

Data Deduplication

Business Continuity

What is Data Deduplication?

Data Deduplication

- Data deduplication is process of **reducing redundant data**.
- Similar to the problem of **compression** and how to reduce and archive data.
- Considerable redundant data could be consume space and energy of **data at rest** on disks.
- Not only the storage of redundant data, consider the communication **bandwidth expense** of **data in transit**.

Files or blocks

Data Deduplication

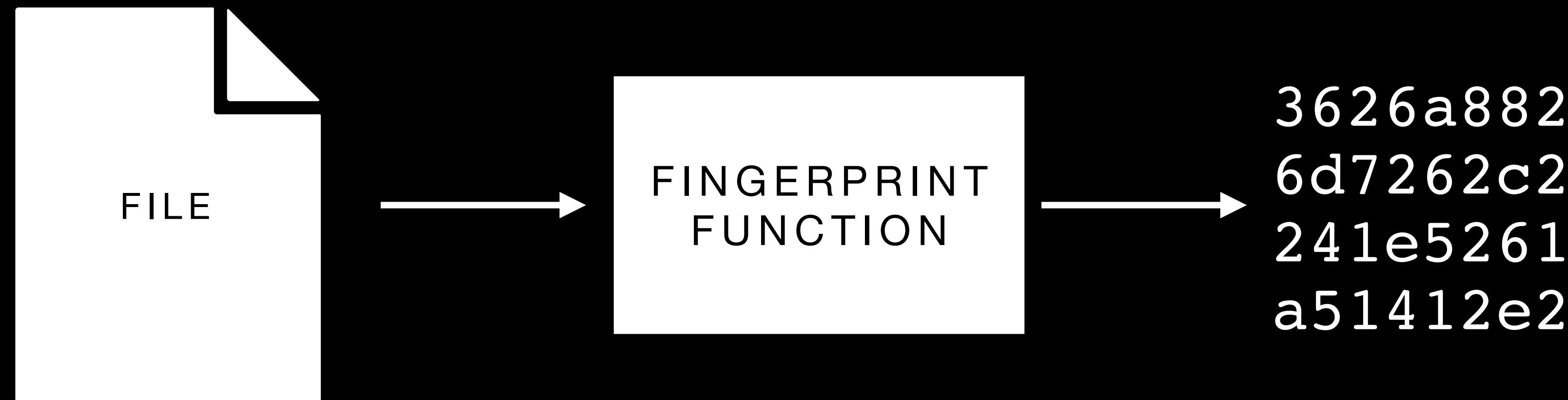
- Data duplication could be in terms of **files** or **blocks**.
- Files could have different filenames, but essentially represent the **same binary data**.
 - Solution: store one file of binary data rather than multiple files and point to it.
- Files could be different, but contain similar **patterns** or **blocks of binary data**.
 - Solution: store one pattern or block of binary data rather than multiple blocks and point to it.

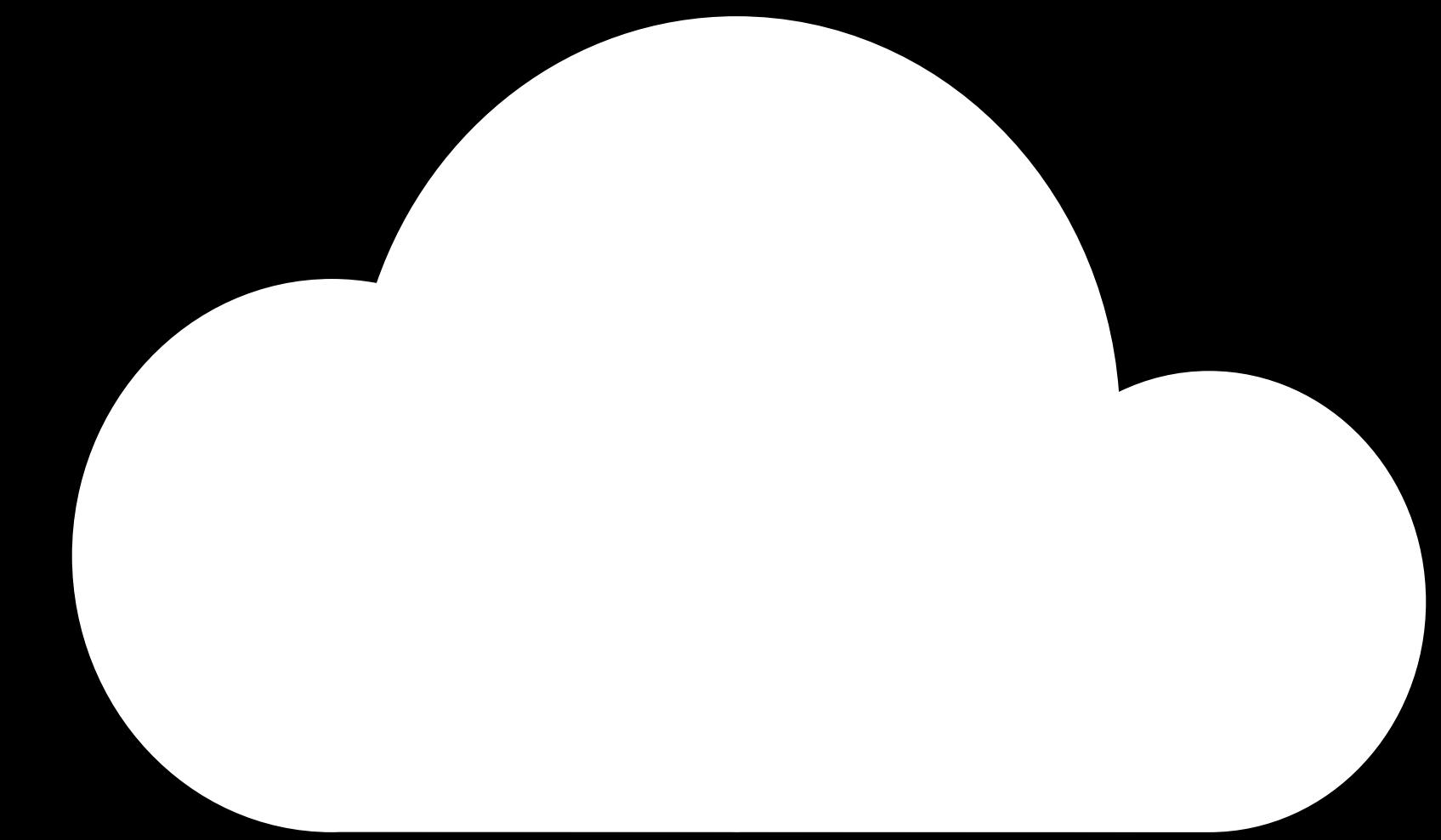
Data Deduplication

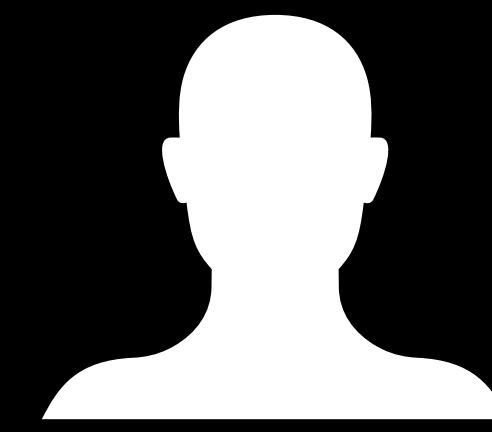
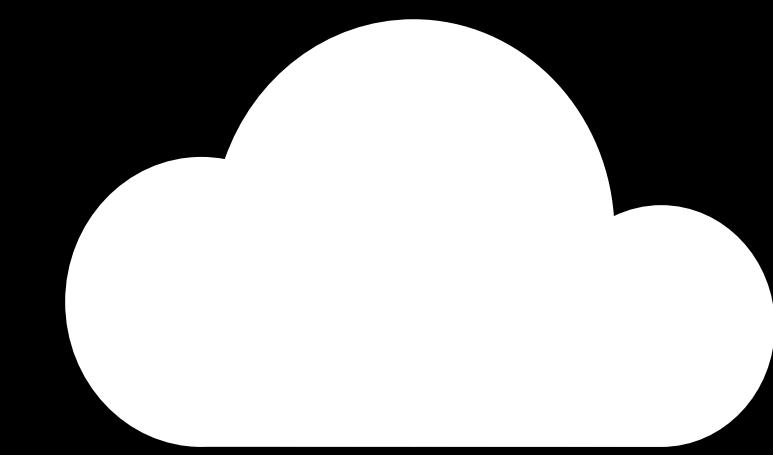
Business Continuity

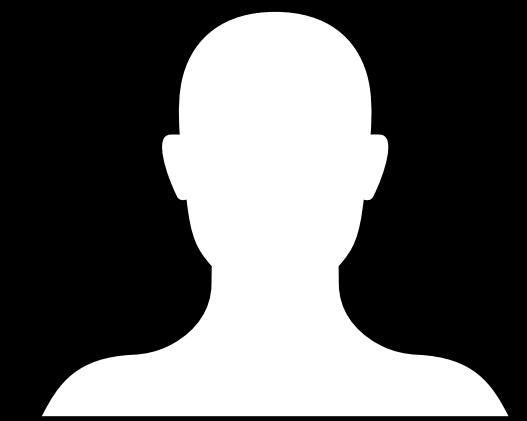
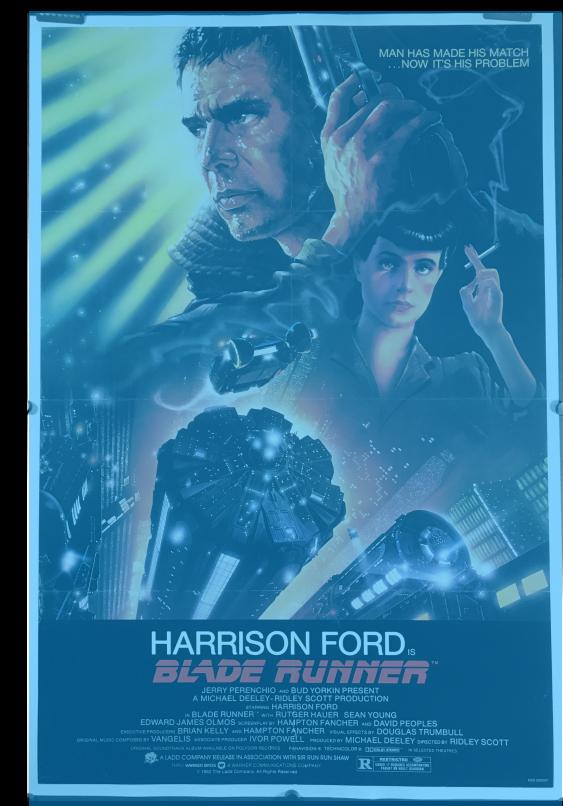
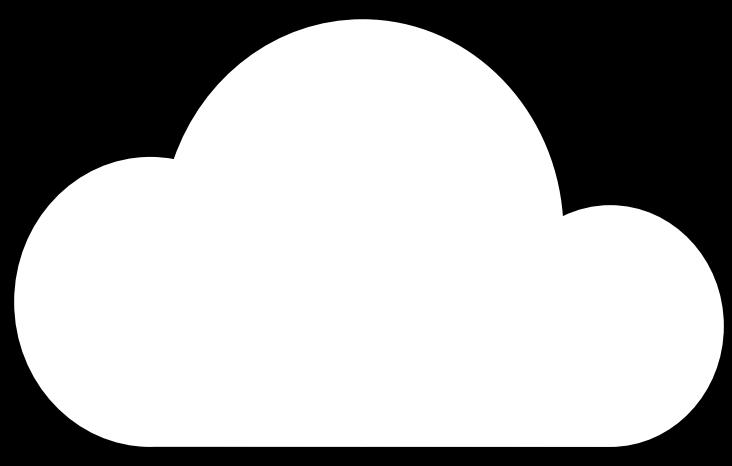
- **Fingerprint** binary data using a one-way hash function.
- Determine if have previously encountered fingerprint.
 - If never encountered, store fingerprint of binary data and store binary data.
 - If encountered and already in store, discard binary data and associate existing binary data with user.

