069
[1-5]	Range of characters	PIN0100[6-9]* finds PIN01006x through PIN01009x

**Examples:**

- SIN\* — all sales invoices
- PIN01\* — all purchase invoices starting with PIN01
- \*Acme\* — anything containing “Acme”

Click **Search tips** below the search box for more examples.

## Finding Documents by Customer Name

1. Click **Additional Filters** to expand
2. Type the customer name in the **Account Name** box (e.g., “Acme Corp”)
3. Results show all documents for that customer

## Finding Documents by Deal Name

If you know the project or opportunity name:

1. Click **Additional Filters** to expand
2. Type the deal name in the **Opportunity Name** box
3. Results show all documents for that deal

**Combined search:** Use multiple filters together to narrow results.

## Finding a Specific Invoice

### Method 1: By invoice number (fastest)

1. Type the invoice number in the Search box (e.g., “SIN001234”)
2. Click the invoice to view details

### Method 2: Through customer

1. Click Additional Filters → type customer name
2. Look for invoices in the results

## Finding Contracts

1. Type keywords like “contract”, “agreement”, or “SOW”
2. Check “**PDF only**” — contracts are usually PDFs
3. Or search by customer name in Additional Filters

## Previewing Documents

Click any document to preview it directly in the browser:

File type	What you see
<b>PDF</b>	Multi-page preview with smooth scrolling
<b>Images</b> (JPG, PNG, GIF, BMP, TIFF, JFIF)	Inline image
<b>Excel</b> (XLSX, XLS)	Table with sheet selection
<b>CSV / TSV</b>	Table preview
<b>HTML / EML emails</b>	Content preview
<b>Outlook emails</b> (.msg)	Headers and body preview
<b>Text files</b>	Code-style preview
<b>Other files</b> (.docx, .zip, etc.)	Download button

No download required for most file types — everything previews in the browser.

## Selecting Documents from Results

After searching, use the **Select a document** dropdown:

- Documents show as: 001 - PIN010063 | invoice\_filename.pdf
- The record name (PIN/SIN) appears first for easy identification
- Click to preview the document

## Viewing the Parent Record

After finding a document, you can view the full record it belongs to:

1. Click **Open parent record** below the search results
2. This switches to DB Viewer mode showing the record’s details, relationships, and all attached documents
3. Click **Back to Explorer** in the sidebar to return to search

## Common Tasks

### All invoices for a customer

1. Click Additional Filters → Account Name: [customer name]
2. Review all attached documents

### All documents for a deal

1. Click Additional Filters → Opportunity Name: [deal name]
2. Review all attached documents

### Purchase invoices (bills) in a range

1. Search: PIN0100[6-9]\*
2. This finds PIN01006x through PIN01009x

### All PDFs containing a keyword

1. Search: \*keyword\*
2. Check “PDF only”

## Understanding the Data

### Account vs Opportunity vs Invoice

Term	Meaning	Example
<b>Account</b>	Customer company	“Acme Corp”
<b>Opportunity</b>	Specific deal or project	“Acme Corp Q2 2024”
<b>Invoice</b>	Billing document	“SIN001234”
<b>Purchase Invoice</b>	Bill from supplier	“PIN010063”

### Why some records have no documents

You might see: “*No invoices found. This is expected for Closed Lost opportunities.*”

This is normal — it means:

- The deal didn’t happen, so no invoice was created
- Or documents were never uploaded to Salesforce

This doesn’t mean data is lost.

## Tips

- **Partial names work** — “Acme” finds “Acme Corp SA”
- **Case doesn’t matter** — “acme” = “ACME” = “Acme”
- **Use wildcards** — PIN\* finds all purchase invoices
- **Too many results?** — Check “PDF only” or use more specific search terms
- **Can’t find something?** — Try wildcard search like \*keyword\*

## Troubleshooting

### “No matches found”

- Check spelling
- Try wildcards: \*partial\*
- Uncheck “PDF only”

### Viewer not loading

- Make sure terminal is still open
- Run `sf view` again

### Can’t find a document

- Try wildcard search: \*filename\*
- Try searching by customer name in Additional Filters
- Document might be attached to a different record

## Quick Reference

Task	Steps
Find by invoice number	Search → type PIN010063
Find invoice range	Search → type PIN0100[6-9]*
Find customer documents	Additional Filters → Account Name
Find deal documents	Additional Filters → Opportunity Name
Preview document	Click document in results
View parent record	Click “Open parent record”

You don’t need technical skills — just type and click!

### 1.3.3 FAQ

Frequently asked questions about sfdump.

#### General

##### What does sfdump do?

sfdump exports all your Salesforce data (files, accounts, opportunities, invoices, etc.) to your local computer so you can access it offline. It’s designed for organizations archiving data before shutting down their Salesforce instance.

##### What gets exported?

- **Files** — All Attachments and Documents (ContentVersion)
- **Data** — Accounts, Contacts, Opportunities, Invoices, and other business records
- **Relationships** — Links between records are preserved

### How long does an export take?

Depends on your data size:

- Small org (< 10,000 files): Under an hour
- Medium org (10,000 - 100,000 files): Several hours
- Large org (100,000+ files): May take a day or more

You can interrupt and resume at any time.

### Running Exports

#### What if the export is interrupted?

Just run `sf dump` again. Files already downloaded are skipped automatically. The export continues from where it stopped.

#### Can I run the export multiple times?

Yes. It's safe to run `sf dump` as many times as needed:

- Already-downloaded files are skipped
- Only missing files are attempted
- No duplicates are created

#### How do I export a specific date range?

sfdump exports all data by default. For filtered exports, use the advanced `sfdump` command with custom SOQL queries.

#### Can I export to a network drive?

Yes. Specify the path when running the export:

```
sf dump --export-dir /path/to/network/drive/export
```

### Missing Files

#### Why are some files missing?

Common reasons:

- **Deleted from Salesforce** — File was removed after metadata was queried
- **No permission** — Your user doesn't have access to the parent record
- **Archived externally** — File was moved to external storage
- **Network issues** — Connection dropped during download

#### Can I recover deleted files?

No. If a file was deleted from Salesforce, it no longer exists and cannot be recovered. The export will note these as "404 Not Found".

