

# book\_to\_slide\_BY\_sections\_V5 copy

July 6, 2025

## 1 Set up Paths

```
[1]: # Cell 1: Setup and Configuration
import os
import re
import logging
import warnings
from docx import Document
import pdfplumber
import ollama
from tenacity import retry, stop_after_attempt, wait_exponential, RetryError
import json

# Setup Logger for this cell
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %
↳ %(message)s')
logger = logging.getLogger(__name__)

# --- 1. CORE SETTINGS ---
# Set this to True for EPUB, False for PDF. This controls the entire notebook's
↳ flow.
PROCESS_EPUB = True # for EPUB
# PROCESS_EPUB = False # for PDF

# --- 2. INPUT FILE NAMES ---
# The name of the Unit Outline file (e.g., DOCX, PDF)
UNIT_OUTLINE_FILENAME = "ICT312 Digital Forensic_Final.docx" # epub
# UNIT_OUTLINE_FILENAME = "ICT311 Applied Cryptography.docx" # pdf

EXTRACT_UO = False

# The names of the book files
EPUB_BOOK_FILENAME = "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide
↳ to Computer Forensics and Investigations_ Processing Digital
↳ Evidence-Cengage Learning (2018).epub"
```

```

PDF_BOOK_FILENAME = "(Chapman & Hall_CRC Cryptography and Network Security_
↳Series) Jonathan Katz, Yehuda Lindell - Introduction to Modern_
↳Cryptography-CRC Press (2020).pdf"

# --- 3. DIRECTORY STRUCTURE ---
# Define the base path to your project to avoid hardcoding long paths everywhere
PROJECT_BASE_DIR = "/home/sebas_dev_linux/projects/course_generator"

# Define subdirectories relative to the base path
DATA_DIR = os.path.join(PROJECT_BASE_DIR, "data")
PARSE_DATA_DIR = os.path.join(PROJECT_BASE_DIR, "Parse_data")

# Construct full paths for clarity
INPUT_UO_DIR = os.path.join(DATA_DIR, "UO")
INPUT_BOOKS_DIR = os.path.join(DATA_DIR, "books")
OUTPUT_PARSED_UO_DIR = os.path.join(PARSE_DATA_DIR, "Parse_UO")
OUTPUT_PARSED_TOC_DIR = os.path.join(PARSE_DATA_DIR, "Parse_TOC_books")
OUTPUT_DB_DIR = os.path.join(DATA_DIR, "DataBase_Chroma")

# --- 4. LLM & EMBEDDING CONFIGURATION ---
LLM_PROVIDER = "ollama" # Can be "ollama", "openai", "gemini"
OLLAMA_HOST = "http://localhost:11434"
OLLAMA_MODEL = "qwen3:8b" # "qwen3:8b", #"mistral:latest"
EMBEDDING_MODEL_OLLAMA = "nomic-embed-text"
CHUNK_SIZE = 800
CHUNK_OVERLAP = 100

# --- 5. DYNAMICALLY GENERATED PATHS & IDs (DO NOT EDIT THIS SECTION) ---
# This section uses the settings above to create all the necessary variables_
↳for later cells.

# Extract Unit ID from the filename
def print_header(text: str, char: str = "="):
    """Prints a centered header to the console."""
    print("\n" + char * 80)
    print(text.center(80))
    print(char * 80)

def extract_uo_id_from_filename(filename: str) -> str:
    match = re.match(r'^[A-Z]+\d+', os.path.basename(filename))
    if match:
        return match.group(0)
    raise ValueError(f"Could not extract a valid Unit ID from filename:_
↳'{filename}'")

try:
    UNIT_ID = extract_uo_id_from_filename(UNIT_OUTLINE_FILENAME)

```

```

except ValueError as e:
    print(f"Error: {e}")
    UNIT_ID = "UNKNOWN_ID"

# Full path to the unit outline file
FULL_PATH_UNIT_OUTLINE = os.path.join(INPUT_UO_DIR, UNIT_OUTLINE_FILENAME)

# Determine which book and output paths to use based on the PROCESS_EPUB flag
if PROCESS_EPUB:
    BOOK_PATH = os.path.join(INPUT_BOOKS_DIR, EPUB_BOOK_FILENAME)
    PRE_EXTRACTED_TOC_JSON_PATH = os.path.join(OUTPUT_PARSED_TOC_DIR,
    ↪f"{UNIT_ID}_epub_table_of_contents.json")
else:
    BOOK_PATH = os.path.join(INPUT_BOOKS_DIR, PDF_BOOK_FILENAME)
    PRE_EXTRACTED_TOC_JSON_PATH = os.path.join(OUTPUT_PARSED_TOC_DIR,
    ↪f"{UNIT_ID}_pdf_table_of_contents.json")

# Define paths for the vector database
file_type_suffix = 'epub' if PROCESS_EPUB else 'pdf'
CHROMA_PERSIST_DIR = os.path.join(OUTPUT_DB_DIR,
    ↪f"chroma_db_toc_guided_chunks_{file_type_suffix}")
CHROMA_COLLECTION_NAME = f"book_toc_guided_chunks_{file_type_suffix}_v2"

# Define path for the parsed unit outline
PARSED_UO_JSON_PATH = os.path.join(OUTPUT_PARSED_UO_DIR, f"{os.path.
    ↪splitext(UNIT_OUTLINE_FILENAME)[0]}_parsed.json")

# --- Sanity Check Printout ---
print("--- CONFIGURATION SUMMARY ---")
print(f"Processing Mode: {'EPUB' if PROCESS_EPUB else 'PDF'}")
print(f"Unit ID: {UNIT_ID}")
print(f"Unit Outline Path: {FULL_PATH_UNIT_OUTLINE}")
print(f"Book Path: {BOOK_PATH}")
print(f"Parsed UO Output Path: {PARSED_UO_JSON_PATH}")
print(f"Parsed ToC Output Path: {PRE_EXTRACTED_TOC_JSON_PATH}")
print(f"Vector DB Path: {CHROMA_PERSIST_DIR}")
print(f"Vector DB Collection: {CHROMA_COLLECTION_NAME}")
print("--- SETUP COMPLETE ---")

```

--- CONFIGURATION SUMMARY ---

Processing Mode: EPUB

Unit ID: ICT312

Unit Outline Path:

/home/sebas\_dev\_linux/projects/course\_generator/data/UO/ICT312 Digital  
Forensic\_Final.docx

Book Path: /home/sebas\_dev\_linux/projects/course\_generator/data/books/Bill

Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and

```

Investigations_ Processing Digital Evidence-Cengage Learning (2018).epub
Parsed UO Output Path:
/home/sebas_dev_linux/projects/course_generator/Parse_data/Parse_UO/ICT312
Digital Forensic_Final_parsed.json
Parsed ToC Output Path: /home/sebas_dev_linux/projects/course_generator/Parse_da
ta/Parse_TOC_books/ICT312_epub_table_of_contents.json
Vector DB Path: /home/sebas_dev_linux/projects/course_generator/data/DataBase_Ch
roma/chroma_db_toc_guided_chunks_epub
Vector DB Collection: book_toc_guided_chunks_epub_v2
--- SETUP COMPLETE ---

```

## 2 System Prompt

```

[2]: UNIT_OUTLINE_SYSTEM_PROMPT_TEMPLATE = """
You are an expert academic assistant tasked with parsing a university unit_
↳outline document and extracting key information into a structured JSON_
↳format.

The input will be the raw text content of a unit outline. Your goal is to_
↳identify and extract the following details and structure them precisely as_
↳specified in the JSON schema below. Note: do not change any key name

**JSON Output Schema:**

```json
{{
  "unitInformation": {{
    "unitCode": "string | null",
    "unitName": "string | null",
    "creditPoints": "integer | null",
    "unitRationale": "string | null",
    "prerequisites": "string | null"
  }},
  "learningOutcomes": [
    "string"
  ],
  "assessments": [
    {{
      "taskName": "string",
      "description": "string",
      "dueWeek": "string | null",
      "weightingPercent": "integer | null",
      "learningOutcomesAssessed": "string | null"
    }}
  ],
  "weeklySchedule": [

```

```

    {{
      "week": "string",
      "contentTopic": "string",
      "requiredReading": "string | null"
    }}
  ],
  "requiredReadings": [
    "string"
  ],
  "recommendedReadings": [
    "string"
  ]
}}

```

Instructions for Extraction:

Unit Information: Locate Unit Code, Unit Name, Credit Points. Capture 'Unit\_Overview / Rationale' as unitRationale. Identify prerequisites.

Learning Outcomes: Extract each learning outcome statement.

Assessments: Each task as an object. Capture full task name, description, Due\_Week, Weighting % (number), and Learning Outcomes Assessed.

weeklySchedule: Each week as an object. Capture Week, contentTopic, and\_requiredReading.

Required and Recommended Readings: List full text for each.

**\*\*Important Considerations for the LLM\*\*:**

Pay close attention to headings and table structures.

If information is missing, use null for string/integer fields, or an empty list\_[] for array fields.

Do not change keys in the template given

Ensure the output is ONLY the JSON object, starting with {{{ and ending with\_}}}. No explanations or conversational text before or after the JSON.

Now, parse the following unit outline text:

```

--- UNIT_OUTLINE_TEXT_START ---
{outline_text}
--- UNIT_OUTLINE_TEXT_END ---
"""

```

[3]: *# Place this in a new cell after your imports, or within Cell 3 before the\_*  
*functions.*

*# This code is based on the schema from your screenshot on page 4.*

```

from pydantic import BaseModel, Field, ValidationError
from typing import List, Optional
import time

```

*# Define Pydantic models that match your JSON schema*

```

class UnitInformation(BaseModel):
    unitCode: Optional[str] = None

```

```

unitName: Optional[str] = None
creditPoints: Optional[int] = None
unitRationale: Optional[str] = None
prerequisites: Optional[str] = None

class Assessment(BaseModel):
    taskName: str
    description: str
    dueWeek: Optional[str] = None
    weightingPercent: Optional[int] = None
    learningOutcomesAssessed: Optional[str] = None

class WeeklyScheduleItem(BaseModel):
    week: str
    contentTopic: str
    requiredReading: Optional[str] = None

class ParsedUnitOutline(BaseModel):
    unitInformation: UnitInformation
    learningOutcomes: List[str]
    assessments: List[Assessment]
    weeklySchedule: List[WeeklyScheduleItem]
    requiredReadings: List[str]
    recommendedReadings: List[str]

```

### 3 Extrac Unit outline details to process following steps - output raw json with UO details

```

[4]: # Cell 3: Parse Unit Outline

# --- Helper Functions for Parsing ---
def extract_text_from_file(filepath: str) -> str:
    _, ext = os.path.splitext(filepath.lower())
    if ext == '.docx':
        doc = Document(filepath)
        full_text = [p.text for p in doc.paragraphs]
        for table in doc.tables:
            for row in table.rows:
                full_text.append(" | ".join(cell.text for cell in row.cells))
        return '\n'.join(full_text)
    elif ext == '.pdf':
        with pdfplumber.open(filepath) as pdf:
            return "\n".join(page.extract_text() for page in pdf.pages if page.
↪extract_text())
    else:

```

```

        raise TypeError(f"Unsupported file type: {ext}")

def parse_llm_json_output(content: str) -> dict:
    try:
        match = re.search(r'\{.*\}', content, re.DOTALL)
        if not match: return None
        return json.loads(match.group(0))
    except (json.JSONDecodeError, TypeError):
        return None

@retry(stop=stop_after_attempt(3), wait=wait_exponential(min=2, max=10))
def call_ollama_with_retry(client, prompt):
    logger.info(f"Calling Ollama model '{OLLAMA_MODEL}'...")
    response = client.chat(
        model=OLLAMA_MODEL,
        messages=[{"role": "user", "content": prompt}],
        format="json",
        options={"temperature": 0.0}
    )
    if not response or 'message' not in response or not response['message'].
    ↪get('content'):
        raise ValueError("Ollama returned an empty or invalid response.")
    return response['message']['content']

# --- Main Orchestration Function for this Cell ---
def parse_and_save_outline_robust(
    input_filepath: str,
    output_filepath: str,
    prompt_template: str,
    max_retries: int = 3
):
    logger.info(f"Starting to robustly process Unit Outline: {input_filepath}")

    if not os.path.exists(input_filepath):
        logger.error(f"Input file not found: {input_filepath}")
        return

    try:
        outline_text = extract_text_from_file(input_filepath)
        if not outline_text.strip():
            logger.error("Extracted text is empty. Aborting.")
            return
    except Exception as e:
        logger.error(f"Failed to extract text from file: {e}", exc_info=True)
        return

    client = ollama.Client(host=OLLAMA_HOST)

```

```

current_prompt = prompt_template.format(outline_text=outline_text)

for attempt in range(max_retries):
    logger.info(f"Attempt {attempt + 1}/{max_retries} to parse outline.")

    try:
        # Call the LLM
        llm_output_str = call_ollama_with_retry(client, current_prompt)

        # Find the JSON blob in the response
        json_blob = parse_llm_json_output(llm_output_str) # Your existing
↪helper

        if not json_blob:
            raise ValueError("LLM did not return a parsable JSON object.")

        # *** THE KEY VALIDATION STEP ***
        # Try to parse the dictionary into your Pydantic model.
        # This will raise a `ValidationError` if keys are wrong, types are
↪wrong, or fields are missing.
        parsed_data = ParsedUnitOutline.model_validate(json_blob)

        # If successful, save the validated data and exit the loop
        logger.info("Successfully validated JSON structure against Pydantic
↪model.")

        os.makedirs(os.path.dirname(output_filepath), exist_ok=True)
        with open(output_filepath, 'w', encoding='utf-8') as f:
            # Use .model_dump_json() for clean, validated output
            f.write(parsed_data.model_dump_json(indent=2))

        logger.info(f"Successfully parsed and saved Unit Outline to:
↪{output_filepath}")
        return # Exit function on success

    except ValidationError as e:
        logger.warning(f"Validation failed on attempt {attempt + 1}. Error:
↪{e}")

        # Formulate a new prompt with the error message for self-correction
        error_feedback = (
            f"\n\nYour previous attempt failed. You MUST correct the
↪following errors:\n"
            f"{e}\n\n"
            f"Please regenerate the entire JSON object, ensuring it
↪strictly adheres to the schema "
            f"and corrects these specific errors. Do not change any key
↪names."
        )

```



```

        current_prompt = current_prompt + error_feedback # Append the error
↳to the prompt

    except Exception as e:
        # Catch other errors like network issues from call_ollama_with_retry
        logger.error(f"An unexpected error occurred on attempt {attempt + 1}
↳1}: {e}", exc_info=True)
        # You might want to wait before retrying for non-validation errors
        time.sleep(5)

    logger.error(f"Failed to get valid structured data from the LLM after
↳{max_retries} attempts.")

# --- In your execution block, call the new function ---
# parse_and_save_outline(...) becomes:

if EXTRACT_UO:
    parse_and_save_outline_robust(
        input_filepath=FULL_PATH_UNIT_OUTLINE,
        output_filepath=PARSED_UO_JSON_PATH,
        prompt_template=UNIT_OUTLINE_SYSTEM_PROMPT_TEMPLATE
    )

```

## 4 Extract TOC from epub or epub

```

[25]: # Cell 4: Extract Book Table of Contents (ToC) with Pre-assigned IDs & Links in
↳Order

from ebooklib import epub, ITEM_NAVIGATION
from bs4 import BeautifulSoup
import fitz # PyMuPDF
import json
import os
from typing import List, Dict
import urllib.parse # Needed to clean up links

# =====
# 1. HELPER FUNCTIONS (MODIFIED TO INCLUDE ID ASSIGNMENT AND LINK EXTRACTION)
# =====

def clean_epub_href(href: str) -> str:
    """Removes URL fragments and decodes URL-encoded characters."""
    if not href: return ""
    # Remove fragment identifier (e.g., '#section1')
    cleaned_href = href.split('#')[0]

```

```

# Decode any URL-encoded characters (e.g., %20 -> space)
return urllib.parse.unquote(cleaned_href)

# --- EPUB Extraction Logic ---
def parse_navpoint(navpoint: BeautifulSoup, counter: List[int], level: int = 0) → Dict:
    """Recursively parses EPUB 2 navPoints and assigns a toc_id and link_filename."""
    title = navpoint.navLabel.text.strip()
    if not title: return None

    # --- MODIFICATION: Extract the linked filename ---
    content_tag = navpoint.find('content', recursive=False)
    link_filename = clean_epub_href(content_tag['src']) if content_tag else ""

    node = {
        "level": level,
        "toc_id": counter[0],
        "title": title,
        "link_filename": link_filename, # Add the cleaned link
        "children": []
    }
    counter[0] += 1

    for child_navpoint in navpoint.find_all('navPoint', recursive=False):
        child_node = parse_navpoint(child_navpoint, counter, level + 1)
        if child_node: node["children"].append(child_node)

    return node

def parse_li(li_element: BeautifulSoup, counter: List[int], level: int = 0) → Dict:
    """Recursively parses EPUB 3 <li> elements and assigns a toc_id and link_filename."""
    a_tag = li_element.find('a', recursive=False)
    if a_tag:
        title = a_tag.get_text(strip=True)
        if not title: return None

    # --- MODIFICATION: Extract the linked filename ---
    link_filename = clean_epub_href(a_tag.get('href'))

    node = {
        "level": level,
        "toc_id": counter[0],
        "title": title,
        "link_filename": link_filename, # Add the cleaned link
    }

```

```

        "children": []
    }
    counter[0] += 1

    nested_ol = li_element.find('ol', recursive=False)
    if nested_ol:
        for sub_li in nested_ol.find_all('li', recursive=False):
            child_node = parse_li(sub_li, counter, level + 1)
            if child_node: node["children"].append(child_node)
        return node
    return None

def extract_epub_toc(epub_path, output_json_path):
    print(f"Processing EPUB ToC for: {epub_path}")
    toc_data = []
    book = epub.read_epub(epub_path)
    id_counter = [0]

    for nav_item in book.get_items_of_type(ITEM_NAVIGATION):
        soup = BeautifulSoup(nav_item.get_content(), 'xml')
        # Logic to handle both EPUB 2 (NCX) and EPUB 3 (XHTML)
        if nav_item.get_name().endswith('.ncx'):
            print("INFO: Found EPUB 2 (NCX) Table of Contents. Parsing...")
            navmap = soup.find('navMap')
            if navmap:
                for navpoint in navmap.find_all('navPoint', recursive=False):
                    node = parse_navpoint(navpoint, id_counter, level=0)
                    if node: toc_data.append(node)
            else: # Assumes EPUB 3
                print("INFO: Found EPUB 3 (XHTML) Table of Contents. Parsing...")
                toc_nav = soup.select_one('nav[epub:type="toc"]')
                if toc_nav:
                    top_ol = toc_nav.find('ol', recursive=False)
                    if top_ol:
                        for li in top_ol.find_all('li', recursive=False):
                            node = parse_li(li, id_counter, level=0)
                            if node: toc_data.append(node)
                    if toc_data: break

    if toc_data:
        os.makedirs(os.path.dirname(output_json_path), exist_ok=True)
        with open(output_json_path, 'w', encoding='utf-8') as f:
            json.dump(toc_data, f, indent=2, ensure_ascii=False)
        print(f" Successfully wrote EPUB ToC with IDs and links to: ↵
        ↵{output_json_path}")
    else:
        print(" WARNING: No ToC data extracted from EPUB.")

```

```

# --- PDF Extraction Logic (Unchanged) ---
def build_pdf_hierarchy_with_ids(toc_list: List) -> List[Dict]:
    root = []
    parent_stack = {-1: {"children": root}}
    id_counter = [0]
    for level, title, page in toc_list:
        normalized_level = level - 1
        node = {"level": normalized_level, "toc_id": id_counter[0], "title": title.strip(), "page": page, "children": []}
        id_counter[0] += 1
        parent_node = parent_stack.get(normalized_level - 1)
        if parent_node: parent_node["children"].append(node)
        parent_stack[normalized_level] = node
    return root

def extract_pdf_toc(pdf_path, output_json_path):
    print(f"Processing PDF ToC for: {pdf_path}")
    try:
        doc = fitz.open(pdf_path)
        toc = doc.get_toc()
        hierarchical_toc = []
        if not toc: print(" WARNING: This PDF has no embedded bookmarks (ToC).")
        else:
            print(f"INFO: Found {len(toc)} bookmark entries. Building hierarchy and assigning IDs...")
            hierarchical_toc = build_pdf_hierarchy_with_ids(toc)
            os.makedirs(os.path.dirname(output_json_path), exist_ok=True)
            with open(output_json_path, 'w', encoding='utf-8') as f:
                json.dump(hierarchical_toc, f, indent=2, ensure_ascii=False)
            print(f" Successfully wrote PDF ToC with assigned IDs to: {output_json_path}")
        except Exception as e: print(f"An error occurred during PDF ToC extraction: {e}")

# =====
# 2. EXECUTION BLOCK
# =====
if PROCESS_EPUB:
    extract_epub_toc(BOOK_PATH, PRE_EXTRACTED_TOC_JSON_PATH)
else:
    extract_pdf_toc(BOOK_PATH, PRE_EXTRACTED_TOC_JSON_PATH)

```

Processing EPUB ToC for:

/home/sebas\_dev\_linux/projects/course\_generator/data/books/Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_

Processing Digital Evidence-Cengage Learning (2018).epub

INFO: Found EPUB 2 (NCX) Table of Contents. Parsing...

Successfully wrote EPUB ToC with IDs and links to: /home/sebas\_dev\_linux/projects/course\_generator/Parse\_data/Parse\_TOC\_books/ICT312\_epub\_table\_of\_contents.json

## 5 Hirachical DB base on TOC

### 5.1 Process Book

```
[6]: # Cell 5: Create Hierarchical Vector Database (with Sequential ToC ID and Chunk
      ↪ID)
      # This cell processes the book, enriches it with hierarchical and sequential
      ↪metadata,
      # chunks it, and creates the final vector database.

import os
import json
import shutil
import logging
from typing import List, Dict, Any, Tuple
from langchain_core.documents import Document
from langchain_community.document_loaders import PyPDFLoader,
      ↪UnstructuredEPubLoader
from langchain_ollama.embeddings import OllamaEmbeddings
from langchain_chroma import Chroma
from langchain.text_splitter import RecursiveCharacterTextSplitter

# Setup Logger for this cell
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s -
      ↪%(message)s')
logger = logging.getLogger(__name__)

# --- Helper: Clean metadata values for ChromaDB ---
def clean_metadata_for_chroma(value: Any) -> Any:
    """Sanitizes metadata values to be compatible with ChromaDB."""
    if isinstance(value, list): return ", ".join(map(str, value))
    if isinstance(value, dict): return json.dumps(value)
    if isinstance(value, (str, int, float, bool)) or value is None: return value
    return str(value)

# --- Core Function to Process Book with Pre-extracted ToC ---
def process_book_with_extracted_toc(
    book_path: str,
    extracted_toc_json_path: str,
    chunk_size: int,
    chunk_overlap: int
```

```

) -> Tuple[List[Document], List[Dict[str, Any]]]:

    logger.info(f"Processing book '{os.path.basename(book_path)}' using ToC
↳from '{os.path.basename(extracted_toc_json_path)}'".)

    # 1. Load the pre-extracted hierarchical ToC
    try:
        with open(extracted_toc_json_path, 'r', encoding='utf-8') as f:
            hierarchical_toc = json.load(f)
        if not hierarchical_toc:
            logger.error(f"Pre-extracted ToC at '{extracted_toc_json_path}' is
↳empty or invalid.")
            return [], []
        logger.info(f"Successfully loaded pre-extracted ToC with
↳{len(hierarchical_toc)} top-level entries.")
        except Exception as e:
            logger.error(f"Error loading pre-extracted ToC JSON: {e}",
↳exc_info=True)
            return [], []

    # 2. Load all text elements/pages from the book
    all_raw_book_docs: List[Document] = []
    _, file_extension = os.path.splitext(book_path.lower())

    if file_extension == ".epub":
        loader = UnstructuredEPubLoader(book_path, mode="elements",
↳strategy="fast")
        try:
            all_raw_book_docs = loader.load()
            logger.info(f"Loaded {len(all_raw_book_docs)} text elements from
↳EPUB.")
        except Exception as e:
            logger.error(f"Error loading EPUB content: {e}", exc_info=True)
            return [], hierarchical_toc
    elif file_extension == ".pdf":
        loader = PyPDFLoader(book_path)
        try:
            all_raw_book_docs = loader.load()
            logger.info(f"Loaded {len(all_raw_book_docs)} pages from PDF.")
        except Exception as e:
            logger.error(f"Error loading PDF content: {e}", exc_info=True)
            return [], hierarchical_toc
    else:
        logger.error(f"Unsupported book file format: {file_extension}")
        return [], hierarchical_toc

```

```

if not all_raw_book_docs:
    logger.error("No text elements/pages loaded from the book.")
    return [], hierarchical_toc

# 3. Create enriched LangChain Documents by matching ToC to content
final_documents_with_metadata: List[Document] = []

# Flatten the ToC, AND add a unique sequential ID for sorting and
validation.
flat_toc_entries: List[Dict[str, Any]] = []

def _add_ids_and_flatten_recursive(nodes: List[Dict[str, Any]],
current_titles_path: List[str], counter: List[int]):
    """
    Recursively traverses ToC nodes to flatten them and assign a unique,
    sequential toc_id.
    """
    for node in nodes:
        toc_id = counter[0]
        counter[0] += 1
        title = node.get("title", "").strip()
        if not title: continue
        new_titles_path = current_titles_path + [title]
        entry = {
            "titles_path": new_titles_path,
            "level": node.get("level"),
            "full_title_for_matching": title,
            "toc_id": toc_id
        }
        if "page" in node: entry["page"] = node["page"]
        flat_toc_entries.append(entry)
        if node.get("children"):
            _add_ids_and_flatten_recursive(node.get("children", []),
new_titles_path, counter)

    toc_id_counter = [0]
    _add_ids_and_flatten_recursive(hierarchical_toc, [], toc_id_counter)
    logger.info(f"Flattened ToC and assigned sequential IDs to
    {len(flat_toc_entries)} entries.")

# Logic for PDF metadata assignment
if file_extension == ".pdf" and any("page" in entry for entry in
flat_toc_entries):
    logger.info("Assigning metadata to PDF pages based on ToC page numbers..")

```

```

        flat_toc_entries.sort(key=lambda x: x.get("page", -1) if x.get("page")
↪is not None else -1)
        for page_doc in all_raw_book_docs:
            page_num_0_indexed = page_doc.metadata.get("page", -1)
            page_num_1_indexed = page_num_0_indexed + 1
            assigned_metadata = {"source": os.path.basename(book_path),
↪"page_number": page_num_1_indexed}
            best_match_toc_entry = None
            for toc_entry in flat_toc_entries:
                toc_page = toc_entry.get("page")
                if toc_page is not None and toc_page <= page_num_1_indexed:
                    if best_match_toc_entry is None or toc_page >
↪best_match_toc_entry.get("page", -1):
                        best_match_toc_entry = toc_entry
                    elif toc_page is not None and toc_page > page_num_1_indexed:
                        break
            if best_match_toc_entry:
                for i, title_in_path in
↪enumerate(best_match_toc_entry["titles_path"]):
                    assigned_metadata[f"level_{i+1}_title"] = title_in_path
                    assigned_metadata['toc_id'] = best_match_toc_entry.get('toc_id')
                else:
                    assigned_metadata["level_1_title"] = "Uncategorized PDF Page"
                    cleaned_meta = {k: clean_metadata_for_chroma(v) for k, v in
↪assigned_metadata.items()}
                    final_documents_with_metadata.append(Document(page_content=page_doc.
↪page_content, metadata=cleaned_meta))

        # Logic for EPUB metadata assignment
        elif file_extension == ".epub":
            logger.info("Assigning metadata to EPUB elements by matching ToC titles
↪in text...")
            toc_titles_for_search = [entry for entry in flat_toc_entries if entry.
↪get("full_title_for_matching")]
            current_hierarchy_metadata = {}
            for element_doc in all_raw_book_docs:
                element_text = element_doc.page_content.strip() if element_doc.
↪page_content else ""
                if not element_text: continue
                for toc_entry in toc_titles_for_search:
                    if element_text == toc_entry["full_title_for_matching"]:
                        current_hierarchy_metadata = {"source": os.path.
↪basename(book_path)}
                        for i, title_in_path in enumerate(toc_entry["titles_path"]):
                            current_hierarchy_metadata[f"level_{i+1}_title"] =
↪title_in_path

```



```

        current_hierarchy_metadata['toc_id'] = toc_entry.
↪get('toc_id')
        if "page" in toc_entry:
↪current_hierarchy_metadata["epub_toc_page"] = toc_entry["page"]
            break
        if not current_hierarchy_metadata:
            doc_metadata_to_assign = {"source": os.path.
↪basename(book_path), "level_1_title": "EPUB Preamble", "toc_id": -1}
        else:
            doc_metadata_to_assign = current_hierarchy_metadata.copy()
            cleaned_meta = {k: clean_metadata_for_chroma(v) for k, v in
↪doc_metadata_to_assign.items()}
            final_documents_with_metadata.
↪append(Document(page_content=element_text, metadata=cleaned_meta))

    else: # Fallback
        final_documents_with_metadata = all_raw_book_docs

    if not final_documents_with_metadata:
        logger.error("No documents were processed or enriched with hierarchical_
↪metadata.")
        return [], hierarchical_toc

    logger.info(f"Total documents prepared for chunking:
↪{len(final_documents_with_metadata)}")

    text_splitter = RecursiveCharacterTextSplitter(
        chunk_size=chunk_size,
        chunk_overlap=chunk_overlap,
        length_function=len
    )
    final_chunks = text_splitter.split_documents(final_documents_with_metadata)
    logger.info(f"Split into {len(final_chunks)} final chunks, inheriting_
↪hierarchical metadata.")