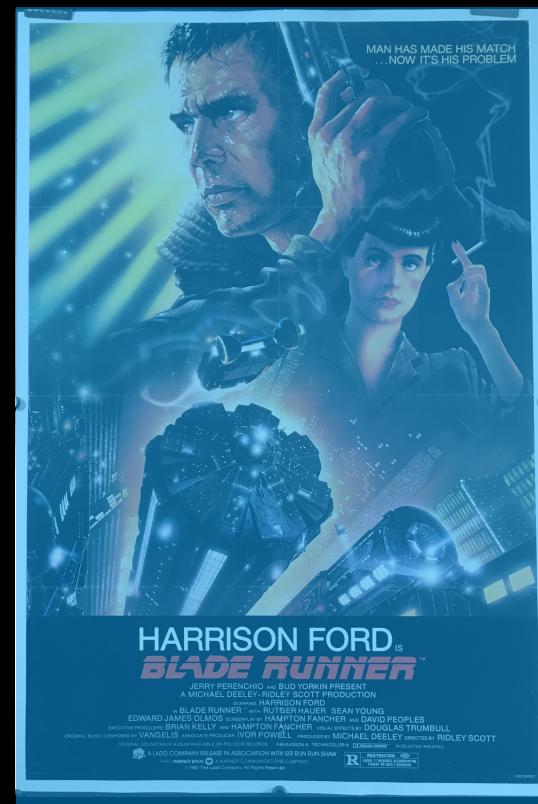
Fingerprint Data Deduplication







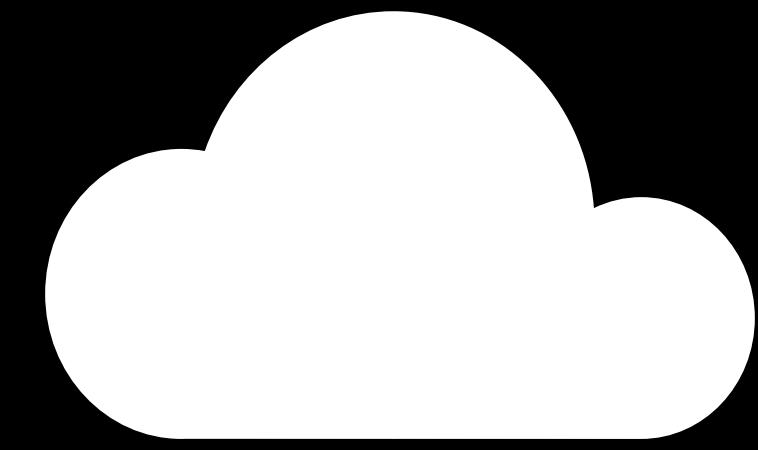
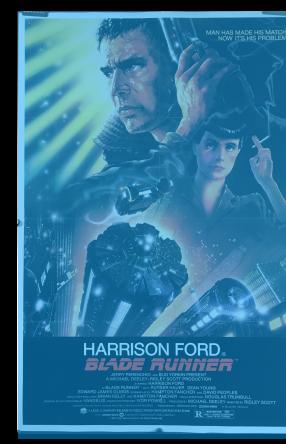




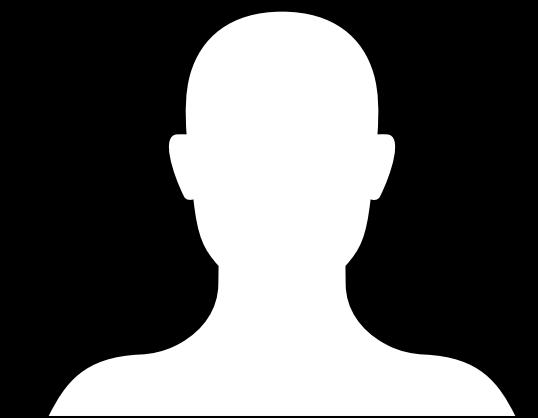
FINGERPRINT
FUNCTION

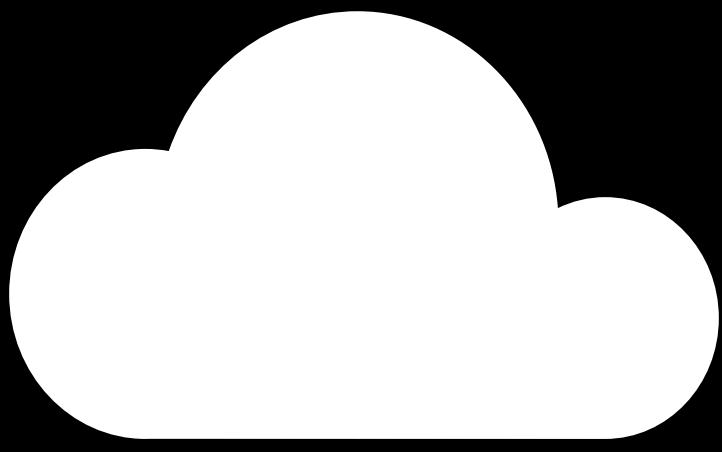


1188f67d
48c9f11a
fb857297
7ef74c5e

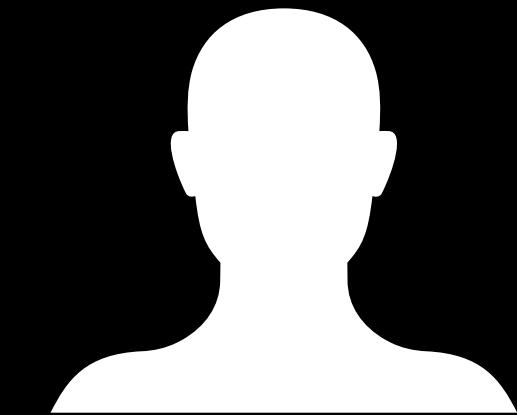
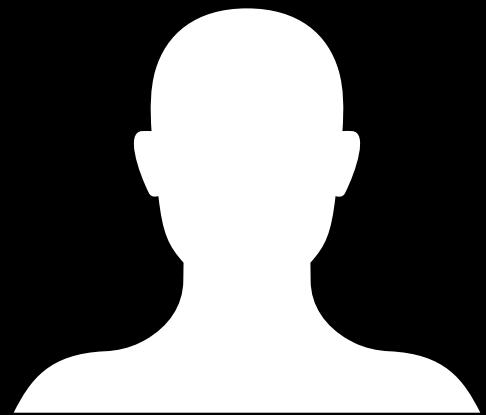


1188f67d
48c9f11a
fb857297
7ef74c5e



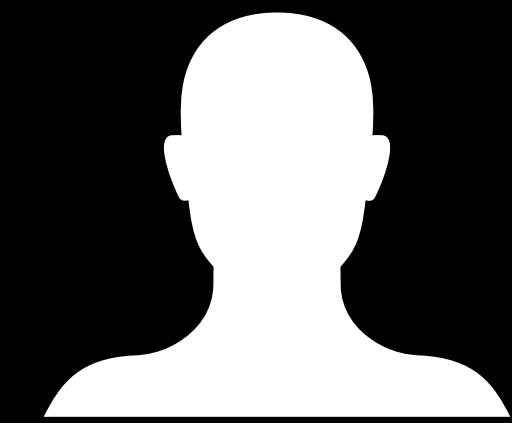
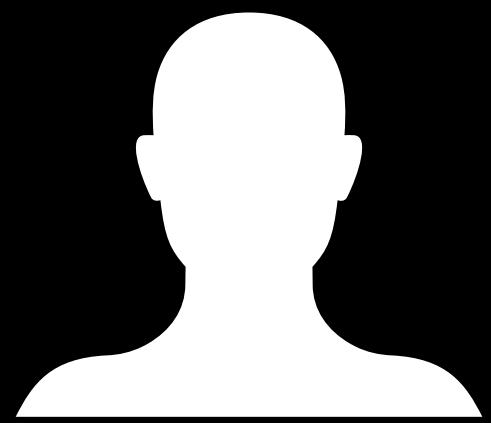


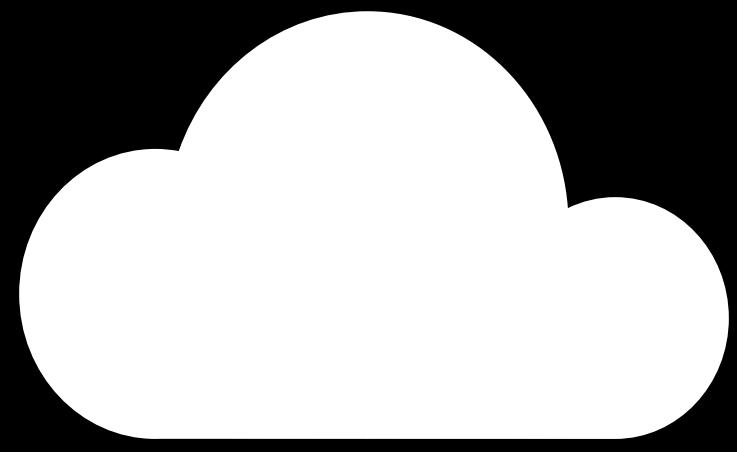
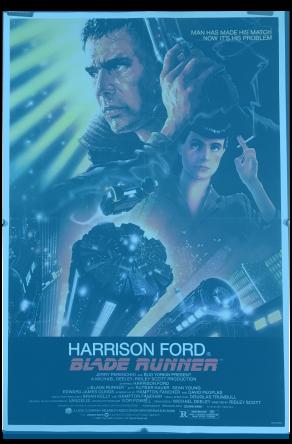
1188f67d
48c9f11a
fb857297
7ef74c5e



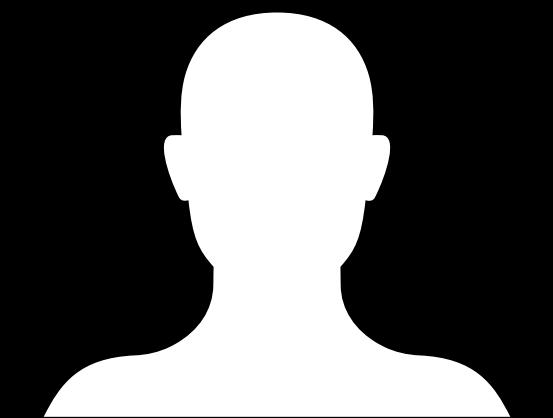
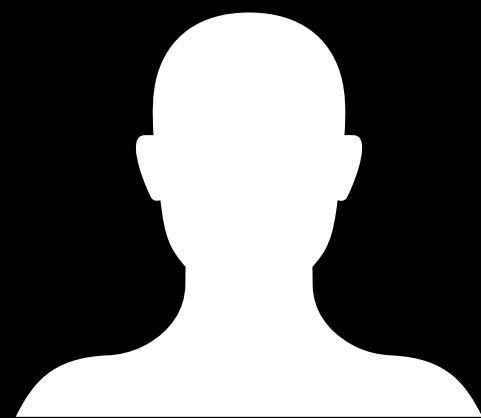
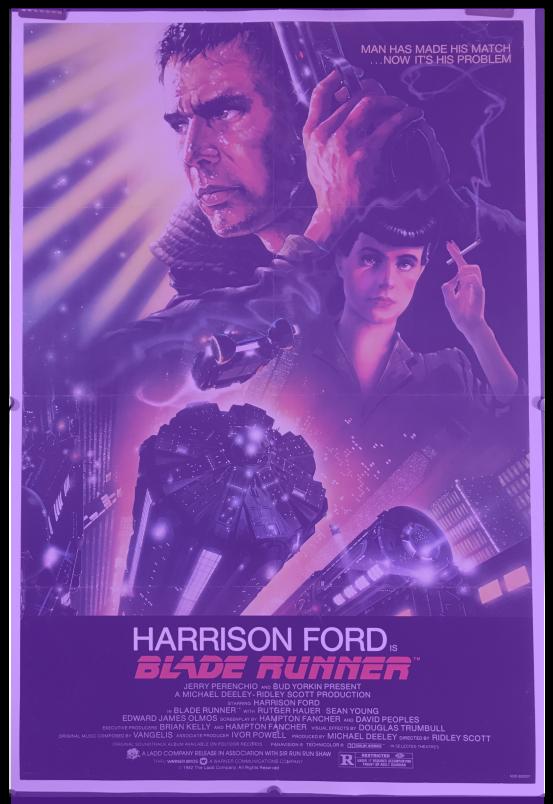