## What does “zero-byte file” mean?

Sometimes Salesforce returns an empty response due to API limits. Run `sf dump` again — the retry logic will re-attempt these downloads.

## Should I worry about a few missing files?

Usually not. A 99%+ complete export is normal. Files get deleted over time as part of normal business operations. Review the missing files list to confirm they’re not critical.

## Viewing Data

### How do I browse my exported data?

Run:

```
sf view
```

This opens a web browser with an interactive viewer where you can search and navigate your data.

### Can multiple people use the viewer?

Yes. Share the Network URL shown when you start the viewer:

```
Network URL: http://192.168.1.100:8503
```

Anyone on your local network can access it.

### Can I search across all documents?

Yes. The viewer opens in Explorer mode where you can search by Account name, Opportunity name, invoice number, or any keyword to find all related documents.

### How do I find a specific invoice?

1. Type the invoice number in the search box (e.g., “SIN001234”)
2. Click the matching result to preview the document
3. Click **Open parent record** to see the full invoice details in DB Viewer

## Reports

### Do I need Pandoc for PDF reports?

Only if you want PDF output. Markdown reports (.md) always work without additional software.

### Where should I store full (unredacted) reports?

Never commit them to Git or share externally. Store them in a private folder:

- Local encrypted drive
- Private network share
- Secure document management system

## What's the difference between redacted and full reports?

Report Type	Contains	Use For
Redacted	[REDACTED] placeholders	External sharing, auditors, documentation
Full	Actual IDs and filenames	Internal IT review, troubleshooting

## Technical

### What credentials do I need?

You need a Salesforce Connected App with:

- Consumer Key (Client ID)
- Consumer Secret
- Your Salesforce username and password

Contact your Salesforce administrator or IT department.

### Where are credentials stored?

In a .env file in your sfdump directory. This file contains sensitive information — don't share it or commit it to Git.

### How much disk space do I need?

Depends on your Salesforce data:

- Check your Salesforce file storage usage for an estimate
- Plan for at least 40GB free space for medium-sized orgs
- Large orgs may need 100GB+

### Can I run exports on a schedule?

Yes. Run `sf dump` from a scheduled task (Windows) or cron job (Mac/Linux). The resume feature ensures interrupted exports continue automatically.

### What if I have a very large org?

Use chunking to split the export:

```
export SFDUMP_FILES_CHUNK_TOTAL=4
export SFDUMP_FILES_CHUNK_INDEX=1
sf dump
```

Run with INDEX=1, 2, 3, 4 to process in parallel or across multiple days.

## After Export

### How do I archive the export?

The export directory contains everything:

```
exports/export-2026-01-25/
└── files/          # Documents
└── files_legacy/   # Attachments
└── csv/            # Data files
└── links/          # Metadata
└── meta/           # Database
```

Copy this entire folder to your archive location.

## Can I access the data without the viewer?

Yes:

- **CSV files** — Open in Excel or any spreadsheet
- **SQLite database** — Query with any SQLite tool
- **Files** — Access directly in the `files/` folder

## How long should I keep the export?

Follow your organization's data retention policy. Exports are self-contained and don't depend on Salesforce, so they can be kept indefinitely.

## Can I import the data back into Salesforce?

sfdump is designed for archival, not migration. For re-importing data, you'd need Salesforce Data Loader or similar tools, plus the CSV files from your export.

## Getting Help

### Where do I report issues?

Open an issue at: <https://github.com/ksteptoe/sfdump/issues>

### Is there a support email?

Contact your IT administrator first. For technical issues, use the GitHub issues page.

### 1.3.4 Database Viewer

Browse your exported Salesforce data in an interactive web interface.

### Launching the Viewer

After running `sf dump`, launch the viewer:

```
sf view
```

This opens a web browser where you can:

- Search for documents by name, invoice number, or customer
- Navigate relationships (Account → Opportunity → Invoice)
- Preview documents inline (PDFs, images, spreadsheets, and more)

The viewer automatically finds your most recent export.

## Opening a Specific Export

To view a specific export:

```
sf view ./exports/export-2026-01-15
```

## Two Views

The viewer has two modes you can switch between:

View	Purpose	Default?
<b>Explorer</b>	Full-width document search across all exported files	Yes (landing page)
<b>DB Viewer</b>	Record browser with sidebar controls and relationship navigation	No (click “DB Viewer” button)

## Switching Between Views

- **Explorer → DB Viewer:** Click the **DB Viewer** button in the top-right corner
- **DB Viewer → Explorer:** Click **Back to Explorer** at the top of the sidebar
- **Explorer → DB Viewer (via record):** Click **Open parent record** on any search result to jump to that record in DB Viewer

### Explorer (Default View)

The Explorer is a full-width document search with no sidebar. This is where most users spend their time.

#### Search by invoice number (PIN/SIN):

1. Type the invoice number in the Search box (e.g., “PIN010063”)
2. Results show the record name first for easy identification

#### Using wildcards:

- **PIN01006\*** — finds PIN010060, PIN010061, etc.
- **SIN\*** — finds all sales invoices
- **PIN0100[6-9]\*** — finds a range (PIN01006x through PIN01009x)

Click **Search tips** for more wildcard examples.

#### Additional filters (click to expand):

- Filter by Account Name
- Filter by Opportunity Name
- Filter by Object Type

#### Filter results:

- Check “PDF only” to show only PDFs
- Results show record\_name first, then file\_name

## DB Viewer

The DB Viewer has a sidebar and two-column layout for browsing records and their relationships:

Area	Purpose
<b>Sidebar</b>	Object selector, search box, filters, navigation breadcrumbs
<b>Left column</b>	Record list, record details, relationships, documents
<b>Right column</b>	Subtree document list and file preview

## Selecting Records

1. Choose an object type from the dropdown (Account, Opportunity, Invoice, etc.)
2. Type in the search box to filter by name
3. Click a record to view its details

## Viewing Record Details

The **Details** tab shows important fields for the selected record:

- Account: Name, Industry, Website, Phone
- Opportunity: Name, Stage, Amount, Close Date
- Invoice: Number, Total, Status

Toggle “Show all fields” in the sidebar to see everything.