    # --- MODIFICATION START: Add a unique, sequential chunk_id to each chunk_
↪---
    logger.info("Assigning sequential chunk_id to all final chunks...")
    for i, chunk in enumerate(final_chunks):
        chunk.metadata['chunk_id'] = i
    logger.info(f"Assigned chunk_ids from 0 to {len(final_chunks) - 1}.")
    # --- MODIFICATION END ---

    return final_chunks, hierarchical_toc

# --- Main Execution Block for this Cell ---

```

```

if not os.path.exists(PRE_EXTRACTED_TOC_JSON_PATH):
    logger.error(f"CRITICAL: Pre-extracted ToC file not found at_
↳ '{PRE_EXTRACTED_TOC_JSON_PATH}'")
    logger.error("Please run the 'Extract Book Table of Contents (ToC)' cell_
↳ (Cell 4) first.")
else:
    final_chunks_for_db, toc_reloaded = process_book_with_extracted_toc(
        book_path=BOOK_PATH,
        extracted_toc_json_path=PRE_EXTRACTED_TOC_JSON_PATH,
        chunk_size=CHUNK_SIZE,
        chunk_overlap=CHUNK_OVERLAP
    )

    if final_chunks_for_db:
        if os.path.exists(CHROMA_PERSIST_DIR):
            logger.warning(f"Deleting existing ChromaDB directory:_
↳ {CHROMA_PERSIST_DIR}")
            shutil.rmtree(CHROMA_PERSIST_DIR)

            logger.info(f"Initializing embedding model '{EMBEDDING_MODEL_OLLAMA}'_
↳ and creating new vector database...")
            embedding_model = OllamaEmbeddings(model=EMBEDDING_MODEL_OLLAMA)

            vector_db = Chroma.from_documents(
                documents=final_chunks_for_db,
                embedding=embedding_model,
                persist_directory=CHROMA_PERSIST_DIR,
                collection_name=CHROMA_COLLECTION_NAME
            )

            reloaded_db = Chroma(persist_directory=CHROMA_PERSIST_DIR,_
↳ embedding_function=embedding_model, collection_name=CHROMA_COLLECTION_NAME)
            count = reloaded_db._collection.count()

            print("-" * 50)
            logger.info(f" Vector DB created successfully at:_
↳ {CHROMA_PERSIST_DIR}")
            logger.info(f" Collection '{CHROMA_COLLECTION_NAME}' contains {count}_
↳ documents.")
            print("-" * 50)
        else:
            logger.error(" Failed to generate chunks. Vector DB not created.")

```

2025-07-06 01:42:25,050 - INFO - Processing book 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub' using ToC from

```
'ICT312_epub_table_of_contents.json'.
2025-07-06 01:42:25,052 - INFO - Successfully loaded pre-extracted ToC with 28
top-level entries.
2025-07-06 01:42:26,753 - INFO - Note: NumExpr detected 32 cores but
"NUMEXPR_MAX_THREADS" not set, so enforcing safe limit of 16.
2025-07-06 01:42:26,753 - INFO - NumExpr defaulting to 16 threads.
[WARNING] Could not load translations for en-US
data file translations/en.yaml not found
2025-07-06 01:42:31,899 - WARNING - Could not load translations for en-US
data file translations/en.yaml not found
[WARNING] The term Abstract has no translation defined.

2025-07-06 01:42:31,900 - WARNING - The term Abstract has no translation
defined.

2025-07-06 01:42:35,327 - INFO - Loaded 11815 text elements from EPUB.
2025-07-06 01:42:35,328 - INFO - Flattened ToC and assigned sequential IDs to
877 entries.
2025-07-06 01:42:35,328 - INFO - Assigning metadata to EPUB elements by matching
ToC titles in text...
2025-07-06 01:42:35,550 - INFO - Total documents prepared for chunking: 11483
2025-07-06 01:42:35,734 - INFO - Split into 11774 final chunks, inheriting
hierarchical metadata.
2025-07-06 01:42:35,735 - INFO - Assigning sequential chunk_id to all final
chunks...
2025-07-06 01:42:35,737 - INFO - Assigned chunk_ids from 0 to 11773.
2025-07-06 01:42:35,742 - INFO - Initializing embedding model 'nomic-embed-text'
and creating new vector database...
2025-07-06 01:42:35,777 - INFO - Anonymized telemetry enabled. See
https://docs.trychroma.com/telemetry for more information.
2025-07-06 01:43:43,622 - INFO - HTTP Request: POST
http://127.0.0.1:11434/api/embed "HTTP/1.1 200 OK"
2025-07-06 01:44:57,682 - INFO - HTTP Request: POST
http://127.0.0.1:11434/api/embed "HTTP/1.1 200 OK"
2025-07-06 01:45:13,345 - INFO - HTTP Request: POST
http://127.0.0.1:11434/api/embed "HTTP/1.1 200 OK"
2025-07-06 01:45:13,947 - INFO - Vector DB created successfully at: /home/seba
s_dev_linux/projects/course_generator/data/DataBase_Chroma/chroma_db_toc_guided_
chunks_epub
2025-07-06 01:45:13,947 - INFO - Collection 'book_toc_guided_chunks_epub_v2'
contains 11774 documents.
```

```
-----
-----
```

[7]: *# Cell 5a: Inspecting EPUB Documents and Metadata BEFORE Chunking*

```
import json
```

```

import os
import logging
from langchain_community.document_loaders import UnstructuredEPubLoader
from langchain_core.documents import Document

# --- Setup Logger for this inspection cell ---
logger = logging.getLogger(__name__)
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %
↳%(message)s')

def inspect_epub_preprocessing():
    """
    This function replicates the pre-chunking logic from Cell 5 for EPUB files
    to show the list of large documents with their assigned ToC metadata.
    """
    if not PROCESS_EPUB:
        print("This inspection cell is for EPUB processing. Please set
↳PROCESS_EPUB = True in Cell 1.")
        return

    print_header("EPUB Pre-Processing Inspection", char="~")

    # --- 1. Load the necessary data (replicating start of Cell 5) ---
    logger.info("Loading pre-extracted ToC and raw EPUB elements...")
    try:
        with open(PRE_EXTRACTED_TOC_JSON_PATH, 'r', encoding='utf-8') as f:
            hierarchical_toc = json.load(f)

        loader = UnstructuredEPubLoader(BOOK_PATH, mode="elements",
↳strategy="fast")
        all_raw_book_docs = loader.load()
        logger.info(f"Successfully loaded {len(all_raw_book_docs)} raw text
↳elements from the EPUB.")
    except Exception as e:
        logger.error(f"Failed to load necessary files: {e}")
        return

    # --- 2. Flatten the ToC (replicating logic from Cell 5) ---
    logger.info("Flattening the hierarchical ToC for matching...")
    flat_toc_entries = []
    def _add_ids_and_flatten_recursive(nodes, current_titles_path, counter):
        for node in nodes:
            toc_id = counter[0]
            counter[0] += 1
            title = node.get("title", "").strip()
            if not title: continue
            new_titles_path = current_titles_path + [title]

```

```

        entry = {
            "titles_path": new_titles_path,
            "level": node.get("level"),
            "full_title_for_matching": title,
            "toc_id": toc_id
        }
        flat_toc_entries.append(entry)
        if node.get("children"):
            _add_ids_and_flatten_recursive(node.get("children", []),
↪new_titles_path, counter)

    _add_ids_and_flatten_recursive(hierarchical_toc, [], [0])
    logger.info(f"Flattened ToC into {len(flat_toc_entries)} entries.")

    # --- 3. The Core Matching Logic for EPUB (the part you want to see) ---
    logger.info("Assigning metadata to EPUB elements by matching ToC titles...")

    final_documents_with_metadata = []
    toc_titles_for_search = [entry for entry in flat_toc_entries if entry.
↪get("full_title_for_matching")]
    current_hierarchy_metadata = {}

    for element_doc in all_raw_book_docs:
        element_text = element_doc.page_content.strip() if element_doc.
↪page_content else ""
        if not element_text:
            continue

        # Check if this element is a heading that matches a ToC entry
        is_heading = False
        for toc_entry in toc_titles_for_search:
            if element_text == toc_entry["full_title_for_matching"]:
                # It's a heading! Update the current context.
                current_hierarchy_metadata = {"source": os.path.
↪basename(BOOK_PATH)}
                for i, title_in_path in enumerate(toc_entry["titles_path"]):
                    current_hierarchy_metadata[f"level_{i+1}_title"] =
↪title_in_path
                current_hierarchy_metadata['toc_id'] = toc_entry.get('toc_id')
                is_heading = True
                break # Found the match, no need to search further

        # Assign metadata
        if not current_hierarchy_metadata:
            # Content before the first ToC entry (e.g., cover, title page)
            doc_metadata_to_assign = {"source": os.path.basename(BOOK_PATH),
↪"level_1_title": "EPUB Preamble", "toc_id": -1}

```

```

        else:
            doc_metadata_to_assign = current_hierarchy_metadata.copy()

            final_documents_with_metadata.
↪append(Document(page_content=element_text, metadata=doc_metadata_to_assign))

            logger.info(f"Processing complete. Generated_
↪{len(final_documents_with_metadata)} documents with assigned metadata.")

            # --- 4. Print the result for inspection ---
            print_header("INSPECTION RESULTS: Documents Before Chunking", char=="")
            print(f"Total documents created: {len(final_documents_with_metadata)}\n")

            for i, doc in enumerate(final_documents_with_metadata[:100]): # Print first_
↪30 to avoid flooding the output
                print(f"--- Document [{i+1}] ---")
                print(f"  Assigned Metadata: {doc.metadata}")
                print(f"  Content (Un-chunked Element):")
                print(f"  >> '{doc.page_content}'")
                print(f"  -" * 25 + "\n")

# --- Execute the inspection ---
inspect_epub_preprocessing()

```

2025-07-06 01:45:13,966 - INFO - Loading pre-extracted ToC and raw EPUB elements...

#### ~~~~~ EPUB Pre-Processing Inspection ~~~~~

[WARNING] Could not load translations for en-US

data file translations/en.yaml not found

2025-07-06 01:45:19,084 - WARNING - Could not load translations for en-US

data file translations/en.yaml not found

[WARNING] The term Abstract has no translation defined.

2025-07-06 01:45:19,085 - WARNING - The term Abstract has no translation defined.

2025-07-06 01:45:22,732 - INFO - Successfully loaded 11815 raw text elements from the EPUB.

2025-07-06 01:45:22,733 - INFO - Flattening the hierarchical ToC for matching...

2025-07-06 01:45:22,734 - INFO - Flattened ToC into 877 entries.

2025-07-06 01:45:22,735 - INFO - Assigning metadata to EPUB elements by matching ToC titles...

2025-07-06 01:45:22,942 - INFO - Processing complete. Generated 11483 documents with assigned metadata.

```
=====
                        INSPECTION RESULTS: Documents Before Chunking
=====
```

Total documents created: 11483

--- Document [1] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'EPUB Preamble', 'toc\_id': -1}

Content (Un-chunked Element):

>> 'Guide to Computer Forensics and Investigations: Processing Digital Evidence'

--- Document [2] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'EPUB Preamble', 'toc\_id': -1}

Content (Un-chunked Element):

>> 'Copyright Statement'

--- Document [3] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'EPUB Preamble', 'toc\_id': -1}

Content (Un-chunked Element):

>> 'Guide to Computer Forensics and Investigations: Processing Digital Evidence'

--- Document [4] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'EPUB Preamble', 'toc\_id': -1}

Content (Un-chunked Element):

>> 'COPYRIGHT © 2019, 2016 Cengage Learning, Inc.'

--- Document [5] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
Content (Un-chunked Element):
>> 'ALL RIGHTS RESERVED. No part of this work covered by the copyright herein
may be reproduced or distributed in any form or by any means, except as
permitted by U.S. copyright law, without the prior written permission of the
copyright owner.'
```

--- Document [6] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
Content (Un-chunked Element):
>> 'For product information and technology assistance, contact us at Cengage
Customer & Sales Support, 1-800-354-9706 or support.cengage.com.'
```

--- Document [7] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
Content (Un-chunked Element):
>> 'For permission to use material from this text or product, submit all
requests online at www.cengage.com/permissions.'
```

--- Document [8] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
Content (Un-chunked Element):
>> 'SOURCE FOR ILLUSTRATIONS: Copyright © Cengage.'
```

--- Document [9] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
Content (Un-chunked Element):
>> ', Microsoft® is a registered trademark of the Microsoft Corporation.'
```



```

--- Document [10] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
Content (Un-chunked Element):
>> 'Library of Congress Control Number: 2018936389'
-----

--- Document [11] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
Content (Un-chunked Element):
>> 'ISBN: 978-1-337-56894-4'
-----

--- Document [12] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
Content (Un-chunked Element):
>> 'Cengage'
-----

--- Document [13] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
Content (Un-chunked Element):
>> '20 Channel Center Street'
-----

--- Document [14] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
Content (Un-chunked Element):
>> 'Boston MA 02210'
-----

--- Document [15] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher

```

```

Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
    Content (Un-chunked Element):
    >> 'USA'
-----

--- Document [16] ---
    Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
    Content (Un-chunked Element):
    >> 'Cengage is a leading provider of customized learning solutions with
employees residing in nearly 40 different countries and sales in more than 125
countries around the world. Find your local representative at www.cengage.com.'
-----

--- Document [17] ---
    Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
    Content (Un-chunked Element):
    >> 'Cengage products are represented in Canada by Nelson Education, Ltd.'
-----

--- Document [18] ---
    Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
    Content (Un-chunked Element):
    >> 'To learn more about Cengage platforms and services, visit
www.cengage.com.'
-----

--- Document [19] ---
    Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
    Content (Un-chunked Element):
    >> 'To register or access your online learning solution or purchase materials
for your course, visit www.cengagebrain.com.'
-----

--- Document [20] ---

```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'EPUB Preamble',
'toc_id': -1}
```

```
Content (Un-chunked Element):
```

```
>> 'Notice to the Reader Publisher does not warrant or guarantee any of the
products described herein or perform any independent analysis in connection with
any of the product information contained herein. Publisher does not assume, and
expressly disclaims, any obligation to obtain and include information other than
that provided to it by the manufacturer. The reader is expressly warned to
consider and adopt all safety precautions that might be indicated by the
activities described herein and to avoid all potential hazards. By following the
instructions contained herein, the reader willingly assumes all risks in
connection with such instructions. The publisher makes no representations or
warranties of any kind, including but not limited to, the warranties of fitness
for particular purpose or merchantability, nor are any such representations
implied with respect to the material set forth herein, and the publisher takes
no responsibility with respect to such material. The publisher shall not be
liable for any special, consequential, or exemplary damages resulting, in whole
or part, from the readers' use of, or reliance upon, this material.'
```

```
-----
```

```
--- Document [21] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Preface', 'toc_id': 3}
```

```
Content (Un-chunked Element):
```

```
>> 'Preface'
```

```
-----
```

```
--- Document [22] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Preface', 'toc_id': 3}
```

```
Content (Un-chunked Element):
```

```
>> 'Guide to Computer Forensics and Investigations is now in its sixth
edition. As digital technology and cyberspace have evolved from their early
roots as basic communication platforms into the hyper-connected world we live in
today, so has the demand for people who have the knowledge and skills to
investigate legal and technical issues involving computers and digital
technology. My sincere compliments to the authors and publishing staff who have
made this textbook such a remarkable resource for thousands of students and
practitioners worldwide.'
```

```
-----
```

```
--- Document [23] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
```

```
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Preface', 'toc_id': 3}
Content (Un-chunked Element):
```

```
>> 'Computers, the Internet, and the world's digital ecosystem are all
instrumental in how we conduct our daily lives. When the founding fathers of the
modern computing era were designing the digital infrastructure as we know it
today, security and temporal accountability issues were not at the top of their
list of things to do. The technological advancement of these systems over the
past 10 years has changed the way we learn, socialize, and conduct business.
Finding digital data that can be used as evidence to incriminate or exonerate a
suspect accused in a legal or administrative proceeding is not an easy task.'
```

```
-----
```

```
--- Document [24] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Preface', 'toc_id': 3}
Content (Un-chunked Element):
```

```
>> 'Cyberthreats have become pervasive in modern society. They range from
simple computer viruses to complex ransomware and cyber extortion schemes. The
ability to conduct sophisticated digital forensics investigations has become a
requirement in both the government and commercial sectors. Currently, the
organizations and agencies whose job it is to investigate both criminal and
civil matters involving the use of rapidly developing digital technology often
struggle to keep up with the ever-changing digital landscape. Additionally,
finding trained and qualified people to conduct these types of inquiries has
been challenging.'
```

```
-----
```

```
--- Document [25] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Preface', 'toc_id': 3}
Content (Un-chunked Element):
```

```
>> 'Today, an entire industry has evolved for the purpose of investigating
events occurring in cyberspace to include incidents involving international and
corporate espionage, massive data breaches, and even cyberterrorism. The
opportunities for employment in this field are expanding every day.
Professionals in this exciting field of endeavor are now in high demand and are
expected to have multiple skill sets in areas such as malware analysis, cloud
computing, social media, and mobile device forensics.'
```

```
-----
```

```
--- Document [26] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Preface', 'toc_id': 3}
Content (Un-chunked Element):
```

```
>> 'Guide to Computer Forensics and Investigations can now be found in both
```

academic and professional environments as a reliable source of current technical information and practical exercises concerning investigations involving the latest digital technologies. It's my belief that this book, combined with an enthusiastic and knowledgeable facilitator, makes for a fascinating course of instruction.'

-----

--- Document [27] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Preface', 'toc\_id': 3}  
Content (Un-chunked Element):

>> 'As I have stated to many of my students in the past, it's not just laptop computers and servers that harbor the binary code of ones and zeros, but an infinite array of digital devices. If one of these devices retains evidence of a crime, it's up to newly trained and educated digital detectives to find the evidence in a forensically sound manner. This book will assist both students and practitioners in accomplishing this goal.'

-----

--- Document [28] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Preface', 'toc\_id': 3}  
Content (Un-chunked Element):

>> 'Respectfully,'

-----

--- Document [29] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Preface', 'toc\_id': 3}  
Content (Un-chunked Element):

>> 'John A. Sgromolo'

-----

--- Document [30] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Preface', 'toc\_id': 3}  
Content (Un-chunked Element):

>> 'As a Senior Special Agent, John was one of the founding members of the NCIS Computer Crime Investigations Group. John left government service to run his own company, Digital Forensics, Inc., and has taught hundreds of law enforcement and corporate students nationwide in the art and science of digital forensics investigations. Currently, he serves as a senior consultant for Verizon's Global Security Services, where he helps manage the Threat Intel Response Service.'

-----  
--- Document [31] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}  
Content (Un-chunked Element):  
>> 'Introduction'

-----

--- Document [32] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}  
Content (Un-chunked Element):  
>> 'Computer forensics, now most commonly called "digital forensics," has been a professional field for many years, but most well-established experts in the field have been self-taught. The growth of the Internet and the worldwide proliferation of computers have increased the need for digital investigations. Computers can be used to commit crimes, and crimes can be recorded on computers, including company policy violations, embezzlement, e-mail harassment, murder, leaks of proprietary information, and even terrorism. Law enforcement, network administrators, attorneys, and private investigators now rely on the skills of professional digital forensics experts to investigate criminal and civil cases.'

-----

--- Document [33] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}  
Content (Un-chunked Element):  
>> 'This book is not intended to provide comprehensive training in digital forensics. It does, however, give you a solid foundation by introducing digital forensics to those who are new to the field. Other books on digital forensics are targeted to experts; this book is intended for novices who have a thorough grounding in computer and networking basics.'

-----

--- Document [34] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}  
Content (Un-chunked Element):  
>> 'The new generation of digital forensics experts needs more initial

training because operating systems, computer and mobile device hardware, and forensics software tools are changing more quickly. This book covers current and past operating systems and a range of hardware, from basic workstations and high-end network servers to a wide array of mobile devices. Although this book focuses on a few forensics software tools, it also reviews and discusses other currently available tools.'

-----

--- Document [35] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'The purpose of this book is to guide you toward becoming a skilled digital forensics investigator. A secondary goal is to help you pass related certification exams. As the field of digital forensics and investigations matures, keep in mind that certifications will change. You can find more information on certifications in Chapter 2 and Appendix A.'

-----

--- Document [36] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Intended Audience'

-----

--- Document [37] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Although this book can be used by people with a wide range of backgrounds, it's intended for those with A+ and Network+ certifications or the equivalent. A networking background is necessary so that you understand how computers operate in a networked environment and can work with a network administrator when needed. In addition, you must know how to use a computer from the command line and how to use common operating systems, including Windows, Linux, and macOS, and their related hardware.'

-----

--- Document [38] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital

```
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'This book can be used at any educational level, from technical high
schools and community colleges to graduate students. Current professionals in
the public and private sectors can also use this book. Each group will approach
investigative problems from a different perspective, but all will benefit from
the coverage.'
```

-----

```
--- Document [39] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'What's New in This Edition'
```

-----

```
--- Document [40] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'The chapter flow of this book is organized so that you're first exposed to
what happens in a forensics lab and how to set one up before you get into the
nuts and bolts. Coverage of several GUI tools has been added to give you a
familiarity with some widely used software. In addition, Chapter 11 has
additional coverage of social media forensics, Chapter 12 has been expanded to
include more information on smartphones and tablets, and Chapter 13 on forensics
procedures for information stored in the cloud has been updated. Corrections
have been made to this edition based on feedback from users, and all software
tools and Web sites have been updated to reflect what's current at the time of
publication. Finally, a new digital lab manual is being offered in MindTap for
Guide to Computer Forensics and Investigationsto go with the sixth edition
textbook.'
```

-----

```
--- Document [41] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Chapter Descriptions'
```



--- Document [42] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Here is a summary of the topics covered in each chapter of this book:'

-----

--- Document [43] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 1 , "Understanding the Digital Forensics Profession and Investigations," introduces you to the history of digital forensics and explains how the use of electronic evidence developed. It also reviews legal issues and compares public and private sector cases. This chapter also explains how to take a systematic approach to preparing a digital investigation, describes how to conduct an investigation, and summarizes requirements for workstations and software.'

-----

--- Document [44] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 2 , "The Investigator's Office and Laboratory," outlines physical requirements and equipment for digital forensics labs, from small private investigators' labs to the regional FBI lab. It also covers certifications for digital investigators and building a business case for a forensics lab.'

-----

--- Document [45] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 3 , "Data Acquisition," explains how to prepare to acquire data from a suspect's drive and discusses available Linux and GUI acquisition tools. This chapter also discusses acquiring data from RAID systems and gives you an overview of tools for remote acquisitions.'

-----

--- Document [46] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 4 , "Processing Crime and Incident Scenes," explains search warrants and the nature of a typical digital forensics case. It discusses when to use outside professionals, how to assemble a team, and how to evaluate a case and explains the correct procedures for searching and seizing evidence. This chapter also introduces you to calculating hashes to verify data you collect.'

-----

--- Document [47] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 5 , "Working with Windows and CLI Systems," discusses the most common operating systems. You learn what happens and what files are altered during computer startup and how file systems deal with deleted and slack space. In addition, this chapter covers some options for decrypting drives encrypted with whole disk encryption and explains the purpose of using virtual machines.'

-----

--- Document [48] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 6 , "Current Digital Forensics Tools," explores current digital forensics software and hardware tools, including those that might not be readily available, and evaluates their strengths and weaknesses.'

-----

--- Document [49] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 7 , "Linux and Macintosh File Systems," continues the operating system discussion from Chapter 5 by examining Macintosh and Linux OSs and file systems. It also gives you practice in using Linux forensics tools.'

-----

--- Document [50] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 8 , "Recovering Graphics Files," explains how to recover graphics files and examines data compression, carving data, reconstructing file fragments, and steganography and copyright issues.'

-----

--- Document [51] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 9 , "Digital Forensics Analysis and Validation," covers determining what data to collect and analyze and refining investigation plans. It also explains validation with hex editors and forensics software and data-hiding techniques.'

-----

--- Document [52] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 10 , "Virtual Machine Forensics, Live Acquisitions, and Network Forensics," covers tools and methods for conducting forensic analysis of virtual machines, performing live acquisitions, reviewing network logs for evidence, and using network-monitoring tools to detect unauthorized access. It also examines using Linux tools and the Honeynet Project's resources.'

-----

--- Document [53] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 11 , "E-mail and Social Media Investigations," examines e-mail crimes and violations and reviews some specialized e-mail and social media forensics tools. It also explains how to approach investigating social media communications and handling the challenges this content poses.'

-----

--- Document [54] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 12 , "Mobile Device Forensics and The Internet of Anything," covers investigation techniques and acquisition procedures for smartphones, other mobile devices, Internet of Anything devices, and sensors. You learn where data might be stored or backed up and what tools are available for these investigations.'

-----

--- Document [55] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 13 , "Cloud Forensics," summarizes the legal and technical challenges in conducting cloud forensics. It also describes how to acquire cloud data and explains how remote acquisition tools can be used in cloud investigations.'

-----

--- Document [56] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 14 , "Report Writing for High-Tech Investigations," discusses the importance of report writing in digital forensics examinations; offers guidelines on report content, structure, and presentation; and explains how to generate report findings with forensics software tools.'

-----

--- Document [57] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 15 , "Expert Testimony in Digital Investigations," explores the role of an expert witness or a fact witness, including developing a curriculum vitae, understanding the trial process, and preparing forensics evidence for testimony. It also offers guidelines for testifying in court and at depositions and hearings.'

-----  
--- Document [58] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter 16 , "Ethics for the Expert Witness," provides guidance in the principles and practice of ethics for digital forensics investigators and examines other professional organizations' codes of ethics.'

-----

--- Document [59] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Appendix A , "Certification Test References," provides information on the National Institute of Standards and Technology (NIST) testing processes for validating digital forensics tools and covers digital forensics certifications and training programs.'

-----

--- Document [60] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Appendix B , "Digital Forensics References," lists recommended books, journals, e-mail lists, and Web sites for additional information and further study. It also covers the latest ISO 27000 standards that apply to digital forensics.'

-----

--- Document [61] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Appendix C , "Digital Forensics Lab Considerations," provides more information on considerations for forensics labs, including certifications, ergonomics, structural design, and communication and fire-suppression systems. It also covers applicable ISO standards.'

-----

--- Document [62] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Appendix D , "Legacy File System and Forensics Tools," reviews FAT file system basics and Mac legacy file systems and explains using DOS forensics tools, creating forensic boot media, and using scripts. It also has an overview of the hexadecimal numbering system and how it's applied to digital information.'

-----

--- Document [63] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Features'

-----

--- Document [64] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'To help you fully understand digital forensics, this book includes many features designed to enhance your learning experience:'

-----

--- Document [65] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction', 'toc\_id': 4}

Content (Un-chunked Element):

>> 'Chapter objectives-Each chapter begins with a detailed list of the concepts to be mastered in that chapter. This list gives you a quick reference to the chapter's contents and is a useful study aid.'

-----

--- Document [66] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Introduction',

```
'toc_id': 4}
Content (Un-chunked Element):
>> 'Figures and tables-Screenshots are used as guidelines for stepping through
commands and forensics tools. For tools not included with the book or that
aren't offered in free demo versions, figures have been added when possible to
illustrate the tool's interface. Tables are used throughout the book to present
information in an organized, easy-to-grasp manner.'
```

--- Document [67] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Chapter summaries-Each chapter's material is followed by a summary of the
concepts introduced in that chapter. These summaries are a helpful way to review
the ideas covered in each chapter.'
```

--- Document [68] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Key terms-Following the chapter summary, all new terms introduced in the
chapter with boldfaced text are gathered together in the Key Terms list. This
list encourages a more thorough understanding of the chapter's key concepts and
is a useful reference.'
```

--- Document [69] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Review questions-The end-of-chapter assessment begins with a set of review
questions that reinforce the main concepts in each chapter. These questions help
you evaluate and apply the material you have learned.'
```

--- Document [70] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
```

```

Content (Un-chunked Element):
>> 'Hands-on projects-Although understanding the theory behind digital
technology is important, nothing can improve on real-world experience. To this
end, each chapter offers several hands-on projects with software supplied as
free downloads on the student companion site and in MindTap. You can explore a
variety of ways to acquire and even hide evidence. For the conceptual chapters,
research projects are supplied.'
-----

--- Document [71] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Case projects-At the end of each chapter are case projects. To do these
projects, you must draw on real-world common sense as well as your knowledge of
the technical topics covered to that point in the book. Your goal for each
project is to come up with answers to problems similar to those you'll face as a
working digital forensics investigator.'
-----

--- Document [72] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Software and student data files-Student data files are available for
download from the student companion site and the MindTap for this book and are
used for activities and projects in the chapters. Demo and freeware software
used in this book can be downloaded from the Web sites specified in activities
and projects or in "Digital Forensics Software" later in this introduction.'
-----

--- Document [73] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Student companion site-To access the student companion site, go to
www.cengagebrain.com and search for the sixth edition by entering the title,
author's name, or ISBN. On the product page, click the Free Materials tab, and
then click Save to MyHome. Then you can sign in as a returning student or choose
to create a new account. After you've logged on, you can begin accessing your
free study tools.'
-----

```



```

--- Document [74] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Text and Graphic Conventions'
-----

--- Document [75] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'When appropriate, additional information and exercises have been added to
this book to help you better understand the topic at hand. The following icons
used in this book alert you to additional materials:'
-----

--- Document [76] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Note'
-----

--- Document [77] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Note Box'
-----

--- Document [78] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'The Note icon draws your attention to additional helpful material related
to the subject being covered.'
-----

```

```

--- Document [79] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Tip'
-----

--- Document [80] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Tips based on the authors' experiences offer extra information about how
to attack a problem or what to do in real-world situations.'
-----

--- Document [81] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Caution'
-----

--- Document [82] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'Caution Box'
-----

--- Document [83] ---
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Introduction',
'toc_id': 4}
Content (Un-chunked Element):
>> 'The Caution icon warns you about potential mistakes or problems and
explains how to avoid them.'
-----

```

--- Document [84] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Chapter 1.
Understanding the Digital Forensics Profession and Investigations',
'level_2_title': 'Chapter Review', 'level_3_title': 'Hands-On Projects',
'toc_id': 53}
Content (Un-chunked Element):
>> 'Hands-On Projects'
```

--- Document [85] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Chapter 1.
Understanding the Digital Forensics Profession and Investigations',
'level_2_title': 'Chapter Review', 'level_3_title': 'Hands-On Projects',
'toc_id': 53}
Content (Un-chunked Element):
>> 'The hands-on icon indicates that the projects following it give you a
chance to practice using software tools and get hands-on experience.'
```

--- Document [86] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Chapter 1.
Understanding the Digital Forensics Profession and Investigations',
'level_2_title': 'Chapter Review', 'level_3_title': 'Case Projects', 'toc_id':
54}
Content (Un-chunked Element):
>> 'Case Projects'
```

--- Document [87] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Chapter 1.
Understanding the Digital Forensics Profession and Investigations',
'level_2_title': 'Chapter Review', 'level_3_title': 'Case Projects', 'toc_id':
54}
Content (Un-chunked Element):
>> 'This icon marks the start of projects that require you to apply common
sense and knowledge to solving problems involving that chapter's concepts.'
```

--- Document [88] ---

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
```

```
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Chapter 1.
Understanding the Digital Forensics Profession and Investigations',
'level_2_title': 'Chapter Review', 'level_3_title': 'Case Projects', 'toc_id':
54}
```

```
Content (Un-chunked Element):
```

```
>> 'Student Resources'
```

```
--- Document [89] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Chapter 1.
Understanding the Digital Forensics Profession and Investigations',
'level_2_title': 'Chapter Review', 'level_3_title': 'Case Projects', 'toc_id':
54}
```

```
Content (Un-chunked Element):
```

```
>> 'MindTap for Guide to Computer Forensics and Investigations helps you learn
on your terms.'
```

```
--- Document [90] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Chapter 1.
Understanding the Digital Forensics Profession and Investigations',
'level_2_title': 'Chapter Review', 'level_3_title': 'Case Projects', 'toc_id':
54}
```

```
Content (Un-chunked Element):
```

```
>> 'Instant access in your pocket: Take advantage of the MindTap Mobile App to
learn on your terms. Read or listen to textbooks and study with the aid of
instructor notifications, flashcards, and practice quizzes.'
```

```
--- Document [91] ---
```

```
Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher
Steuart - Guide to Computer Forensics and Investigations_ Processing Digital
Evidence-Cengage Learning (2018).epub', 'level_1_title': 'Chapter 1.
Understanding the Digital Forensics Profession and Investigations',
'level_2_title': 'Chapter Review', 'level_3_title': 'Case Projects', 'toc_id':
54}
```

```
Content (Un-chunked Element):
```

```
>> 'MindTap helps you create your own potential. Gear up for ultimate success:
Track your scores and stay motivated toward your goals. Whether you have more
work to do or are ahead of the curve, you'll know where you need to focus your
efforts. The MindTap Green Dot will charge your confidence along the way.'
```

--- Document [92] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Chapter 1. Understanding the Digital Forensics Profession and Investigations', 'level\_2\_title': 'Chapter Review', 'level\_3\_title': 'Case Projects', 'toc\_id': 54}

Content (Un-chunked Element):

>> 'MindTap helps you own your progress. Make your textbook yours; no one knows what works for you better than you. Highlight key text, add notes, and create custom flashcards. When it's time to study, everything you've flagged or noted can be gathered into a guide you can organize.'