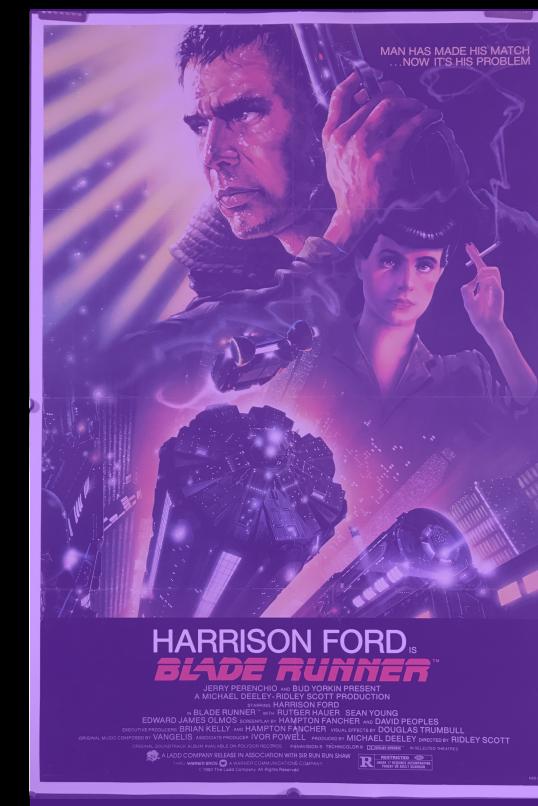
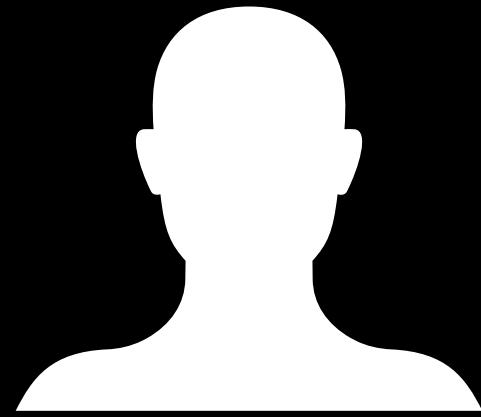
1188f67d
48c9f11a
fb857297
7ef74c5e





1188f67d
48c9f11a
fb857297
7ef74c5e





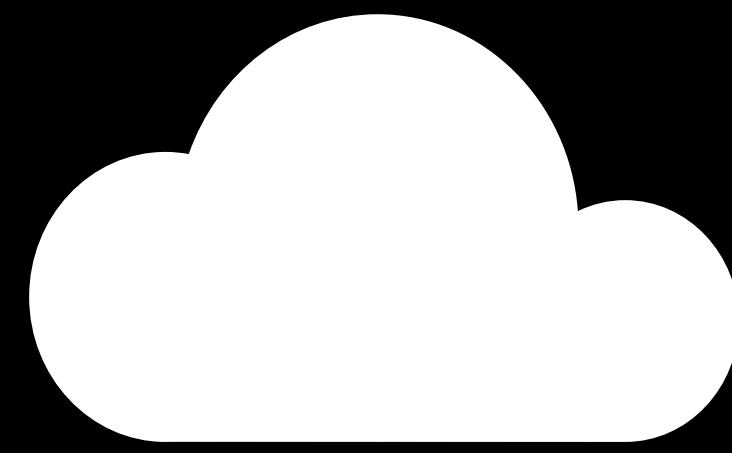
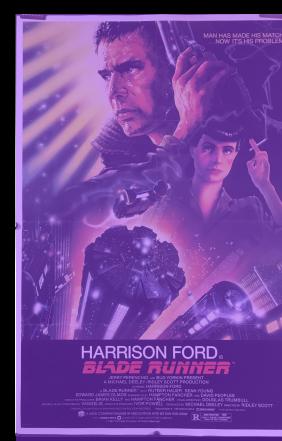
FINGERPRINT
FUNCTION



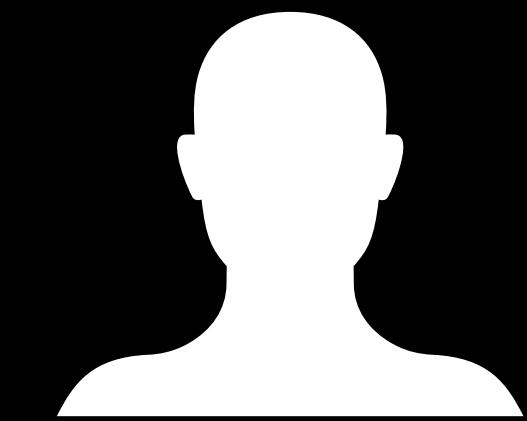
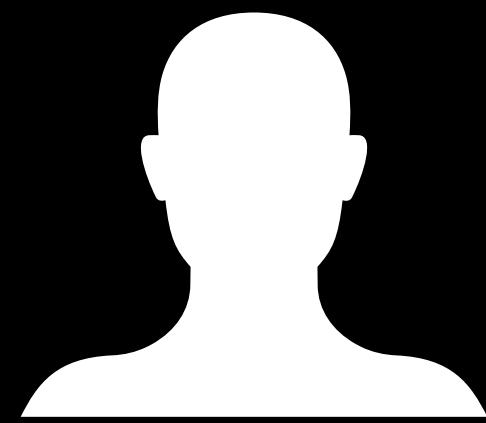
2893a22d
322c221a
1b2213de
2221a21a

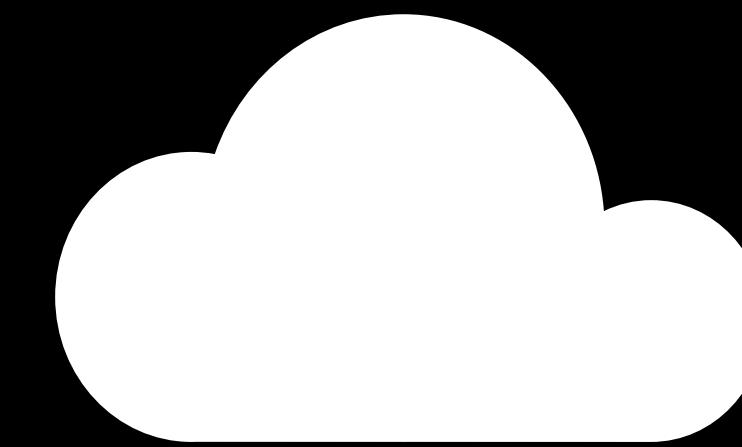
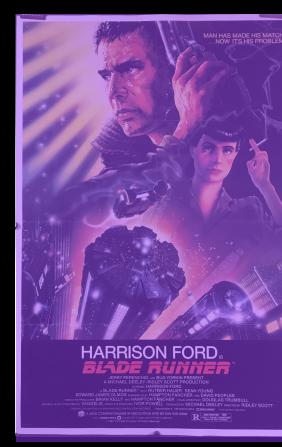
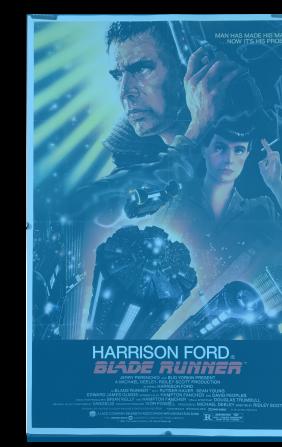


1188f67d
48c9f11a
fb857297
7ef74c5e

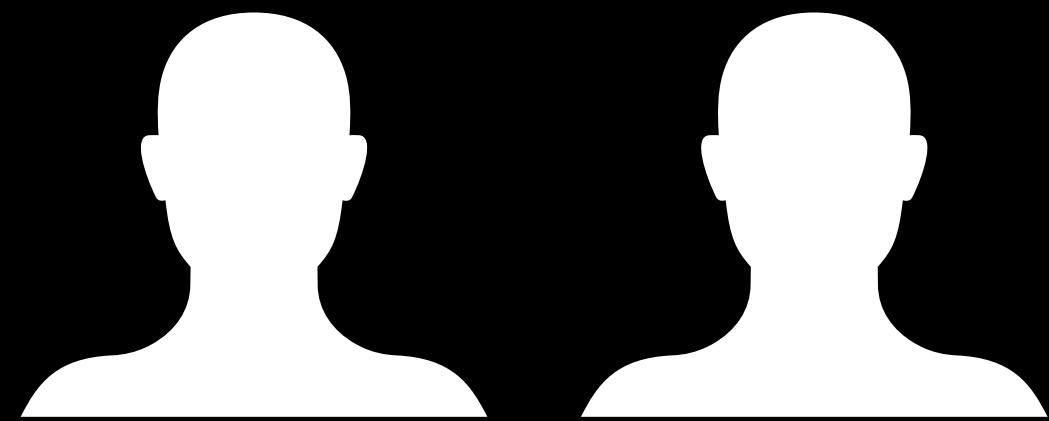


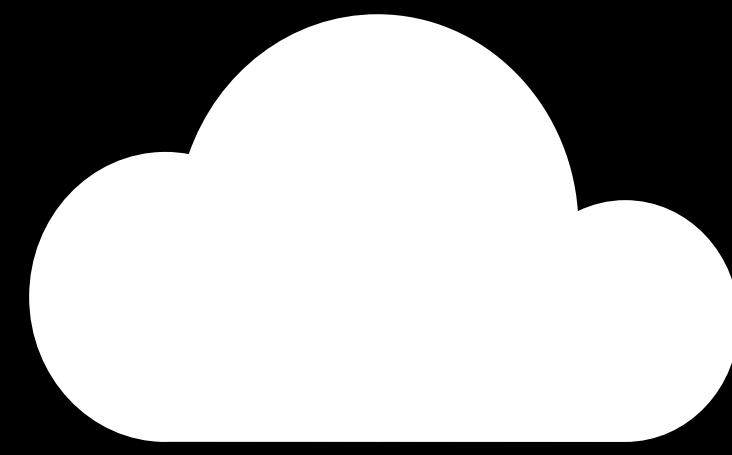
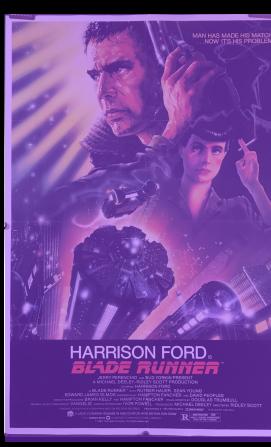
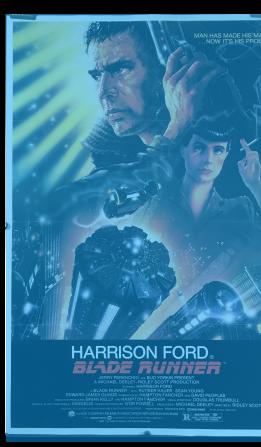
1188f67d	2893a22d
48c9f11a	322c221a
fb857297	1b2213de
7ef74c5e	2221a21a