## Navigating Relationships

The **Children** tab (in DB Viewer) shows related records. This is how you drill down through your data.

### Example navigation:

1. Start at an Account (e.g., “Acme Corp”)
2. Click **Children** tab → expand **Opportunities**
3. Select an Opportunity → click **Open**
4. Click **Children** tab → expand **Invoices**
5. View all invoices for that deal

### Navigation features:

- **Breadcrumbs** — see your full path (Account → Opportunity → Invoice)
- **Back button** — return to previous record
- **Reset** — return to the starting point

## Finding Documents

### From the Explorer

The easiest way to find documents — just search:

1. Type a customer name, invoice number, or keyword
2. Results show all matching documents across the entire export

3. Click any result to preview it

## From a Record (DB Viewer)

Select any record in DB Viewer and click the **Documents** tab to see:

- Documents attached to the current record
- Documents from parent records in your navigation path

**Example:** Viewing an Invoice shows documents from the Invoice, its parent Opportunity, and the parent Account.

## Jumping Between Views

When viewing a document in Explorer, click **Open parent record** to jump to DB Viewer with that record selected. This lets you explore the record's relationships and other attached documents.

## Previewing Documents

Click any document to preview it. The viewer supports inline preview for many file types:

File type	Preview
<b>PDF</b>	Multi-page inline preview with smooth scrolling
<b>Images</b> (JPG, PNG, GIF, BMP, TIFF, JFIF)	Inline image preview
<b>Excel</b> (XLSX, XLS, XLTX, XLSM)	Table preview with sheet selection
<b>CSV / TSV</b>	Table preview
<b>HTML / HTM</b>	Source preview
<b>EML emails</b> (.eml)	Headers and body preview
<b>Outlook emails</b> (.msg)	Headers and body preview
<b>Text files</b>	Code-style preview
<b>Other files</b> (.docx, .pptx, .zip, etc.)	Download button

All previewable file types also include a download button.

## Common Tasks

### Find All Documents for a Customer

1. Type the customer/account name in the Search box
2. Review all related documents

### Review Invoices for a Deal

1. Click **DB Viewer** to switch to record browser
2. Select **Account** → search for customer
3. Navigate to **Children** → **Opportunities**
4. Open the specific opportunity
5. Navigate to **Children** → **Invoices**
6. Browse all invoices and attached documents

## Find a Specific Invoice

1. Type the invoice number in the Search box (e.g., “SIN001234”)
2. Click **Open parent record** to view the full invoice details in DB Viewer

## Tips

- **Search is case-insensitive** — “acme” finds “Acme Corp”
- **Partial matches work** — “beta” finds “Beta Industries Ltd”
- **Documents include parents** — viewing a child record shows parent documents too
- **Empty relationships are explained** — contextual messages explain why data may be missing

## Troubleshooting

### Viewer won’t start:

- Make sure you’ve run `sf dump` first
- Check that the export directory exists

### No documents found:

- Try a partial name match
- Uncheck “PDF only” to see all file types
- Verify files were exported (check `files/` directory)

### Slow performance with large exports:

- Use specific searches instead of browsing all records
- Filter by object type or file type

## Next Steps

- *Finding Documents* — Simplified guide for end users
- *FAQ* — Common questions

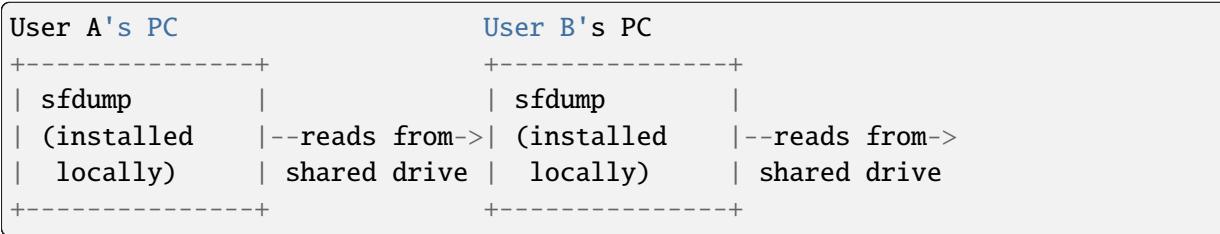
### 1.3.5 Shared Network Drive Setup

This guide explains how to set up sfdump so that **multiple users** can browse the same Salesforce export using the viewer, with the exported data stored on a shared network drive.

## How It Works

Each user installs sfdump on their own PC and runs the viewer locally. The exported data (files and database) lives on a shared network drive that everyone can access. No server or Docker is needed.

```
Shared Drive (\server\sfdump-export\)
|-- files/           <- Downloaded Salesforce files
|-- csv/            <- Exported object data
|-- meta/
|   +- sfdata.db    <- SQLite database (viewer uses this)
+- indexes/         <- Search indexes
```



## Step 1: Run the Export (Admin Only)

One person (the admin) runs the full export. This only needs to happen once.

### Export to the Shared Drive

If the shared drive is already mapped (e.g., S:\):

```
sf dump -d S:\sfdump-export
```

Or export locally first, then copy:

```
sf dump -d ./exports/export-2026-01-26
```

Then copy the entire export folder to the shared drive:

```
xcopy /E /I exports\export-2026-01-26 S:\sfdump-export
```

### Verify the Database Exists

Check that the SQLite database was created:

```
dir S:\sfdump-export\meta\sfdata.db
```

If it doesn't exist, build it manually:

```
sfdump build-db -d S:\sfdump-export
```

---

## Step 2: Share the Export Folder

Make the export folder available on a network drive or shared folder that all users can access. Common options:

Method	Example Path
Mapped network drive	S:\sfdump-export
UNC path	\\"fileserver\\shared\\sfdump-export
OneDrive/SharePoint synced folder	C:\\Users\\Name\\OneDrive - Company\\sfdump-export

**Important:** Users need **read access** to the shared folder. Write access is not required for viewing.

---

### Step 3: Each User Installs sfdump

Each user installs sfdump on their own PC using the standard installer. They do **not** need Salesforce credentials — they only need sfdump installed to run the viewer.

