JCL SORT on PS File: Removing Duplicates and Sorting Records

Mastering JCL SORT is essential for efficient data management in mainframe environments. This presentation covers practical techniques for cleaning and organizing Physical Sequential (PS) files.

Introduction to PS Files and ISPF

Physical Sequential (PS) Files: The most basic dataset type on mainframes. Records are stored one after another in the order they are written.

- Simple and efficient for sequential processing.
- Commonly used for logs, reports, and input/output for batch jobs.

ISPF 3.2 (Dataset Utility): An Interactive System Productivity Facility (ISPF) option used for managing datasets.

- Create, delete, rename, and browse datasets.
- Manual file creation for testing and development.

The Data Challenge: Problem Statement

1

Initial State

We have a **Physical Sequential (PS) file** containing 100 records.

2

Record Format

Each record is 80 bytes long.

The **primary key** for these records resides in columns 13-20.

3

Data Issues

The file contains **10 duplicate records**, resulting in only 90 unique entries.

Records are currently in **random order**.

Our Objectives

Eliminate Duplicates

Ensure each record in the final output is unique based on its content.

Sort by Primary Key

Arrange all unique records in ascending order based on the primary key (columns 13-20).

Creating the PS File Manually

Before processing, we need a sample PS file. This can be done using ISPF option 3.2.

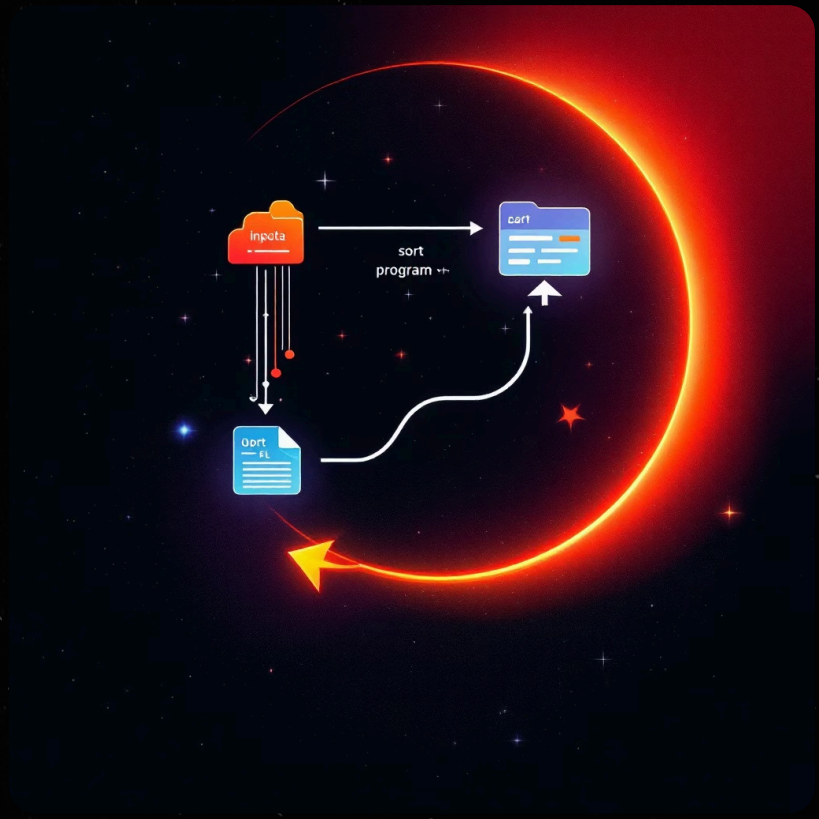
- Navigate to the **ISPF Primary Option Menu** and select option 3 (Utilities).
- Then select option 2 (Dataset).
- Specify the dataset name (e.g., `'YOUR.TEST.PSFILE'`) and select option 'A' (Allocate New Dataset).

Set **Record Format (RECFM)** to FB (Fixed Block), **Logical Record Length (LRECL)** to 80, and **Block Size (BLKSIZE)** to an appropriate multiple of 80 (e.g., 0 for system determination).

Understanding JCL SORT

JCL SORT is a powerful utility program on mainframes for sorting, merging, copying, and manipulating data in datasets.

- **Efficiency:** Optimized for large volumes of data.
- **Flexibility:** Supports various data types and complex sorting criteria.
- **Control Statements:** Uses specific control statements (e.g., SORT, OUTFIL, INREC, OMIT) to define operations.



It's integral to many batch processing workflows for data preparation.