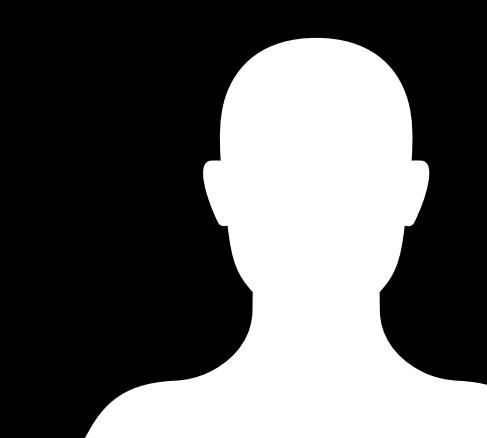
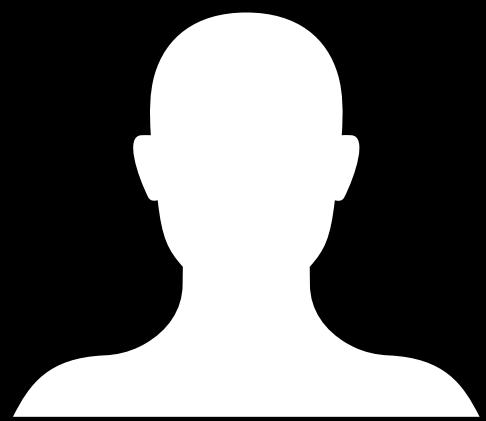
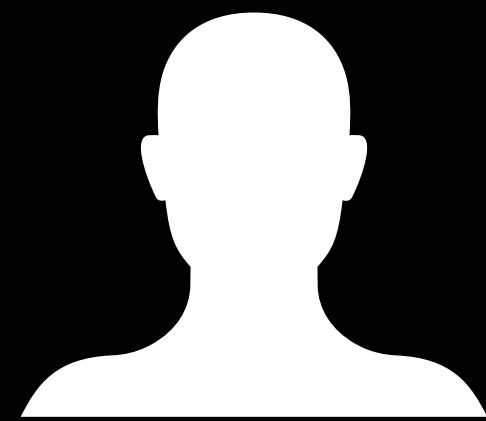
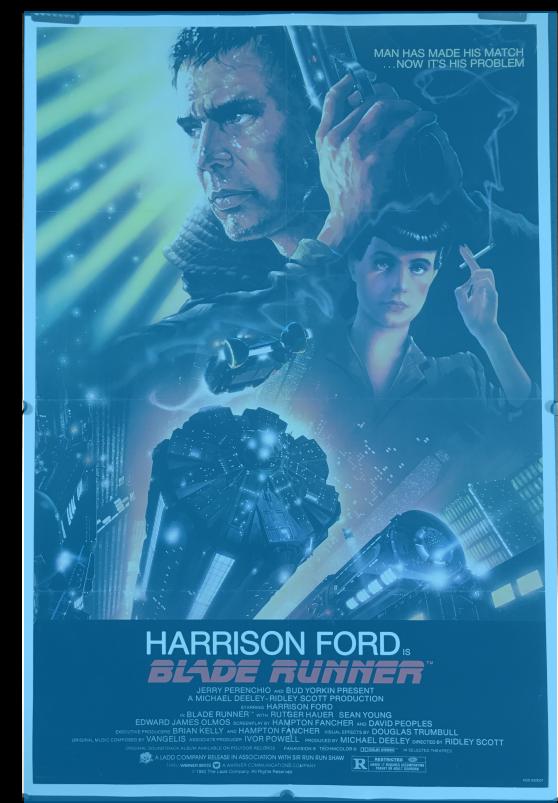


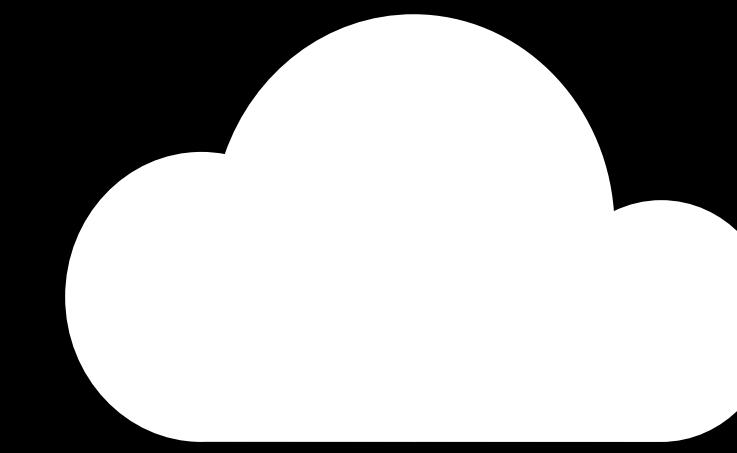
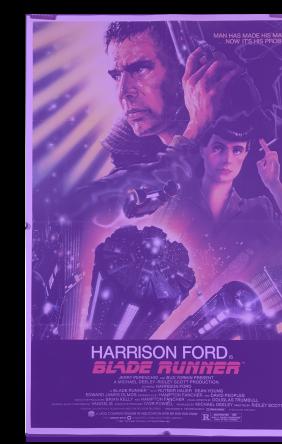
1188f67d	2893a22d
48c9f11a	322c221a
fb857297	1b2213de
7ef74c5e	2221a21a



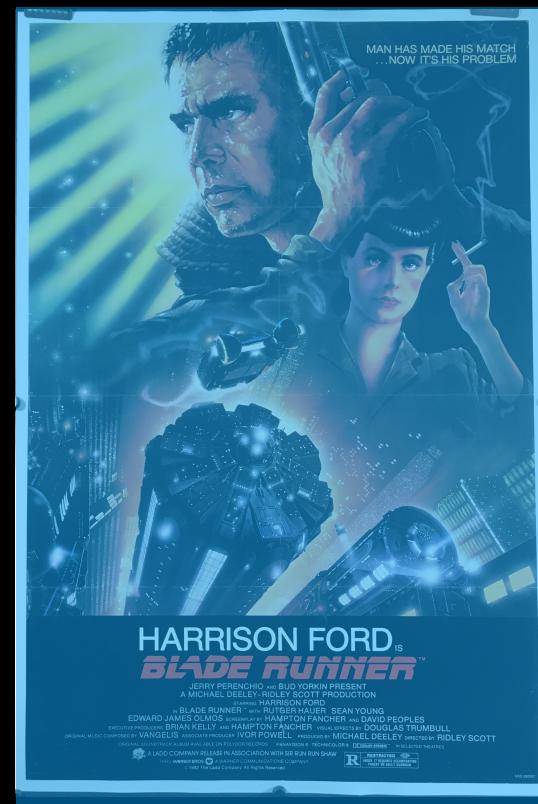


1188f67d	2893a22d
48c9f11a	322c221a
fb857297	1b2213de
7ef74c5e	2221a21a





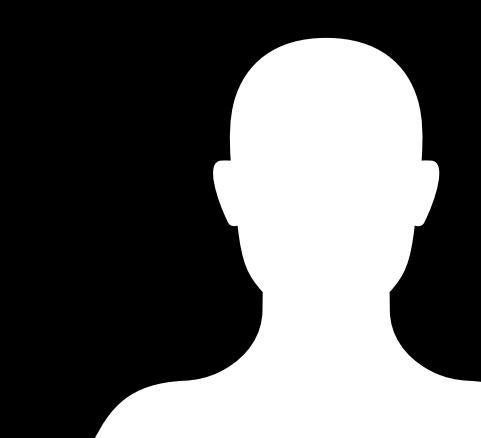
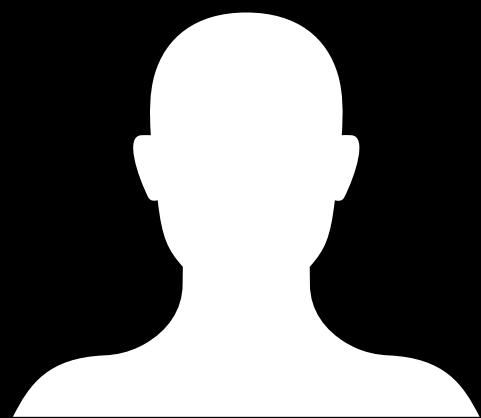
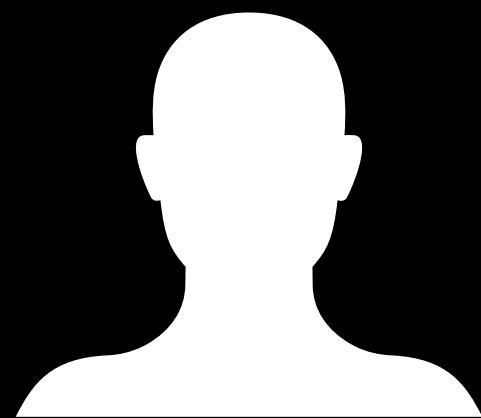
1188f67d 2893a22d
48c9f11a 322c221a
fb857297 1b2213de
7ef74c5e 2221a21a

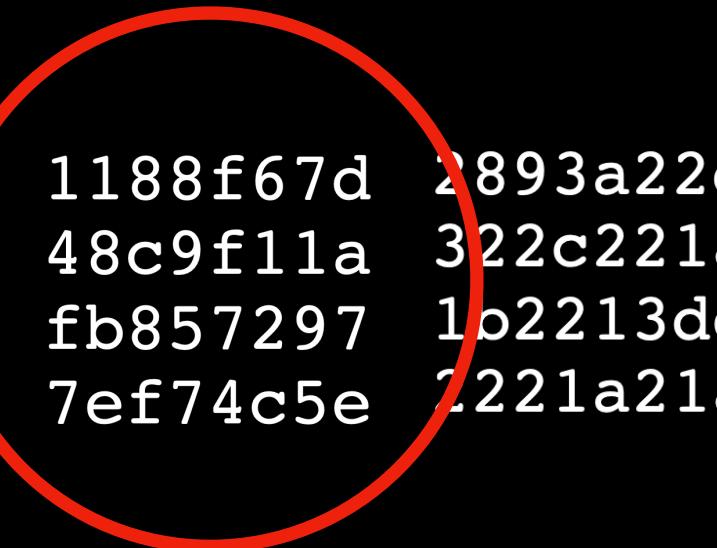
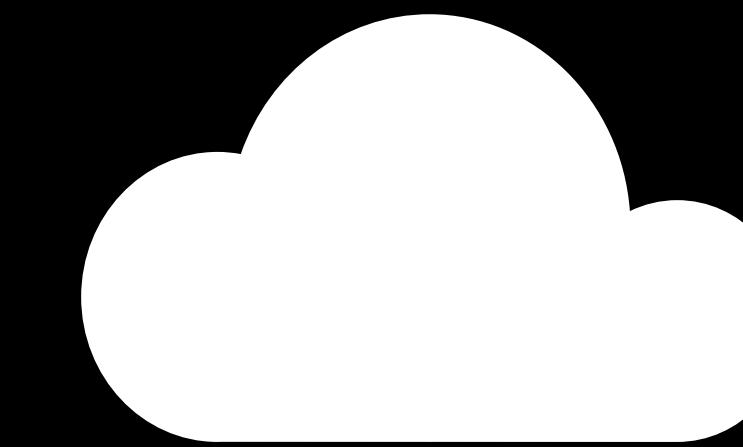
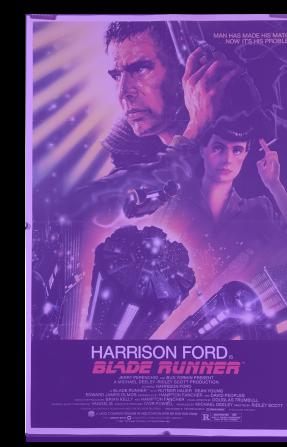
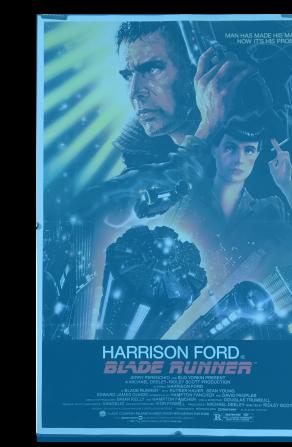