#### Windows Installation

Open PowerShell and run:

```
irm https://raw.githubusercontent.com/ksteptoe/sfdump/main/bootstrap.ps1 | iex
```

See *Installation* for details and troubleshooting.

#### Skip Credential Setup

When the setup wizard asks for Salesforce credentials, users can **skip this step** — credentials are only needed for exporting data, not for viewing.

---

### Step 4: Launch the Viewer

Each user opens a terminal and runs:

```
sfdump db-viewer -d S:\sfdump-export
```

Replace S:\sfdump-export with whatever path the shared folder is at on their machine.

This opens a browser to `http://localhost:8501` where they can search records, browse relationships, and preview documents.

#### Using a UNC Path

If the drive isn't mapped, use the full UNC path:

```
sfdump db-viewer -d "\\\fileserver\shared\sfdump-export"
```

#### Creating a Shortcut (Optional)

To make it easy for users, create a .bat file they can double-click:

1. Create a file called `View Salesforce Data.bat`
2. Add this content:

```
@echo off
cd /d %USERPROFILE%\sfdump
call .\venv\Scripts\activate
sfdump db-viewer -d "S:\sfdump-export"
pause
```

3. Place it on the user's desktop or in the shared folder
-

### Updating the Export

When you need to refresh the data with a newer export:

1. Run a new export: `sf dump -d ./exports/export-2026-02-13`
2. Copy the new export to the shared drive, replacing the old data
3. Users restart their viewer to pick up the changes (Ctrl+C then re-run the command)

No reinstallation needed — users just restart the viewer.

---

### Troubleshooting

#### “SQLite DB not found”

- Check that `meta\sfdata.db` exists in the shared folder
- Run `sfdump build-db -d S:\sfdump-export` to rebuild it

#### Viewer is slow

- Network latency can affect performance when the database is on a remote drive
- For better performance, users can copy `meta\sfdata.db` to their local machine and use `sfdump db-viewer --db C:\local\sfdata.db` — but document preview will still need the shared drive for the actual files

#### “Permission denied” errors

- Verify the user has read access to the shared folder
- On Windows, check that the network drive is mapped or the UNC path is accessible

#### Multiple users at the same time

- Each user runs their own local Streamlit process, so there are no conflicts
- SQLite handles concurrent read access without issues
- Users do not interfere with each other

#### Streamlit not installed

- If users see “Streamlit is not installed”, they need to activate their virtual environment first:

```
cd %USERPROFILE%\sfdump
.venv\Scripts\activate
sfdump db-viewer -d S:\sfdump-export
```

### Tips

- **Read-only is fine** — the viewer never writes to the export folder
- **One export, many users** — any number of users can view the same data simultaneously
- **No Salesforce credentials needed** — users only need sfdump installed, not connected to Salesforce
- **Keep the terminal open** — closing the terminal stops the viewer

### 1.3.6 Installation

This guide explains how to install **sfdump** on your computer.

#### Requirements

- **Windows 10 or 11** (recommended) — macOS and Linux also supported
- **40 GB+ free disk space** for Salesforce exports
- **Salesforce credentials** — contact your IT department for:
  - Connected App credentials (Client ID and Secret)
  - Your Salesforce username and password

#### Windows Installation

##### Step 1: Open PowerShell

1. Press the **Windows key** on your keyboard
2. Type **PowerShell**
3. Click on **Windows PowerShell** (the blue icon)

##### Step 2: Run the Installer

Copy and paste this command into PowerShell, then press **Enter**:

```
irm https://raw.githubusercontent.com/ksteptoe/sfdump/main/bootstrap.ps1 | iex
```

##### Step 3: Follow the Setup Wizard

The installer will:

1. Download sfdump to your home folder (`C:\Users\YourName\sfdump`)
2. Install Python if needed (no admin rights required)
3. Ask for your Salesforce credentials
4. Create a `.env` configuration file

Answer the prompts to complete setup.

#### Troubleshooting

**“Running scripts is disabled”** — If you see this error, use this command instead:

```
powershell -ExecutionPolicy Bypass -Command `
"irm https://raw.githubusercontent.com/ksteptoe/sfdump/main/bootstrap.ps1 | 
iex"
```

**Already installed?** — The installer will detect existing installations and offer to update.

#### macOS / Linux Installation

For macOS or Linux users:

```
curl -LO https://github.com/ksteptoe/sfdump/archive/refs/heads/main.zip  
unzip main.zip  
cd sfdump-main  
make bootstrap
```

## Next Steps

Once installed, continue to [Getting Started](#) for credential setup and your first export.

### 1.3.7 Exporting Files

How to export files from Salesforce and what to expect.

#### Basic Export

To export all files from Salesforce:

```
sf dump
```

This downloads:

- **Documents** (ContentVersion) — Files uploaded to Salesforce Files, Notes, and Libraries
- **Attachments** — Legacy files attached directly to records

No additional commands needed. The export handles everything automatically.

#### What Gets Downloaded

Type	Description	Location
Documents	Modern Salesforce Files (PDFs, images, etc.)	files/
Attachments	Legacy attachments on records	files_legacy/
Metadata	File details and parent record links	links/

#### Resume and Retry

**Interrupted export?** Just run `sf dump` again.

- Files already downloaded are skipped automatically
- Failed downloads are retried
- Progress continues from where it stopped

This is safe to run multiple times — it won't duplicate files or waste API calls.

