

# File Hash Deduplication Testing

---

This document outlines the testing performed to verify the file hash deduplication functionality in the PSScript platform.

## Overview

---

The PSScript platform uses file hash deduplication to prevent duplicate scripts from being uploaded to the database. When a script is uploaded, the system calculates a hash of the file content and checks if a script with the same hash already exists in the database. If a match is found, the upload is rejected with a 409 Conflict response.

# Test Results

---

## Test 1: Upload Original Script

- **Script:** `test-script.ps1`
- **Hash:** `0db47317f1f1f0867103b3df0bf6a360`
- **Result:** Successfully uploaded with ID 11
- **Conclusion:** The system correctly accepts a new script and stores its hash in the database.

## Test 2: Upload Duplicate Script

- **Script:** `test-script.ps1` (same content)
- **Hash:** `0db47317f1f1f0867103b3df0bf6a360`
- **Result:** Rejected with 409 Conflict, referencing existing script ID 11
- **Conclusion:** The system correctly identifies duplicate content based on file hash.

## Test 3: Upload Modified Script

- **Script:** `test-script-unique.ps1` (new content)
- **Hash:** `7f2a8ffa3577ad81591e7f24d53bd642`
- **Result:** Successfully uploaded with ID 20
- **Conclusion:** The system correctly accepts a script with different content.

## Test 4: Upload Duplicate of Modified Script

- **Script:** `test-script-unique.ps1` (same content as Test 3)
- **Hash:** `7f2a8ffa3577ad81591e7f24d53bd642`
- **Result:** Rejected with 409 Conflict, referencing existing script ID 20
- **Conclusion:** The system correctly identifies duplicate content of the second script.

## Test 5: Upload Further Modified Script

- **Script:** `test-script-unique-modified.ps1` (modified version of Test 3)
- **Hash:** `43294651b4d5318371cc5d803cfb5746`

- **Result:** Successfully uploaded with ID 21
- **Conclusion:** The system correctly accepts a script with different content, even if it's similar to an existing script.

### **Test 6: Upload Duplicate of Further Modified Script**

- **Script:** test-script-unique-modified.ps1 (same content as Test 5)
- **Hash:** 43294651b4d5318371cc5d803cfb5746
- **Result:** Rejected with 409 Conflict, referencing existing script ID 21
- **Conclusion:** The system correctly identifies duplicate content of the third script.

## Implementation Details

---

The file hash deduplication is implemented in the following components:

1. **File Integrity Utility:** Calculates the MD5 hash of the file content.
2. **Script Controller:** Checks if a script with the same hash exists before saving.
3. **Database Schema:** Stores the file hash in the `scripts` table.

# Conclusion

---

The file hash deduplication functionality is working correctly. The system successfully:

1. Calculates a unique hash for each script based on its content.
2. Stores the hash in the database when a new script is uploaded.
3. Checks for duplicate hashes before accepting new uploads.
4. Rejects uploads with duplicate content, providing a reference to the existing script.

This feature prevents duplicate scripts from cluttering the database and helps maintain data integrity.