FINGERPRINT
FUNCTION

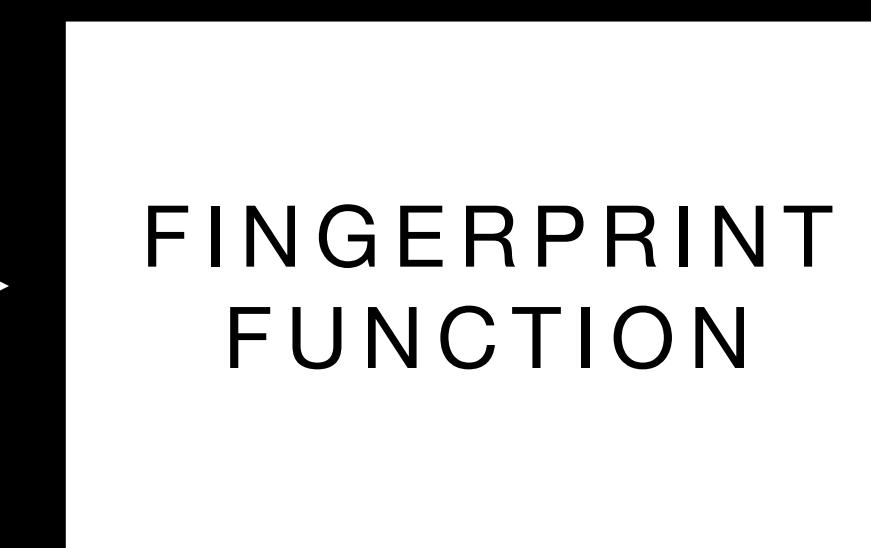
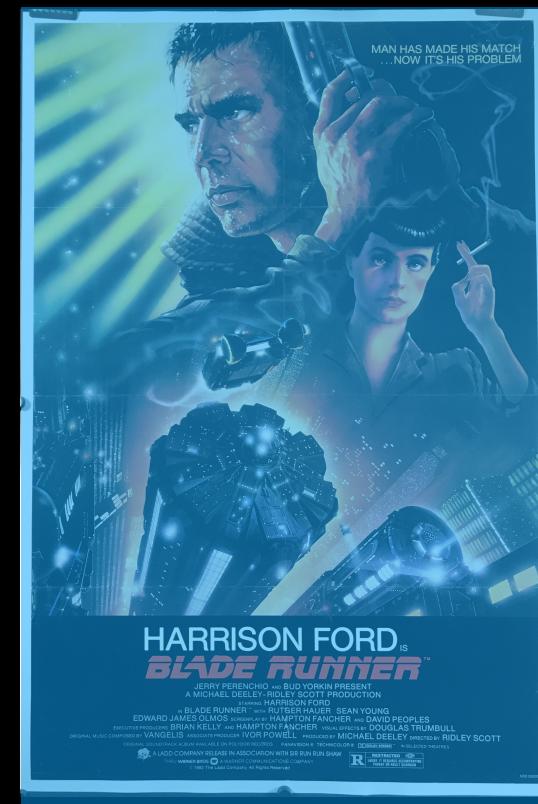


1188f67d
48c9f11a
fb857297
7ef74c5e

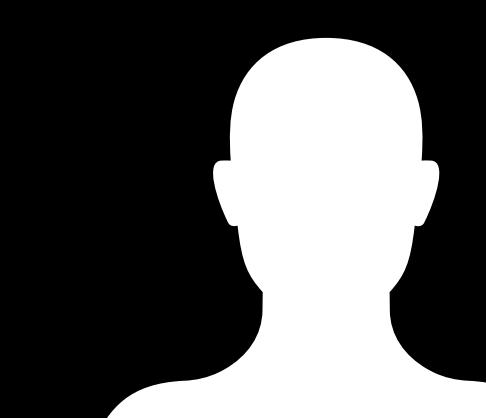
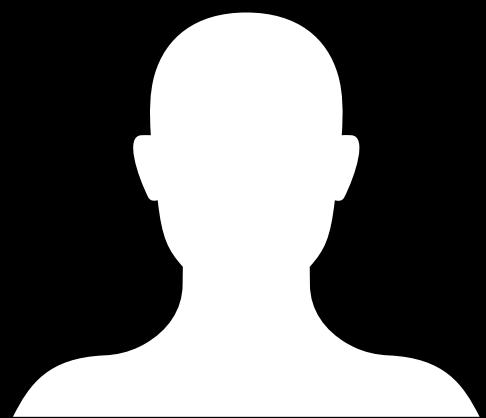
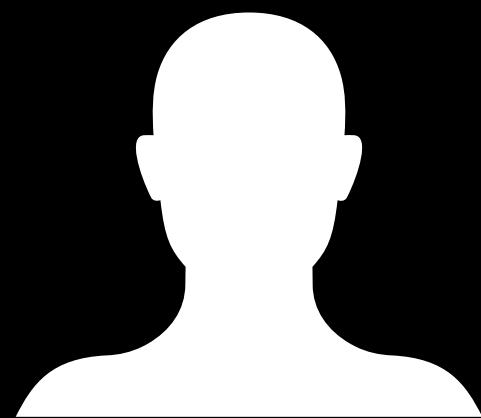


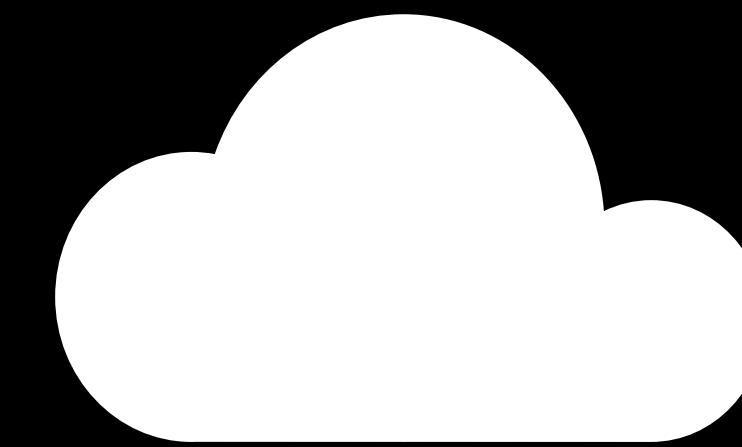
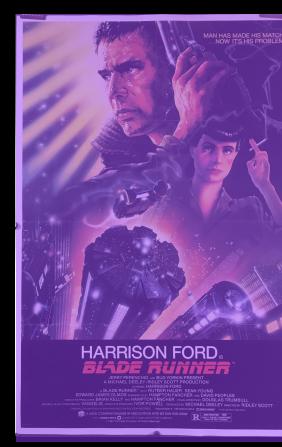
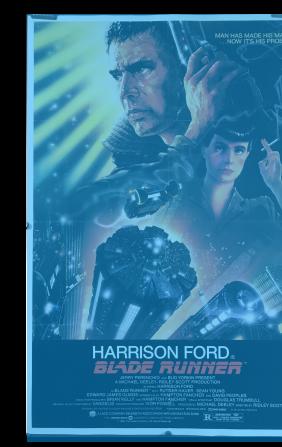


1188f67d 2893a22d
48c9f11a 322c221a
fb857297 1b2213de
7ef74c5e 2221a21a

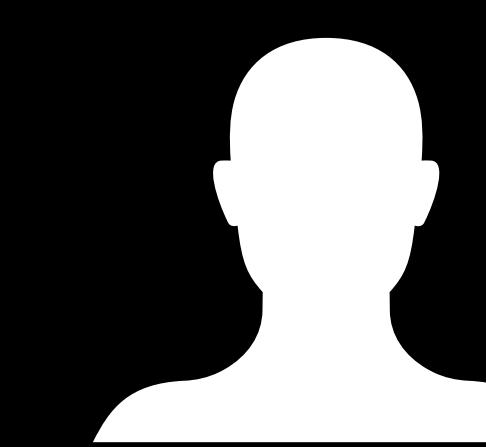
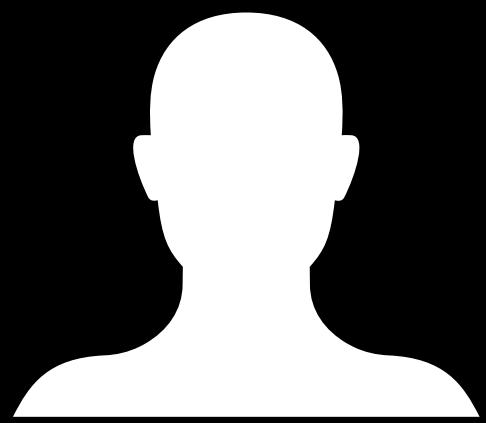
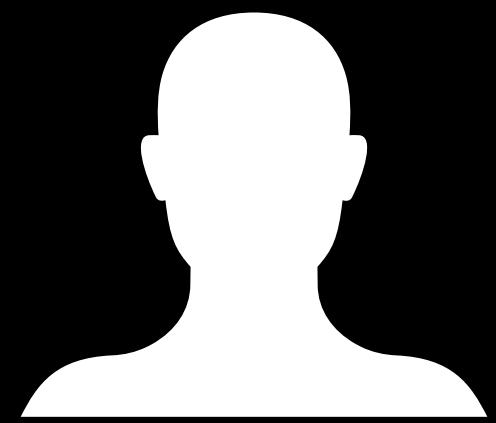


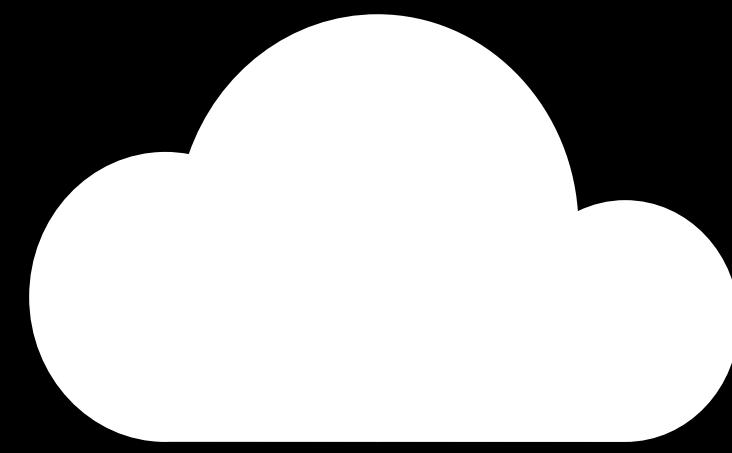
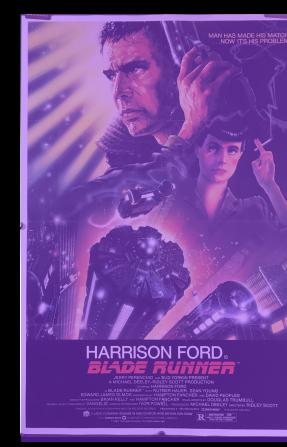
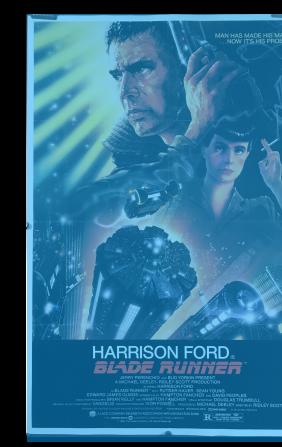
1188f67d
48c9f11a
fb857297
7ef74c5e