#### Checking Export Status

To see what's been exported:

```
sf status
```

This shows:

- Available exports and their dates
- Number of files downloaded
- Number of objects exported

- Whether the database is ready

## Output Structure

After export, your folder looks like this:

```
exports/export-2026-01-25/
├── files/          # Documents (ContentVersion)
├── files_legacy/   # Attachments
├── csv/            # Object data (Account.csv, etc.)
└── links/          # Metadata and indexes
    ├── content_versions.csv
    ├── attachments.csv
    ...
└── meta/
    ├── sfdata.db      # SQLite database for viewer
    └── master_documents_index.csv
```

## Viewing Exported Files

To browse your exported files:

```
sf view
```

This opens a web interface where you can search by Account or Opportunity and see all related documents.

## Advanced Options

These options are for power users with large Salesforce orgs.

### Download Order

To download newest files first (useful for large exports):

```
export SFDDUMP_FILES_ORDER=desc
sf dump
```

### Chunking for Large Orgs

For very large Salesforce orgs, split the export across multiple runs:

```
# Run 1 of 4
export SFDDUMP_FILES_CHUNK_TOTAL=4
export SFDDUMP_FILES_CHUNK_INDEX=1
sf dump

# Run 2 of 4
export SFDDUMP_FILES_CHUNK_INDEX=2
sf dump
```

This divides files into chunks so you can run exports in parallel or spread them over multiple days.

## Verbose Output

To see detailed progress for each file:

```
sf dump -v
```

## Next Steps

- *Verifying and Retrying* — Manual verification options
- *Generating Reports* — Create reports of missing files

### 1.3.8 Verifying and Retrying

How sfdump ensures all your files are downloaded completely.

## Automatic Verification

When you run `sf dump`, verification happens automatically:

1. Downloads all files from Salesforce
2. Checks that each file was saved correctly
3. Retries any failed downloads
4. Reports final status

You don't need to run separate verify or retry commands.

## Checking Results

At the end of an export, you'll see a summary:

```
=====
Export Summary
=====

Location: /home/user/sfdump/exports/export-2026-01-25

Files:
  Expected: 12,847
  Downloaded: 12,845
  Missing: 2
  Complete: 99.98%

Status: NEARLY COMPLETE - 2 files could not be retrieved
        (These may have been deleted from Salesforce)
```

## What the Status Means

Status	Meaning
<b>COMPLETE</b>	All files downloaded successfully
<b>NEARLY COMPLETE</b>	99%+ downloaded, missing files may be deleted from Salesforce
<b>INCOMPLETE</b>	Run <code>sf dump</code> again to continue downloading

## Re-running the Export

If files are missing, simply run:

```
sf dump
```

This is safe to run multiple times:

- Already-downloaded files are skipped
- Only missing files are attempted
- Progress is preserved between runs

## Why Files Might Be Missing

Some files cannot be downloaded. Common reasons:

Reason	Explanation
<b>Deleted in Salesforce</b>	File was removed after metadata was queried
<b>Permission restricted</b>	Your user doesn't have access to the parent record
<b>Archived content</b>	File was moved to external storage
<b>Network interruption</b>	Connection dropped during download

Files deleted from Salesforce cannot be recovered — they no longer exist.

## Viewing Missing Files

To see which files are missing, check the master index:

```
exports/export-2026-01-25/meta/master_documents_index.csv
```

Files with an empty local\_path column were not downloaded.

## Advanced: Manual Verification

For detailed control, use the full sfdump commands:

### Verify Only

```
sfdump verify-files --export-dir exports/export-2026-01-25/files
```

Creates:

- links/attachments\_missing.csv
- links/content\_versions\_missing.csv

### Retry Only

```
sfdump retry-missing --export-dir exports/export-2026-01-25/files -v
```

Attempts to re-download files listed in the missing CSVs.

### Analyze Missing Files

```
sfdump analyze-missing --export-dir exports/export-2026-01-25
```

Shows breakdown of why files are missing (permissions, deleted, etc.).

### Next Steps

- *Generating Reports* — Create reports for audit or compliance
- *FAQ* — Common questions about missing files

## 1.3.9 Generating Reports

Create reports for audits, compliance, or IT review.

### When to Generate Reports

Reports are useful for:

- **Audit trails** — Document what was exported and what's missing
- **Compliance** — Prove data retention requirements are met
- **IT/CFO review** — Share export status with stakeholders
- **Handover** — Document the export for future reference

### Basic Report

After running `sf dump`, generate a report of any missing files:

```
sfdump report-missing --export-dir exports/export-2026-01-25 --out missing_report
```

This creates:

- `missing_report.md` — Markdown report you can view in any text editor

### PDF Report

To generate a PDF (requires Pandoc installed):

```
sfdump report-missing --export-dir exports/export-2026-01-25 --out missing_report --pdf
```

This creates both `missing_report.md` and `missing_report.pdf`.

### Redacted Reports

For reports you'll share externally or commit to a repository, use redaction:

```
sfdump report-missing --export-dir exports/export-2026-01-25 --out missing_report --redact
```

Redaction hides sensitive information:

Hidden	Example
Salesforce IDs	001xxx...xxx → [REDACTED]
Filenames	Contract_Acme.pdf → [REDACTED]
URLs	Full Salesforce URLs removed

Use redacted reports for external sharing. Keep full reports internal only.

## Report Contents

A typical report includes:

- **Summary** — Total files expected, downloaded, missing
- **Missing files list** — Details of files that couldn't be downloaded
- **Failure reasons** — Why each file failed (deleted, permissions, etc.)
- **Recommendations** — Next steps to resolve issues

## Example Workflow

```
# 1. Run export
sf dump

# 2. Generate internal report (full details)
sfdump report-missing --export-dir exports/export-2026-01-25 --out internal_
↪report --pdf

# 3. Generate external report (redacted)
sfdump report-missing --export-dir exports/export-2026-01-25 --out external_
↪report --redact --pdf
```

## Next Steps

- *Interpreting Reports* — Understanding report contents
- *FAQ* — Common questions about exports and compliance

### 1.3.10 Interpreting Reports

How to read and understand the export reports generated by sfdump.