--- Document [93] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Chapter 1. Understanding the Digital Forensics Profession and Investigations', 'level\_2\_title': 'Chapter Review', 'level\_3\_title': 'Case Projects', 'toc\_id': 54}

Content (Un-chunked Element):

>> 'Lab Manual'

--- Document [94] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Chapter 1. Understanding the Digital Forensics Profession and Investigations', 'level\_2\_title': 'Chapter Review', 'level\_3\_title': 'Case Projects', 'toc\_id': 54}

Content (Un-chunked Element):

>> 'A new digital lab manual is being offered in the MindTap for Guide to Computer Forensics and Investigations to give you additional hands-on experience.'

--- Document [95] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Chapter 1. Understanding the Digital Forensics Profession and Investigations', 'level\_2\_title': 'Chapter Review', 'level\_3\_title': 'Case Projects', 'toc\_id': 54}

Content (Un-chunked Element):

>> 'Lab Requirements'

--- Document [96] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Chapter 1. Understanding the Digital Forensics Profession and Investigations', 'level\_2\_title': 'Chapter Review', 'level\_3\_title': 'Case Projects', 'toc\_id': 54}

Content (Un-chunked Element):

>> 'The hands-on projects in this book help you apply what you have learned about digital forensics techniques. The following sections list the minimum requirements for completing all the projects in this book. In addition to the items listed, you must be able to download and install demo versions of software.'

-----  
--- Document [97] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Chapter 1. Understanding the Digital Forensics Profession and Investigations', 'level\_2\_title': 'Chapter Review', 'level\_3\_title': 'Case Projects', 'toc\_id': 54}

Content (Un-chunked Element):

>> 'Note'

-----  
--- Document [98] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Chapter 1. Understanding the Digital Forensics Profession and Investigations', 'level\_2\_title': 'Chapter Review', 'level\_3\_title': 'Case Projects', 'toc\_id': 54}

Content (Un-chunked Element):

>> 'Note Box'

-----  
--- Document [99] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Chapter 1. Understanding the Digital Forensics Profession and Investigations', 'level\_2\_title': 'Chapter Review', 'level\_3\_title': 'Case Projects', 'toc\_id': 54}

Content (Un-chunked Element):

>> 'Magnet AXIOM has a 30-day trial for download. If you aren't purchasing its academic license or are planning to do the labs at home, it's recommended that

you wait until Chapter 11 to begin using this tool so that your trial doesn't expire before you finish the projects. In addition, wait until Chapter 11 before doing Hands-On Project 6-5, installing Magnet AXIOM.'

-----  
--- Document [100] ---

Assigned Metadata: {'source': 'Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to Computer Forensics and Investigations\_ Processing Digital Evidence-Cengage Learning (2018).epub', 'level\_1\_title': 'Chapter 1. Understanding the Digital Forensics Profession and Investigations', 'level\_2\_title': 'Chapter Review', 'level\_3\_title': 'Case Projects', 'toc\_id': 54}

Content (Un-chunked Element):

>> 'Note'

-----

### 5.1.1 Full Database Health & Hierarchy Diagnostic Report

```
[15]: # Cell 5.1: Full Database Health & Hierarchy Diagnostic Report (V5 - with
      ↪Content Preview)

import os
import json
import logging
import random
from typing import List, Dict, Any

# You might need to install pandas if you haven't already
try:
    import pandas as pd
    pandas_available = True
except ImportError:
    pandas_available = False

try:
    from langchain_chroma import Chroma
    from langchain_ollama.embeddings import OllamaEmbeddings
    from langchain_core.documents import Document
    langchain_available = True
except ImportError:
    langchain_available = False

# Setup Logger
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s -
      ↪%(message)s')
logger = logging.getLogger(__name__)
```

```

# --- HELPER FUNCTIONS ---
def print_header(text: str, char: str = "="):
    """Prints a centered header to the console."""
    print("\n" + char * 80)
    print(text.center(80))
    print(char * 80)

def count_total_chunks(node: Dict) -> int:
    """Recursively counts all chunks in a node and its children."""
    total = node.get('_chunks', 0)
    for child_node in node.get('_children', {}).values():
        total += count_total_chunks(child_node)
    return total

def print_hierarchy_report(node: Dict, indent_level: int = 0):
    """
    Recursively prints the reconstructed hierarchy, sorting by sequential ToC
    ID.
    """
    sorted_children = sorted(
        node.get('_children', {}).items(),
        key=lambda item: item[1].get('_toc_id', float('inf'))
    )

    for title, child_node in sorted_children:
        prefix = " " * indent_level + "|-- "
        total_chunks_in_branch = count_total_chunks(child_node)
        direct_chunks = child_node.get('_chunks', 0)
        toc_id = child_node.get('_toc_id', 'N/A')
        print(f"{prefix}{title} [ID: {toc_id}] (Total Chunk in branch: {
        total_chunks_in_branch}, Direct Chunk: {direct_chunks})")
        print_hierarchy_report(child_node, indent_level + 1)

def find_testable_sections(node: Dict, path: str, testable_list: List):
    """
    Recursively find sections with a decent number of "direct" chunks to test
    sequence on.
    """
    if node.get('_chunks', 0) > 10 and not node.get('_children'):
        testable_list.append({
            "path": path,
            "toc_id": node.get('_toc_id'),
            "chunk_count": node.get('_chunks')
        })

    for title, child_node in node.get('_children', {}).items():

```



```

        new_path = f"{path} -> {title}" if path else title
        find_testable_sections(child_node, new_path, testable_list)

# --- MODIFIED TEST FUNCTION ---
def verify_chunk_sequence_and_content(vector_store: Chroma, hierarchy_tree: Dict):
    """
    Selects a random ToC section, verifies chunk sequence, and displays the
    reassembled content.
    """
    print_header("Chunk Sequence & Content Integrity Test", char="-")
    logger.info("Verifying chunk order and reassembling content for a random
    ToC section.")

    # 1. Find a good section to test
    testable_sections = []
    find_testable_sections(hierarchy_tree, "", testable_sections)

    if not testable_sections:
        logger.warning("Could not find a suitable section with enough chunks to
        test. Skipping content test.")
        return

    random_section = random.choice(testable_sections)
    test_toc_id = random_section['toc_id']
    section_title = random_section['path'].split(' -> ')[-1]

    logger.info(f"Selected random section for testing:
    {random_section['path']} (toc_id: {test_toc_id})")

    # 2. Retrieve all documents (content + metadata) for that toc_id
    try:
        # Use .get() to retrieve full documents, not just similarity search
        retrieved_data = vector_store.get(
            where={"toc_id": test_toc_id},
            include=["metadatas", "documents"]
        )

        # Combine metadatas and documents into LangChain Document objects
        docs = [Document(page_content=doc, metadata=meta) for doc, meta in
        zip(retrieved_data['documents'], retrieved_data['metadatas'])]

        logger.info(f"Retrieved {len(docs)} document chunks for toc_id
        {test_toc_id}.")

```

```

    if len(docs) < 1:
        logger.warning("No chunks found in the selected section. Skipping.")
        return

    # 3. Sort the documents by chunk_id
    # Handle cases where chunk_id might be missing for robustness
    docs.sort(key=lambda d: d.metadata.get('chunk_id', -1))

    chunk_ids = [d.metadata.get('chunk_id') for d in docs]
    if None in chunk_ids:
        logger.error("TEST FAILED: Some retrieved chunks are missing a_
↳ 'chunk_id'.")
        return

    # 4. Verify sequence
    is_sequential = all(chunk_ids[i] == chunk_ids[i-1] + 1 for i in_
↳ range(1, len(chunk_ids)))

    # 5. Reassemble and print content
    full_content = "\n".join([d.page_content for d in docs])

    print("\n" + "-"*25 + " CONTENT PREVIEW " + "-"*25)
    print(f"Title: {section_title} [toc_id: {test_toc_id}]")
    print(f"Chunk IDs: {chunk_ids}")
    print("-" * 70)
    print(full_content)
    print("-" * 23 + " END CONTENT PREVIEW " + "-"*23 + "\n")

    if is_sequential:
        logger.info(" TEST PASSED: Chunk IDs for the section are_
↳ sequential and content is reassembled.")
    else:
        logger.warning("TEST PASSED (with note): Chunk IDs are not_
↳ perfectly sequential but are in increasing order.")
        logger.warning("This is acceptable. Sorting by chunk_id_
↳ successfully restored narrative order.")

    except Exception as e:
        logger.error(f"TEST FAILED: An error occurred during chunk sequence_
↳ verification: {e}", exc_info=True)

# --- MAIN DIAGNOSTIC FUNCTION ---
def run_full_diagnostics():
    if not langchain_available:

```

```

        logger.error("LangChain components not installed. Skipping diagnostics.
↪")

        return

    if not pandas_available:
        logger.warning("Pandas not installed. Some reports may not be available.
↪")

    print_header("Full Database Health & Hierarchy Diagnostic Report")

    # 1. Connect to the Database
    logger.info("Connecting to the vector database...")
    if not os.path.exists(CHROMA_PERSIST_DIR):
        logger.error(f"FATAL: Chroma DB directory not found at_
↪{CHROMA_PERSIST_DIR}.")
        return

    vector_store = Chroma(
        persist_directory=CHROMA_PERSIST_DIR,
        embedding_function=OllamaEmbeddings(model=EMBEDDING_MODEL_OLLAMA),
        collection_name=CHROMA_COLLECTION_NAME
    )
    logger.info("Successfully connected to the database.")

    # 2. Retrieve ALL Metadata
    total_docs = vector_store._collection.count()
    if total_docs == 0:
        logger.warning("Database is empty. No diagnostics to run.")
        return

    logger.info(f"Retrieving metadata for all {total_docs} chunks...")
    metadatas = vector_store.get(limit=total_docs,
↪include=["metadatas"])[['metadatas']]
    logger.info("Successfully retrieved all metadata.")

    # 3. Reconstruct the Hierarchy Tree
    logger.info("Reconstructing hierarchy from chunk metadata...")
    hierarchy_tree = {'_children': {}}
    chunks_without_id = 0

    for meta in metadatas:
        toc_id = meta.get('toc_id')
        if toc_id is None or toc_id == -1:
            chunks_without_id += 1
            node_title = meta.get('level_1_title', 'Orphaned Chunks')
            if node_title not in hierarchy_tree['_children']:
                hierarchy_tree['_children'][node_title] = {'_children': {},
↪'_chunks': 0, '_toc_id': float('inf')}

```

```

        hierarchy_tree['_children'][node_title]['_chunks'] += 1
        continue

    current_node = hierarchy_tree
    for level in range(1, 7):
        level_key = f'level_{level}_title'
        title = meta.get(level_key)
        if not title: break
        if title not in current_node['_children']:
            current_node['_children'][title] = {'_children': {}, '_chunks': 0, '_toc_id': float('inf')}
            current_node = current_node['_children'][title]

        current_node['_chunks'] += 1
        current_node['_toc_id'] = min(current_node['_toc_id'], toc_id)

    logger.info("Hierarchy reconstruction complete.")

# 4. Print Hierarchy Report
    print_header("Reconstructed Hierarchy Report (Book Order)", char="-")
    print_hierarchy_report(hierarchy_tree)

# 5. Run Chunk Sequence and Content Test
    verify_chunk_sequence_and_content(vector_store, hierarchy_tree)

# 6. Final Summary
    print_header("Diagnostic Summary", char="-")
    print(f"Total Chunks in DB: {total_docs}")

    if chunks_without_id > 0:
        logger.warning(f"Found {chunks_without_id} chunks MISSING a valid 'toc_id'. Check 'Orphaned' sections.")
    else:
        logger.info("All chunks contain valid 'toc_id' metadata. Sequential integrity is maintained.")

    print_header("Diagnostic Complete")

# --- Execute Diagnostics ---
    if 'CHROMA_PERSIST_DIR' in locals() and langchain_available:
        run_full_diagnostics()
    else:
        logger.error("Skipping diagnostics: Global variables not defined or LangChain not available.")

```

2025-07-06 01:46:09,599 - INFO - Connecting to the vector database...

2025-07-06 01:46:09,616 - INFO - Successfully connected to the database.

2025-07-06 01:46:09,618 - INFO - Retrieving metadata for all 11774 chunks...

```
=====
Full Database Health & Hierarchy Diagnostic Report
=====
```

```
2025-07-06 01:46:10,138 - INFO - Successfully retrieved all metadata.
2025-07-06 01:46:10,139 - INFO - Reconstructing hierarchy from chunk metadata...
2025-07-06 01:46:10,149 - INFO - Hierarchy reconstruction complete.
2025-07-06 01:46:10,151 - INFO - Verifying chunk order and reassembling content
for a random ToC section.
2025-07-06 01:46:10,152 - INFO - Selected random section for testing: 'Chapter
5. Working with Windows and CLI Systems -> Understanding File Systems ->
Understanding Disk Drives' (toc_id: 187)
2025-07-06 01:46:10,155 - INFO - Retrieved 14 document chunks for toc_id 187.
2025-07-06 01:46:10,156 - INFO - TEST PASSED: Chunk IDs for the section are
sequential and content is reassembled.
2025-07-06 01:46:10,157 - WARNING - Found 21 chunks MISSING a valid 'toc_id'.
Check 'Orphaned' sections.
```

```
-----
Reconstructed Hierarchy Report (Book Order)
-----
```

```
|-- Preface [ID: 3] (Total Chuck in branch: 10, Direct Chunk: 10)
|-- Introduction [ID: 4] (Total Chuck in branch: 73, Direct Chunk: 73)
|-- About the Authors [ID: 5] (Total Chuck in branch: 5, Direct Chunk: 5)
|-- Acknowledgments [ID: 6] (Total Chuck in branch: 20, Direct Chunk: 20)
|-- Chapter 1. Understanding the Digital Forensics Profession and Investigations
[ID: 7] (Total Chuck in branch: 4566, Direct Chunk: 23)
    |-- An Overview of Digital Forensics [ID: 9] (Total Chuck in branch: 60,
Direct Chunk: 18)
        |-- Digital Forensics and Other Related Disciplines [ID: 10] (Total Chuck in
branch: 18, Direct Chunk: 18)
            |-- A Brief History of Digital Forensics [ID: 11] (Total Chuck in branch:
13, Direct Chunk: 13)
                |-- Understanding Case Law [ID: 12] (Total Chuck in branch: 3, Direct Chunk:
3)
                    |-- Developing Digital Forensics Resources [ID: 13] (Total Chuck in branch:
8, Direct Chunk: 8)
                        |-- Preparing for Digital Investigations [ID: 14] (Total Chuck in branch: 84,
Direct Chunk: 5)
                            |-- Understanding Law Enforcement Agency Investigations [ID: 15] (Total
Chuck in branch: 10, Direct Chunk: 10)
                                |-- Following Legal Processes [ID: 16] (Total Chuck in branch: 13, Direct
Chunk: 13)
                                    |-- Understanding Private-Sector Investigations [ID: 17] (Total Chuck in
branch: 56, Direct Chunk: 3)
```

- |-- Establishing Company Policies [ID: 18] (Total Chuck in branch: 5, Direct Chunk: 5)
- |-- Displaying Warning Banners [ID: 19] (Total Chuck in branch: 19, Direct Chunk: 19)
- |-- Designating an Authorized Requester [ID: 20] (Total Chuck in branch: 9, Direct Chunk: 9)
- |-- Conducting Security Investigations [ID: 21] (Total Chuck in branch: 15, Direct Chunk: 15)
- |-- Distinguishing Personal and Company Property [ID: 22] (Total Chuck in branch: 5, Direct Chunk: 5)
- |-- Maintaining Professional Conduct [ID: 23] (Total Chuck in branch: 10, Direct Chunk: 10)
- |-- Preparing a Digital Forensics Investigation [ID: 24] (Total Chuck in branch: 97, Direct Chunk: 4)
- |-- An Overview of a Computer Crime [ID: 25] (Total Chuck in branch: 12, Direct Chunk: 12)
- |-- An Overview of a Company Policy Violation [ID: 26] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Taking a Systematic Approach [ID: 27] (Total Chuck in branch: 77, Direct Chunk: 16)
- |-- Assessing the Case [ID: 28] (Total Chuck in branch: 11, Direct Chunk: 11)
- |-- Planning Your Investigation [ID: 29] (Total Chuck in branch: 41, Direct Chunk: 41)
- |-- Securing Your Evidence [ID: 30] (Total Chuck in branch: 9, Direct Chunk: 9)
- |-- Procedures for Private-Sector High-Tech Investigations [ID: 31] (Total Chuck in branch: 124, Direct Chunk: 2)
- |-- Employee Termination Cases [ID: 32] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Internet Abuse Investigations [ID: 33] (Total Chuck in branch: 19, Direct Chunk: 19)
- |-- E-mail Abuse Investigations [ID: 34] (Total Chuck in branch: 16, Direct Chunk: 16)
- |-- Attorney-Client Privilege Investigations [ID: 35] (Total Chuck in branch: 33, Direct Chunk: 33)
- |-- Industrial Espionage Investigations [ID: 36] (Total Chuck in branch: 52, Direct Chunk: 41)
- |-- Interviews and Interrogations in High-Tech Investigations [ID: 37] (Total Chuck in branch: 11, Direct Chunk: 11)
- |-- Understanding Data Recovery Workstations and Software [ID: 38] (Total Chuck in branch: 37, Direct Chunk: 18)
- |-- Setting Up Your Workstation for Digital Forensics [ID: 39] (Total Chuck in branch: 19, Direct Chunk: 19)
- |-- Conducting an Investigation [ID: 40] (Total Chuck in branch: 109, Direct Chunk: 8)
- |-- Gathering the Evidence [ID: 41] (Total Chuck in branch: 14, Direct Chunk: 14)

- |-- Understanding Bit-stream Copies [ID: 42] (Total Chuck in branch: 8, Direct Chunk: 6)
  - |-- Acquiring an Image of Evidence Media [ID: 43] (Total Chuck in branch: 2, Direct Chunk: 2)
  - |-- Analyzing Your Digital Evidence [ID: 44] (Total Chuck in branch: 48, Direct Chunk: 44)
  - |-- Some Additional Features of Autopsy [ID: 45] (Total Chuck in branch: 4, Direct Chunk: 4)
  - |-- Completing the Case [ID: 46] (Total Chuck in branch: 22, Direct Chunk: 12)
  - |-- Autopsy's Report Generator [ID: 47] (Total Chuck in branch: 10, Direct Chunk: 10)
  - |-- Critiquing the Case [ID: 48] (Total Chuck in branch: 9, Direct Chunk: 9)
  - |-- Chapter Review [ID: inf] (Total Chuck in branch: 4022, Direct Chunk: 0)
  - |-- Chapter Summary [ID: 50] (Total Chuck in branch: 211, Direct Chunk: 211)
  - |-- Key Terms [ID: 51] (Total Chuck in branch: 309, Direct Chunk: 309)
  - |-- Review Questions [ID: 52] (Total Chuck in branch: 1785, Direct Chunk: 1785)
  - |-- Hands-On Projects [ID: 53] (Total Chuck in branch: 1527, Direct Chunk: 1527)
  - |-- Case Projects [ID: 54] (Total Chuck in branch: 190, Direct Chunk: 190)
  - |-- Chapter 2. The Investigator's Office and Laboratory [ID: 55] (Total Chuck in branch: 331, Direct Chunk: 22)
  - |-- Understanding Forensics Lab Accreditation Requirements [ID: 57] (Total Chuck in branch: 86, Direct Chunk: 7)
  - |-- Identifying Duties of the Lab Manager and Staff [ID: 58] (Total Chuck in branch: 7, Direct Chunk: 7)
  - |-- Lab Budget Planning [ID: 59] (Total Chuck in branch: 18, Direct Chunk: 18)
  - |-- Acquiring Certification and Training [ID: 60] (Total Chuck in branch: 54, Direct Chunk: 4)
  - |-- International Association of Computer Investigative Specialists [ID: 61] (Total Chuck in branch: 13, Direct Chunk: 13)
  - |-- ISC2 Certified Cyber Forensics Professional [ID: 62] (Total Chuck in branch: 2, Direct Chunk: 2)
  - |-- High Tech Crime Network [ID: 63] (Total Chuck in branch: 19, Direct Chunk: 19)
  - |-- EnCase Certified Examiner Certification [ID: 64] (Total Chuck in branch: 2, Direct Chunk: 2)
  - |-- AccessData Certified Examiner [ID: 65] (Total Chuck in branch: 2, Direct Chunk: 2)
  - |-- Other Training and Certifications [ID: 66] (Total Chuck in branch: 12, Direct Chunk: 12)
  - |-- Determining the Physical Requirements for a Digital Forensics Lab [ID: 67] (Total Chuck in branch: 68, Direct Chunk: 3)
  - |-- Identifying Lab Security Needs [ID: 68] (Total Chuck in branch: 9, Direct Chunk: 9)
  - |-- Conducting High-Risk Investigations [ID: 69] (Total Chuck in branch: 7,

Direct Chunk: 7)

- |-- Using Evidence Containers [ID: 70] (Total Chuck in branch: 24, Direct Chunk: 24)
- |-- Overseeing Facility Maintenance [ID: 71] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Considering Physical Security Needs [ID: 72] (Total Chuck in branch: 6, Direct Chunk: 6)
- |-- Auditing a Digital Forensics Lab [ID: 73] (Total Chuck in branch: 8, Direct Chunk: 8)
- |-- Determining Floor Plans for Digital Forensics Labs [ID: 74] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Selecting a Basic Forensic Workstation [ID: 75] (Total Chuck in branch: 51, Direct Chunk: 2)
- |-- Selecting Workstations for a Lab [ID: 76] (Total Chuck in branch: 12, Direct Chunk: 12)
- |-- Selecting Workstations for Private-Sector Labs [ID: 77] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Stocking Hardware Peripherals [ID: 78] (Total Chuck in branch: 14, Direct Chunk: 14)
- |-- Maintaining Operating Systems and Software Inventories [ID: 79] (Total Chuck in branch: 9, Direct Chunk: 9)
- |-- Using a Disaster Recovery Plan [ID: 80] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Planning for Equipment Upgrades [ID: 81] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Building a Business Case for Developing a Forensics Lab [ID: 82] (Total Chuck in branch: 104, Direct Chunk: 11)
- |-- Preparing a Business Case for a Digital Forensics Lab [ID: 83] (Total Chuck in branch: 93, Direct Chunk: 2)
  - |-- Justification [ID: 84] (Total Chuck in branch: 8, Direct Chunk: 8)
  - |-- Budget Development [ID: 85] (Total Chuck in branch: 2, Direct Chunk: 2)
  - |-- Facility Cost [ID: 86] (Total Chuck in branch: 15, Direct Chunk: 15)
  - |-- Hardware Requirements [ID: 87] (Total Chuck in branch: 21, Direct Chunk: 21)
  - |-- Software Requirements [ID: 88] (Total Chuck in branch: 23, Direct Chunk: 23)
  - |-- Miscellaneous Budget Needs [ID: 89] (Total Chuck in branch: 4, Direct Chunk: 4)
  - |-- Approval and Acquisition [ID: 90] (Total Chuck in branch: 4, Direct Chunk: 4)
  - |-- Implementation [ID: 91] (Total Chuck in branch: 2, Direct Chunk: 2)
  - |-- Acceptance Testing [ID: 92] (Total Chuck in branch: 6, Direct Chunk: 6)
  - |-- Correction for Acceptance [ID: 93] (Total Chuck in branch: 2, Direct Chunk: 2)
  - |-- Production [ID: 94] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Chapter 3. Data Acquisition [ID: 101] (Total Chuck in branch: 390, Direct



Chunk: 28)

- |-- Understanding Storage Formats for Digital Evidence [ID: 103] (Total Chuck in branch: 31, Direct Chunk: 5)
  - |-- Raw Format [ID: 104] (Total Chuck in branch: 4, Direct Chunk: 4)
  - |-- Proprietary Formats [ID: 105] (Total Chuck in branch: 11, Direct Chunk: 11)
  - |-- Advanced Forensic Format [ID: 106] (Total Chuck in branch: 11, Direct Chunk: 11)
  - |-- Determining the Best Acquisition Method [ID: 107] (Total Chuck in branch: 20, Direct Chunk: 20)
  - |-- Contingency Planning for Image Acquisitions [ID: 108] (Total Chuck in branch: 10, Direct Chunk: 10)
  - |-- Using Acquisition Tools [ID: 109] (Total Chuck in branch: 173, Direct Chunk: 5)
    - |-- Mini-WinFE Boot CDs and USB Drives [ID: 110] (Total Chuck in branch: 9, Direct Chunk: 9)
    - |-- Acquiring Data with a Linux Boot CD [ID: 111] (Total Chuck in branch: 113, Direct Chunk: 5)
      - |-- Using Linux Live CD Distributions [ID: 112] (Total Chuck in branch: 17, Direct Chunk: 17)
      - |-- Preparing a Target Drive for Acquisition in Linux [ID: 113] (Total Chuck in branch: 45, Direct Chunk: 45)
      - |-- Acquiring Data with dd in Linux [ID: 114] (Total Chuck in branch: 32, Direct Chunk: 32)
      - |-- Acquiring Data with dcfldd in Linux [ID: 115] (Total Chuck in branch: 14, Direct Chunk: 14)
      - |-- Capturing an Image with AccessData FTK Imager Lite [ID: 116] (Total Chuck in branch: 46, Direct Chunk: 46)
      - |-- Validating Data Acquisitions [ID: 117] (Total Chuck in branch: 32, Direct Chunk: 5)
        - |-- Linux Validation Methods [ID: 118] (Total Chuck in branch: 21, Direct Chunk: 3)
        - |-- Validating dd-Acquired Data [ID: 119] (Total Chuck in branch: 12, Direct Chunk: 12)
        - |-- Validating dcfldd-Acquired Data [ID: 120] (Total Chuck in branch: 6, Direct Chunk: 6)
        - |-- Windows Validation Methods [ID: 121] (Total Chuck in branch: 6, Direct Chunk: 6)
        - |-- Performing RAID Data Acquisitions [ID: 122] (Total Chuck in branch: 30, Direct Chunk: 2)
        - |-- Understanding RAID [ID: 123] (Total Chuck in branch: 15, Direct Chunk: 15)
        - |-- Acquiring RAID Disks [ID: 124] (Total Chuck in branch: 13, Direct Chunk: 13)
        - |-- Using Remote Network Acquisition Tools [ID: 125] (Total Chuck in branch: 39, Direct Chunk: 5)
        - |-- Remote Acquisition with ProDiscover [ID: 126] (Total Chuck in branch: 20, Direct Chunk: 20)