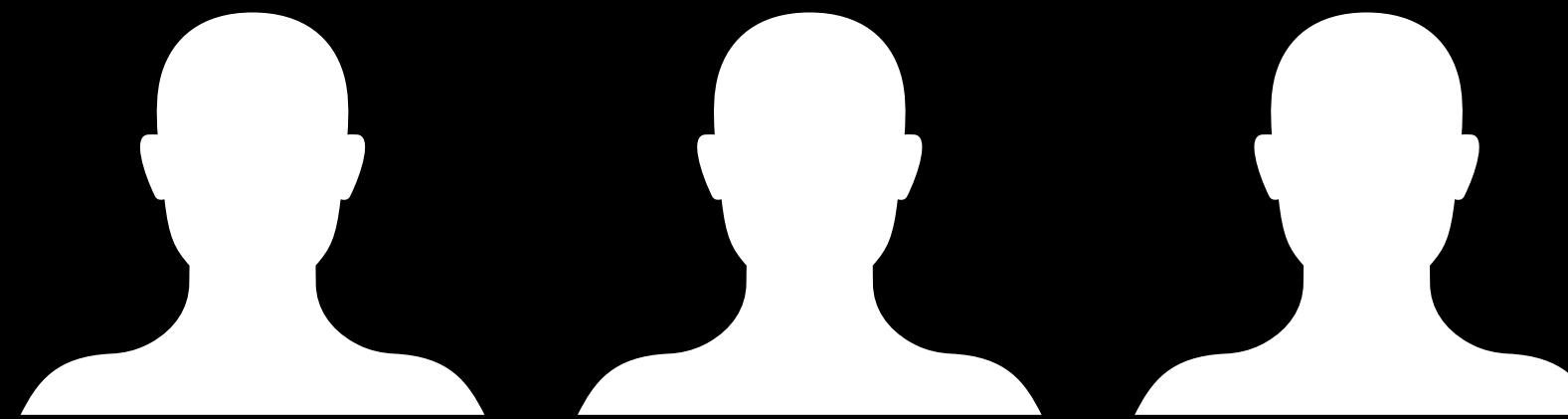


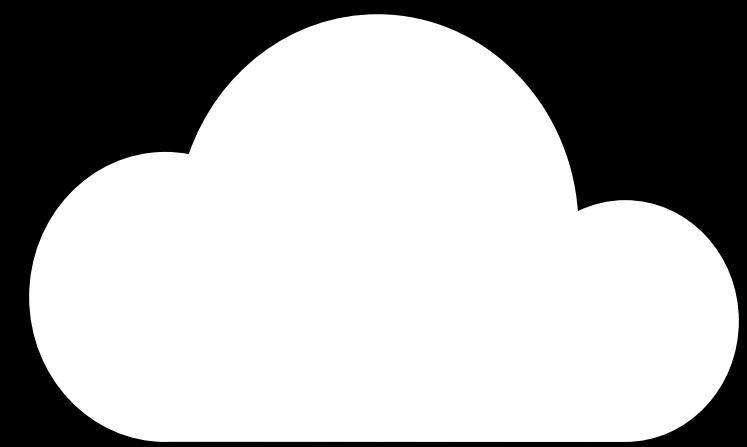
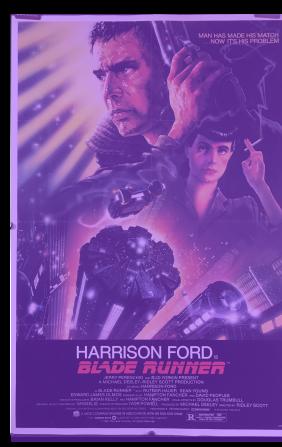
1188f67d	2893a22d
48c9f11a	322c221a
fb857297	1b2213de
7ef74c5e	2221a21a



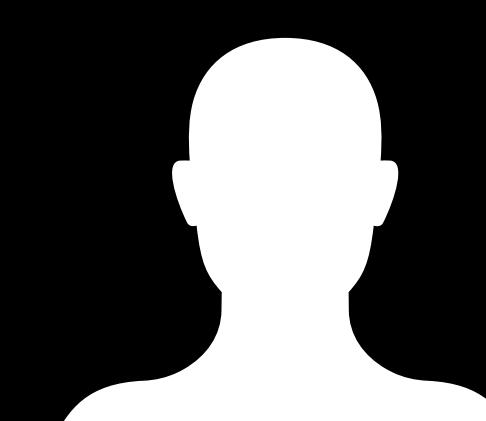
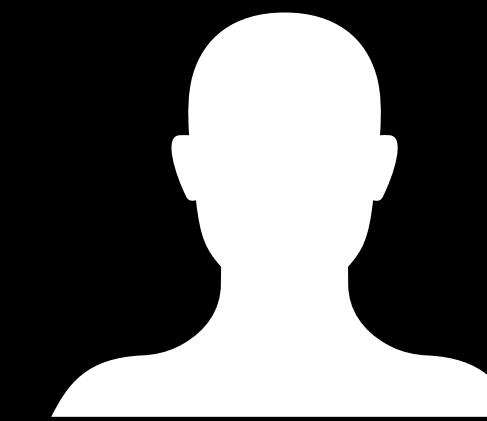
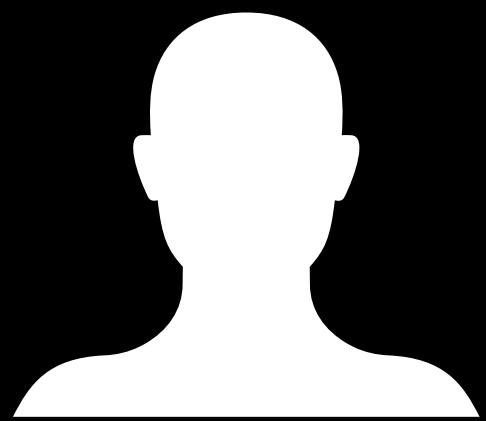
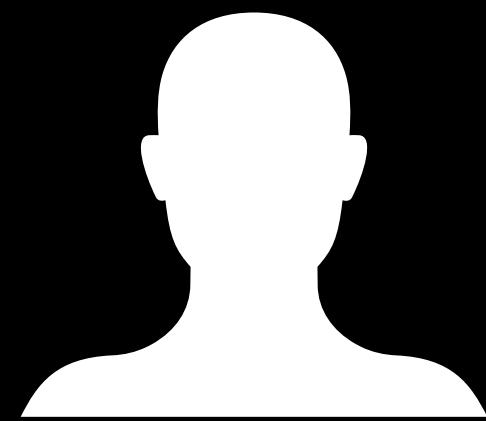
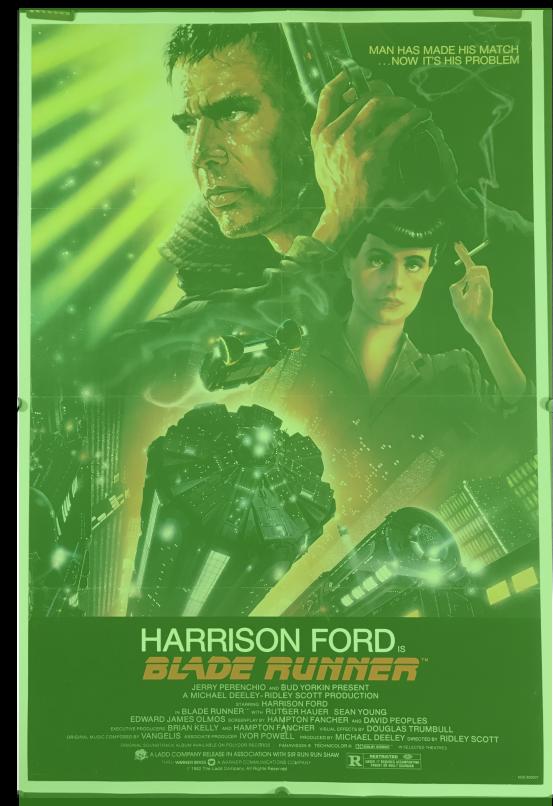


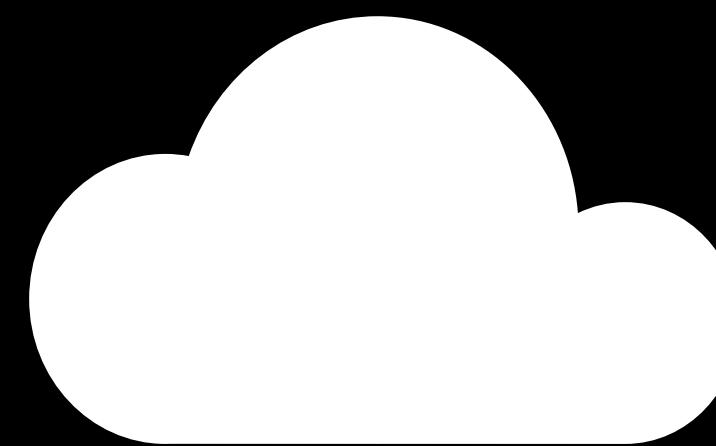
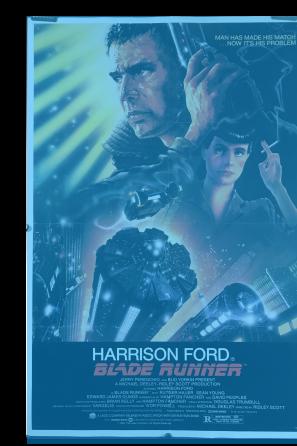
1188f67d	2893a22d
48c9f11a	322c221a
fb857297	1b2213de
7ef74c5e	2221a21a



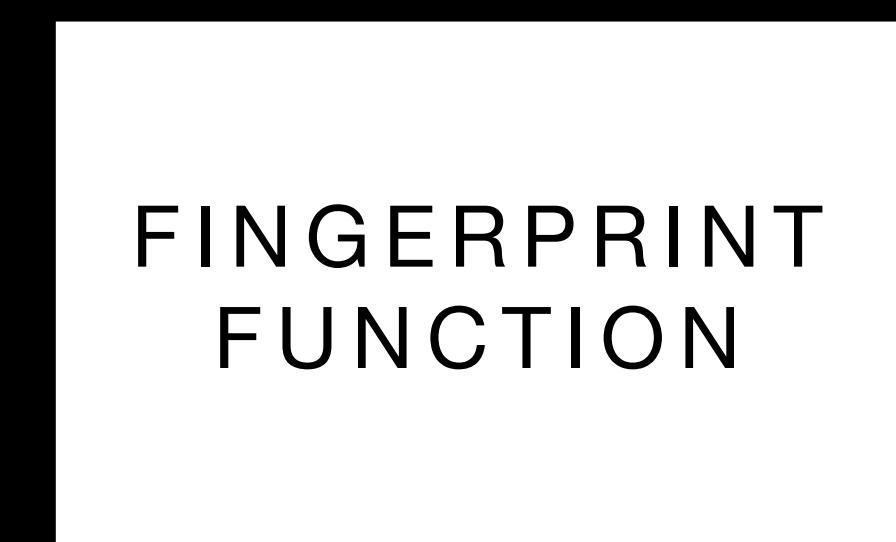
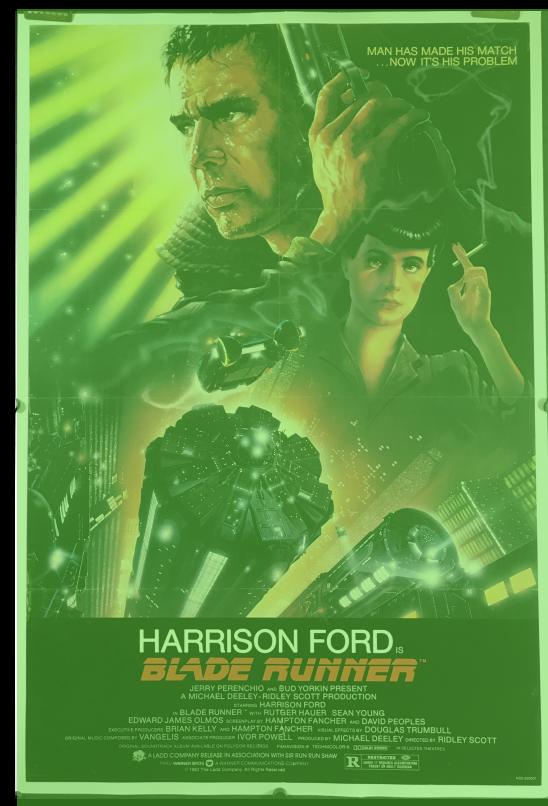