## Report Overview

After running `sf dump` and generating a report, you'll have a document that summarizes your export. This guide explains what each section means.

## Executive Summary

The top of the report shows key numbers:

Metric	Meaning
<b>Total Attachments</b>	Legacy files found in Salesforce
<b>Total Documents</b>	Modern files (ContentVersion) found
<b>Downloaded</b>	Files successfully saved locally
<b>Missing</b>	Files that couldn't be downloaded
<b>Recovered</b>	Files recovered during retry
<b>Unrecoverable</b>	Files that no longer exist

**Example:****Export Summary**

```
=====
Total Files:      12,847
Downloaded:     12,845 (99.98%)
Missing:          2 (0.02%)
```

**What the Numbers Mean**

- 99-100% downloaded** — Excellent. Export is complete or nearly complete.
- 95-99% downloaded** — Good. A few files may have been deleted from Salesforce.
- Below 95%** — Review the missing files. May need to re-run `sf dump`.

**Missing Files Section**

Lists each file that couldn't be downloaded:

Column	Meaning
<b>File ID</b>	Salesforce ID (or [REDACTED] in redacted reports)
<b>Filename</b>	Original filename (or [REDACTED])
<b>Parent Record</b>	What record the file was attached to
<b>Retry Status</b>	Whether recovery was attempted
<b>Error</b>	Why it failed

**Common Error Messages**

Error	Meaning	Action
<b>404 Not Found</b>	File deleted from Salesforce	None — file no longer exists
<b>403 Forbidden</b>	No permission to access	Contact Salesforce admin
<b>Connection timeout</b>	Network issue during download	Run <code>sf dump</code> again
<b>Zero bytes received</b>	API returned empty file	Run <code>sf dump</code> again

**Impact by Record Type**

Shows which business records are affected by missing files:

**Missing Files by Record Type**

```
=====
Opportunity:      1 file
```

(continues on next page)

(continued from previous page)

Account:	0 files
Invoice:	1 file

This helps you understand the business impact:

- Missing Opportunity files might affect deal history
- Missing Invoice files might affect financial records
- Missing Account files might affect customer documentation

## Redacted vs Full Reports

### Redacted Reports

Safe for external sharing. Sensitive information is hidden:

```
File: [REDACTED]
Parent: [REDACTED]
ID: ATTACHMENT_001
```

Use redacted reports for:

- Sharing with auditors
- Including in documentation
- Committing to repositories

### Full Reports

Contains actual Salesforce IDs and filenames:

```
File: Contract_AcmeCorp_2024.pdf
Parent: 0064J00000ABC123 (Opportunity)
ID: 00P4J00000XYZ789
```

Use full reports for:

- Internal IT review
- Troubleshooting specific files
- Salesforce admin investigation

**Never share full reports externally.**

## Reading the Recommendations

Reports end with actionable recommendations:

### “Export Complete”

```
Status: COMPLETE
All files downloaded successfully. No action required.
```

Your export is done. Archive it safely.

## “Nearly Complete”

Status: NEARLY COMPLETE

2 files could **not** be retrieved. These may have been deleted **from Salesforce**.

Recommendation: Review missing files [list](#). If acceptable, archive the export.

A few files are missing but this is often normal — files get deleted over time.

## “Incomplete”

Status: INCOMPLETE

847 files could **not** be downloaded.

Recommendation: Run '[sf dump](#)' again to retry failed downloads.

Significant files are missing. Re-run the export to recover them.

## Example Report

### SALESFORCE EXPORT REPORT

Generated: 2026-01-25

### EXECUTIVE SUMMARY

Total Files: 12,847  
Downloaded: 12,845 (99.98%)  
Missing: 2 (0.02%)

### MISSING FILES

1. [REDACTED] - 404 Not Found  
Parent: Opportunity [REDACTED]
2. [REDACTED] - 404 Not Found  
Parent: Invoice [REDACTED]

### IMPACT BY RECORD TYPE

Opportunity: 1 file  
Invoice: 1 file

### RECOMMENDATION

Status: NEARLY COMPLETE

2 files could **not** be retrieved **and** appear to have been deleted **from Salesforce**. This **is** a 99.98% complete export.

Recommended action: Archive the export. The missing files no longer exist **in** Salesforce **and** cannot be recovered.

## Next Steps

- *FAQ* — Common questions about missing files
- *Generating Reports* — How to create reports

2

---

# Developer Guide

Welcome to the **sfdump Developer Guide**, providing a complete technical overview of how the system works internally.

## 2.1 Architecture Overview

sfdump is built around a modular architecture designed to safely export all Salesforce files, avoid API limits, and ensure completeness.

### 2.1.1 Key Architectural Goals

- Reliable bulk export of Attachments and ContentVersions
- Automatic resume without re-downloading existing files
- Protection against Salesforce API limits
- Deep verification and retry mechanisms
- Automatic redacted/unredacted reporting
- Clean CLI and Makefile-driven interface

### 2.1.2 High-Level Components

- `api.py` — Salesforce REST API wrapper
- `files.py` — File download logic (Attachments + ContentVersions)
- `dumper.py` — Export orchestration
- `verify.py` — Completeness checker
- `retry.py` — Retrying failed downloads
- `analyze.py` — Parent-object analysis
- `report.py` — Markdown/PDF report generator
- `cli.py` — User-facing command-line interface
- `Makefile.export` — Batch orchestration

### 2.1.3 Export Lifecycle

1. Query metadata
2. Apply chunking/order
3. Download files
4. Build indexes
5. Verify completeness
6. Retry missing files

## 7. Produce audit-ready reports

sfdump ensures no step loses state, allowing users to safely re-run workflows.

## 2.2 Modules and Data Flow

This page documents how each module interacts.