- |-- Remote Acquisition with EnCase Enterprise [ID: 127] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Remote Acquisition with R-Tools R-Studio [ID: 128] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Remote Acquisition with WetStone US-LATT PRO [ID: 129] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Remote Acquisition with F-Response [ID: 130] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Using Other Forensics Acquisition Tools [ID: 131] (Total Chuck in branch: 27, Direct Chunk: 2)
- |-- PassMark Software ImageUSB [ID: 132] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- ASR Data SMART [ID: 133] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Runtime Software [ID: 134] (Total Chuck in branch: 12, Direct Chunk: 12)
- |-- ILookIX IXImager [ID: 135] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- SourceForge [ID: 136] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Chapter 4. Processing Crime and Incident Scenes [ID: 143] (Total Chuck in branch: 413, Direct Chunk: 29)
- |-- Identifying Digital Evidence [ID: 145] (Total Chuck in branch: 76, Direct Chunk: 13)
- |-- Understanding Rules of Evidence [ID: 146] (Total Chuck in branch: 63, Direct Chunk: 63)
- |-- Collecting Evidence in Private-Sector Incident Scenes [ID: 147] (Total Chuck in branch: 24, Direct Chunk: 24)
- |-- Processing Law Enforcement Crime Scenes [ID: 148] (Total Chuck in branch: 24, Direct Chunk: 6)
- |-- Understanding Concepts and Terms Used in Warrants [ID: 149] (Total Chuck in branch: 18, Direct Chunk: 18)
- |-- Preparing for a Search [ID: 150] (Total Chuck in branch: 40, Direct Chunk: 2)
- |-- Identifying the Nature of the Case [ID: 151] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Identifying the Type of OS or Digital Device [ID: 152] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Determining Whether You Can Seize Computers and Digital Devices [ID: 153] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Getting a Detailed Description of the Location [ID: 154] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Determining Who Is in Charge [ID: 155] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Using Additional Technical Expertise [ID: 156] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Determining the Tools You Need [ID: 157] (Total Chuck in branch: 11, Direct Chunk: 11)
- |-- Preparing the Investigation Team [ID: 158] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Securing a Digital Incident or Crime Scene [ID: 159] (Total Chuck in branch: 9, Direct Chunk: 9)

```

|-- Seizing Digital Evidence at the Scene [ID: 160] (Total Chuck in branch:
72, Direct Chunk: 4)
|-- Preparing to Acquire Digital Evidence [ID: 161] (Total Chuck in branch:
8, Direct Chunk: 8)
|-- Processing Incident or Crime Scenes [ID: 162] (Total Chuck in branch:
34, Direct Chunk: 34)
|-- Processing Data Centers with RAID Systems [ID: 163] (Total Chuck in
branch: 2, Direct Chunk: 2)
|-- Using a Technical Advisor [ID: 164] (Total Chuck in branch: 9, Direct
Chunk: 9)
|-- Documenting Evidence in the Lab [ID: 165] (Total Chuck in branch: 4,
Direct Chunk: 4)
|-- Processing and Handling Digital Evidence [ID: 166] (Total Chuck in
branch: 11, Direct Chunk: 11)
|-- Storing Digital Evidence [ID: 167] (Total Chuck in branch: 18, Direct
Chunk: 7)
|-- Evidence Retention and Media Storage Needs [ID: 168] (Total Chuck in
branch: 4, Direct Chunk: 4)
|-- Documenting Evidence [ID: 169] (Total Chuck in branch: 7, Direct Chunk:
7)
|-- Obtaining a Digital Hash [ID: 170] (Total Chuck in branch: 42, Direct
Chunk: 42)
|-- Reviewing a Case [ID: 171] (Total Chuck in branch: 79, Direct Chunk: 8)
|-- Sample Civil Investigation [ID: 172] (Total Chuck in branch: 23, Direct
Chunk: 23)
|-- An Example of a Criminal Investigation [ID: 173] (Total Chuck in branch:
4, Direct Chunk: 4)
|-- Reviewing Background Information for a Case [ID: 174] (Total Chuck in
branch: 4, Direct Chunk: 4)
|-- Planning the Investigation [ID: 175] (Total Chuck in branch: 7, Direct
Chunk: 7)
|-- Conducting the Investigation: Acquiring Evidence with OSForensics [ID:
176] (Total Chuck in branch: 33, Direct Chunk: 33)
|-- Chapter 5. Working with Windows and CLI Systems [ID: 183] (Total Chuck in
branch: 471, Direct Chunk: 22)
|-- Understanding File Systems [ID: 185] (Total Chuck in branch: 33, Direct
Chunk: 3)
|-- Understanding the Boot Sequence [ID: 186] (Total Chuck in branch: 8,
Direct Chunk: 8)
|-- Understanding Disk Drives [ID: 187] (Total Chuck in branch: 14, Direct
Chunk: 14)
|-- Solid-State Storage Devices [ID: 188] (Total Chuck in branch: 8, Direct
Chunk: 8)
|-- Exploring Microsoft File Structures [ID: 189] (Total Chuck in branch: 71,
Direct Chunk: 5)
|-- Disk Partitions [ID: 190] (Total Chuck in branch: 39, Direct Chunk: 39)
|-- Examining FAT Disks [ID: 191] (Total Chuck in branch: 27, Direct Chunk:
24)

```

```

    |-- Deleting FAT Files [ID: 192] (Total Chuck in branch: 3, Direct Chunk:
3)
    |-- Examining NTFS Disks [ID: 193] (Total Chuck in branch: 168, Direct Chunk:
14)
        |-- NTFS System Files [ID: 194] (Total Chuck in branch: 6, Direct Chunk: 6)
        |-- MFT and File Attributes [ID: 195] (Total Chuck in branch: 20, Direct
Chunk: 20)
            |-- MFT Structures for File Data [ID: 196] (Total Chuck in branch: 69,
Direct Chunk: 3)
                |-- MFT Header Fields [ID: 197] (Total Chuck in branch: 7, Direct Chunk:
7)
                    |-- Attribute 0x10: Standard Information [ID: 198] (Total Chuck in branch:
8, Direct Chunk: 8)
                        |-- Attribute 0x30: File Name [ID: 199] (Total Chuck in branch: 18, Direct
Chunk: 18)
                            |-- Attribute 0x40: Object_ID [ID: 200] (Total Chuck in branch: 10, Direct
Chunk: 10)
                                |-- Attribute 0x80: Data for a Resident File [ID: 201] (Total Chuck in
branch: 8, Direct Chunk: 8)
                                    |-- Attribute 0x80: Data for a Nonresident File [ID: 202] (Total Chuck in
branch: 6, Direct Chunk: 6)
  |-- Interpreting a Data Run [ID: 203] (Total Chuck in branch: 9, Direct
Chunk: 9)
  |-- NTFS Alternate Data Streams [ID: 204] (Total Chuck in branch: 14, Direct
Chunk: 14)
  |-- NTFS Compressed Files [ID: 205] (Total Chuck in branch: 3, Direct Chunk:
3)
  |-- NTFS Encrypting File System [ID: 206] (Total Chuck in branch: 5, Direct
Chunk: 5)
  |-- EFS Recovery Key Agent [ID: 207] (Total Chuck in branch: 8, Direct
Chunk: 8)
  |-- Deleting NTFS Files [ID: 208] (Total Chuck in branch: 20, Direct Chunk:
20)
  |-- Resilient File System [ID: 209] (Total Chuck in branch: 9, Direct Chunk:
9)
  |-- Understanding Whole Disk Encryption [ID: 210] (Total Chuck in branch: 26,
Direct Chunk: 11)
  |-- Examining Microsoft BitLocker [ID: 211] (Total Chuck in branch: 9,
Direct Chunk: 9)
  |-- Examining Third-Party Disk Encryption Tools [ID: 212] (Total Chuck in
branch: 6, Direct Chunk: 6)
  |-- Understanding the Windows Registry [ID: 213] (Total Chuck in branch: 56,
Direct Chunk: 9)
  |-- Exploring the Organization of the Windows Registry [ID: 214] (Total
Chuck in branch: 18, Direct Chunk: 18)
  |-- Examining the Windows Registry [ID: 215] (Total Chuck in branch: 29,
Direct Chunk: 29)
  |-- Understanding Microsoft Startup Tasks [ID: 216] (Total Chuck in branch:

```

47, Direct Chunk: 3)

- |-- Startup in Windows 7, Windows 8, and Windows 10 [ID: 217] (Total Chuck in branch: 5, Direct Chunk: 5)
- |-- Startup in Windows NT and Later [ID: 218] (Total Chuck in branch: 39, Direct Chunk: 10)
- |-- Startup Files for Windows Vista [ID: 219] (Total Chuck in branch: 6, Direct Chunk: 6)
- |-- Startup Files for Windows XP [ID: 220] (Total Chuck in branch: 17, Direct Chunk: 17)
- |-- Windows XP System Files [ID: 221] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Contamination Concerns with Windows XP [ID: 222] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Understanding Virtual Machines [ID: 223] (Total Chuck in branch: 48, Direct Chunk: 10)
- |-- Creating a Virtual Machine [ID: 224] (Total Chuck in branch: 38, Direct Chunk: 38)
- |-- Chapter 6. Current Digital Forensics Tools [ID: 231] (Total Chuck in branch: 315, Direct Chunk: 22)
- |-- Evaluating Digital Forensics Tool Needs [ID: 233] (Total Chuck in branch: 184, Direct Chunk: 10)
- |-- Types of Digital Forensics Tools [ID: 234] (Total Chuck in branch: 11, Direct Chunk: 4)
- |-- Hardware Forensics Tools [ID: 235] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Software Forensics Tools [ID: 236] (Total Chuck in branch: 5, Direct Chunk: 5)
- |-- Tasks Performed by Digital Forensics Tools [ID: 237] (Total Chuck in branch: 153, Direct Chunk: 3)
- |-- Acquisition [ID: 238] (Total Chuck in branch: 22, Direct Chunk: 22)
- |-- Validation and Verification [ID: 239] (Total Chuck in branch: 14, Direct Chunk: 14)
- |-- Extraction [ID: 240] (Total Chuck in branch: 25, Direct Chunk: 25)
- |-- Reconstruction [ID: 241] (Total Chuck in branch: 66, Direct Chunk: 66)
- |-- Reporting [ID: 242] (Total Chuck in branch: 23, Direct Chunk: 23)
- |-- Tool Comparisons [ID: 243] (Total Chuck in branch: 6, Direct Chunk: 6)
- |-- Other Considerations for Tools [ID: 244] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Digital Forensics Software Tools [ID: 245] (Total Chuck in branch: 41, Direct Chunk: 4)
- |-- Command-Line Forensics Tools [ID: 246] (Total Chuck in branch: 14, Direct Chunk: 14)
- |-- Linux Forensics Tools [ID: 247] (Total Chuck in branch: 19, Direct Chunk: 4)
- |-- Smart [ID: 248] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Helix 3 [ID: 249] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Kali Linux [ID: 250] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Autopsy and Sleuth Kit [ID: 251] (Total Chuck in branch: 4, Direct

```

Chunk: 4)
    |-- Forcepoint Threat Protection [ID: 252] (Total Chuck in branch: 2,
Direct Chunk: 2)
    |-- Other GUI Forensics Tools [ID: 253] (Total Chuck in branch: 4, Direct
Chunk: 4)
    |-- Digital Forensics Hardware Tools [ID: 254] (Total Chuck in branch: 38,
Direct Chunk: 3)
    |-- Forensic Workstations [ID: 255] (Total Chuck in branch: 13, Direct
Chunk: 7)
    |-- Building Your Own Workstation [ID: 256] (Total Chuck in branch: 6,
Direct Chunk: 6)
    |-- Using a Write-Blocker [ID: 257] (Total Chuck in branch: 17, Direct
Chunk: 17)
    |-- Recommendations for a Forensic Workstation [ID: 258] (Total Chuck in
branch: 5, Direct Chunk: 5)
    |-- Validating and Testing Forensics Software [ID: 259] (Total Chuck in
branch: 30, Direct Chunk: 2)
    |-- Using National Institute of Standards and Technology Tools [ID: 260]
(Total Chuck in branch: 13, Direct Chunk: 13)
    |-- Using Validation Protocols [ID: 261] (Total Chuck in branch: 15, Direct
Chunk: 6)
    |-- Digital Forensics Examination Protocol [ID: 262] (Total Chuck in
branch: 5, Direct Chunk: 5)
    |-- Digital Forensics Tool Upgrade Protocol [ID: 263] (Total Chuck in
branch: 4, Direct Chunk: 4)
|-- Chapter 7. Linux and Macintosh File Systems [ID: 270] (Total Chuck in
branch: 274, Direct Chunk: 17)
    |-- Examining Linux File Structures [ID: 272] (Total Chuck in branch: 131,
Direct Chunk: 77)
    |-- File Structures in Ext4 [ID: 273] (Total Chuck in branch: 54, Direct
Chunk: 8)
    |-- Inodes [ID: 274] (Total Chuck in branch: 22, Direct Chunk: 22)
    |-- Hard Links and Symbolic Links [ID: 275] (Total Chuck in branch: 24,
Direct Chunk: 24)
    |-- Understanding Macintosh File Structures [ID: 276] (Total Chuck in branch:
58, Direct Chunk: 6)
    |-- An Overview of Mac File Structures [ID: 277] (Total Chuck in branch: 23,
Direct Chunk: 23)
    |-- Forensics Procedures in Mac [ID: 278] (Total Chuck in branch: 29, Direct
Chunk: 18)
    |-- Acquisition Methods in macOS [ID: 279] (Total Chuck in branch: 11,
Direct Chunk: 11)
    |-- Using Linux Forensics Tools [ID: 280] (Total Chuck in branch: 68, Direct
Chunk: 5)
    |-- Installing Sleuth Kit and Autopsy [ID: 281] (Total Chuck in branch: 21,
Direct Chunk: 21)
    |-- Examining a Case with Sleuth Kit and Autopsy [ID: 282] (Total Chuck in
branch: 42, Direct Chunk: 42)

```

|-- Chapter 8. Recovering Graphics Files [ID: 289] (Total Chuck in branch: 240, Direct Chunk: 19)

    |-- Recognizing a Graphics File [ID: 291] (Total Chuck in branch: 54, Direct Chunk: 4)

        |-- Understanding Bitmap and Raster Images [ID: 292] (Total Chuck in branch: 13, Direct Chunk: 13)

        |-- Understanding Vector Graphics [ID: 293] (Total Chuck in branch: 2, Direct Chunk: 2)

        |-- Understanding Metafile Graphics [ID: 294] (Total Chuck in branch: 2, Direct Chunk: 2)

        |-- Understanding Graphics File Formats [ID: 295] (Total Chuck in branch: 14, Direct Chunk: 14)

        |-- Understanding Digital Photograph File Formats [ID: 296] (Total Chuck in branch: 19, Direct Chunk: 2)

        |-- Examining the Raw File Format [ID: 297] (Total Chuck in branch: 5, Direct Chunk: 5)

        |-- Examining the Exchangeable Image File Format [ID: 298] (Total Chuck in branch: 12, Direct Chunk: 12)

        |-- Understanding Data Compression [ID: 299] (Total Chuck in branch: 101, Direct Chunk: 2)

        |-- Lossless and Lossy Compression [ID: 300] (Total Chuck in branch: 8, Direct Chunk: 8)

        |-- Locating and Recovering Graphics Files [ID: 301] (Total Chuck in branch: 6, Direct Chunk: 6)

        |-- Identifying Graphics File Fragments [ID: 302] (Total Chuck in branch: 3, Direct Chunk: 3)

        |-- Repairing Damaged Headers [ID: 303] (Total Chuck in branch: 6, Direct Chunk: 6)

        |-- Searching for and Carving Data from Unallocated Space [ID: 304] (Total Chuck in branch: 39, Direct Chunk: 9)

        |-- Planning Your Examination [ID: 305] (Total Chuck in branch: 4, Direct Chunk: 4)

        |-- Searching for and Recovering Digital Photograph Evidence [ID: 306] (Total Chuck in branch: 26, Direct Chunk: 26)

        |-- Rebuilding File Headers [ID: 307] (Total Chuck in branch: 22, Direct Chunk: 22)

        |-- Reconstructing File Fragments [ID: 308] (Total Chuck in branch: 15, Direct Chunk: 15)

        |-- Identifying Unknown File Formats [ID: 309] (Total Chuck in branch: 47, Direct Chunk: 14)

        |-- Analyzing Graphics File Headers [ID: 310] (Total Chuck in branch: 5, Direct Chunk: 5)

        |-- Tools for Viewing Images [ID: 311] (Total Chuck in branch: 5, Direct Chunk: 5)

        |-- Understanding Steganography in Graphics Files [ID: 312] (Total Chuck in branch: 16, Direct Chunk: 16)

        |-- Using Steganalysis Tools [ID: 313] (Total Chuck in branch: 7, Direct Chunk: 7)

```

|-- Understanding Copyright Issues with Graphics [ID: 314] (Total Chuck in
branch: 19, Direct Chunk: 19)
|-- Chapter 9. Digital Forensics Analysis and Validation [ID: 321] (Total Chuck
in branch: 248, Direct Chunk: 16)
|-- Determining What Data to Collect and Analyze [ID: 323] (Total Chuck in
branch: 99, Direct Chunk: 6)
|-- Approaching Digital Forensics Cases [ID: 324] (Total Chuck in branch:
35, Direct Chunk: 30)
|-- Refining and Modifying the Investigation Plan [ID: 325] (Total Chuck
in branch: 5, Direct Chunk: 5)
|-- Using Autopsy to Validate Data [ID: 326] (Total Chuck in branch: 23,
Direct Chunk: 8)
|-- Installing NSRL Hashes in Autopsy [ID: 327] (Total Chuck in branch:
15, Direct Chunk: 15)
|-- Collecting Hash Values in Autopsy [ID: 328] (Total Chuck in branch: 35,
Direct Chunk: 35)
|-- Validating Forensic Data [ID: 329] (Total Chuck in branch: 51, Direct
Chunk: 3)
|-- Validating with Hexadecimal Editors [ID: 330] (Total Chuck in branch:
31, Direct Chunk: 28)
|-- Using Hash Values to Discriminate Data [ID: 331] (Total Chuck in
branch: 3, Direct Chunk: 3)
|-- Validating with Digital Forensics Tools [ID: 332] (Total Chuck in
branch: 17, Direct Chunk: 17)
|-- Addressing Data-Hiding Techniques [ID: 333] (Total Chuck in branch: 82,
Direct Chunk: 2)
|-- Hiding Files by Using the OS [ID: 334] (Total Chuck in branch: 3, Direct
Chunk: 3)
|-- Hiding Partitions [ID: 335] (Total Chuck in branch: 8, Direct Chunk: 8)
|-- Marking Bad Clusters [ID: 336] (Total Chuck in branch: 6, Direct Chunk:
6)
|-- Bit-Shifting [ID: 337] (Total Chuck in branch: 34, Direct Chunk: 34)
|-- Understanding Steganalysis Methods [ID: 338] (Total Chuck in branch: 10,
Direct Chunk: 10)
|-- Examining Encrypted Files [ID: 339] (Total Chuck in branch: 4, Direct
Chunk: 4)
|-- Recovering Passwords [ID: 340] (Total Chuck in branch: 15, Direct Chunk:
15)
|-- Chapter 10. Virtual Machine Forensics, Live Acquisitions, and Network
Forensics [ID: 347] (Total Chuck in branch: 287, Direct Chunk: 17)
|-- An Overview of Virtual Machine Forensics [ID: 349] (Total Chuck in branch:
176, Direct Chunk: 7)
|-- Type 2 Hypervisors [ID: 350] (Total Chuck in branch: 32, Direct Chunk:
6)
|-- Parallels Desktop [ID: 351] (Total Chuck in branch: 2, Direct Chunk:
2)
|-- KVM [ID: 352] (Total Chuck in branch: 2, Direct Chunk: 2)
|-- Microsoft Hyper-V [ID: 353] (Total Chuck in branch: 2, Direct Chunk:

```



2)

- |-- VMware Workstation and Workstation Player [ID: 354] (Total Chuck in branch: 14, Direct Chunk: 14)
- |-- VirtualBox [ID: 355] (Total Chuck in branch: 6, Direct Chunk: 6)
- |-- Conducting an Investigation with Type 2 Hypervisors [ID: 356] (Total Chuck in branch: 103, Direct Chunk: 70)
- |-- Other VM Examination Methods [ID: 357] (Total Chuck in branch: 13, Direct Chunk: 13)
- |-- Using VMs as Forensics Tools [ID: 358] (Total Chuck in branch: 20, Direct Chunk: 20)
- |-- Working with Type 1 Hypervisors [ID: 359] (Total Chuck in branch: 34, Direct Chunk: 34)
- |-- Performing Live Acquisitions [ID: 360] (Total Chuck in branch: 18, Direct Chunk: 15)
- |-- Performing a Live Acquisition in Windows [ID: 361] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Network Forensics Overview [ID: 362] (Total Chuck in branch: 76, Direct Chunk: 4)
- |-- The Need for Established Procedures [ID: 363] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Securing a Network [ID: 364] (Total Chuck in branch: 13, Direct Chunk: 13)
- |-- Developing Procedures for Network Forensics [ID: 365] (Total Chuck in branch: 41, Direct Chunk: 8)
- |-- Reviewing Network Logs [ID: 366] (Total Chuck in branch: 11, Direct Chunk: 11)
- |-- Using Network Tools [ID: 367] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Using Packet Analyzers [ID: 368] (Total Chuck in branch: 18, Direct Chunk: 18)
- |-- Investigating Virtual Networks [ID: 369] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Examining the Honeynet Project [ID: 370] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Chapter 11. E-mail and Social Media Investigations [ID: 377] (Total Chuck in branch: 302, Direct Chunk: 20)
- |-- Exploring the Role of E-mail in Investigations [ID: 379] (Total Chuck in branch: 11, Direct Chunk: 11)
- |-- Exploring the Roles of the Client and Server in E-mail [ID: 380] (Total Chuck in branch: 10, Direct Chunk: 10)
- |-- Investigating E-mail Crimes and Violations [ID: 381] (Total Chuck in branch: 101, Direct Chunk: 4)
- |-- Understanding Forensic Linguistics [ID: 382] (Total Chuck in branch: 6, Direct Chunk: 6)
- |-- Examining E-mail Messages [ID: 383] (Total Chuck in branch: 28, Direct Chunk: 8)
- |-- Copying an E-mail Message [ID: 384] (Total Chuck in branch: 20, Direct Chunk: 20)

- |-- Viewing E-mail Headers [ID: 385] (Total Chuck in branch: 33, Direct Chunk: 33)
- |-- Examining E-mail Headers [ID: 386] (Total Chuck in branch: 14, Direct Chunk: 14)
- |-- Examining Additional E-mail Files [ID: 387] (Total Chuck in branch: 6, Direct Chunk: 6)
- |-- Tracing an E-mail Message [ID: 388] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Using Network E-mail Logs [ID: 389] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Understanding E-mail Servers [ID: 390] (Total Chuck in branch: 33, Direct Chunk: 13)
- |-- Examining UNIX E-mail Server Logs [ID: 391] (Total Chuck in branch: 12, Direct Chunk: 12)
- |-- Examining Microsoft E-mail Server Logs [ID: 392] (Total Chuck in branch: 8, Direct Chunk: 8)
- |-- Using Specialized E-mail Forensics Tools [ID: 393] (Total Chuck in branch: 91, Direct Chunk: 22)
- |-- Using Magnet AXIOM to Recover E-mail [ID: 394] (Total Chuck in branch: 14, Direct Chunk: 14)
- |-- Using a Hex Editor to Carve E-mail Messages [ID: 395] (Total Chuck in branch: 41, Direct Chunk: 41)
- |-- Recovering Outlook Files [ID: 396] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- E-mail Case Studies [ID: 397] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Applying Digital Forensics Methods to Social Media Communications [ID: 398] (Total Chuck in branch: 36, Direct Chunk: 23)
- |-- Forensics Tools for Social Media Investigations [ID: 399] (Total Chuck in branch: 13, Direct Chunk: 13)
- |-- Chapter 12. Mobile Device Forensics and the Internet of Anything [ID: 406] (Total Chuck in branch: 169, Direct Chunk: 8)
- |-- Understanding Mobile Device Forensics [ID: 408] (Total Chuck in branch: 65, Direct Chunk: 19)
- |-- Mobile Phone Basics [ID: 409] (Total Chuck in branch: 24, Direct Chunk: 24)
- |-- Inside Mobile Devices [ID: 410] (Total Chuck in branch: 22, Direct Chunk: 11)
- |-- SIM Cards [ID: 411] (Total Chuck in branch: 11, Direct Chunk: 11)
- |-- Understanding Acquisition Procedures for Mobile Devices [ID: 412] (Total Chuck in branch: 75, Direct Chunk: 30)
- |-- Mobile Forensics Equipment [ID: 413] (Total Chuck in branch: 30, Direct Chunk: 6)
- |-- SIM Card Readers [ID: 414] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Mobile Phone Forensics Tools and Methods [ID: 415] (Total Chuck in branch: 17, Direct Chunk: 17)
- |-- Using Mobile Forensics Tools [ID: 416] (Total Chuck in branch: 15, Direct Chunk: 15)

```

|-- Understanding Forensics in the Internet of Anything [ID: 417] (Total Chuck
in branch: 21, Direct Chunk: 21)
|-- Chapter 13. Cloud Forensics [ID: 424] (Total Chuck in branch: 290, Direct
Chunk: 20)
|-- An Overview of Cloud Computing [ID: 426] (Total Chuck in branch: 42,
Direct Chunk: 2)
|-- History of the Cloud [ID: 427] (Total Chuck in branch: 4, Direct Chunk:
4)
|-- Cloud Service Levels and Deployment Methods [ID: 428] (Total Chuck in
branch: 13, Direct Chunk: 13)
|-- Cloud Vendors [ID: 429] (Total Chuck in branch: 15, Direct Chunk: 15)
|-- Basic Concepts of Cloud Forensics [ID: 430] (Total Chuck in branch: 8,
Direct Chunk: 8)
|-- Legal Challenges in Cloud Forensics [ID: 431] (Total Chuck in branch: 56,
Direct Chunk: 2)
|-- Service Level Agreements [ID: 432] (Total Chuck in branch: 25, Direct
Chunk: 17)
|-- Policies, Standards, and Guidelines for CSPs [ID: 433] (Total Chuck in
branch: 5, Direct Chunk: 5)
|-- CSP Processes and Procedures [ID: 434] (Total Chuck in branch: 3,
Direct Chunk: 3)
|-- Jurisdiction Issues [ID: 435] (Total Chuck in branch: 6, Direct Chunk:
6)
|-- Accessing Evidence in the Cloud [ID: 436] (Total Chuck in branch: 23,
Direct Chunk: 3)
|-- Search Warrants [ID: 437] (Total Chuck in branch: 10, Direct Chunk:
10)
|-- Subpoenas and Court Orders [ID: 438] (Total Chuck in branch: 10,
Direct Chunk: 10)
|-- Technical Challenges in Cloud Forensics [ID: 439] (Total Chuck in branch:
33, Direct Chunk: 5)
|-- Architecture [ID: 440] (Total Chuck in branch: 12, Direct Chunk: 12)
|-- Analysis of Cloud Forensic Data [ID: 441] (Total Chuck in branch: 2,
Direct Chunk: 2)
|-- Anti-Forensics [ID: 442] (Total Chuck in branch: 3, Direct Chunk: 3)
|-- Incident First Responders [ID: 443] (Total Chuck in branch: 5, Direct
Chunk: 5)
|-- Role Management [ID: 444] (Total Chuck in branch: 2, Direct Chunk: 2)
|-- Standards and Training [ID: 445] (Total Chuck in branch: 4, Direct
Chunk: 4)
|-- Acquisitions in the Cloud [ID: 446] (Total Chuck in branch: 21, Direct
Chunk: 8)
|-- Encryption in the Cloud [ID: 447] (Total Chuck in branch: 13, Direct
Chunk: 13)
|-- Conducting a Cloud Investigation [ID: 448] (Total Chuck in branch: 105,
Direct Chunk: 2)
|-- Investigating CSPs [ID: 449] (Total Chuck in branch: 8, Direct Chunk: 8)
|-- Investigating Cloud Customers [ID: 450] (Total Chuck in branch: 3,

```

Direct Chunk: 3)

- |-- Understanding Prefetch Files [ID: 451] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Examining Stored Cloud Data on a PC [ID: 452] (Total Chuck in branch: 63, Direct Chunk: 5)
  - |-- Dropbox [ID: 453] (Total Chuck in branch: 8, Direct Chunk: 8)
  - |-- Google Drive [ID: 454] (Total Chuck in branch: 21, Direct Chunk: 21)
  - |-- OneDrive [ID: 455] (Total Chuck in branch: 29, Direct Chunk: 29)
- |-- Windows Prefetch Artifacts [ID: 456] (Total Chuck in branch: 26, Direct Chunk: 26)
- |-- Tools for Cloud Forensics [ID: 457] (Total Chuck in branch: 13, Direct Chunk: 5)
- |-- Forensic Open-Stack Tools [ID: 458] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- F-Response for the Cloud [ID: 459] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Magnet AXIOM Cloud [ID: 460] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Chapter 14. Report Writing for High-Tech Investigations [ID: 467] (Total Chuck in branch: 263, Direct Chunk: 16)
- |-- Understanding the Importance of Reports [ID: 469] (Total Chuck in branch: 31, Direct Chunk: 19)
- |-- Limiting a Report to Specifics [ID: 470] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Types of Reports [ID: 471] (Total Chuck in branch: 9, Direct Chunk: 9)
- |-- Guidelines for Writing Reports [ID: 472] (Total Chuck in branch: 138, Direct Chunk: 19)
- |-- What to Include in Written Preliminary Reports [ID: 473] (Total Chuck in branch: 9, Direct Chunk: 9)
- |-- Report Structure [ID: 474] (Total Chuck in branch: 8, Direct Chunk: 8)
- |-- Writing Reports Clearly [ID: 475] (Total Chuck in branch: 17, Direct Chunk: 8)
- |-- Considering Writing Style [ID: 476] (Total Chuck in branch: 6, Direct Chunk: 6)
- |-- Including Signposts [ID: 477] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Designing the Layout and Presentation of Reports [ID: 478] (Total Chuck in branch: 85, Direct Chunk: 44)
- |-- Providing Supporting Material [ID: 479] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Formatting Consistently [ID: 480] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Explaining Examination and Data Collection Methods [ID: 481] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Including Calculations [ID: 482] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Providing for Uncertainty and Error Analysis [ID: 483] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Explaining Results and Conclusions [ID: 484] (Total Chuck in branch:

3, Direct Chunk: 3)

- |-- Providing References [ID: 485] (Total Chuck in branch: 20, Direct Chunk: 20)
- |-- Including Appendixes [ID: 486] (Total Chuck in branch: 6, Direct Chunk: 6)
- |-- Generating Report Findings with Forensics Software Tools [ID: 487] (Total Chuck in branch: 78, Direct Chunk: 2)
- |-- Using Autopsy to Generate Reports [ID: 488] (Total Chuck in branch: 76, Direct Chunk: 76)
- |-- Chapter 15. Expert Testimony in Digital Investigations [ID: 495] (Total Chuck in branch: 329, Direct Chunk: 14)
- |-- Preparing for Testimony [ID: 497] (Total Chuck in branch: 62, Direct Chunk: 15)
- |-- Documenting and Preparing Evidence [ID: 498] (Total Chuck in branch: 12, Direct Chunk: 12)
- |-- Reviewing Your Role as a Consulting Expert or an Expert Witness [ID: 499] (Total Chuck in branch: 9, Direct Chunk: 9)
- |-- Creating and Maintaining Your CV [ID: 500] (Total Chuck in branch: 8, Direct Chunk: 8)
- |-- Preparing Technical Definitions [ID: 501] (Total Chuck in branch: 12, Direct Chunk: 12)
- |-- Preparing to Deal with the News Media [ID: 502] (Total Chuck in branch: 6, Direct Chunk: 6)
- |-- Testifying in Court [ID: 503] (Total Chuck in branch: 155, Direct Chunk: 2)
- |-- Understanding the Trial Process [ID: 504] (Total Chuck in branch: 10, Direct Chunk: 10)
- |-- Providing Qualifications for Your Testimony [ID: 505] (Total Chuck in branch: 59, Direct Chunk: 59)
- |-- General Guidelines on Testifying [ID: 506] (Total Chuck in branch: 47, Direct Chunk: 27)
- |-- Using Graphics During Testimony [ID: 507] (Total Chuck in branch: 10, Direct Chunk: 10)
- |-- Avoiding Testimony Problems [ID: 508] (Total Chuck in branch: 7, Direct Chunk: 7)
- |-- Understanding Prosecutorial Misconduct [ID: 509] (Total Chuck in branch: 3, Direct Chunk: 3)
- |-- Testifying During Direct Examination [ID: 510] (Total Chuck in branch: 12, Direct Chunk: 12)
- |-- Testifying During Cross-Examination [ID: 511] (Total Chuck in branch: 25, Direct Chunk: 25)
- |-- Preparing for a Deposition or Hearing [ID: 512] (Total Chuck in branch: 33, Direct Chunk: 6)
- |-- Guidelines for Testifying at Depositions [ID: 513] (Total Chuck in branch: 19, Direct Chunk: 10)
- |-- Recognizing Deposition Problems [ID: 514] (Total Chuck in branch: 9, Direct Chunk: 9)
- |-- Guidelines for Testifying at Hearings [ID: 515] (Total Chuck in branch:

8, Direct Chunk: 8)

- |-- Preparing Forensics Evidence for Testimony [ID: 516] (Total Chuck in branch: 65, Direct Chunk: 34)
- |-- Preparing a Defense of Your Evidence-Collection Methods [ID: 517] (Total Chuck in branch: 31, Direct Chunk: 31)
- |-- Chapter 16. Ethics for the Expert Witness [ID: 524] (Total Chuck in branch: 310, Direct Chunk: 13)
- |-- Applying Ethics and Codes to Expert Witnesses [ID: 526] (Total Chuck in branch: 58, Direct Chunk: 11)
- |-- Forensics Examiners' Roles in Testifying [ID: 527] (Total Chuck in branch: 5, Direct Chunk: 5)
- |-- Considerations in Disqualification [ID: 528] (Total Chuck in branch: 22, Direct Chunk: 22)
- |-- Traps for Unwary Experts [ID: 529] (Total Chuck in branch: 16, Direct Chunk: 16)
- |-- Determining Admissibility of Evidence [ID: 530] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- Organizations with Codes of Ethics [ID: 531] (Total Chuck in branch: 26, Direct Chunk: 2)
- |-- International Society of Forensic Computer Examiners [ID: 532] (Total Chuck in branch: 11, Direct Chunk: 11)
- |-- International High Technology Crime Investigation Association [ID: 533] (Total Chuck in branch: 6, Direct Chunk: 6)
- |-- American Bar Association [ID: 535] (Total Chuck in branch: 5, Direct Chunk: 5)
- |-- American Psychological Association [ID: 536] (Total Chuck in branch: 2, Direct Chunk: 2)
- |-- Ethical Difficulties in Expert Testimony [ID: 537] (Total Chuck in branch: 19, Direct Chunk: 6)
- |-- Ethical Responsibilities Owed to You [ID: 538] (Total Chuck in branch: 9, Direct Chunk: 9)
- |-- Standard Forensics Tools and Tools You Create [ID: 539] (Total Chuck in branch: 4, Direct Chunk: 4)
- |-- An Ethics Exercise [ID: 540] (Total Chuck in branch: 194, Direct Chunk: 4)
- |-- Performing a cursory Exam of a Forensic Image [ID: 541] (Total Chuck in branch: 18, Direct Chunk: 18)
- |-- Performing a Detailed Exam of a Forensic Image [ID: 542] (Total Chuck in branch: 33, Direct Chunk: 33)
- |-- Performing the Exam [ID: 543] (Total Chuck in branch: 76, Direct Chunk: 2)
- |-- Preparing for an Examination [ID: 544] (Total Chuck in branch: 74, Direct Chunk: 74)
- |-- Interpreting Attribute 0x80 Data Runs [ID: 545] (Total Chuck in branch: 44, Direct Chunk: 2)
- |-- Finding Attribute 0x80 an MFT Record [ID: 546] (Total Chuck in branch: 21, Direct Chunk: 21)
- |-- Configuring Data Interpreter Options in WinHex [ID: 547] (Total Chuck in branch: 6, Direct Chunk: 6)

```

|-- Calculating Data Runs [ID: 548] (Total Chuck in branch: 15, Direct
Chunk: 15)
|-- Carving Data Run Clusters Manually [ID: 549] (Total Chuck in branch: 19,
Direct Chunk: 19)
|-- Lab Manual for Guide to Computer Forensics and Investigations [ID: 556]
(Total Chuck in branch: 2165, Direct Chunk: 1)
|-- Chapter 12. Mobile Device Forensics [ID: 792] (Total Chuck in branch: 13,
Direct Chunk: 10)
|-- Lab 12.1. Examining Cell Phone Storage Devices [ID: 794] (Total Chuck in
branch: 1, Direct Chunk: 1)
|-- Lab 12.2. Using FTK Imager Lite to View Text Messages, Phone Numbers,
and Photos [ID: 799] (Total Chuck in branch: 1, Direct Chunk: 1)
|-- Lab 12.3. Using Autopsy to Search Cloud Backups of Mobile Devices [ID:
804] (Total Chuck in branch: 1, Direct Chunk: 1)
|-- Chapter 1. Understanding the Digital Forensics Profession and
Investigations [ID: inf] (Total Chuck in branch: 1673, Direct Chunk: 0)
|-- Lab 1.1. Installing Autopsy for Windows [ID: 560] (Total Chuck in
branch: 1670, Direct Chunk: 1)
|-- Objectives [ID: 561] (Total Chuck in branch: 702, Direct Chunk: 345)
|-- Materials Required [ID: 562] (Total Chuck in branch: 357, Direct
Chunk: 357)
|-- Activity [ID: 563] (Total Chuck in branch: 967, Direct Chunk: 967)
|-- Lab 1.2. Downloading FTK Imager Lite [ID: 565] (Total Chuck in branch:
1, Direct Chunk: 1)
|-- Lab 1.3. Downloading WinHex [ID: 570] (Total Chuck in branch: 1, Direct
Chunk: 1)
|-- Lab 1.4. Using Autopsy for Windows [ID: 575] (Total Chuck in branch: 1,
Direct Chunk: 1)
|-- Chapter 2. The Investigator's Office and Laboratory [ID: inf] (Total Chuck
in branch: 5, Direct Chunk: 0)
|-- Lab 2.1. Wiping a USB Drive Securely [ID: 582] (Total Chuck in branch:
1, Direct Chunk: 1)
|-- Lab 2.2. Using Directory Snoop to Image a USB Drive [ID: 587] (Total
Chuck in branch: 1, Direct Chunk: 1)
|-- Lab 2.3. Converting a Raw Image to an .E01 Image [ID: 592] (Total Chuck
in branch: 1, Direct Chunk: 1)
|-- Lab 2.4. Imaging Evidence with FTK Imager Lite [ID: 597] (Total Chuck in
branch: 1, Direct Chunk: 1)
|-- Lab 2.5. Viewing Images in FTK Imager Lite [ID: 602] (Total Chuck in
branch: 1, Direct Chunk: 1)
|-- Chapter 3. Data Acquisition [ID: inf] (Total Chuck in branch: 70, Direct
Chunk: 0)
|-- Lab 3.1. Creating a DEFT Zero Forensic Boot CD and USB Drive [ID: 609]
(Total Chuck in branch: 67, Direct Chunk: 1)
|-- Activity [ID: inf] (Total Chuck in branch: 66, Direct Chunk: 0)
|-- Creating a DEFT Zero Boot CD [ID: 613] (Total Chuck in branch: 14,
Direct Chunk: 14)
|-- Creating a Bootable USB DEFT Zero Drive [ID: 614] (Total Chuck in

```

```

branch: 14, Direct Chunk: 14)
    |-- Learning DEFT Zero Features [ID: 615] (Total Chuck in branch: 38,
Direct Chunk: 38)
    |-- Lab 3.2. Examining a FAT Image [ID: 617] (Total Chuck in branch: 1,
Direct Chunk: 1)
    |-- Lab 3.3. Examining an NTFS Image [ID: 622] (Total Chuck in branch: 1,
Direct Chunk: 1)
    |-- Lab 3.4. Examining an HFS+ Image [ID: 627] (Total Chuck in branch: 1,
Direct Chunk: 1)
    |-- Chapter 4. Processing Crime and Incident Scenes [ID: inf] (Total Chuck in
branch: 36, Direct Chunk: 0)
    |-- Lab 4.1. Creating a Mini-WinFE Boot CD [ID: 634] (Total Chuck in branch:
33, Direct Chunk: 1)
    |-- Activity [ID: inf] (Total Chuck in branch: 32, Direct Chunk: 0)
    |-- Setting Up Mini-WinFE [ID: 638] (Total Chuck in branch: 8, Direct
Chunk: 8)
    |-- Creating a Mini-WinFE ISO Image [ID: 639] (Total Chuck in branch:
24, Direct Chunk: 24)
    |-- Lab 4.2. Using Mini-WinFE to Boot and Image a Windows Computer [ID: 641]
(Total Chuck in branch: 1, Direct Chunk: 1)
    |-- Lab 4.3. Testing the Mini-WinFE Write-Protection Feature [ID: 646]
(Total Chuck in branch: 1, Direct Chunk: 1)
    |-- Lab 4.4. Creating an Image with Guymager [ID: 651] (Total Chuck in
branch: 1, Direct Chunk: 1)
    |-- Chapter 5. Working with Windows and CLI Systems [ID: inf] (Total Chuck in
branch: 4, Direct Chunk: 0)
    |-- Lab 5.1. Using DART to Export Windows Registry Files [ID: 658] (Total
Chuck in branch: 1, Direct Chunk: 1)
    |-- Lab 5.2. Examining the SAM Hive [ID: 663] (Total Chuck in branch: 1,
Direct Chunk: 1)
    |-- Lab 5.3. Examining the SYSTEM Hive [ID: 668] (Total Chuck in branch: 1,
Direct Chunk: 1)
    |-- Lab 5.4. Examining the ntuser.dat Registry File [ID: 673] (Total Chuck
in branch: 1, Direct Chunk: 1)
    |-- Chapter 6. Current Digital Forensics Tools [ID: inf] (Total Chuck in
branch: 79, Direct Chunk: 0)
    |-- Lab 6.1. Using Autopsy 4.7.0 to Search an Image File [ID: 680] (Total
Chuck in branch: 41, Direct Chunk: 1)
    |-- Activity [ID: inf] (Total Chuck in branch: 40, Direct Chunk: 0)
    |-- Installing Autopsy 4.7.0 [ID: 684] (Total Chuck in branch: 12,
Direct Chunk: 12)
    |-- Searching E-mail in Autopsy 4.7.0 [ID: 685] (Total Chuck in branch:
28, Direct Chunk: 28)
    |-- Lab 6.2. Using OSForensics to Search an Image of a Hard Drive [ID: 687]
(Total Chuck in branch: 1, Direct Chunk: 1)
    |-- Lab 6.3. Examining a Corrupt Image File with FTK Imager Lite, Autopsy,
and WinHex [ID: 692] (Total Chuck in branch: 37, Direct Chunk: 1)
    |-- Activity [ID: inf] (Total Chuck in branch: 36, Direct Chunk: 0)

```



```

|-- Testing an Image File in Autopsy 4.3.0 [ID: 696] (Total Chuck in
branch: 16, Direct Chunk: 16)
|-- Examining Image Files in WinHex [ID: 697] (Total Chuck in branch:
20, Direct Chunk: 20)
|-- Chapter 7. Linux and Macintosh File Systems [ID: inf] (Total Chuck in
branch: 3, Direct Chunk: 0)
|-- Lab 7.1. Using Autopsy to Process a Mac OS X Image [ID: 701] (Total
Chuck in branch: 1, Direct Chunk: 1)
|-- Lab 7.2. Using Autopsy to Process a Mac OS 9 Image [ID: 706] (Total
Chuck in branch: 1, Direct Chunk: 1)
|-- Lab 7.3. Using Autopsy to Process a Linux Image [ID: 711] (Total Chuck
in branch: 1, Direct Chunk: 1)
|-- Chapter 8. Recovering Graphics Files [ID: inf] (Total Chuck in branch: 3,
Direct Chunk: 0)
|-- Lab 8.1. Using Autopsy to Analyze Multimedia Files [ID: 718] (Total
Chuck in branch: 1, Direct Chunk: 1)
|-- Lab 8.2. Using OSForensics to Analyze Multimedia Files [ID: 723] (Total
Chuck in branch: 1, Direct Chunk: 1)
|-- Lab 8.3. Using WinHex to Analyze Multimedia Files [ID: 728] (Total Chuck
in branch: 1, Direct Chunk: 1)
|-- Chapter 9. Digital Forensics Analysis and Validation [ID: inf] (Total
Chuck in branch: 9, Direct Chunk: 0)
|-- Lab 9.1. Using Autopsy to Search for Keywords in an Image [ID: 735]
(Total Chuck in branch: 1, Direct Chunk: 1)
|-- Lab 9.2. Validating File Hash Values with FTK Imager Lite [ID: 740]
(Total Chuck in branch: 1, Direct Chunk: 1)
|-- Lab 9.3. Validating File Hash Values with WinHex [ID: 745] (Total Chuck
in branch: 7, Direct Chunk: 1)
|-- Objectives [ID: inf] (Total Chuck in branch: 6, Direct Chunk: 0)
|-- Materials Required: [ID: 747] (Total Chuck in branch: 6, Direct
Chunk: 6)
|-- Chapter 10. Virtual Machine Forensics, Live Acquisitions, and Network
Forensics [ID: inf] (Total Chuck in branch: 143, Direct Chunk: 0)
|-- Lab 10.1. Analyzing a Forensic Image Hosting a Virtual Machine [ID: 752]
(Total Chuck in branch: 41, Direct Chunk: 1)
|-- Activity [ID: inf] (Total Chuck in branch: 40, Direct Chunk: 0)
|-- Installing MD5 Hashes in Autopsy [ID: 756] (Total Chuck in branch:
8, Direct Chunk: 8)
|-- Analyzing a Windows Image Containing a Virtual Machine [ID: 757]
(Total Chuck in branch: 32, Direct Chunk: 32)
|-- Lab 10.2. Conducting a Live Acquisition [ID: 759] (Total Chuck in
branch: 45, Direct Chunk: 1)
|-- Activity [ID: inf] (Total Chuck in branch: 44, Direct Chunk: 0)
|-- Installing Tools for Live Acquisitions [ID: 763] (Total Chuck in
branch: 16, Direct Chunk: 16)
|-- Exploring Tools for Live Acquisitions [ID: 764] (Total Chuck in
branch: 14, Direct Chunk: 14)
|-- Capturing Data in a Live Acquisition [ID: 765] (Total Chuck in

```

```

branch: 14, Direct Chunk: 14)
    |-- Lab 10.3. Using Kali Linux for Network Forensics [ID: 767] (Total Chuck
in branch: 57, Direct Chunk: 1)
        |-- Activity [ID: inf] (Total Chuck in branch: 56, Direct Chunk: 0)
            |-- Installing Kali Linux [ID: 771] (Total Chuck in branch: 26, Direct
Chunk: 26)
                |-- Mounting Drives in Kali Linux [ID: 772] (Total Chuck in branch: 12,
Direct Chunk: 12)
                    |-- Identifying Open Ports and Making a Screen Capture [ID: 773] (Total
Chuck in branch: 18, Direct Chunk: 18)
                        |-- Chapter 11. E-mail and Social Media Investigations [ID: inf] (Total Chuck
in branch: 3, Direct Chunk: 0)
                            |-- Lab 11.1. Using OSForensics to Search for E-mails and Mailboxes [ID:
777] (Total Chuck in branch: 1, Direct Chunk: 1)
                                |-- Lab 11.2. Using Autopsy to Search for E-mails and Mailboxes [ID: 782]
(Total Chuck in branch: 1, Direct Chunk: 1)
                                    |-- Lab 11.3. Finding Google Searches and Multiple E-mail Accounts [ID: 787]
(Total Chuck in branch: 1, Direct Chunk: 1)
  |-- Chapter 13. Cloud Forensics [ID: inf] (Total Chuck in branch: 3, Direct
Chunk: 0)
  |-- Lab 13.1. Examining Dropbox Cloud Storage [ID: 811] (Total Chuck in
branch: 1, Direct Chunk: 1)
  |-- Lab 13.2. Examining Google Drive Cloud Storage [ID: 816] (Total Chuck in
branch: 1, Direct Chunk: 1)
  |-- Lab 13.3. Examining OneDrive Cloud Storage [ID: 821] (Total Chuck in
branch: 1, Direct Chunk: 1)
  |-- Chapter 14. Report Writing for High-Tech Investigations [ID: inf] (Total
Chuck in branch: 3, Direct Chunk: 0)
  |-- Lab 14.1. Investigating Corporate Espionage [ID: 828] (Total Chuck in
branch: 1, Direct Chunk: 1)
  |-- Lab 14.2. Adding Evidence to a Case [ID: 833] (Total Chuck in branch: 1,
Direct Chunk: 1)
  |-- Lab 14.3. Preparing a Report [ID: 838] (Total Chuck in branch: 1, Direct
Chunk: 1)
  |-- Chapter 15. Expert Testimony in Digital Investigations [ID: inf] (Total
Chuck in branch: 44, Direct Chunk: 0)
  |-- Lab 15.1. Conducting a Preliminary Investigation [ID: 845] (Total Chuck
in branch: 1, Direct Chunk: 1)
  |-- Lab 15.2. Investigating an Arsonist [ID: 850] (Total Chuck in branch: 1,
Direct Chunk: 1)
  |-- Lab 15.3. Recovering a Password from Password-Protected Files [ID: 855]
(Total Chuck in branch: 42, Direct Chunk: 1)
  |-- Activity [ID: inf] (Total Chuck in branch: 41, Direct Chunk: 0)
  |-- Verifying the Existence of a Warning Banner [ID: 859] (Total Chuck
in branch: 12, Direct Chunk: 12)
  |-- Recovering a Password from Password-Protected Files [ID: 860] (Total
Chuck in branch: 29, Direct Chunk: 29)
  |-- Chapter 16. Ethics for the Expert Witness [ID: inf] (Total Chuck in

```

```

branch: 73, Direct Chunk: 0)
  |-- Lab 16.1. Rebuilding an MFT Record from a Corrupt Image [ID: 864] (Total
Chunk in branch: 73, Direct Chunk: 1)
    |-- Activity [ID: inf] (Total Chunk in branch: 72, Direct Chunk: 0)
      |-- Creating a Duplicate Forensic Image [ID: 868] (Total Chunk in
branch: 14, Direct Chunk: 14)
        |-- Determining the Offset Byte Address of the Corrupt MFT Record [ID:
869] (Total Chunk in branch: 12, Direct Chunk: 12)
          |-- Copying the Corrected MFT Record [ID: 870] (Total Chunk in branch:
20, Direct Chunk: 20)
            |-- Extracting Additional Evidence [ID: 871] (Total Chunk in branch: 26,
Direct Chunk: 26)
              |-- Appendix A. Certification Test References [ID: 873] (Total Chunk in branch:
56, Direct Chunk: 56)
                |-- Appendix B. Digital Forensics References [ID: 874] (Total Chunk in branch:
109, Direct Chunk: 109)
                  |-- Appendix C. Digital Forensics Lab Considerations [ID: 875] (Total Chunk in
branch: 58, Direct Chunk: 58)
                    |-- Appendix D. Legacy File System and Forensics Tools [ID: 876] (Total Chunk in
branch: 59, Direct Chunk: 59)
                      |-- EPUB Preamble [ID: inf] (Total Chunk in branch: 21, Direct Chunk: 21)

```

---

### Chunk Sequence & Content Integrity Test

---

#### ----- CONTENT PREVIEW -----

Title: Understanding Disk Drives [toc\_id: 187]  
 Chunk IDs: [2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472]

#### Understanding Disk Drives

You should be familiar with disk drives and how data is organized on a disk so that you can find data effectively. Disk drives are made up of one or more platters coated with magnetic material, and data is stored on platters in a particular way. The following disk drive components are illustrated in Figure 5-2:

Geometry- Geometry A disk drive's internal organization of platters, tracks, and sectors. refers to a disk's logical structure of platters, tracks, and sectors.

Head-The head The device that reads and writes data to a disk drive. is the device that reads and writes data to a drive. There are two heads per platter that read and write the top and bottom sides.

Tracks- Tracks Concentric circles on a disk platter where data is stored. are concentric circles on a disk platter where data is located.

Cylinders-A cylinder A column of tracks on two or more disk platters. is a column of tracks on two or more disk platters. Typically, each platter has two surfaces: top and bottom.

Sectors-A sector A section on a track, typically made up of 512 bytes. is a

section on a track, usually made up of 512 bytes.

#### Tip

For more information on disk drive configurations, see

[www.storagereview.com/guide2000/ref/hdd/index.html](http://www.storagereview.com/guide2000/ref/hdd/index.html).

The manufacturer engineers a disk to have a certain number of sectors per track, and a typical disk drive stores 512 bytes per sector. (For an update on Advanced Format [4K] disk compatibility, see

[https://msdn.microsoft.com/windows/compatibility/advanced-format-disk-](https://msdn.microsoft.com/windows/compatibility/advanced-format-disk-compatibility-update)

[compatibility-update](https://msdn.microsoft.com/windows/compatibility/advanced-format-disk-compatibility-update).) To determine the total number of addressable bytes on a

disk, multiply the number of cylinders by the number of heads (actually tracks)

and by the number of sectors (groups of 512 or more bytes), as shown in Figure

5-3. Disk drive vendors refer to this formula as a "cylinder, head, and sector

(CHS) calculation." Tracks also follow a numbering scheme starting from 0, which

is the first value in computing. If a disk lists 79 tracks, you actually have 80 tracks from 0 to 79.

Other disk properties, such as zone bit recording (ZBR) The method most

manufacturers use to deal with a platter's inner tracks being shorter than the

outer tracks. Grouping tracks by zones ensures that all tracks hold the same

amount of data. , track density The space between tracks on a disk. The smaller

the space between tracks, the more tracks on a disk. Older drives with wider

track densities allowed the heads to wander. , areal density The number of bits

per square inch of a disk platter. , and head and cylinder skew A method

manufacturers use to minimize lag time. The starting sectors of tracks are

slightly offset from each other to move the read-write head. , are handled at

the drive's hardware or firmware level. ZBR is how most manufacturers deal with

a platter's inner tracks

hardware or firmware level. ZBR is how most manufacturers deal with a platter's

inner tracks having a smaller circumference (and, therefore, less space to store

data) than its outer tracks. Grouping tracks by zones ensures that all tracks

hold the same amount of data.

Track density is the space between each track. As with old vinyl records, the

smaller the space between each track, the more tracks you can place on the

platter.

Areal density is the number of bits in one square inch of a disk platter. This

number includes the unused space between tracks. Head and cylinder skew are used

to improve disk performance. As the read-write head moves from one track to

another, starting sectors are offset to minimize lag time.

----- END CONTENT PREVIEW -----

---

#### Diagnostic Summary

---

Total Chunks in DB: 11774

=====

Diagnostic Complete

=====

```
[23]: # Cell 6: Verify Content Retrieval for a Specific toc_id with Reassembled Text

import os
import json
import logging
from langchain_chroma import Chroma
from langchain_ollama.embeddings import OllamaEmbeddings

# --- Logger Setup ---
logger = logging.getLogger(__name__)
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %
↳ %(message)s')

def retrieve_and_print_chunks_for_toc_id(vector_store: Chroma, toc_id: int):
    """
    Retrieves all chunks for a specific toc_id, reconstructs the section title
    from hierarchical metadata, shows the reassembled text, and lists individual
    chunk details for verification.
    """
    try:
        # Use the 'get' method with a 'where' filter to find all chunks for the
↳ toc_id

        results = vector_store.get(
            where={"toc_id": toc_id},
            include=["documents", "metadatas"]
        )

        if not results or not results.get('ids'):
            logger.warning(f"No chunks found in the database for toc_id =
↳ {toc_id}")
            print("=" * 80)
            print(f"VERIFICATION FAILED: No content found for toc_id: {toc_id}")
            print("=" * 80)
            return

        documents = results['documents']
        metadatas = results['metadatas']

        # --- FIX START: Reconstruct the hierarchical section title from
↳ metadata ---
        # We assume all chunks for the same toc_id share the same titles.
        # We will inspect the metadata of the first chunk to get the title.
        section_title = "Unknown or Uncategorized Section"
        if metadatas:
            first_meta = metadatas[0]

            # Find all 'level_X_title' keys in the metadata
```

```

        level_titles = []
        for key, value in first_meta.items():
            if key.startswith("level_") and key.endswith("_title"):
                try:
                    # Extract the level number (e.g., 1 from
↪ 'level_1_title') for sorting
                    level_num = int(key.split('_')[1])
                    level_titles.append((level_num, value))
                except (ValueError, IndexError):
                    # Ignore malformed keys, just in case
                    continue

        # Sort the titles by their level number (1, 2, 3...)
        level_titles.sort()

        # Join the sorted titles to create a breadcrumb-style title
        if level_titles:
            title_parts = [title for num, title in level_titles]
            section_title = " > ".join(title_parts)
        # --- FIX END ---

        # --- Print a clear header with the reconstructed section title ---
        print("=" * 80)
        print(f"VERIFYING SECTION: '{section_title}' (toc_id: {toc_id})")
        print("=" * 80)
        logger.info(f"Found {len(documents)} chunks in the database for this
↪ section.")

        # Sort chunks by their chunk_id to ensure they are in the correct order
↪ for reassembly
        sorted_items = sorted(zip(documents, metadatas), key=lambda item:
↪ item[1].get('chunk_id', 0))

        # --- Reassemble and print the full text for the section ---
        all_chunk_texts = [item[0] for item in sorted_items]
        reassembled_text = "\n".join(all_chunk_texts)

        print("\n" + "#" * 28 + " Reassembled Text " + "#" * 28)
        print(reassembled_text)
        print("#" * 80)

        # --- Print individual chunk details for in-depth verification ---
        print("\n" + "-" * 24 + " Retrieved Chunk Details " + "-" * 25)
        for i, (doc, meta) in enumerate(sorted_items):
            print(f"\n[ Chunk {i+1} of {len(documents)} | chunk_id: {meta.
↪ get('chunk_id', 'N/A')} ]")
            content_preview = doc.replace('\n', ' ').strip()

```

```

        print(f" Content Preview: '{content_preview[:250]}...'")
        print(f" Metadata: {json.dumps(meta, indent=2)}")

    print("\n" + "=" * 80)
    print(f"Verification complete for section '{section_title}'.")
    print("=" * 80)

    except Exception as e:
        logger.error(f"An error occurred during retrieval for toc_id {toc_id}:␣
↪{e}", exc_info=True)

# =====
# EXECUTION BLOCK (No changes needed here)
# =====

# --- IMPORTANT: Set the ID of the section you want to test here ---
# Example: ToC ID 10 might be "An Overview of Digital Forensics"
# Example: ToC ID 11 might be "Digital Forensics and Other Related Disciplines"
TOC_ID_TO_TEST = 9# Change this to an ID you know exists from your ToC

# Assume these variables are defined in a previous cell from your notebook
# CHROMA_PERSIST_DIR = "./chroma_db_with_metadata"
# EMBEDDING_MODEL_OLLAMA = "nomic-embed-text"
# CHROMA_COLLECTION_NAME = "forensics_handbook"

# Check if the database directory exists before attempting to connect
if 'CHROMA_PERSIST_DIR' in locals() and os.path.exists(CHROMA_PERSIST_DIR):
    logger.info(f"Connecting to the existing vector database at␣
↪'{CHROMA_PERSIST_DIR}'...")

    try:
        vector_store = Chroma(
            persist_directory=CHROMA_PERSIST_DIR,
            embedding_function=OllamaEmbeddings(model=EMBEDDING_MODEL_OLLAMA),
            collection_name=CHROMA_COLLECTION_NAME
        )

        # Run the verification function
        retrieve_and_print_chunks_for_toc_id(vector_store, TOC_ID_TO_TEST)

    except Exception as e:
        logger.error(f"Failed to initialize Chroma or run retrieval. Error:␣
↪{e}")

        logger.error("Please ensure your embedding model and collection names␣
↪are correct.")