1188f67d	2893a22d
48c9f11a	322c221a
fb857297	1b2213de
7ef74c5e	2221a21a

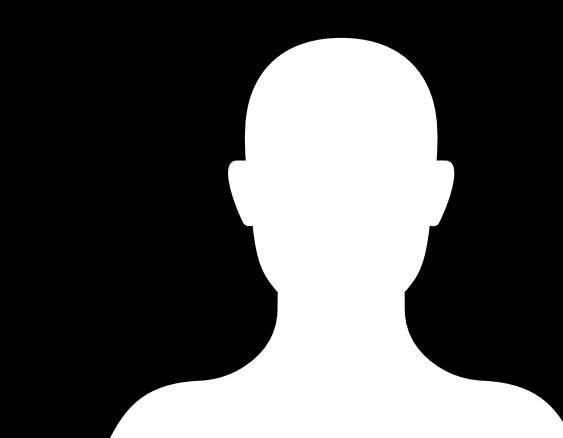
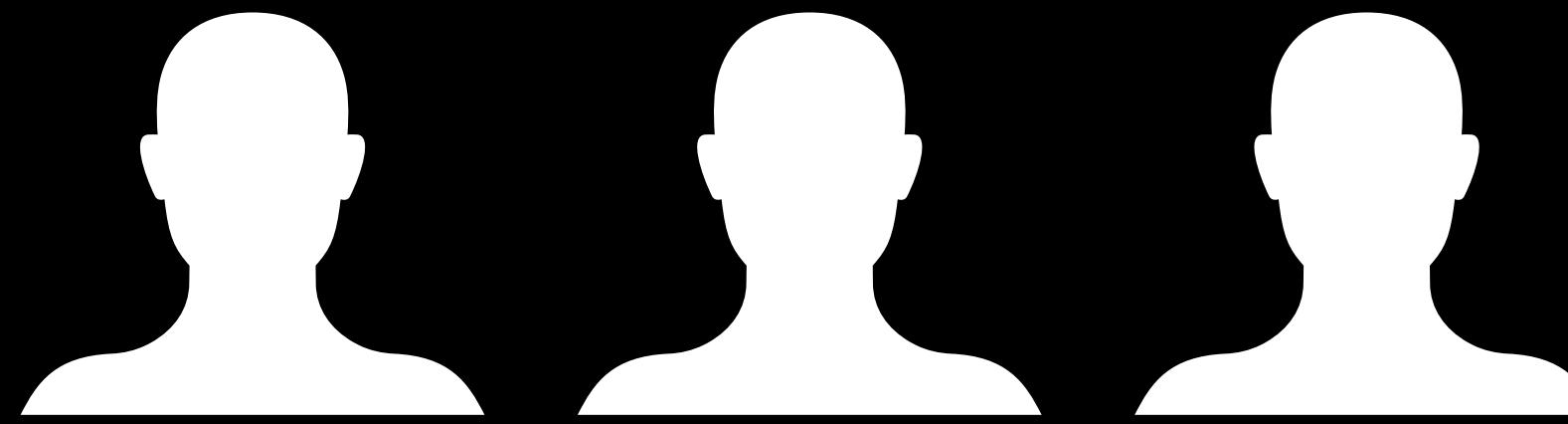


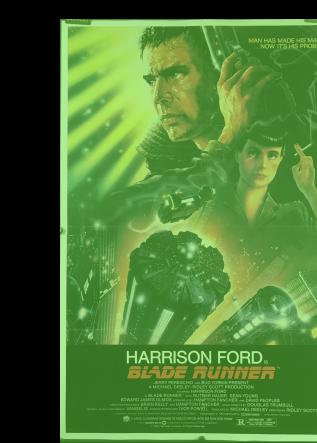
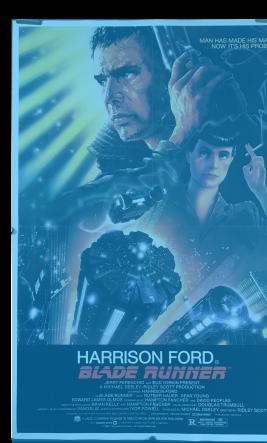


1188f67d 2893a22d
48c9f11a 322c221a
fb857297 1b2213de
7ef74c5e 2221a21a

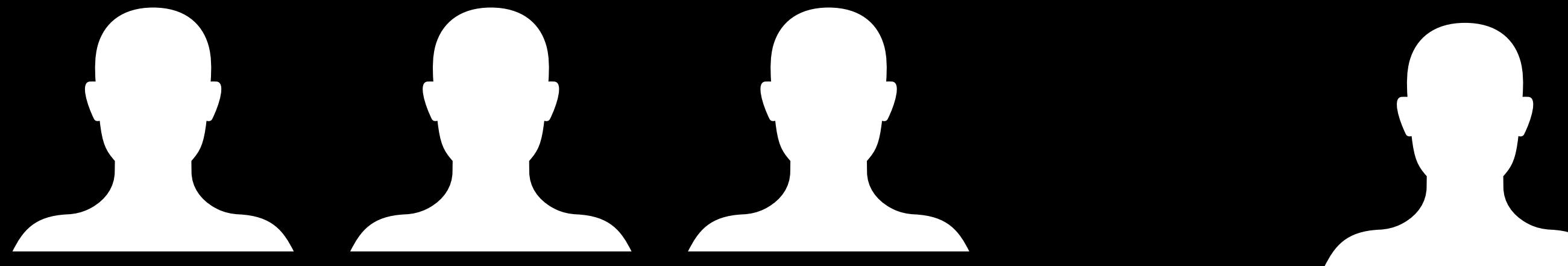


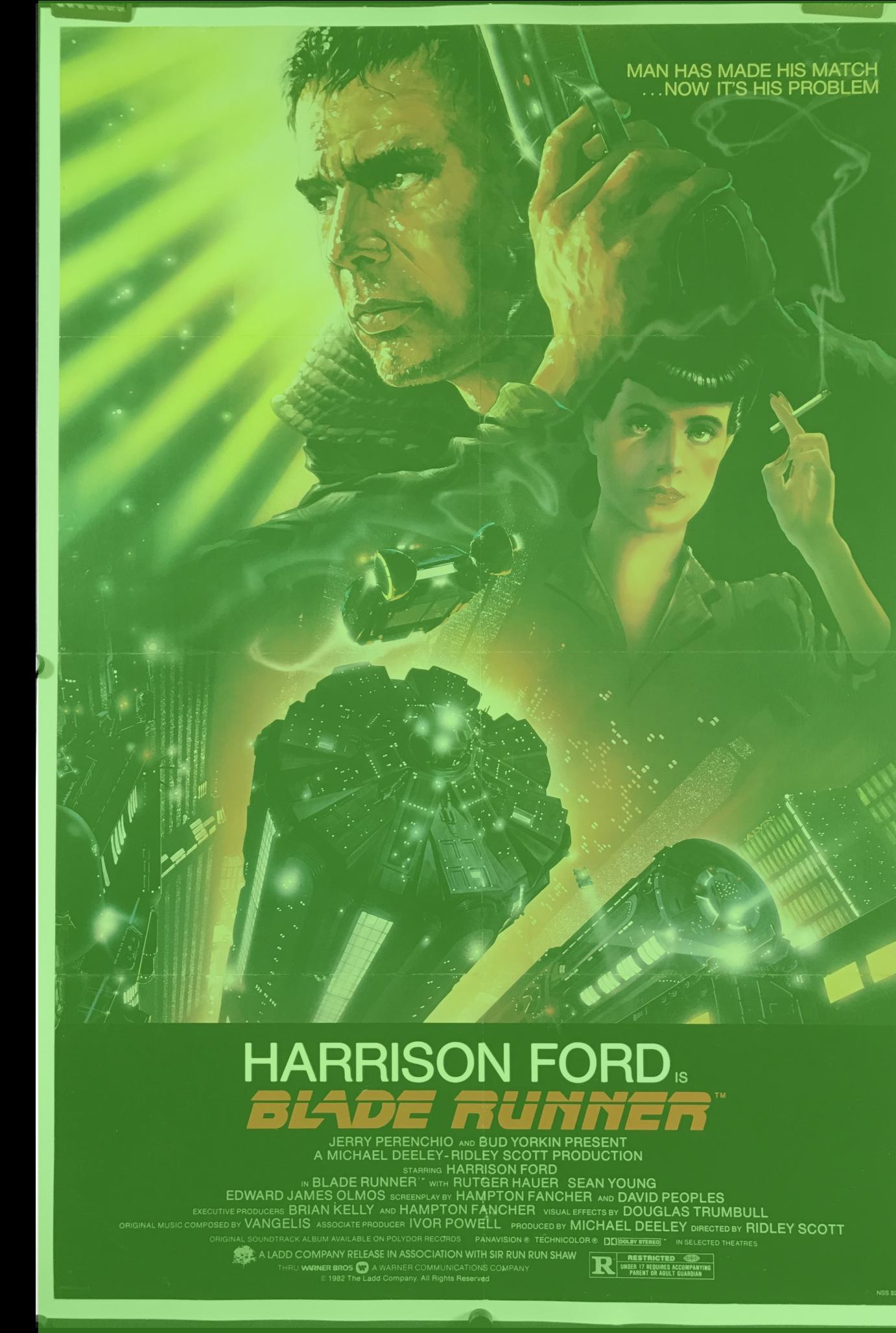
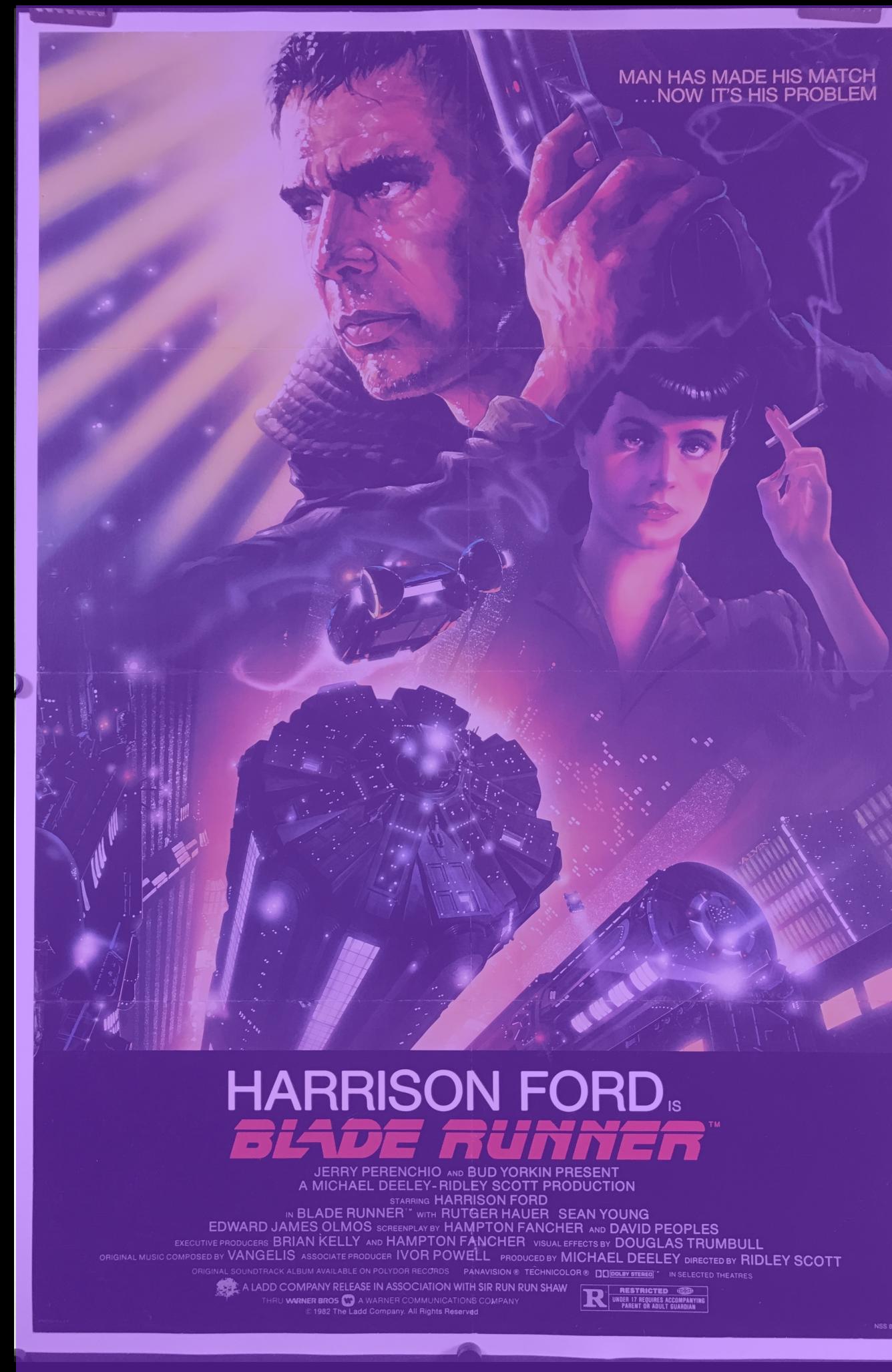
3626a882
6d7262c2
241e5261
a51412e2