### 2.2.1 Module Breakdown

#### `api.py`

- Authenticates with Salesforce
- Wraps REST calls with retries
- Logs all HTTP interactions

#### `files.py`

- Queries Attachments and ContentVersions
- Applies chunking
- Downloads binaries
- Writes metadata CSVs

#### `dumper.py`

Central orchestrator for the full export.

#### `verify.py`

- Scans filesystem
- Compares against metadata CSVs
- Produces missing-file CSVs

#### `retry.py`

- Attempts to re-download missing files
- Writes retry CSVs

#### `analyze.py`

- Maps failed files to parent records
- Identifies impacted business objects

#### `report.py`

- Assembles all diagnostics into Markdown/PDF
- Handles redaction
- Injects logos

## cli.py

- Ties everything together
- Provides user-friendly commands

### 2.2.2 Data Flow Diagram (Text)

```
Salesforce → api.py → files.py → dumps
                                ↓
                                verify.py
                                ↓
                                retry.py
                                ↓
                                analyze.py
                                ↓
                                report.py
```

## 2.3 Limits, Chunking and Resume Strategy

Salesforce imposes strict API limits. sfdump includes a robust system to avoid hitting them.

### 2.3.1 Resume Logic

If a file exists locally:

- It is skipped
- Saves API calls
- Enables safe reruns

### 2.3.2 Ordering

Use descending order to recover from interrupted runs:

```
export SFDUMP_FILES_ORDER=desc
```

### 2.3.3 Chunking

Divide work into slices to avoid API throttling:

```
export SFDUMP_FILES_CHUNK_TOTAL=4
export SFDUMP_FILES_CHUNK_INDEX=2
```

Each chunk receives:

- subset of ContentVersions
- subset of Attachments

Chunk flow:

1. Query all metadata
2. Slice locally
3. Download only that slice

### 2.3.4 Limit Failure Behaviour

If Salesforce returns:

- REQUEST\_LIMIT\_EXCEEDED
- 0-byte response bodies
- Connection abort during bulk download

You can always rerun safely.

### 2.3.5 Recommended Pattern

1. Run chunked export
2. Verify
3. Retry
4. Generate report

This ensures completeness.

## 2.4 Indexing Logic

sfdump builds indexes connecting each file to its parent business object.

### 2.4.1 Why Index Files?

- Finance audit trails
- HR compliance
- Opportunity/Account reconstruction
- Facilitates redaction anonymisation

### 2.4.2 Index Sources

- Attachment.ParentId relationships
- ContentDocument → ContentVersion → Linked records
- Custom parent objects (e.g. PSA, HR, FinanceForce)

### 2.4.3 Label Fields

Some objects have special label fields:

```
INDEX_LABEL_FIELDS = {
    "SalesforceInvoice": "InvoiceNumber",
    "SalesforceContract": "BillingCompany",
    "SalesforceQuote": "SalesforceContractId",
}
```

### 2.4.4 Output Files

Generated under:

```
links/<Object>_files_index.csv
```

Each contains:

- Parent record ID
- Parent descriptive name
- File ID
- File name
- File path

### 2.4.5 Use in Reports

Indexing feeds:

- redaction mapping
- parent impact reporting

- hierarchical summaries

## 2.5 CLI Reference

This document lists all sfdump CLI commands.

### 2.5.1 sfdump verify-files

Check for missing or zero-byte files.

### 2.5.2 sfdump retry-missing

Retry files listed in missing CSVs.

### 2.5.3 sfdump analyze-missing

Map missing files to parent records.

### 2.5.4 sfdump report-missing

Generate Markdown/PDF reports (redacted or full).

### 2.5.5 New Flags

#### --redact

Hides sensitive IDs and filenames.

### Verbosity

-v	→ INFO
-vv	→ DEBUG

### 2.5.6 Environment Flags

- SFDUMP\_FILES\_ORDER
- SFDUMP\_FILES\_CHUNK\_TOTAL
- SFDUMP\_FILES\_CHUNK\_INDEX

## 2.6 Redaction System

sfdump includes a deterministic redaction engine for safe sharing.

### 2.6.1 What Is Redacted?

- Attachment IDs
- Parent IDs
- Filenames
- URLs

### 2.6.2 Deterministic Mapping

Anonymised IDs are stable within a run:

ATTACHMENT_1 → maps to real attachment X
PARENT_1 → maps to real parent Y

### 2.6.3 Where Redaction Applies

- Reports
- Diagnostic tables
- Parent impact summary

## 2.6.4 Where It Does NOT Apply

- Local filesystem
- Internal reports
- Retry/verify CSVs

## 2.6.5 How to Enable

```
sfdump report-missing --redact
```

## 2.7 Testing and CI

Guidelines for automated validation.

### 2.7.1 Unit Tests

Recommended tests:

- API wrapper behaviour
- Chunk slicing logic
- Resume detection
- Missing-file detection
- Retry behaviour
- Redaction mapping
- Index building

### 2.7.2 CI Recommendations

- Run full export against a test org with synthetic attachments
- Validate report generation
- Ensure redacted output contains no IDs
- Ensure unredacted output is correctly excluded via `.gitignore`

3

---

# Admin & Finance Documentation

This section contains guides and references intended for Finance, Audit, and internal administrative users.

## 3.1 Contents

### 3.1.1 Finance User Guide

This guide explains how finance teams should use sfdump outputs.

#### What Finance Needs

- All financial attachments
- All ContentVersions linked to invoices, POs, receipts
- Evidence of completeness

- A redacted audit trail for documentation

## Key Steps

1. Run the export (usually done by IT)
2. Verify all financial objects have complete attachment sets
3. Review missing-file reports
4. Provide redacted report for auditors

### 3.1.2 Audit Completeness

How to demonstrate that Salesforce data has been fully extracted.