JCL SORT Code for Deduping & Sorting

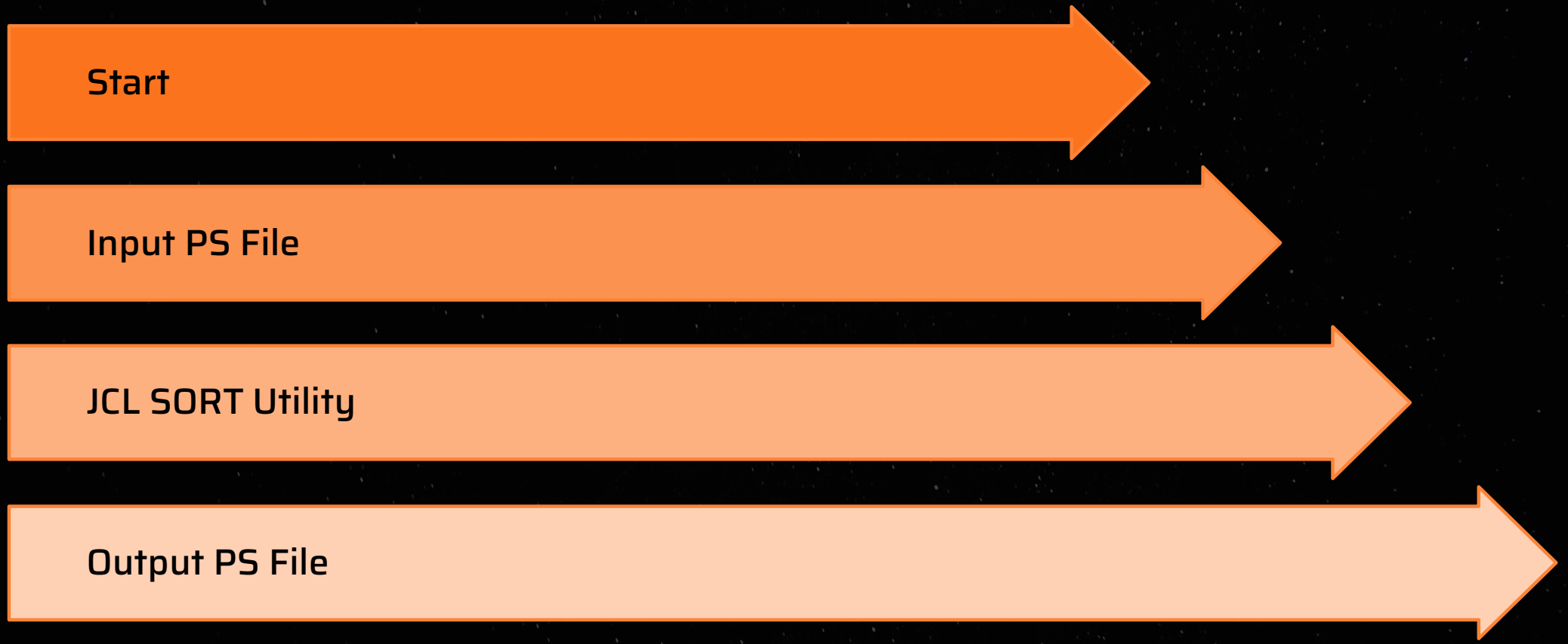
This JCL snippet uses DFSORT (IBM's sort utility) to achieve our objectives.

```
//SORTSAMP JOB (ACCOUNT,ROOM),'DEDUP_SORT',CLASS=A,MSGCLASS=X
//STEP01   EXEC PGM=SORT
//SYSOUT   DD SYSOUT=*
//SORTIN   DD DSN=<YOUR.TEST.PSFILE>,DISP=SHR
//SORTOUT  DD DSN=<YOUR.DEDUP.SORTED.PSFILE>,
//            DISP=(NEW,CATLG,DELETE),
//            DCB=(RECFM=FB,LRECL=80,BLKSIZE=0),
//            SPACE=(CYL,(1,1),RLSE)
//SYSIN    DD *
  SORT FIELDS=(13,8,CH,A)
  SUM FIELDS=NONE
/*
//
```

SORT FIELDS=(13,8,CH,A): Sorts from position 13 for 8 bytes, character format, ascending order.

SUM FIELDS=NONE: This is the key to deduplication. It instructs DFSORT to retain only one record for each unique sort key, effectively removing duplicates.

Execution Flow



The JCL job directs the DFSORT program to read the specified input PS file, apply the sorting and deduplication logic, and write the processed data to a new output PS file.

Output Example: Before & After

Original PS File (Random Order, Duplicates)

```
Record 1  ItemB   Key00005
Record 2  ItemA   Key00001
Record 3  ItemC   Key00010
Record 4  ItemA   Key00001
Record 5  ItemB   Key00005
Record 6  ItemZ   Key00003
```

Sorted PS File (Unique, Sorted by Key)

```
Record 2  ItemA   Key00001
Record 6  ItemZ   Key00003
Record 1  ItemB   Key00005
Record 3  ItemC   Key00010
```

Notice how **Key00001** and **Key00005** duplicates are removed, and all records are ordered by their key.

Key Takeaways & Conclusion

1 JCL SORT Power

A fundamental tool for data manipulation, crucial for mainframe batch processing.

2 Efficient Deduping

SUM FIELDS=NONE provides a simple yet powerful method for eliminating duplicate records.

3 Data Integrity

Sorting and deduplication ensure data accuracy and improve processing efficiency for downstream applications.

Mastering these JCL SORT techniques is vital for any computer engineering student working with mainframe data.