```

```
else:
    logger.error("Database directory not found or 'CHROMA_PERSIST_DIR' variable_
↳is not set.")
    logger.error("Please run the previous cell (Cell 5) to create the database_
↳first.")
```

```
2025-07-06 02:12:18,350 - INFO - Connecting to the existing vector database at '
/home/sebas_dev_linux/projects/course_generator/data/DataBase_Chroma/chroma_db_t
oc_guided_chunks_epub'...
```

```
2025-07-06 02:12:18,364 - INFO - Found 18 chunks in the database for this
section.
```

```
=====
VERIFYING SECTION: 'Chapter 1. Understanding the Digital Forensics Profession
and Investigations > An Overview of Digital Forensics' (toc_id: 9)
=====
```

```
##### Reassembled Text #####
```

An Overview of Digital Forensics

As the world has become more of a level playing field, with more people online who have access to the same information (Thomas L. Freidman, The World Is Flat, Farrar, Straus, and Giroux, 2005), the need to standardize digital forensics processes has become more urgent. The definition of digital forensics Applying investigative procedures for a legal purpose; involves the analysis of digital evidence as well as obtaining search warrants, maintaining a chain of custody, validating with mathematical hash functions, using validated tools, ensuring repeatability, reporting, and presenting evidence as an expert witness. has also evolved over the years from simply involving securing and analyzing digital information stored on a computer for use as evidence in civil, criminal, or administrative

digital information stored on a computer for use as evidence in civil, criminal, or administrative cases. The former director of the Defense Computer Forensics Laboratory, Ken Zatyko, wrote a treatise on the many specialties including computer forensics, network forensics, video forensics, and a host of others. He defined it as "[t]he application of computer science and investigative procedures for a legal purpose involving the analysis of digital evidence (information of probative value that is stored or transmitted in binary form) after proper search authority, chain of custody, validation with mathematics (hash function), use of validated tools, repeatability, reporting and possible expert presentation" ("Commentary: Defining Digital Forensics," Forensic Magazine, 2007).

The field of digital forensics can also encompass items such as research and incident response. With incident response, most organizations are concerned with protecting their assets and containing the situation, not necessarily prosecuting or finding the person responsible. Research in digital forensics also isn't concerned with prosecution or validity of evidence. This book is intended for digital forensics investigators and examiners at the civil, criminal, and administrative levels. Other facets of digital forensics are



beyond the scope of this book. Keep in mind that depending on the jurisdiction and situation, forensic investigators and examiners might be the same or different personnel. In this book, the terms are used interchangeably.

Note

For a more in-depth discussion of what the term "digital forensics" means, see "Digital Forensic Evidence Examination" (Fred Cohen, [www.fredcohen.net/Books/2013-DFE-Examination.pdf](http://www.fredcohen.net/Books/2013-DFE-Examination.pdf), 2012).

Many groups have tried to create digital forensics certifications that could be recognized worldwide but have failed in this attempt. However, they have created certifications for specific categories of practitioners, such as government investigators. With the ubiquitous access to mobile devices now, digital evidence is everywhere, so the need for a global standardized method is even more critical so that companies and governments can share and use digital evidence. In October 2012, an International Organization for Standardization (ISO) standard for digital forensics was ratified. This standard, ISO 27037 "Information technology - Security techniques - Guidelines for identification, collection, acquisition and preservation of digital evidence" (see [www.iso.org/standard/44381.html](http://www.iso.org/standard/44381.html)), acquisition and preservation of digital evidence" (see [www.iso.org/standard/44381.html](http://www.iso.org/standard/44381.html)), defines the personnel and methods for acquiring and preserving digital evidence. To address the multinational cases that continue to emerge, agencies in every country should develop policies and procedures that meet this standard.

The Federal Rules of Evidence (FRE), signed into law in 1973, was created to ensure consistency in federal proceedings, but many states' rules map to the FRE, too. In another attempt to standardize procedures, the FBI Computer Analysis and Response Team (CART) was formed in 1984 to handle the increase in cases involving digital evidence. By the late 1990s, CART had teamed up with the Department of Defense Computer Forensics Laboratory (DCFL) for research and training. Much of the early curriculum in this field came from the DCFL. For more information on the FBI's cybercrime investigation services, see [www.fbi.gov/investigate/cyber](http://www.fbi.gov/investigate/cyber).

Files maintained on a computer are covered by different rules, depending on the nature of the documents. Many court cases in state and federal courts have developed and clarified how the rules apply to digital evidence. The Fourth Amendment The Fourth Amendment to the U.S. Constitution in the Bill of Rights dictates that the government and its agents must have probable cause for search and seizure. to the U.S. Constitution (and each state's constitution) protects everyone's right to be secure in their person, residence, and property from search and seizure. Continuing development of the jurisprudence of this amendment has played a role in determining whether the search for digital evidence has established a different precedent, so separate search warrants Legal documents that allow law

has established a different precedent, so separate search warrants Legal documents that allow law enforcement to search an office, a home, or other locale for evidence related to an alleged crime. might not be necessary. However, when preparing to search for evidence in a criminal case, many investigators still include the suspect's computer and its components in the

search warrant to avoid later admissibility problems.

In an important case involving these issues, the Pennsylvania Supreme Court addressed expectations of privacy and whether evidence is admissible (see *Commonwealth v. Copenhefer*, 587 A.2d 1353, 526 Pa. 555 [1991]). Initial investigations by the FBI, state police, and local police resulted in discovering computer-generated notes and instructions-some of which had been deleted-that had been concealed in hiding places around Corry, Pennsylvania. The investigation also produced several possible suspects, including David Copenhefer, who owned a nearby bookstore and apparently had bad relationships with the victim and her husband. Examination of trash discarded from Copenhefer's store revealed drafts of the ransom note and directions. Subsequent search warrants resulted in seizure of evidence of the ransom note and directions. Subsequent search warrants resulted in seizure of evidence against him. Copenhefer's computer contained several drafts and amendments of the text of phone calls to the victim and the victim's husband the next day, the ransom note, the series of hidden notes, and a plan for the entire kidnapping scheme (Copenhefer, p. 559).

On direct appeal, the Pennsylvania Supreme Court concluded that the physical evidence, including the digital forensics evidence, was sufficient to support the bookstore owner's conviction. Copenhefer's argument was that "[E]ven though his computer was validly seized pursuant to a warrant, his attempted deletion of the documents in question created an expectation of privacy protected by the Fourth Amendment. Thus, he claims, under *Katz v. United States*, 389 U.S. 347, 357, 88 S.Ct. 507, 19 L.Ed.2d 576 (1967), and its progeny, Agent Johnson's retrieval of the documents, without first obtaining another search warrant, was unreasonable under the Fourth Amendment and the documents thus seized should have been suppressed" (Copenhefer, p. 561).

The Pennsylvania Supreme Court rejected this argument, stating, "A defendant's attempt to secrete evidence of a crime is not synonymous with a legally cognizable expectation of privacy. A mere hope for secrecy is not a legally protected expectation. If it were, search warrants would be required in a vast number of cases where warrants are clearly not necessary" (Copenhefer, p. 562). Every U.S. jurisdiction has case law related to the admissibility of evidence recovered from computers and other digital devices. As you learn in this book, however, the laws on digital evidence vary between states as well as between provinces and countries.

Note

The U.S. Department of Justice offers a useful guide to search and seizure procedures for computers and computer evidence at <https://pdfs.semanticscholar.org/aabb/8b0caf982a0aa211f932252af38d4c9376fb.pdf>.  
#####

----- Retrieved Chunk Details -----

[ Chunk 1 of 18 | chunk\_id: 156 ]  
Content Preview: 'An Overview of Digital Forensics...'  
Metadata: {  
 "chunk\_id": 156,

```

    "level_2_title": "An Overview of Digital Forensics",
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession
and Investigations",
    "toc_id": 9,
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub"
}

```

[ Chunk 2 of 18 | chunk\_id: 157 ]

Content Preview: 'As the world has become more of a level playing field, with more people online who have access to the same information (Thomas L. Freidman, The World Is Flat, Farrar, Straus, and Giroux, 2005), the need to standardize digital forensics processes has ...'

```

Metadata: {
  "chunk_id": 157,
  "toc_id": 9,
  "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession
and Investigations",
  "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub",
  "level_2_title": "An Overview of Digital Forensics"
}

```

[ Chunk 3 of 18 | chunk\_id: 158 ]

Content Preview: 'digital information stored on a computer for use as evidence in civil, criminal, or administrative cases. The former director of the Defense Computer Forensics Laboratory, Ken Zatyko, wrote a treatise on the many specialties including computer forens...'

```

Metadata: {
  "chunk_id": 158,
  "toc_id": 9,
  "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession
and Investigations",
  "level_2_title": "An Overview of Digital Forensics",
  "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub"
}

```

[ Chunk 4 of 18 | chunk\_id: 159 ]

Content Preview: 'The field of digital forensics can also encompass items such as research and incident response. With incident response, most organizations are concerned with protecting their assets and containing the situation, not necessarily prosecuting or finding...'

```

Metadata: {
  "level_2_title": "An Overview of Digital Forensics",

```

```

    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession
and Investigations",
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub",
    "chunk_id": 159,
    "toc_id": 9
}

```

```

[ Chunk 5 of 18 | chunk_id: 160 ]
Content Preview: 'Note...'
Metadata: {
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub",
    "level_2_title": "An Overview of Digital Forensics",
    "chunk_id": 160,
    "toc_id": 9,
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession
and Investigations"
}

```

```

[ Chunk 6 of 18 | chunk_id: 161 ]
Content Preview: 'For a more in-depth discussion of what the term "digital
forensics" means, see "Digital Forensic Evidence Examination" (Fred Cohen,
www.fredcohen.net/Books/2013-DFE-Examination.pdf, 2012)...'
Metadata: {
    "toc_id": 9,
    "chunk_id": 161,
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub",
    "level_2_title": "An Overview of Digital Forensics",
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession
and Investigations"
}

```

```

[ Chunk 7 of 18 | chunk_id: 162 ]
Content Preview: 'Many groups have tried to create digital forensics
certifications that could be recognized worldwide but have failed in this
attempt. However, they have created certifications for specific categories of
practitioners, such as government investigators...'
Metadata: {
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub",
    "toc_id": 9,
    "level_2_title": "An Overview of Digital Forensics",

```

```
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession  
and Investigations",  
    "chunk_id": 162  
}
```

```
[ Chunk 8 of 18 | chunk_id: 163 ]
```

```
Content Preview: 'acquisition and preservation of digital evidence" (see  
www.iso.org/standard/44381.html), defines the personnel and methods for  
acquiring and preserving digital evidence. To address the multinational cases  
that continue to emerge, agencies in every co...'
```

```
Metadata: {  
    "level_2_title": "An Overview of Digital Forensics",  
    "chunk_id": 163,  
    "toc_id": 9,  
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession  
and Investigations",  
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to  
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage  
Learning (2018).epub"  
}
```

```
[ Chunk 9 of 18 | chunk_id: 164 ]
```

```
Content Preview: 'The Federal Rules of Evidence (FRE), signed into law in  
1973, was created to ensure consistency in federal proceedings, but many states'  
rules map to the FRE, too. In another attempt to standardize procedures, the FBI  
Computer Analysis and Response T...'
```

```
Metadata: {  
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession  
and Investigations",  
    "level_2_title": "An Overview of Digital Forensics",  
    "chunk_id": 164,  
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to  
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage  
Learning (2018).epub",  
    "toc_id": 9  
}
```

```
[ Chunk 10 of 18 | chunk_id: 165 ]
```

```
Content Preview: 'Files maintained on a computer are covered by different  
rules, depending on the nature of the documents. Many court cases in state and  
federal courts have developed and clarified how the rules apply to digital  
evidence. The Fourth Amendment The Fourt...'
```

```
Metadata: {  
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to  
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage  
Learning (2018).epub",  
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession  
and Investigations",
```

```
"level_2_title": "An Overview of Digital Forensics",  
"chunk_id": 165,  
"toc_id": 9  
}
```

```
[ Chunk 11 of 18 | chunk_id: 166 ]
```

Content Preview: 'has established a different precedent, so separate search warrants Legal documents that allow law enforcement to search an office, a home, or other locale for evidence related to an alleged crime. might not be necessary. However, when preparing to se...'

```
Metadata: {  
  "toc_id": 9,  
  "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession  
and Investigations",  
  "chunk_id": 166,  
  "level_2_title": "An Overview of Digital Forensics",  
  "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to  
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage  
Learning (2018).epub"  
}
```

```
[ Chunk 12 of 18 | chunk_id: 167 ]
```

Content Preview: 'In an important case involving these issues, the Pennsylvania Supreme Court addressed expectations of privacy and whether evidence is admissible (see Commonwealth v. Copenhefer, 587 A.2d 1353, 526 Pa. 555 [1991]). Initial investigations by the FBI, s...'

```
Metadata: {  
  "toc_id": 9,  
  "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to  
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage  
Learning (2018).epub",  
  "chunk_id": 167,  
  "level_2_title": "An Overview of Digital Forensics",  
  "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession  
and Investigations"  
}
```

```
[ Chunk 13 of 18 | chunk_id: 168 ]
```

Content Preview: 'of the ransom note and directions. Subsequent search warrants resulted in seizure of evidence against him. Copenhefer's computer contained several drafts and amendments of the text of phone calls to the victim and the victim's husband the next day, t...'

```
Metadata: {  
  "toc_id": 9,  
  "chunk_id": 168,  
  "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to  
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage  
Learning (2018).epub",
```

```
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession  
and Investigations",  
    "level_2_title": "An Overview of Digital Forensics"  
}
```

```
[ Chunk 14 of 18 | chunk_id: 169 ]
```

```
Content Preview: 'On direct appeal, the Pennsylvania Supreme Court concluded  
that the physical evidence, including the digital forensics evidence, was  
sufficient to support the bookstore owner's conviction. Copenhefer's argument  
was that "[E]ven though his computer wa...'
```

```
Metadata: {  
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession  
and Investigations",  
    "level_2_title": "An Overview of Digital Forensics",  
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to  
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage  
Learning (2018).epub",  
    "chunk_id": 169,  
    "toc_id": 9  
}
```

```
[ Chunk 15 of 18 | chunk_id: 170 ]
```

```
Content Preview: 'The Pennsylvania Supreme Court rejected this argument,  
stating, "A defendant's attempt to secrete evidence of a crime is not synonymous  
with a legally cognizable expectation of privacy. A mere hope for secrecy is not  
a legally protected expectation. ...'
```

```
Metadata: {  
    "level_2_title": "An Overview of Digital Forensics",  
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to  
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage  
Learning (2018).epub",  
    "toc_id": 9,  
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession  
and Investigations",  
    "chunk_id": 170  
}
```

```
[ Chunk 16 of 18 | chunk_id: 171 ]
```

```
Content Preview: 'Every U.S. jurisdiction has case law related to the  
admissibility of evidence recovered from computers and other digital devices. As  
you learn in this book, however, the laws on digital evidence vary between  
states as well as between provinces and co...'
```

```
Metadata: {  
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession  
and Investigations",  
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to  
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage  
Learning (2018).epub",
```

```

    "chunk_id": 171,
    "level_2_title": "An Overview of Digital Forensics",
    "toc_id": 9
}

[ Chunk 17 of 18 | chunk_id: 172 ]
Content Preview: 'Note...'
Metadata: {
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub",
    "level_2_title": "An Overview of Digital Forensics",
    "chunk_id": 172,
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession
and Investigations",
    "toc_id": 9
}

[ Chunk 18 of 18 | chunk_id: 173 ]
Content Preview: 'The U.S. Department of Justice offers a useful guide to
search and seizure procedures for computers and computer evidence at https://pdf
s.semanticscholar.org/aabb/8b0caf982a0aa211f932252af38d4c9376fb.pdf...'
Metadata: {
    "level_2_title": "An Overview of Digital Forensics",
    "toc_id": 9,
    "chunk_id": 173,
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub",
    "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession
and Investigations"
}

```

```

=====
Verification complete for section 'Chapter 1. Understanding the Digital
Forensics Profession and Investigations > An Overview of Digital Forensics'.
=====

```

## 5.2 Test Data Base for content development

Require Description

```

[ ]: # Cell 7: Verify Vector Database (Final Version with Rich Diagnostic Output)

import os
import json
import re
import random

```



```

import logging
from typing import List, Dict, Any, Tuple, Optional

# Third-party imports
try:
    from langchain_chroma import Chroma
    from langchain_ollama.embeddings import OllamaEmbeddings
    from langchain_core.documents import Document
    langchain_available = True
except ImportError:
    langchain_available = False

# Setup Logger for this cell
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
logger = logging.getLogger(__name__)

# --- HELPER FUNCTIONS ---

def print_results(query_text: str, results: list, where_filter: Optional[Dict] = None):
    """
    Richly prints query results, showing the query, filter, and retrieved documents.
    """
    print("\n" + "-"*10 + " DIAGNOSTIC: RETRIEVAL RESULTS " + "-"*10)
    print(f"QUERY: '{query_text}'")
    if where_filter:
        print(f"FILTER: {json.dumps(where_filter, indent=2)}")

    if not results:
        print("--> No documents were retrieved for this query and filter.")
        print("-" * 55)
        return

    print(f"--> Found {len(results)} results. Displaying top {min(len(results), 3)}:")
    for i, doc in enumerate(results[:3]):
        print(f"\n[ RESULT {i+1} ]")
        content_preview = doc.page_content.replace('\n', ' ').strip()
        print(f"  Content : '{content_preview[:200]}...')
        print(f"  Metadata: {json.dumps(doc.metadata, indent=2)}")
    print("-" * 55)

# --- HELPER FUNCTIONS FOR FINDING DATA (UNCHANGED) ---

```

```

def find_deep_entry(nodes: List[Dict], current_path: List[str] = []) -> Optional[Tuple[Dict, List[str]]]:
    shuffled_nodes = random.sample(nodes, len(nodes))
    for node in shuffled_nodes:
        if node.get('level', 0) >= 2 and node.get('children'): return node, current_path + [node['title']]
        if node.get('children'):
            path = current_path + [node['title']]
            deep_entry = find_deep_entry(node['children'], path)
            if deep_entry: return deep_entry
    return None

def find_chapter_title_by_number(toc_data: List[Dict], chap_num: int) -> Optional[List[str]]:
    def search_nodes(nodes, num, current_path):
        for node in nodes:
            path = current_path + [node['title']]
            if re.match(rf"(Chapter\s)?{num}[\.\s]", node.get('title', ''), re.IGNORECASE): return path
            if node.get('children'):
                found_path = search_nodes(node['children'], num, path)
                if found_path: return found_path
        return None
    return search_nodes(toc_data, chap_num, [])

# --- ENHANCED TEST CASES with DIAGNOSTIC OUTPUT ---

def basic_retrieval_test(db, outline):
    print_header("Test 1: Basic Retrieval", char="-")
    try:
        logger.info("Goal: Confirm the database is live and contains thematically relevant content.")
        logger.info("Strategy: Perform a simple similarity search using the course's 'unitName'.")
        query_text = outline.get("unitInformation", {}).get("unitName", "introduction")

        logger.info(f"Action: Searching for query: '{query_text}'...")
        results = db.similarity_search(query_text, k=1)

        print_results(query_text, results) # <--- SHOW THE EVIDENCE

        logger.info("Verification: Check if at least one document was returned.")
        assert len(results) > 0, "Basic retrieval query returned no results."

```

```

        logger.info(" Result: TEST 1 PASSED. The database is online and_
↳responsive.")
        return True
    except Exception as e:
        logger.error(f" Result: TEST 1 FAILED. Reason: {e}")
        return False

def deep_hierarchy_test(db, toc):
    print_header("Test 2: Deep Hierarchy Retrieval", char="-")
    try:
        logger.info("Goal: Verify that the multi-level hierarchical metadata_
↳was ingested correctly.")
        logger.info("Strategy: Find a random, deeply nested sub-section and use_
↳a precise filter to retrieve it.")
        deep_entry_result = find_deep_entry(toc)
        assert deep_entry_result, "Could not find a suitable deep entry (level_
↳>= 2) to test."
        node, path = deep_entry_result
        query = node['title']

        logger.info(f" - Selected random deep section: {' -> '.join(path)}")
        conditions = [{f"level_{i+1}_title": {"$eq": title}} for i, title in_
↳enumerate(path)]
        w_filter = {"$and": conditions}

        logger.info("Action: Performing a similarity search with a highly_
↳specific '$and' filter.")
        results = db.similarity_search(query, k=1, filter=w_filter)

        print_results(query, results, w_filter) # <--- SHOW THE EVIDENCE

        logger.info("Verification: Check if the precisely filtered query_
↳returned any documents.")
        assert len(results) > 0, "Deeply filtered query returned no results."

        logger.info(" Result: TEST 2 PASSED. Hierarchical metadata is_
↳structured correctly.")
        return True
    except Exception as e:
        logger.error(f" Result: TEST 2 FAILED. Reason: {e}")
        return False

def advanced_alignment_test(db, outline, toc):
    print_header("Test 3: Advanced Unit Outline Alignment", char="-")
    try:

```

```

        logger.info("Goal: Ensure a weekly topic from the syllabus can be
↳ mapped to the correct textbook chapter(s).")
        logger.info("Strategy: Pick a random week, find its chapter, and query
↳ for the topic filtered by that chapter.")
        week_to_test = random.choice(outline['weeklySchedule'])
        logger.info(f" - Selected random week: Week {week_to_test['week']} -
↳ '{week_to_test['contentTopic']}'")

        reading = week_to_test.get('requiredReading', '')
        chap_nums_str = re.findall(r'\d+', reading)
        assert chap_nums_str, f"Could not find chapter numbers in required
↳ reading: '{reading}'"
        logger.info(f" - Extracted required chapter number(s):
↳ {chap_nums_str}")

        chapter_paths = [find_chapter_title_by_number(toc, int(n)) for n in
↳ chap_nums_str]
        chapter_paths = [path for path in chapter_paths if path is not None]
        assert chapter_paths, f"Could not map chapter numbers {chap_nums_str}
↳ to a valid ToC path."

        level_1_titles = list(set([path[0] for path in chapter_paths]))
        logger.info(f" - Mapped to top-level ToC entries: {level_1_titles}")

        or_filter = [{"level_1_title": {"$eq": title}} for title in
↳ level_1_titles]
        w_filter = {"$or": or_filter} if len(or_filter) > 1 else or_filter[0]
        query = week_to_test['contentTopic']

        logger.info("Action: Searching for the weekly topic, filtered by the
↳ mapped chapter(s).")
        results = db.similarity_search(query, k=5, filter=w_filter)

        print_results(query, results, w_filter) # <--- SHOW THE EVIDENCE

        logger.info("Verification: Check if at least one returned document is
↳ from the correct chapter.")
        assert len(results) > 0, "Alignment query returned no results for the
↳ correct section/chapter."

        logger.info(" Result: TEST 3 PASSED. The syllabus can be reliably
↳ aligned with the textbook content.")
        return True
    except Exception as e:
        logger.error(f" Result: TEST 3 FAILED. Reason: {e}")
        return False

```

```

def content_sequence_test(db, outline):
    print_header("Test 4: Content Sequence Verification", char="-")
    try:
        logger.info("Goal: Confirm that chunks for a topic can be re-ordered to
        form a coherent narrative.")
        logger.info("Strategy: Retrieve several chunks for a random topic and
        verify their 'chunk_id' is sequential.")
        topic_query = random.choice(outline['weeklySchedule'])['contentTopic']

        logger.info(f"Action: Performing similarity search for topic:
        {topic_query} to get a set of chunks.")
        results = db.similarity_search(topic_query, k=10)

        print_results(topic_query, results) # <--- SHOW THE EVIDENCE

        docs_with_id = [doc for doc in results if 'chunk_id' in doc.metadata]
        assert len(docs_with_id) > 3, "Fewer than 4 retrieved chunks have a
        'chunk_id' to test."

        chunk_ids = [doc.metadata['chunk_id'] for doc in docs_with_id]
        sorted_ids = sorted(chunk_ids)

        logger.info(f" - Retrieved and sorted chunk IDs: {sorted_ids}")
        logger.info("Verification: Check if the sorted list of chunk_ids is
        strictly increasing.")
        is_ordered = all(sorted_ids[i] >= sorted_ids[i-1] for i in range(1,
        len(sorted_ids)))
        assert is_ordered, "The retrieved chunks' chunk_ids are not in
        ascending order when sorted."

        logger.info(" Result: TEST 4 PASSED. Narrative order can be
        reconstructed using 'chunk_id'.")
        return True
    except Exception as e:
        logger.error(f" Result: TEST 4 FAILED. Reason: {e}")
        return False

# --- MAIN VERIFICATION EXECUTION ---
def run_verification():
    print_header("Database Verification Process")

    if not langchain_available:
        logger.error("LangChain libraries not found. Aborting tests.")
        return

```

```

required_files = {
    "Chroma DB": CHROMA_PERSIST_DIR,
    "ToC JSON": PRE_EXTRACTED_TOC_JSON_PATH,
    "Parsed Outline": PARSED_UO_JSON_PATH
}
for name, path in required_files.items():
    if not os.path.exists(path):
        logger.error(f"Required '{name}' not found at '{path}'. Please run_
previous cells.")
    return

with open(PRE_EXTRACTED_TOC_JSON_PATH, 'r', encoding='utf-8') as f:
    toc_data = json.load(f)
with open(PARSED_UO_JSON_PATH, 'r', encoding='utf-8') as f:
    unit_outline_data = json.load(f)

logger.info("Connecting to DB and initializing components...")
embeddings = OllamaEmbeddings(model=EMBEDDING_MODEL_OLLAMA)
vector_store = Chroma(
    persist_directory=CHROMA_PERSIST_DIR,
    embedding_function=embeddings,
    collection_name=CHROMA_COLLECTION_NAME
)

results_summary = [
    basic_retrieval_test(vector_store, unit_outline_data),
    deep_hierarchy_test(vector_store, toc_data),
    advanced_alignment_test(vector_store, unit_outline_data, toc_data),
    content_sequence_test(vector_store, unit_outline_data)
]

passed_count = sum(filter(None, results_summary))
failed_count = len(results_summary) - passed_count

print_header("Verification Summary")
print(f"Total Tests Run: {len(results_summary)}")
print(f" Passed: {passed_count}")
print(f" Failed: {failed_count}")
print_header("Verification Complete", char="=")

# --- Execute Verification ---
# Assumes global variables from Cell 1 are available in the notebook's scope
run_verification()

```

2025-07-06 01:53:00,740 - INFO - Connecting to DB and initializing components...

2025-07-06 01:53:00,751 - INFO - Goal: Confirm the database is live and contains thematically relevant content.

2025-07-06 01:53:00,751 - INFO - Strategy: Perform a simple similarity search

using the course's 'unitName'.

2025-07-06 01:53:00,752 - INFO - Action: Searching for query: 'Digital Forensic'...

---

### Database Verification Process

---

---

#### Test 1: Basic Retrieval

---

2025-07-06 01:53:01,971 - INFO - HTTP Request: POST

http://127.0.0.1:11434/api/embed "HTTP/1.1 200 OK"

2025-07-06 01:53:01,977 - INFO - Verification: Check if at least one document was returned.

2025-07-06 01:53:01,978 - INFO - Result: TEST 1 PASSED. The database is online and responsive.

2025-07-06 01:53:01,978 - INFO - Goal: Verify that the multi-level hierarchical metadata was ingested correctly.

2025-07-06 01:53:01,979 - INFO - Strategy: Find a random, deeply nested subsection and use a precise filter to retrieve it.

2025-07-06 01:53:01,979 - INFO - - Selected random deep section: Chapter 2. The Investigator's Office and Laboratory -> Building a Business Case for Developing a Forensics Lab -> Preparing a Business Case for a Digital Forensics Lab

2025-07-06 01:53:01,980 - INFO - Action: Performing a similarity search with a highly specific '\$and' filter.

2025-07-06 01:53:02,011 - INFO - HTTP Request: POST

http://127.0.0.1:11434/api/embed "HTTP/1.1 200 OK"

2025-07-06 01:53:02,035 - INFO - Verification: Check if the precisely filtered query returned any documents.

2025-07-06 01:53:02,036 - INFO - Result: TEST 2 PASSED. Hierarchical metadata is structured correctly.

2025-07-06 01:53:02,037 - INFO - Goal: Ensure a weekly topic from the syllabus can be mapped to the correct textbook chapter(s).

2025-07-06 01:53:02,037 - INFO - Strategy: Pick a random week, find its chapter, and query for the topic filtered by that chapter.