#### Required Evidence

- content\_versions.csv
- attachments.csv
- \*\_files\_index.csv for finance objects
- Missing-file report

#### Verifying Finance Objects

Check:

- c2g\_codaInvoice\_c
- c2g\_codaPurchaseInvoice\_c
- ffvat\_VatReturn\_c
- ffps\_po\_PurchaseOrder\_c

#### Auditor Checklist

- Are all ContentVersions present?
- Are all attachments present?
- Have retries succeeded?
- Are parent objects consistent?

### 3.1.3 Internal vs Redacted Reports

#### Internal Reports

Contain:

- Full IDs
- Filenames
- Salesforce URLs

Must be stored outside Git.

#### Redacted Reports

Contain:

- ATTACHMENT\_n

- PARENT\_n
- [REDACTED] filenames

Safe for:

- Sphinx docs
- Partner communication
- General distribution

### How to Generate

Internal:

```
sfdump report-missing --out ../internal_reports/... --pdf
```

Redacted:

```
sfdump report-missing --out docs/missing_report --redact
```

### 3.1.4 Retention Recommendations

Guidance for record retention after Salesforce offboarding.

### Recommended Storage

- ZIP archive of full export
- Missing-file reports
- Redacted audit reports
- Index CSVs

### Duration

Follow local legal requirements:

- Finance: 6–10 years
- HR: 7+ years
- Tax: 6 years (UK)

### Access Control

- Keep internal reports confidential
- Only distribute redacted versions externally

### 3.1.5 Audit Results & Data Completeness (Template)

**Warning: This template contains no proprietary data.** Actual results will be included at build time from `_generated/`.

### Overview

The export verification pipeline produces three primary artefacts:

1. `attachments_missing.csv`
2. `attachments_missing_retry.csv`
3. `missing_file_analysis.md`

These allow Finance and Audit teams to evaluate completeness.

## Runtime Inclusion

If real outputs exist under:

```
docs/_generated/audit/
```

they will be included automatically:

```
:::{include} ../../_generated/audit/missing_file_analysis.md
:::
```

## 3.1.6 Audit Completeness — Actual Results

**Warning:** Real data is included from `_generated/audit/`, not stored in the repo.

### Summary Statistics

#### 3.1.7 Audit Summary Statistics

Generated: 2026-01-23T18:08:45.209506 UTC

Metric	Value
Total attachments discovered	<b>39917</b>
Successfully downloaded (initial run)	<b>39917</b>
Missing before retry	<b>0</b>
Recovered on retry	<b>0</b>
Still missing after retry	<b>0</b>
Final successfully available	<b>39917</b>
Initial success rate	<b>100.00%</b>
Overall success rate after retry	<b>100.00%</b>

Note: Counts and percentages are derived from the export metadata and missing-file analysis CSVs. They are suitable for internal Finance/Audit documentation but do not expose individual file names.

### Missing File Analysis

No detailed missing file analysis was available at the time of generation.

#### 3.1.8 Salesforce Offboarding – CFO Detailed Report

This chapter summarises the status of the Salesforce file export and highlights any residual risk areas for Finance and Audit.

Note: This version is redacted for distribution outside the core finance and IT teams. No Salesforce IDs or file names are included.

### 1. Scope and data sets reviewed

The report is based on the exported file indexes produced by the `sfdump` tooling, including:

- The master index of exported ContentVersion records (where available).
- The list of attachments that could not be retrieved.
- The list of attachments queued for a retry export, where applicable.

## 2. High-level export status

- Total file index rows analysed: **not available** (content\_versions.csv not found for this export).
- Attachments not retrieved: **0**
- Attachments queued for retry: **0**

In broad terms, the export has captured the vast majority of files required for finance and audit purposes. The residual items are concentrated in the missing and retry lists described below.

## 3. Residual risk – missing attachments

All attachments referenced in the indexes were successfully retrieved. There is no known residual risk from missing files.

## 4. Retry and remediation plan

There is currently no active retry queue. Any further recovery attempts would need to be performed manually on a case-by-case basis in Salesforce.

## 5. Recommended position for Finance and Audit

In summary:

- The exported data set is sufficient to support statutory accounts, management reporting and future audit enquiries.
- A clearly defined list of residual missing attachments has been produced, quantifying the gap and allowing it to be documented as part of Finance's working papers.
- A pragmatic retry and remediation plan is available, but the cost and time of further recovery should be weighed against the relatively small volume of outstanding items.

If required, Finance can reference this chapter directly in due diligence or audit packs as evidence of the structured offboarding process and the limited residual risk.

### 3.1.9 CFO Detailed Audit Report

This chapter is automatically generated from the latest sfdump CFO extraction run.

To regenerate this report:

```
sfdump cfo-generate-docs --export-dir <dir> --redact
```

---

### Generated Report

#### 3.1.10 CFO Forensic Audit Report

##### Summary

- Missing attachments found: **0**
- Retry attempts recorded: **0**
- Files recovered on retry: **0**
- Still missing after retry: **0**

## **Missing Attachments**

No missing attachments detected.

## **Retry Results**

No retry data available.

4

---

# **Reference**

## **4.1 Changelog**

### **4.1.1 November 2025 — Major Release**

- Added limit-safe chunking
- Added resume-based skip-existing logic
- Added verify-files command
- Added retry-missing command
- Added analyze-missing command
- Added full reporting pipeline (Markdown + PDF)
- Added deterministic redaction engine
- Added anonymised label mapping
- Added logo-injection support
- Added admin/finance documentation
- Migrated entire docs system to Markdown (MyST)
- Modernised Sphinx configuration

### **4.1.2 Earlier Versions**

(Include historical release notes as needed.)

## **4.2 Contributing**

We welcome contributions.

### **4.2.1 Guidelines**

- Follow PEP8 (enforced via pre-commit)
- Use Ruff for linting
- Write unit tests for new features
- Document new CLI commands
- Do not commit unredacted internal reports
- Ensure Sphinx documentation builds cleanly

## **4.2.2 Pull Requests**

Must include:

- Description of feature
- Tests (if applicable)
- Updated docs (for user or developer guides)

## **4.3 License**

This project is licensed under the MIT License.

See the full text in the repository root LICENSE file.