1188f67d	2893a22d	3626a882
48c9f11a	322c221a	6d7262c2
fb857297	1b2213de	241e5261
7ef74c5e	2221a21a	a51412e2









HARRISON FORD is
BLADE RUNNER™

JERRY PERENCHIO AND BUD YORKIN PRESENT
A MICHAEL DEELEY-RIDLEY SCOTT PRODUCTION

STARRING HARRISON FORD

IN BLADE RUNNER™ WITH RUTGER HAUER SEAN YOUNG
EDWARD JAMES OLMO'S SCREENPLAY BY HAMPTON FANCHER AND DAVID PEOPLES
EXECUTIVE PRODUCERS BRIAN KELLY AND HAMPTON FANCHER VISUAL EFFECTS BY DOUGLAS TRUMBULL
VANGELIS ASSOCIATE PRODUCER IVOR POWELL PRODUCED BY MICHAEL DEELEY DIRECTED BY RIDLEY SCOTT

ORIGINAL SOUNDTRACK ALBUM AVAILABLE ON POLYDOR RECORDS PANAVISION® TECHNICOLOR® DOLBY STEREO® IN SELECTED THEATRES

THRU WARNER BROS A WARNER COMMUNICATIONS COMPANY

©1982 The Ladd Company. All Rights Reserved

R RESTRICTED
UNDER 17 REQUIRES ACCOMPANYING
PARENT OR ADULT GUARDIAN

NOS 600007





FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



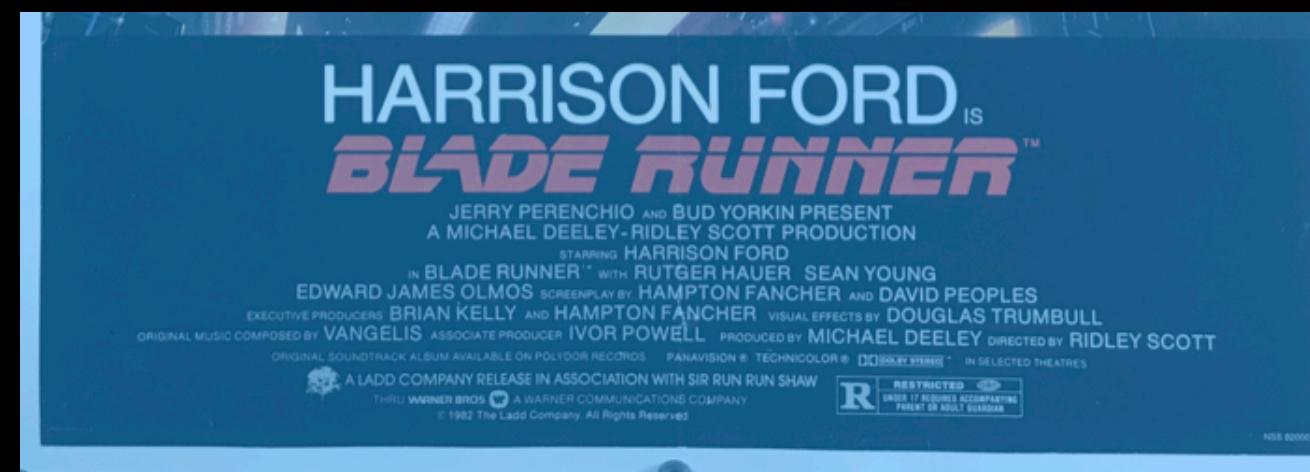
FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



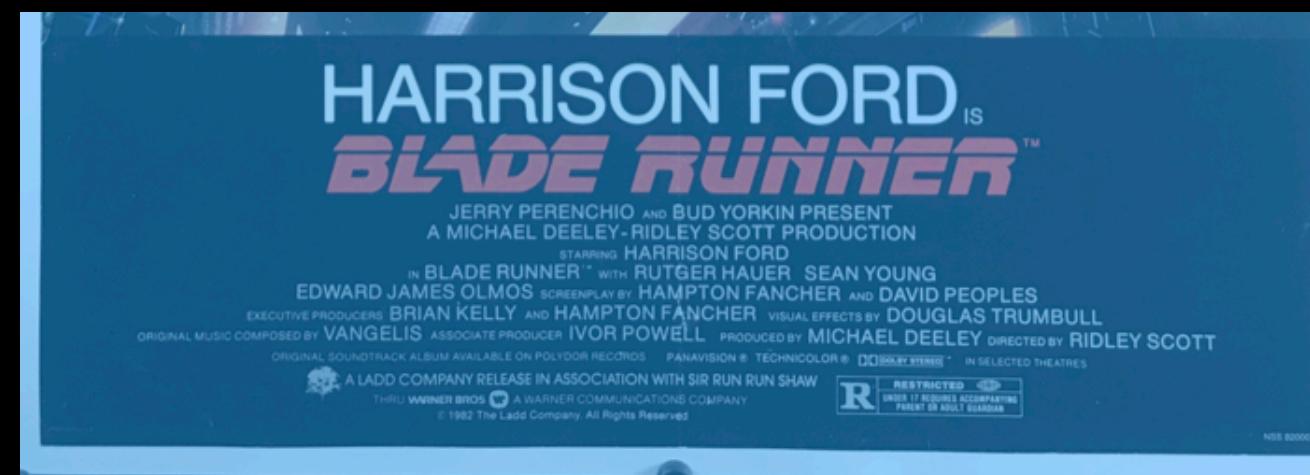
FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



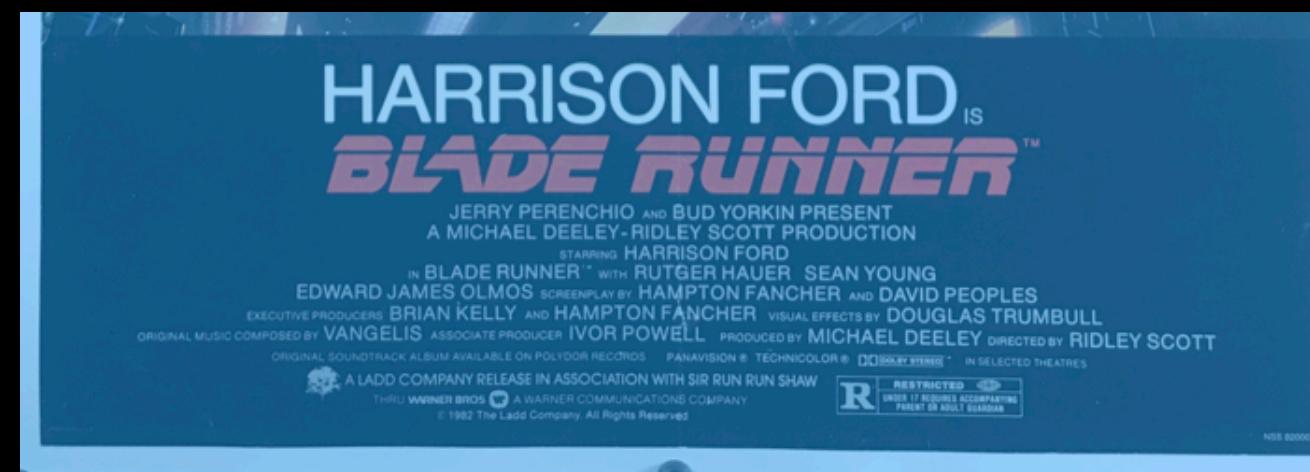
FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



FINGERPRINT
FUNCTION

3626a882
6d7262c2
241e5261
a51412e2



Data Deduplication

Business Continuity

- **Fingerprint** binary data using a one-way hash function from user.
- Determine if have previously encountered fingerprint.
 - If never encountered, store fingerprint of binary data and store binary data.
 - If encountered and already in store, discard binary data and associate existing binary data with user.

Data Deduplication

Business Continuity

- The process **reduces** the need to store lots of redundant data.
- Ensures enterprise is **not storing lots of duplicates** of data on its systems.
- **Reduces the expense in transmitting** all the data between the different parts of the system.
- The next step is to consider where the **process will actually occur**.

Source vs Target

Data Deduplication

- **Source-based deduplication** occurs on the client or source device before transmission.
 - Requires **source device to communicate** with the server to determine if the data already exists.
 - Requires specific client software solution.
- **Target-based deduplication** occurs on the actual server coordinating storage.
 - Efficiencies can be made without making any significant alterations to the source device, e.g. client software.

Privacy and Security concerns

Privacy and security concerns

Privacy and Security concerns

- Potential for attackers to gain **insight** into the data being stored by enterprises and employees in cloud computing.
- Consider two specific attacks, that is **identifying a specific file** and determining the **existence of a file**.
- Consider the potential as well from an implementation perspective of attackers stealing the **fingerprint**.

Privacy and security concerns

Privacy and Security concerns

- Need to determine the **ownership** or **existence** of a specific file.
- Consider an **indecent image** of child pornography or even a celebrity image.
- Alice can use **deduplication** to see if a specific binary data exists on a given service.
- Alice could obtain a court order to demand the cloud service provider reveal the **identity** of the owner, Bob.

Salary Attack

Privacy and Security concerns

- Consider the number of **standard letters** that an enterprise generates and dispatches.
- Standard letters have **small alterations** - consider PIN reminder letters, pay or medical letters.
- Alice could brute-force **generate** such letters, altering the specifics of them to determine the existence of such letters.
- Reasonable attack if the alterations are relatively restricted, becomes difficult if they are varied.

Solutions

Privacy and Security concerns

- Enterprises and employees could **encrypt data** to prevent it from being deduplicated.
- Depending on encryption approach, concerns from cloud service providers that it makes it difficult to **benefit** from the nature of the cloud.
 - Ignore small files and instead **focus on larger files** that would still result in efficiency savings.
 - Consider the reality that keeping **rarer files private** is arguably more important than common files.

Data Deduplication

Business Continuity