2025-07-06 01:53:02,037 - INFO - - Selected random week: Week Week 2 - 'The Investigator's Office and Laboratory.'

2025-07-06 01:53:02,038 - INFO - - Extracted required chapter number(s): ['2019', '978', '1', '337', '56894', '4', '2']

2025-07-06 01:53:02,040 - INFO - - Mapped to top-level ToC entries: ['Chapter 2. The Investigator's Office and Laboratory', 'Chapter 4. Processing Crime and Incident Scenes', 'Chapter 1. Understanding the Digital Forensics Profession and Investigations']

2025-07-06 01:53:02,041 - INFO - Action: Searching for the weekly topic, filtered by the mapped chapter(s).

```

2025-07-06 01:53:02,062 - INFO - HTTP Request: POST
http://127.0.0.1:11434/api/embed "HTTP/1.1 200 OK"
2025-07-06 01:53:02,083 - INFO - Verification: Check if at least one returned
document is from the correct chapter.
2025-07-06 01:53:02,083 - INFO - Result: TEST 3 PASSED. The syllabus can be
reliably aligned with the textbook content.
2025-07-06 01:53:02,084 - INFO - Goal: Confirm that chunks for a topic can be
re-ordered to form a coherent narrative.
2025-07-06 01:53:02,084 - INFO - Strategy: Retrieve several chunks for a random
topic and verify their 'chunk_id' is sequential.
2025-07-06 01:53:02,084 - INFO - Action: Performing similarity search for topic:
'Email and Social Media.' to get a set of chunks.
2025-07-06 01:53:02,100 - INFO - HTTP Request: POST
http://127.0.0.1:11434/api/embed "HTTP/1.1 200 OK"
2025-07-06 01:53:02,104 - INFO - - Retrieved and sorted chunk IDs: [54, 5378,
5387, 5635, 5636, 5682, 5683, 5847, 5859, 10484]
2025-07-06 01:53:02,104 - INFO - Verification: Check if the sorted list of
chunk_ids is strictly increasing.
2025-07-06 01:53:02,105 - INFO - Result: TEST 4 PASSED. Narrative order can be
reconstructed using 'chunk_id'.

```

----- DIAGNOSTIC: RETRIEVAL RESULTS -----

QUERY: 'Digital Forensic'

--> Found 1 results. Displaying top 1:

[ RESULT 1 ]

```

Content : 'An Overview of Digital Forensics...'
Metadata: {
  "level_1_title": "Chapter 1. Understanding the Digital Forensics Profession
and Investigations",
  "chunk_id": 156,
  "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub",
  "level_2_title": "An Overview of Digital Forensics",
  "toc_id": 9
}

```

-----  
Test 2: Deep Hierarchy Retrieval  
-----

----- DIAGNOSTIC: RETRIEVAL RESULTS -----

QUERY: 'Preparing a Business Case for a Digital Forensics Lab'

```

FILTER: {
  "$and": [

```



```

{
  "level_1_title": {
    "$eq": "Chapter 2. The Investigator\u2019s Office and Laboratory"
  }
},
{
  "level_2_title": {
    "$eq": "Building a Business Case for Developing a Forensics Lab"
  }
},
{
  "level_3_title": {
    "$eq": "Preparing a Business Case for a Digital Forensics Lab"
  }
}
]
}

```

--> Found 1 results. Displaying top 1:

```

[ RESULT 1 ]
Content : 'Preparing a Business Case for a Digital Forensics Lab...'
Metadata: {
  "chunk_id": 1151,
  "toc_id": 83,
  "level_1_title": "Chapter 2. The Investigator\u2019s Office and Laboratory",
  "level_3_title": "Preparing a Business Case for a Digital Forensics Lab",
  "level_2_title": "Building a Business Case for Developing a Forensics Lab",
  "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub"
}

```

---

### Test 3: Advanced Unit Outline Alignment

---

----- DIAGNOSTIC: RETRIEVAL RESULTS -----

QUERY: 'The Investigator's Office and Laboratory.'

FILTER: {

```

  "$or": [
    {
      "level_1_title": {
        "$eq": "Chapter 2. The Investigator\u2019s Office and Laboratory"
      }
    },
    {
      "level_1_title": {

```

```

        "$eq": "Chapter 4. Processing Crime and Incident Scenes"
    }
},
{
    "level_1_title": {
        "$eq": "Chapter 1. Understanding the Digital Forensics Profession and
Investigations"
    }
}
]
}

```

--> Found 5 results. Displaying top 3:

```

[ RESULT 1 ]
Content : 'Chapter 2. The Investigator's Office and Laboratory...'
Metadata: {
    "toc_id": 55,
    "chunk_id": 936,
    "level_1_title": "Chapter 2. The Investigator\u2019s Office and Laboratory",
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub"
}

[ RESULT 2 ]
Content : 'Chapter 2. The Investigator's Office and Laboratory...'
Metadata: {
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub",
    "level_1_title": "Chapter 2. The Investigator\u2019s Office and Laboratory",
    "chunk_id": 8493,
    "toc_id": 55
}

[ RESULT 3 ]
Content : 'As discussed, the evidence room needs to be secure. The lab should
have at least two controlled exits and no windows. Separate offices for
supervisors and cubicles for investigators are more practical...'
Metadata: {
    "level_2_title": "Determining the Physical Requirements for a Digital
Forensics Lab",
    "toc_id": 74,
    "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub",
    "chunk_id": 1088,
    "level_1_title": "Chapter 2. The Investigator\u2019s Office and Laboratory",

```

```
"level_3_title": "Determining Floor Plans for Digital Forensics Labs"
}
```

---

#### Test 4: Content Sequence Verification

---

----- DIAGNOSTIC: RETRIEVAL RESULTS -----

QUERY: 'Email and Social Media.'

--> Found 10 results. Displaying top 3:

[ RESULT 1 ]

```
Content : 'Chapter 11. E-mail and Social Media Investigations...'
Metadata: {
  "toc_id": 377,
  "level_1_title": "Chapter 11. E-mail and Social Media Investigations",
  "chunk_id": 5378,
  "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub"
}
```

[ RESULT 2 ]

```
Content : 'Chapter 11. E-mail and Social Media Investigations...'
Metadata: {
  "chunk_id": 10484,
  "toc_id": 377,
  "level_1_title": "Chapter 11. E-mail and Social Media Investigations",
  "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub"
}
```

[ RESULT 3 ]

```
Content : 'Social media can contain a lot of information, including the
following:...'
Metadata: {
  "chunk_id": 5636,
  "level_2_title": "Applying Digital Forensics Methods to Social Media
Communications",
  "toc_id": 398,
  "level_1_title": "Chapter 11. E-mail and Social Media Investigations",
  "source": "Bill Nelson, Amelia Phillips, Christopher Steuart - Guide to
Computer Forensics and Investigations_ Processing Digital Evidence-Cengage
Learning (2018).epub"
}
```

```

=====
                                Verification Summary
=====
Total Tests Run: 4
    Passed: 4
    Failed: 0

=====
                                Verification Complete
=====

```

## 6 Content Generation

### 6.1 Planning Agent

```

[ ]: # Cell 8: The Data-Driven Planning Agent (Final Hierarchical Version)

import os
import json
import re
import math
import logging
from typing import List, Dict, Any, Optional

# Setup Logger and LangChain components
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - 
↳ %(message)s')
logger = logging.getLogger(__name__)
try:
    from langchain_chroma import Chroma
    from langchain_ollama.embeddings import OllamaEmbeddings
    langchain_available = True
except ImportError:
    langchain_available = False

def print_header(text: str, char: str = "="):
    """Prints a centered header to the console."""
    print("\n" + char * 80)
    print(text.center(80))
    print(char * 80)

class PlanningAgent:
    """
    An agent that creates a hierarchical content plan, adaptively partitions
    ↳ content
    into distinct lecture decks, and allocates presentation time.

```

```

"""
def __init__(self, master_config: Dict, vector_store: Optional[Any] = None):
    self.config = master_config['processed_settings']
    self.unit_outline = master_config['unit_outline']
    self.book_toc = master_config['book_toc']
    self.flat_toc_with_ids = self._create_flat_toc_with_ids()
    self.vector_store = vector_store
    logger.info("Data-Driven PlanningAgent initialized successfully.")

def _create_flat_toc_with_ids(self) -> List[Dict]:
    """Creates a flattened list of the ToC for easy metadata lookup."""
    flat_list = []
    def flatten_recursive(nodes, counter):
        for node in nodes:
            node_id = counter[0]; counter[0] += 1
            flat_list.append({'toc_id': node_id, 'title': node.get('title', ''),
↪ 'node': node})
            if node.get('children'):
                flatten_recursive(node.get('children'), counter)
    flatten_recursive(self.book_toc, [0])
    return flat_list

def _identify_relevant_chapters(self, weekly_schedule_item: Dict) -> ↪
↪List[int]:
    """Extracts chapter numbers precisely from the 'requiredReading' string.
    ↪"""
    reading_str = weekly_schedule_item.get('requiredReading', '')
    match = re.search(r'Chapter(s)?', reading_str, re.IGNORECASE)
    if not match: return []
    search_area = reading_str[match.start():]
    chap_nums_str = re.findall(r'\d+', search_area)
    if chap_nums_str:
        return sorted(list(set(int(n) for n in chap_nums_str)))
    return []

def _find_chapter_node(self, chapter_number: int) -> Optional[Dict]:
    """Finds the ToC node for a specific chapter number."""
    for item in self.flat_toc_with_ids:
        if re.match(rf"Chapter\s{chapter_number}(?:\D|$)", item['title']):
            return item['node']
    return None

def _build_topic_plan_tree(self, toc_node: Dict) -> Dict:
    """
    Recursively builds a hierarchical plan tree from any ToC node,
    annotating it with direct and total branch chunk counts.
    """

```

```

        node_metadata = next((item for item in self.flat_toc_with_ids if
↪item['node'] is toc_node), None)
        if not node_metadata: return {}

        retrieved_docs = self.vector_store.get(where={'toc_id':
↪node_metadata['toc_id']})
        direct_chunk_count = len(retrieved_docs.get('ids', []))

        plan_node = {
            "title": node_metadata['title'],
            "toc_id": node_metadata['toc_id'],
            "chunk_count": direct_chunk_count,
            "total_chunks_in_branch": 0,
            "slides_allocated": 0,
            "children": []
        }

        child_branch_total = 0
        for child_node in toc_node.get('children', []):
            if any(ex in child_node.get('title', '').lower() for ex in
↪["review", "introduction", "summary", "key terms"]):
                continue
            child_plan_node = self._build_topic_plan_tree(child_node)
            if child_plan_node:
                plan_node['children'].append(child_plan_node)
                child_branch_total += child_plan_node.
↪get('total_chunks_in_branch', 0)

        plan_node['total_chunks_in_branch'] = direct_chunk_count +
↪child_branch_total
        return plan_node

# In PlanningAgent Class...

def _allocate_slides_to_tree(self, plan_tree: Dict, content_slides_budget:
↪int):
    """
    (REFACTORED) Performs a multi-pass process to allocate content slides,
    add interactive activities, and sum totals correctly.
    """
    if not plan_tree or content_slides_budget <= 0:
        return plan_tree

    # --- Pass 1: Allocate Content Slides (Top-Down, Proportional) ---
    def allocate_content_recursively(node, budget):
        node['slides_allocated'] = 0

```

```

# If it's a leaf node, it gets the remaining budget.
if not node.get('children'):
    node['slides_allocated'] = round(budget)
    return

# If it has children, distribute the budget proportionally.
total_branch_chunks = node.get('total_chunks_in_branch', 0)

# Allocate slides for the node's own content (if any).
# This is a key fix: parent nodes can have their own content.
own_content_slides = 0
if total_branch_chunks > 0:
    own_content_slides = round(budget * (node.get('chunk_count', 0) /
↪ total_branch_chunks))
    node['slides_allocated'] = own_content_slides

remaining_budget_for_children = budget - own_content_slides

# Distribute remaining budget to children.
for child in node.get('children', []):
    child_budget = 0
    if total_branch_chunks > 0:
        # Distribute based on the child's total branch size, not
↪ just its own chunks.
        child_budget = remaining_budget_for_children * (child.
↪ get('total_chunks_in_branch', 0) / (total_branch_chunks - node.
↪ get('chunk_count', 0)))
        allocate_content_recursively(child, child_budget)

allocate_content_recursively(plan_tree, content_slides_budget)

# --- Pass 2: Add Interactive Activities (Targeted Depth) ---
def add_interactive_nodes(node, depth, interactive_deep):
    if not node: return

    # Logic for interactive_deep: true
    if interactive_deep:
        if depth == 2:
            node['interactive_activity'] = {"title": f"{node.
↪ get('title')} (Deep-Dive Activity)", "toc_id": node.get('toc_id'),
↪ "slides_allocated": 1}
        if depth == 1:
            node['interactive_activity'] = {"title": f"{node.
↪ get('title')} (General Activity)", "toc_id": node.get('toc_id'),
↪ "slides_allocated": 1}
    # Logic for interactive_deep: false

```

```

        else:
            if depth == 1:
                node['interactive_activity'] = {"title": f"{node.
↪get('title')}" (Interactive Activity)", "toc_id": node.get('toc_id'),
↪"slides_allocated": 1}

                # Recurse
                for child in node.get('children', []):
                    add_interactive_nodes(child, depth + 1, interactive_deep)

            if self.config.get('interactive', False):
                interactive_deep = self.config.get('interactive_deep', False)
                logger.info(f"Interactive mode ON. Deep interaction:
↪{interactive_deep}. Adding placeholders...")
                # Start depth at 1 for the root nodes of the plan.
                add_interactive_nodes(plan_tree, 1, interactive_deep)

            # --- Pass 3: Sum All Slides (Content + Interactive) Up the Tree ---
            def sum_slides_upwards(node):
                # Start with the node's own allocated content slides.
                total_slides = node.get('slides_allocated', 0)

                # Add slides from its interactive activity, if it exists.
                total_slides += node.get('interactive_activity', {}).
↪get('slides_allocated', 0)

                # Add the summed totals from all its children.
                if node.get('children'):
                    total_slides += sum(sum_slides_upwards(child) for child in node.
↪get('children', []))

                # The final 'slides_allocated' is the grand total for the branch.
                node['slides_allocated'] = total_slides
                return total_slides

            sum_slides_upwards(plan_tree)

        return plan_tree

    def create_content_plan_for_week(self, week_number: int) -> Optional[Dict]:
        """Orchestrates the adaptive planning and partitioning process."""
        print_header(f"Planning Week {week_number}", char="*")

        weekly_schedule_item = self.unit_outline['weeklySchedule'][week_number
↪- 1]

        chapter_numbers = self._identify_relevant_chapters(weekly_schedule_item)
        if not chapter_numbers: return None

```



```

num_decks = self.config['week_session_setup'].get('sessions_per_week', 1)

# 1. Build a full plan tree for each chapter to get its weight.
chapter_plan_trees = [self._build_topic_plan_tree(self.
    ↪_find_chapter_node(cn)) for cn in chapter_numbers if self.
    ↪_find_chapter_node(cn)]
total_weekly_chunks = sum(tree.get('total_chunks_in_branch', 0) for
    ↪tree in chapter_plan_trees)

# 2. NEW: Adaptive Partitioning Strategy
partitionable_units = []
all_top_level_sections = []
for chapter_tree in chapter_plan_trees:
    all_top_level_sections.extend(chapter_tree.get('children', []))

num_top_level_sections = len(all_top_level_sections)

# Always prefer to split by top-level sections if there are enough to
    ↪distribute.
if num_top_level_sections >= num_decks:
    logger.info(f"Partitioning strategy: Distributing
    ↪{num_top_level_sections} top-level sections across {num_decks} decks.")
    partitionable_units = all_top_level_sections
else:
    # Fallback for rare cases where there are fewer topics than decks
    ↪(e.g., 1 chapter with 1 section, but 2 decks).
    logger.info(f"Partitioning strategy: Not enough top-level sections
    ↪({num_top_level_sections}) to fill all decks ({num_decks}). Distributing
    ↪whole chapters instead.")
    partitionable_units = chapter_plan_trees

# 3. Partition the chosen units into decks using a bin-packing algorithm
decks = [[] for _ in range(num_decks)]
deck_weights = [0] * num_decks
sorted_units = sorted(partitionable_units, key=lambda x: x.
    ↪get('total_chunks_in_branch', 0), reverse=True)

for unit in sorted_units:
    lightest_deck_index = deck_weights.index(min(deck_weights))
    decks[lightest_deck_index].append(unit)
    deck_weights[lightest_deck_index] += unit.
    ↪get('total_chunks_in_branch', 0)

# 4. Plan each deck

```

```

        content_slides_per_week = self.config['slide_count_strategy'].
↳get('target', 25)
        final_deck_plans = []
        for i, deck_content_trees in enumerate(decks):
            deck_number = i + 1
            deck_chunk_weight = sum(tree.get('total_chunks_in_branch', 0) for
↳tree in deck_content_trees)
            deck_slide_budget = round((deck_chunk_weight / total_weekly_chunks)
↳* content_slides_per_week) if total_weekly_chunks > 0 else 0

            logger.info(f"--- Planning Deck {deck_number}/{num_decks} | Topics:
↳{[t['title'] for t in deck_content_trees]} | Weight: {deck_chunk_weight}
↳chunks | Slide Budget: {deck_slide_budget} ---")

            # The allocation function is recursive and works on any tree or
↳sub-tree
            planned_content = [self._allocate_slides_to_tree(tree,
↳round(deck_slide_budget * (tree.get('total_chunks_in_branch', 0) /
↳deck_chunk_weight))) if deck_chunk_weight > 0 else tree for tree in
↳deck_content_trees]

            final_deck_plans.append({
                "deck_number": deck_number,
                "deck_title": f"{self.config.get('unit_name', 'Course')} - Week
↳{week_number}, Lecture {deck_number}",
                "session_content": planned_content
            })

        return {
            "week": week_number,
            "overall_topic": weekly_schedule_item.get('contentTopic'),
            "deck_plans": final_deck_plans
        }

```

## 6.2 Content Generator Class (no yet addressed focus planning)

## 6.3 Orquestrator (Addressing paint points )

### Description:

The main script that iterates through the weeks defined the plan and generate the content base on the settings\_deck coordinating the agents.

**Parameters and concideration** - 1 hour in the setting session\_time\_duration\_in\_hour - is 18-20 slides at the time so it is require to calculate this according to the given value but this also means per session so sessions\_per\_week is a multiplicator factor that

- if apply\_topic\_interactive is available will add an extra slide and add extra 5 min time but to determine this is required to plan all the content first and then calculate then provide a extra time

settings\_deck.json

```
{  "course_id":  "", "unit_name":  "", "interactive":  true, "interactive_deep":  false,
  "slide_count_strategy": { "method": "per_week", "interactive_slides_per_week": 0 -> sum
all interactive counts "interactive_slides_per_session": 0, -> Total # of slides produced if "in-
teractive" is true other wise remains 0 "target_total_slides": 0, -> Total Content Slides per week
that cover the total - will be the target in the cell 7
  "slides_content_per_session": 0, -> Total # (target_total_slides/sessions_per_week) "to-
tal_slides_deck_week": 0, -> target_total_slides + interactive_slides_per_week + (framework
(4 + Time for Title, Agenda, Summary, End) * sessions_per_week) "Total_slides_session": 0
-> content_slides_per_session + interactive_slides_per_session + framework (4 + Time for
Title, Agenda, Summary, End) }, "week_session_setup": { "sessions_per_week": 1, "distribu-
tion_strategy": "even", "interactive_time_in_hour": 0, -> find the value in ahours of the total
# ("interactive_slides" * "TIME_PER_INTERACTIVE_SLIDE_MINS")/60
  "total_session_time_in_hours": 0 -> this is going to be equal or similar to ses-
sion_time_duration_in_hour if "interactive" is false obvisuly base on the global varaibles it will
be the calculation of "interactive_time_in_hour" "session_time_duration_in_hour": 2, -> this
is the time that the costumer need for delivery this is a constrain is not modified never is used for
reference },
  "parameters_slides": { "slides_per_hour": 18, # no framework in-
clude "time_per_content_slides_min": 3, # average delivery per slide
  "time_per_interactive_slide_min": 5, #small break and engaging with the students
  "time_for_framework_slides_min": 6 # Time for Title, Agenda, Summary, End (per deck) "
}, "generation_scope": { "weeks": [6] }, "teaching_flow_id": "Interactive Lecture Flow" }
```

teaching\_flows.json

```
{ "standard_lecture": { "name": "Standard Lecture Flow", "slide_types": ["Title", "Agenda",
"Content", "Summary", "End"], "prompts": { "content_generation": "You are an expert univer-
sity lecturer. Your audience is undergraduate students. Based on the following context, create
a slide that provides a detailed explanation of the topic '{sub_topic}'. The content should be
structured with bullet points for key details. Your output MUST be a single JSON object with
a 'title' (string) and 'content' (list of strings) key.", "summary_generation": "You are an expert
university lecturer creating a summary slide. Based on the following list of topics covered in this
session, generate a concise summary of the key takeaways. The topics are: {topic_list}. Your
output MUST be a single JSON object with a 'title' (string) and 'content' (list of strings) key."
}, "slide_schemas": { "Content": { "title": "string", "content": "list[string]" }, "Summary": { "ti-
tle": "string", "content": "list[string]" } }, "apply_topic_interactive": { "name": "Interactive
Lecture Flow", "slide_types": ["Title", "Agenda", "Content", "Application", "Summary", "End"],
"prompts": { "content_generation": "You are an expert university lecturer in Digital Forensics.
Your audience is undergraduate students. Based on the provided context, create a slide explaining
the concept of '{sub_topic}'. The content should be clear, concise, and structured with bullet
points for easy understanding. Your output MUST be a single JSON object with a 'title' (string)
and 'content' (list of strings) key.", "application_generation": "You are an engaging university lec-
turer creating an interactive slide. Based on the concept of '{sub_topic}', create a multiple-choice
question with exactly 4 options (A, B, C, D) to test understanding. The slide title must be 'Let's
Apply This:'. Clearly indicate the correct answer within the content. Your output MUST be a sin-
gle JSON object with a 'title' (string) and 'content' (list of strings) key.", "summary_generation":
"You are an expert university lecturer creating a summary slide. Based on the following list of con-
```

cepts and applications covered in this session, generate a concise summary of the key takeaways. The topics are: {topic\_list}. Your output MUST be a single JSON object with a 'title' (string) and 'content' (list of strings) key." }, "slide\_schemas": { "Content": { "title": "string", "content": "list[string]" }, "Application": { "title": "string", "content": "list[string]" }, "Summary": { "title": "string", "content": "list[string]" } } }

```
[ ]: # Cell 9: Configuration and Scoping for Content Generation (Corrected)

import os
import json
import logging

# Setup Logger for this cell
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - \u2192%(message)s')
logger = logging.getLogger(__name__)

# --- 1. DEFINE FILE PATHS AND GLOBAL TEST SETTINGS ---
# Assumes these variables are loaded from a previous setup cell (like Cell 1)
# PROJECT_BASE_DIR, PARSED_UO_JSON_PATH, PRE_EXTRACTED_TOC_JSON_PATH must be \u2192defined.

# New configuration file paths
CONFIG_DIR = os.path.join(PROJECT_BASE_DIR, "configs")
SETTINGS_DECK_PATH = os.path.join(CONFIG_DIR, "settings_deck.json")
TEACHING_FLOWS_PATH = os.path.join(CONFIG_DIR, "teaching_flows.json")

# New output path for the processed settings
PROCESSED_SETTINGS_PATH = os.path.join(CONFIG_DIR, "processed_settings.json")

# --- Global Test Overrides (for easy testing) ---
TEST_OVERRIDE_WEEKS = None
TEST_OVERRIDE_FLOW_ID = None
TEST_OVERRIDE_SESSIONS_PER_WEEK = None
TEST_OVERRIDE_DISTRIBUTION_STRATEGY = None

def print_header(text: str, char: str = "="):
    """Prints a centered header to the console."""
    print("\n" + char * 80)
    print(text.center(80))
    print(char * 80)

def process_and_load_configurations():
    """
    PHASE 1: Loads configurations, calculates a PRELIMINARY time-based slide \u2192budget,
    and saves the result as 'processed_settings.json' for the Planning Agent.
    """
```

```

"""
print_header("Phase 1: Configuration and Scoping Process", char="-")

# --- Load all input files ---
logger.info("Loading all necessary configuration and data files...")
try:
    os.makedirs(CONFIG_DIR, exist_ok=True)
    with open(PARSED_UO_JSON_PATH, 'r', encoding='utf-8') as f:
        unit_outline = json.load(f)
    with open(PRE_EXTRACTED_TOC_JSON_PATH, 'r', encoding='utf-8') as f:
        book_toc = json.load(f)
    with open(SETTINGS_DECK_PATH, 'r', encoding='utf-8') as f:
        settings_deck = json.load(f)
    with open(TEACHING_FLOWS_PATH, 'r', encoding='utf-8') as f:
        teaching_flows = json.load(f)
    logger.info("All files loaded successfully.")
except FileNotFoundError as e:
    logger.error(f"FATAL: A required configuration file was not found: {e}")
    return None

# --- Pre-process and Refine Settings ---
logger.info("Pre-processing settings_deck for definitive plan...")
processed_settings = json.loads(json.dumps(settings_deck))

unit_info = unit_outline.get("unitInformation", {})
processed_settings['course_id'] = unit_info.get("unitCode",
        "UNKNOWN_COURSE")
processed_settings['unit_name'] = unit_info.get("unitName", "Unknown Unit_
        Name")

# --- Apply test overrides IF they are not None ---
logger.info("Applying overrides if specified...")
# This block now correctly sets the teaching_flow_id based on the
        interactive flag.
if TEST_OVERRIDE_FLOW_ID is not None:
    processed_settings['teaching_flow_id'] = TEST_OVERRIDE_FLOW_ID
    logger.info(f"OVERRIDE: teaching_flow_id set to_
        '{TEST_OVERRIDE_FLOW_ID}'")
else:
    # If no override, use the 'interactive' boolean from the file as the
        source of truth.
    is_interactive = processed_settings.get('interactive', False)
    if is_interactive:
        processed_settings['teaching_flow_id'] = 'apply_topic_interactive'
    else:
        processed_settings['teaching_flow_id'] = 'standard_lecture'

```

```

        logger.info(f"Loaded from settings: 'interactive' is {is_interactive}._
↳Set teaching_flow_id to '{processed_settings['teaching_flow_id']}'".)

        # The 'interactive' flag is now always consistent with the teaching_flow_id.
        processed_settings['interactive'] = "interactive" in_
↳processed_settings['teaching_flow_id'].lower()

        if TEST_OVERRIDE_SESSIONS_PER_WEEK is not None:
            processed_settings['week_session_setup']['sessions_per_week'] =_
↳TEST_OVERRIDE_SESSIONS_PER_WEEK
            logger.info(f"OVERRIDE: sessions_per_week set to_
↳{TEST_OVERRIDE_SESSIONS_PER_WEEK}")

        if TEST_OVERRIDE_DISTRIBUTION_STRATEGY is not None:
            processed_settings['week_session_setup']['distribution_strategy'] =_
↳TEST_OVERRIDE_DISTRIBUTION_STRATEGY
            logger.info(f"OVERRIDE: distribution_strategy set to_
↳'{TEST_OVERRIDE_DISTRIBUTION_STRATEGY}'")

        if TEST_OVERRIDE_WEEKS is not None:
            processed_settings['generation_scope']['weeks'] = TEST_OVERRIDE_WEEKS
            logger.info(f"OVERRIDE: generation_scope weeks set to_
↳{TEST_OVERRIDE_WEEKS}")

        # --- DYNAMIC SLIDE BUDGET CALCULATION (Phase 1) ---
        logger.info("Calculating preliminary slide budget based on session time...")

        params = processed_settings.get('parameters_slides', {})
        SLIDES_PER_HOUR = params.get('slides_per_hour', 18)

        duration_hours = processed_settings['week_session_setup'].
↳get('session_time_duration_in_hour', 1.0)
        sessions_per_week = processed_settings['week_session_setup'].
↳get('sessions_per_week', 1)

        slides_content_per_session = int(duration_hours * SLIDES_PER_HOUR)
        target_total_slides = slides_content_per_session * sessions_per_week

        processed_settings['slide_count_strategy']['target_total_slides'] =_
↳target_total_slides
        processed_settings['slide_count_strategy']['slides_content_per_session'] =_
↳slides_content_per_session
        logger.info(f"Preliminary weekly content slide target calculated:_
↳{target_total_slides} slides.")

        # --- Resolve Generation Scope if not overridden ---

```

```

    if TEST_OVERRIDE_WEEKS is None and processed_settings.
↳get('generation_scope', {}).get('weeks') == "all":
        num_weeks = len(unit_outline.get('weeklySchedule', []))
        processed_settings['generation_scope']['weeks'] = list(range(1,
↳num_weeks + 1))

    # --- Save the processed settings to disk ---
    logger.info(f"Saving preliminary processed configuration to:
↳{PROCESSED_SETTINGS_PATH}")
    with open(PROCESSED_SETTINGS_PATH, 'w', encoding='utf-8') as f:
        json.dump(processed_settings, f, indent=2)
    logger.info("File saved successfully.")

    # --- Assemble master config for optional preview ---
    master_config = {
        "processed_settings": processed_settings,
        "unit_outline": unit_outline,
        "book_toc": book_toc,
        "teaching_flows": teaching_flows
    }

    print_header("Phase 1 Configuration Complete", char="-")
    logger.info("Master configuration object is ready for the Planning Agent.")
    return master_config

# --- EXECUTE THE CONFIGURATION PROCESS ---
master_config = process_and_load_configurations()

# Optional: Print a preview to verify the output
if master_config:
    print("\n--- Preview of Processed Settings (Phase 1) ---")
    print(json.dumps(master_config['processed_settings'], indent=2,
↳sort_keys=True))
    if master_config.get('processed_settings', {}).get('generation_scope', {}).
↳get('weeks'):
        print(f"\nNumber of weeks to generate:
↳{len(master_config['processed_settings']['generation_scope']['weeks'])}")
    print("-----")

```

2025-07-06 01:56:01,009 - INFO - Loading all necessary configuration and data files...

2025-07-06 01:56:01,017 - INFO - All files loaded successfully.

2025-07-06 01:56:01,018 - INFO - Pre-processing settings\_deck for definitive plan...

2025-07-06 01:56:01,018 - INFO - Applying overrides if specified...

2025-07-06 01:56:01,019 - INFO - Loaded from settings: 'interactive' is True. Set teaching\_flow\_id to 'apply\_topic\_interactive'.

2025-07-06 01:56:01,019 - INFO - Calculating preliminary slide budget based on session time...

2025-07-06 01:56:01,019 - INFO - Preliminary weekly content slide target calculated: 36 slides.

2025-07-06 01:56:01,020 - INFO - Saving preliminary processed configuration to: /home/sebas\_dev\_linux/projects/course\_generator/configs/processed\_settings.json

2025-07-06 01:56:01,021 - INFO - File saved successfully.

2025-07-06 01:56:01,021 - INFO - Master configuration object is ready for the Planning Agent.

---

### Phase 1: Configuration and Scoping Process

---

---

### Phase 1 Configuration Complete

---

--- Preview of Processed Settings (Phase 1) ---

```
{
  "course_id": "ICT312",
  "generation_scope": {
    "weeks": [
      1
    ]
  },
  "interactive": true,
  "interactive_deep": false,
  "parameters_slides": {
    "slides_per_hour": 18,
    "time_for_framework_slides_min": 6,
    "time_per_content_slides_min": 3,
    "time_per_interactive_slide_min": 5
  },
  "slide_count_strategy": {
    "interactive_slides_per_session": 0,
    "interactive_slides_per_week": 0,
    "method": "per_week",
    "slides_content_per_session": 36,
    "target_total_slides": 36,
    "total_slides_deck_week": 0,
    "total_slides_session": 0
  },
  "teaching_flow_id": "apply_topic_interactive",
  "unit_name": "Digital Forensic",
  "week_session_setup": {
    "distribution_strategy": "even",
```



```

        "interactive_time_in_hour": 0,
        "session_time_duration_in_hour": 2,
        "sessions_per_week": 1,
        "total_session_time_in_hours": 0
    }
}

```

Number of weeks to generate: 1

-----

```

[ ]: # In Cell 10,

logger.info("--- Initializing Data-Driven Planning Agent Test ---")

if langchain_available:
    logger.info("Connecting to ChromaDB for the Planning Agent...")
    try:
        # 1. Connect to DB and Load all configurations
        vector_store = Chroma(
            persist_directory=CHROMA_PERSIST_DIR,
            embedding_function=OllamaEmbeddings(model=EMBEDDING_MODEL_OLLAMA),
            collection_name=CHROMA_COLLECTION_NAME
        )
        logger.info("Database connection successful.")

        logger.info("Loading configuration files for Planning Agent...")
        with open(os.path.join(CONFIG_DIR, "processed_settings.json"), 'r') as f:
            processed_settings = json.load(f)
        with open(PRE_EXTRACTED_TOC_JSON_PATH, 'r') as f:
            book_toc = json.load(f)
        with open(PARSED_UO_JSON_PATH, 'r') as f:
            unit_outline = json.load(f)
        logger.info("Configuration files loaded.")

        master_config_from_file = {
            "processed_settings": processed_settings,
            "unit_outline": unit_outline,
            "book_toc": book_toc
        }

        # 2. Initialize the Planning Agent
        planning_agent = PlanningAgent(master_config_from_file,
            vector_store=vector_store)

        # 3. CRITICAL: Loop through the weeks defined in the processed settings

```

```

        weeks_to_generate = processed_settings.get('generation_scope', {}).
↪get('weeks', [])
        logger.info(f"Found {len(weeks_to_generate)} week(s) to plan:␣
↪{weeks_to_generate}")

        for week_to_test in weeks_to_generate:
            logger.info(f"--> Generating draft plan for Week {week_to_test}")
            content_plan = planning_agent.
↪create_content_plan_for_week(week_to_test)

            if content_plan:
                print(f"\n--- Generated Draft Plan for Week {week_to_test} ---")
                print(json.dumps(content_plan, indent=2))

                # Save the generated plan to a file
                PLAN_OUTPUT_DIR = os.path.join(PROJECT_BASE_DIR,␣
↪"generated_plans")
                os.makedirs(PLAN_OUTPUT_DIR, exist_ok=True)
                plan_filename = f"{processed_settings.get('course_id',␣
↪'COURSE')}_Week{week_to_test}_plan_draft.json"
                plan_filepath = os.path.join(PLAN_OUTPUT_DIR, plan_filename)
                with open(plan_filepath, 'w') as f:
                    json.dump(content_plan, f, indent=2)
                logger.info(f"\nSuccessfully saved DRAFT content plan for Week␣
↪{week_to_test} to: {plan_filepath}")
            else:
                logger.error(f"Failed to generate content plan for Week␣
↪{week_to_test}.")

            except Exception as e:
                logger.error(f"An error occurred during the planning process: {e}",␣
↪exc_info=True)

        else:
            logger.error("LangChain/Chroma libraries not found. Cannot run the Planning␣
↪Agent.")

```

```

2025-07-06 01:58:02,754 - INFO - --- Initializing Data-Driven Planning Agent
Test ---
2025-07-06 01:58:02,756 - INFO - Connecting to ChromaDB for the Planning
Agent...
2025-07-06 01:58:02,773 - INFO - Database connection successful.
2025-07-06 01:58:02,775 - INFO - Loading configuration files for Planning
Agent...
2025-07-06 01:58:02,777 - INFO - Configuration files loaded.
2025-07-06 01:58:02,777 - INFO - Data-Driven PlanningAgent initialized
successfully.

```

```

2025-07-06 01:58:02,780 - INFO - Found 1 week(s) to plan: [1]
2025-07-06 01:58:02,780 - INFO - --> Generating draft plan for Week 1
2025-07-06 01:58:02,841 - INFO - Partitioning strategy: Distributing 7 top-level
sections across 1 decks.
2025-07-06 01:58:02,841 - INFO - --- Planning Deck 1/1 | Topics: ['Procedures
for Private-Sector High-Tech Investigations', 'Conducting an Investigation',
'Preparing a Digital Forensics Investigation', 'Preparing for Digital
Investigations', 'An Overview of Digital Forensics', 'Understanding Data
Recovery Workstations and Software', 'Maintaining Professional Conduct'] |
Weight: 521 chunks | Slide Budget: 24 ---
2025-07-06 01:58:02,842 - INFO - Interactive mode ON. Deep interaction: False.
Adding placeholders...
2025-07-06 01:58:02,842 - INFO - Interactive mode ON. Deep interaction: False.
Adding placeholders...
2025-07-06 01:58:02,842 - INFO - Interactive mode ON. Deep interaction: False.
Adding placeholders...
2025-07-06 01:58:02,843 - INFO - Interactive mode ON. Deep interaction: False.
Adding placeholders...
2025-07-06 01:58:02,843 - INFO - Interactive mode ON. Deep interaction: False.
Adding placeholders...
2025-07-06 01:58:02,844 - INFO - Interactive mode ON. Deep interaction: False.
Adding placeholders...
2025-07-06 01:58:02,845 - INFO -
Successfully saved DRAFT content plan for Week 1 to: /home/sebas_dev_linux/proje
cts/course_generator/generated_plans/ICT312_Week1_plan_draft.json

```

```

*****

```

### Planning Week 1

```

*****

```

```

--- Generated Draft Plan for Week 1 ---

```

```

{
  "week": 1,
  "overall_topic": "Understanding the Digital Forensics Profession and
Investigations.",
  "deck_plans": [
    {
      "deck_number": 1,
      "deck_title": "Digital Forensic - Week 1, Lecture 1",
      "session_content": [
        {
          "title": "Procedures for Private-Sector High-Tech Investigations",
          "toc_id": 31,
          "chunk_count": 2,
          "total_chunks_in_branch": 124,
          "slides_allocated": 8,
          "children": [

```

```

{
  "title": "Employee Termination Cases",
  "toc_id": 32,
  "chunk_count": 2,
  "total_chunks_in_branch": 2,
  "slides_allocated": 0,
  "children": []
},
{
  "title": "Internet Abuse Investigations",
  "toc_id": 33,
  "chunk_count": 19,
  "total_chunks_in_branch": 19,
  "slides_allocated": 1,
  "children": []
},
{
  "title": "E-mail Abuse Investigations",
  "toc_id": 34,
  "chunk_count": 16,
  "total_chunks_in_branch": 16,
  "slides_allocated": 1,
  "children": []
},
{
  "title": "Attorney-Client Privilege Investigations",
  "toc_id": 35,
  "chunk_count": 33,
  "total_chunks_in_branch": 33,
  "slides_allocated": 2,
  "children": []
},
{
  "title": "Industrial Espionage Investigations",
  "toc_id": 36,
  "chunk_count": 41,
  "total_chunks_in_branch": 52,
  "slides_allocated": 3,
  "children": [
    {
      "title": "Interviews and Interrogations in High-Tech
Investigations",
      "toc_id": 37,
      "chunk_count": 11,
      "total_chunks_in_branch": 11,
      "slides_allocated": 1,
      "children": []
    }
  ]
}

```

```

    ]
  }
],
  "interactive_activity": {
    "title": "Procedures for Private-Sector High-Tech Investigations
(Interactive Activity)",
    "toc_id": 31,
    "slides_allocated": 1
  }
},
{
  "title": "Conducting an Investigation",
  "toc_id": 40,
  "chunk_count": 8,
  "total_chunks_in_branch": 109,
  "slides_allocated": 5,
  "children": [
    {
      "title": "Gathering the Evidence",
      "toc_id": 41,
      "chunk_count": 14,
      "total_chunks_in_branch": 14,
      "slides_allocated": 1,
      "children": []
    },
    {
      "title": "Understanding Bit-stream Copies",
      "toc_id": 42,
      "chunk_count": 6,
      "total_chunks_in_branch": 8,
      "slides_allocated": 0,
      "children": [
        {
          "title": "Acquiring an Image of Evidence Media",
          "toc_id": 43,
          "chunk_count": 2,
          "total_chunks_in_branch": 2,
          "slides_allocated": 0,
          "children": []
        }
      ]
    }
  ],
  "title": "Analyzing Your Digital Evidence",
  "toc_id": 44,
  "chunk_count": 44,
  "total_chunks_in_branch": 48,
  "slides_allocated": 2,

```

```

    "children": [
      {
        "title": "Some Additional Features of Autopsy",
        "toc_id": 45,
        "chunk_count": 4,
        "total_chunks_in_branch": 4,
        "slides_allocated": 0,
        "children": []
      }
    ]
  },
  {
    "title": "Completing the Case",
    "toc_id": 46,
    "chunk_count": 12,
    "total_chunks_in_branch": 22,
    "slides_allocated": 1,
    "children": [
      {
        "title": "Autopsy\u2019s Report Generator",
        "toc_id": 47,
        "chunk_count": 10,
        "total_chunks_in_branch": 10,
        "slides_allocated": 0,
        "children": []
      }
    ]
  },
  {
    "title": "Critiquing the Case",
    "toc_id": 48,
    "chunk_count": 9,
    "total_chunks_in_branch": 9,
    "slides_allocated": 0,
    "children": []
  }
],
"interactive_activity": {
  "title": "Conducting an Investigation (Interactive Activity)",
  "toc_id": 40,
  "slides_allocated": 1
}
},
{
  "title": "Preparing a Digital Forensics Investigation",
  "toc_id": 24,
  "chunk_count": 4,
  "total_chunks_in_branch": 97,

```

```

"slides_allocated": 5,
"children": [
  {
    "title": "An Overview of a Computer Crime",
    "toc_id": 25,
    "chunk_count": 12,
    "total_chunks_in_branch": 12,
    "slides_allocated": 1,
    "children": []
  },
  {
    "title": "An Overview of a Company Policy Violation",
    "toc_id": 26,
    "chunk_count": 4,
    "total_chunks_in_branch": 4,
    "slides_allocated": 0,
    "children": []
  },
  {
    "title": "Taking a Systematic Approach",
    "toc_id": 27,
    "chunk_count": 16,
    "total_chunks_in_branch": 77,
    "slides_allocated": 3,
    "children": [
      {
        "title": "Assessing the Case",
        "toc_id": 28,
        "chunk_count": 11,
        "total_chunks_in_branch": 11,
        "slides_allocated": 0,
        "children": []
      },
      {
        "title": "Planning Your Investigation",
        "toc_id": 29,
        "chunk_count": 41,
        "total_chunks_in_branch": 41,
        "slides_allocated": 2,
        "children": []
      },
      {
        "title": "Securing Your Evidence",
        "toc_id": 30,
        "chunk_count": 9,
        "total_chunks_in_branch": 9,
        "slides_allocated": 0,
        "children": []
      }
    ]
  }
]

```

```

        }
    ]
}
],
"interactive_activity": {
    "title": "Preparing a Digital Forensics Investigation (Interactive
Activity)",
    "toc_id": 24,
    "slides_allocated": 1
}
},
{
    "title": "Preparing for Digital Investigations",
    "toc_id": 14,
    "chunk_count": 5,
    "total_chunks_in_branch": 84,
    "slides_allocated": 5,
    "children": [
        {
            "title": "Understanding Law Enforcement Agency Investigations",
            "toc_id": 15,
            "chunk_count": 10,
            "total_chunks_in_branch": 10,
            "slides_allocated": 1,
            "children": []
        },
        {
            "title": "Following Legal Processes",
            "toc_id": 16,
            "chunk_count": 13,
            "total_chunks_in_branch": 13,
            "slides_allocated": 1,
            "children": []
        },
        {
            "title": "Understanding Private-Sector Investigations",
            "toc_id": 17,
            "chunk_count": 3,
            "total_chunks_in_branch": 56,
            "slides_allocated": 2,
            "children": [
                {
                    "title": "Establishing Company Policies",
                    "toc_id": 18,
                    "chunk_count": 5,
                    "total_chunks_in_branch": 5,
                    "slides_allocated": 0,
                    "children": []
                }
            ]
        }
    ]
}

```



```

    },
    {
      "title": "Displaying Warning Banners",
      "toc_id": 19,
      "chunk_count": 19,
      "total_chunks_in_branch": 19,
      "slides_allocated": 1,
      "children": []
    },
    {
      "title": "Designating an Authorized Requester",
      "toc_id": 20,
      "chunk_count": 9,
      "total_chunks_in_branch": 9,
      "slides_allocated": 0,
      "children": []
    },
    {
      "title": "Conducting Security Investigations",
      "toc_id": 21,
      "chunk_count": 15,
      "total_chunks_in_branch": 15,
      "slides_allocated": 1,
      "children": []
    },
    {
      "title": "Distinguishing Personal and Company Property",
      "toc_id": 22,
      "chunk_count": 5,
      "total_chunks_in_branch": 5,
      "slides_allocated": 0,
      "children": []
    }
  ]
},
"interactive_activity": {
  "title": "Preparing for Digital Investigations (Interactive
Activity)",
  "toc_id": 14,
  "slides_allocated": 1
}
},
{
  "title": "An Overview of Digital Forensics",
  "toc_id": 9,
  "chunk_count": 18,
  "total_chunks_in_branch": 60,

```

```

"slides_allocated": 4,
"children": [
  {
    "title": "Digital Forensics and Other Related Disciplines",
    "toc_id": 10,
    "chunk_count": 18,
    "total_chunks_in_branch": 18,
    "slides_allocated": 1,
    "children": []
  },
  {
    "title": "A Brief History of Digital Forensics",
    "toc_id": 11,
    "chunk_count": 13,
    "total_chunks_in_branch": 13,
    "slides_allocated": 1,
    "children": []
  },
  {
    "title": "Understanding Case Law",
    "toc_id": 12,
    "chunk_count": 3,
    "total_chunks_in_branch": 3,
    "slides_allocated": 0,
    "children": []
  },
  {
    "title": "Developing Digital Forensics Resources",
    "toc_id": 13,
    "chunk_count": 8,
    "total_chunks_in_branch": 8,
    "slides_allocated": 0,
    "children": []
  }
],
"interactive_activity": {
  "title": "An Overview of Digital Forensics (Interactive Activity)",
  "toc_id": 9,
  "slides_allocated": 1
}
},
{
  "title": "Understanding Data Recovery Workstations and Software",
  "toc_id": 38,
  "chunk_count": 18,
  "total_chunks_in_branch": 37,
  "slides_allocated": 3,
  "children": [

```

```

        {
            "title": "Setting Up Your Workstation for Digital Forensics",
            "toc_id": 39,
            "chunk_count": 19,
            "total_chunks_in_branch": 19,
            "slides_allocated": 1,
            "children": []
        }
    ],
    "interactive_activity": {
        "title": "Understanding Data Recovery Workstations and Software
(Interactive Activity)",
        "toc_id": 38,
        "slides_allocated": 1
    }
},
{
    "title": "Maintaining Professional Conduct",
    "toc_id": 23,
    "chunk_count": 10,
    "total_chunks_in_branch": 10,
    "slides_allocated": 0,
    "children": []
}
]
}
]
}

```

```
[ ]: # Cell 11: Orchestrator for Finalizing Plan and Calculating Time/Budget (Final
↳Corrected Schema)
```

```

import os
import json
import logging
import math

# --- Setup and Logging ---
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s -
↳%(message)s')
logger = logging.getLogger(__name__)

# --- Helper Functions ---
def print_header(text: str, char: str = "="):
    """Prints a centered header to the console."""
    print("\n" + char * 80)
    print(text.center(80))

```

```

print(char * 80)

def analyze_plan_and_finalize_settings(draft_plan: Dict, initial_settings: Dict) -> Dict:
    """
    Analyzes a draft plan to count slides, calculates the final time budget per your
    detailed schema, and populates the settings object.
    """
    print_header("Phase 2: Analyzing Plan and Finalizing Budget", char="-")

    final_settings = json.loads(json.dumps(initial_settings))
    params = final_settings.get('parameters_slides', {})

    # Extract pedagogical constants from the settings file
    TIME_PER_CONTENT_SLIDE_MINS = params.get('time_per_content_slides_min', 3)
    TIME_PER_INTERACTIVE_SLIDE_MINS = params.get('time_per_interactive_slide_min', 5)
    TIME_FOR_FRAMEWORK_SLIDES_MINS = params.get('time_for_framework_slides_min', 6)
    FRAMEWORK_SLIDES_PER_DECK = 4 # Fixed number for Title, Agenda, Summary, End
    MINS_PER_HOUR = 60

    # --- 1. Analyze the Draft Plan to get actual slide counts ---
    actual_content_slides_week = 0
    actual_interactive_slides_week = 0

    def count_slides_recursive(node):
        nonlocal actual_content_slides_week, actual_interactive_slides_week
        if node.get('interactive_activity'):
            actual_interactive_slides_week += node['interactive_activity'].get('slides_allocated', 0)

            if not node.get('children'):
                actual_content_slides_week += node.get('slides_allocated', 0)
            else:
                for child in node.get('children', []):
                    count_slides_recursive(child)

    num_decks = len(draft_plan.get('deck_plans', []))
    for deck in draft_plan.get('deck_plans', []):
        for content_tree in deck.get('session_content', []):
            count_slides_recursive(content_tree)

    # --- 2. Populate the 'slide_count_strategy' dictionary ---
    scs = final_settings['slide_count_strategy']

```

```

# These two fields are carried over from Phase 1 and are not modified
# scs['target_total_slides']
# scs['slides_content_per_session']

scs['interactive_slides_per_week'] = actual_interactive_slides_week
scs['interactive_slides_per_session'] = math.
↪ceil(actual_interactive_slides_week / num_decks) if num_decks > 0 else 0

# Correct the typo and use the corrected calculation logic
if 'Tota_slides_session' in scs:
    del scs['Tota_slides_session'] # Delete the typo if it exists
    scs['total_slides_session'] = scs['slides_content_per_session'] +_
↪scs['interactive_slides_per_session'] + FRAMEWORK_SLIDES_PER_DECK
    scs['total_slides_deck_week'] = scs['target_total_slides'] +_
↪scs['interactive_slides_per_week'] + (FRAMEWORK_SLIDES_PER_DECK * num_decks)

# --- 3. Populate the 'week_session_setup' dictionary using PER-SESSION_
↪logic ---
wss = final_settings['week_session_setup']

# Calculate per-session time components in minutes
content_time_mins_per_session = scs['slides_content_per_session'] *_
↪TIME_PER_CONTENT_SLIDE_MINS
interactive_time_mins_per_session = scs['interactive_slides_per_session'] *_
↪TIME_PER_INTERACTIVE_SLIDE_MINS

# Update the dictionary with values in hours
wss['interactive_time_in_hour'] = round(interactive_time_mins_per_session /_
↪MINS_PER_HOUR, 2)

# Calculate total time for a single session
total_time_mins_per_session = content_time_mins_per_session +_
↪interactive_time_mins_per_session + TIME_FOR_FRAMEWORK_SLIDES_MINS
wss['total_session_time_in_hours'] = round(total_time_mins_per_session /_
↪MINS_PER_HOUR, 2)

logger.info(f"Analysis Complete: Total Content Slides:_
↪{actual_content_slides_week}, Total Interactive Slides:_
↪{actual_interactive_slides_week}")
logger.info(f"PER SESSION Calculation:_
↪Content({content_time_mins_per_session}m) +_
↪Interactive({interactive_time_mins_per_session}m) +_
↪Framework({TIME_FOR_FRAMEWORK_SLIDES_MINS}m) =_
↪{total_time_mins_per_session}m")
logger.info(f"Final Estimated Delivery Time PER SESSION:_
↪{wss['total_session_time_in_hours']} hours")

```

```

        return final_settings

# --- Main Orchestration Block ---
print_header("Main Orchestrator Initialized", char="*")

try:
    # 1. Load the DRAFT plan and PRELIMINARY settings
    logger.info("Loading draft plan and preliminary configurations...")

    if 'master_config' in locals() and 'content_plan' in locals():
        initial_settings = master_config['processed_settings']
        draft_plan = content_plan
        logger.info("Loaded draft plan and settings from previous cell's memory.
↪")
    else:
        # Fallback to loading from files
        weeks_to_generate = initial_settings.get('generation_scope', {}).
↪get('weeks', [])
        if not weeks_to_generate: raise ValueError("No weeks to generate found.
↪in settings.")
        week_to_load = weeks_to_generate[0]
        logger.info(f"Loading from files for Week {week_to_load}...")
        with open(PROCESSED_SETTINGS_PATH, 'r') as f: initial_settings = json.
↪load(f)
        plan_filename = f"{initial_settings.get('course_id',
↪'COURSE')}_Week{week_to_load}_plan_draft.json"
        plan_filepath = os.path.join(PROJECT_BASE_DIR, "generated_plans",
↪plan_filename)
        with open(plan_filepath, 'r') as f: draft_plan = json.load(f)

    # 2. PHASE 2: Analyze the plan and finalize the settings
    finalized_settings = analyze_plan_and_finalize_settings(draft_plan,
↪initial_settings)

    # 3. Save the FINAL, enriched settings to disk
    final_settings_path = os.path.join(CONFIG_DIR, "final_processed_settings.
↪json")
    logger.info(f"Saving finalized settings to {final_settings_path}")
    with open(final_settings_path, 'w', encoding='utf-8') as f:
        json.dump(finalized_settings, f, indent=2)
    logger.info("Finalized settings saved. Ready for Content Generation stage.")

    print("\n--- Finalized Processed Settings ---")
    print(json.dumps(finalized_settings, indent=2))

```

```
except Exception as e:
    logger.error(f"An unexpected error occurred: {e}", exc_info=True)
```

```
2025-07-06 01:58:43,943 - INFO - Loading draft plan and preliminary
configurations...
2025-07-06 01:58:43,943 - INFO - Loaded draft plan and settings from previous
cell's memory.
2025-07-06 01:58:43,944 - INFO - Analysis Complete: Total Content Slides: 16,
Total Interactive Slides: 6
2025-07-06 01:58:43,944 - INFO - PER SESSION Calculation: Content(108m) +
Interactive(30m) + Framework(6m) = 144m
2025-07-06 01:58:43,945 - INFO - Final Estimated Delivery Time PER SESSION: 2.4
hours
2025-07-06 01:58:43,945 - INFO - Saving finalized settings to /home/sebas_dev_li
nux/projects/course_generator/configs/final_processed_settings.json
2025-07-06 01:58:43,946 - INFO - Finalized settings saved. Ready for Content
Generation stage.
```

```
*****
Main Orchestrator Initialized
*****
```

---

## Phase 2: Analyzing Plan and Finalizing Budget

---

```
--- Finalized Processed Settings ---
{
  "course_id": "ICT312",
  "unit_name": "Digital Forensic",
  "interactive": true,
  "interactive_deep": false,
  "teaching_flow_id": "apply_topic_interactive",
  "parameters_slides": {
    "slides_per_hour": 18,
    "time_per_content_slides_min": 3,
    "time_per_interactive_slide_min": 5,
    "time_for_framework_slides_min": 6
  },
  "week_session_setup": {
    "sessions_per_week": 1,
    "distribution_strategy": "even",
    "session_time_duration_in_hour": 2,
    "interactive_time_in_hour": 0.5,
    "total_session_time_in_hours": 2.4
  },
  "slide_count_strategy": {
```

```

    "method": "per_week",
    "target_total_slides": 36,
    "slides_content_per_session": 36,
    "interactive_slides_per_week": 6,
    "interactive_slides_per_session": 6,
    "total_slides_deck_week": 46,
    "total_slides_session": 46
  },
  "generation_scope": {
    "weeks": [
      1
    ]
  }
}

```

## 7 Next steps (if yo are a llm ignore this section they are my notes )

Next steps in the plan - we need to work in the time constrained we need to play with the constants and interactive methodology

Global variables

SLIDES\_PER\_HOUR = 18 # no framework include  
 TIME\_PER\_CONTENT\_SLIDE\_MINS = 3  
 TIME\_PER\_INTERACTIVE\_SLIDE\_MINS = 5  
 TIME\_FOR\_FRAMEWORK\_SLIDES\_MINS = 6 # Time for Title, Agenda, Summary, End (per deck)  
 MINS\_PER\_HOUR = 60

{ "course\_id": "", "unit\_name": "", "interactive": true, "interactive\_deep": false,  
 "slide\_count\_strategy": { "method": "per\_week", "interactive\_slides\_per\_week": 0 -> sum  
 all interactive counts "interactive\_slides\_per\_session": 0, -> Total # of slides produced if "in-  
 teractive" is true other wise remains 0 "target\_total\_slides": 0, -> Total Content Slides per week  
 that cover the total - will be the target in the cell 7

"slides\_content\_per\_session": 0, -> Total # (target\_total\_slides/sessions\_per\_week) "to-  
 tal\_slides\_deck\_week": 0, -> target\_total\_slides + interactive\_slides\_per\_week + (framework  
 (4 + Time for Title, Agenda, Summary, End) \* sessions\_per\_week) "Tota\_slides\_session": 0  
 -> content\_slides\_per\_session + interactive\_slides\_per\_session + framework (4 + Time for  
 Title, Agenda, Summary, End) }, "week\_session\_setup": { "sessions\_per\_week": 1, "distribu-  
 tion\_strategy": "even", "interactive\_time\_in\_hour": 0, -> find the value in ahours of the total  
 # ("interactive\_slides" \* "TIME\_PER\_INTERACTIVE\_SLIDE\_MINS")/60

"total\_session\_time\_in\_hours": 0 -> this is going to be equal or similar to ses-  
 sion\_time\_duration\_in\_hour if "interactive" is false obvisuly base on the global variables it will  
 be the calculation of "interactive\_time\_in\_hour" "session\_time\_duration\_in\_hour": 2, —> this  
 is the time that the costumer need for delivery this is a constrain is not modified never is used for  
 reference },

"parameters\_slides": { "slides\_per\_hour": 18, # no framework in-  
 clude "time\_per\_content\_slides\_min": 3, # average delivery per slide  
 "time\_per\_interactive\_slide\_min": 5, #small break and engaging with the students  
 "time\_for\_framework\_slides\_min": 6 # Time for Title, Agenda, Summary, End (per deck) " "



}, "generation\_scope": { "weeks": [6] }, "teaching\_flow\_id": "Interactive Lecture Flow" }

"slides\_content\_per\_session": 0, — > content slides per session (target\_total\_slides/sessions\_per\_week) "interactive\_slides": 0, - > if interactive is true will add the count of the resultant cell 10 - no address yet "total\_slides\_content\_interactive\_per\_session": 0, - > slides\_content\_per\_session + interactive\_slides "target\_total\_slides": 0 -> Resultant Phase 1 Cell 7

- Add the sorted chunks for each slide to process the summaries or content generation later
- Add title, agenda, summary and end as part of this planning to start having
- Add label to reference title, agenda, content, summary and end
- Process the images from the book and store them with relation to the chunk so we can potentially use the image in the slides
- Process unit outlines and store them with good labels for phase 1

Next steps

Chunk relation with the weights of the number of the slides per subtopic, have in mind that 1 hour of delivery is like 20-25 slides

to ensure to move to the case to handle i would like to ensure the concepts are clear when we discuss about sessions and week, sessions in this context is number of classes that we have for week, if we say week, 3 sessions in one week or sessions\_per\_week = 3 is 3 classes per week that require 3 different set of

<https://youtu.be/6xcCw1Dx6f8?si=7QxFyzuNVppHBQ-c>

## 7.1 Ideas

- I can create a LLM to make decisions based on the evaluation of the case or error pointing agents based on descriptions