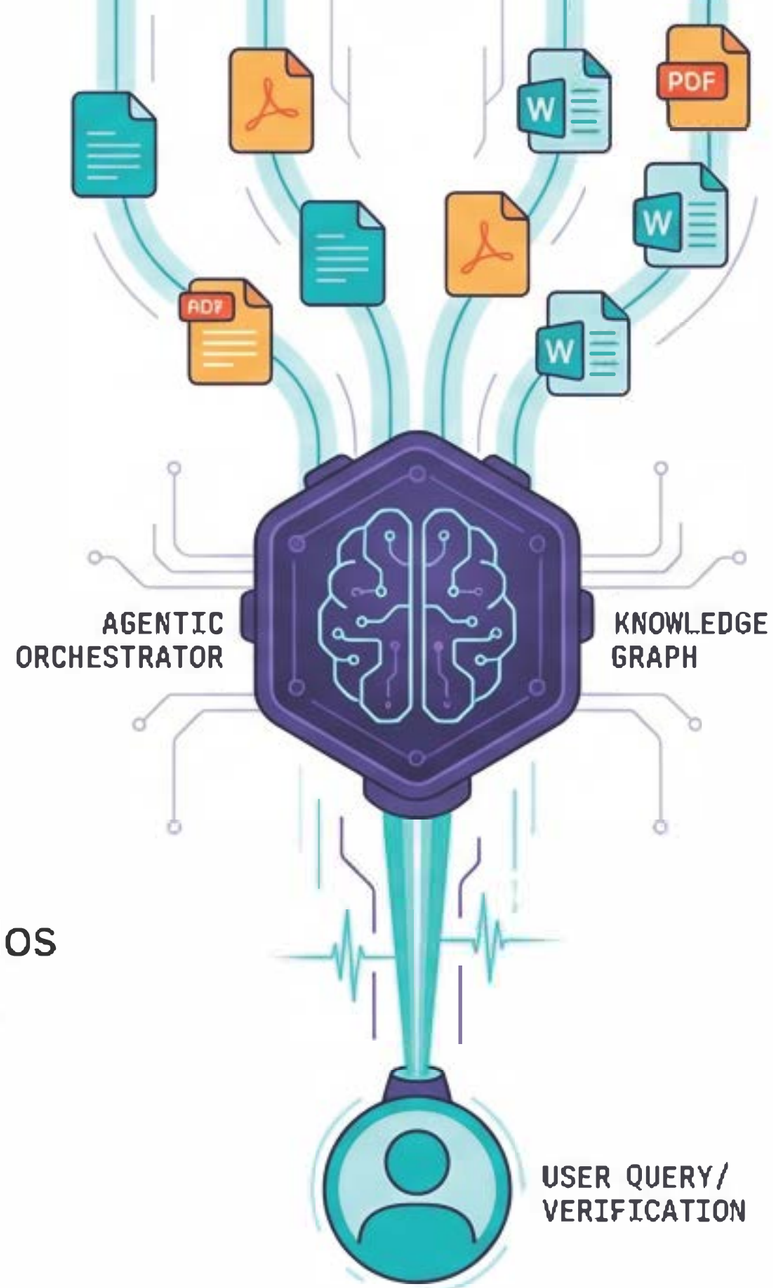
Multi-Division Knowledge Management System

Agentic RAG Architecture with Clickable Page-Level Citations



- **Cross-Division Intelligence**: Unifies HR, Finance, and Legal silos
- **Dual-Agent Orchestration**: Router (Claude Haiku) + Research (Claude 3.5 Sonnet)
- **Verifiable Accuracy**: Inline citations with direct PDF anchors
- **Serverless Infrastructure**: AWS Lambda, Bedrock, Pinecone

Design Philosophy: The Annotated Visual Specification

Logical / Sequential

The "Why"

BLOCK 1: NARRATIVE STRUCTURE

BLOCK 2: CONTEXT SETTING

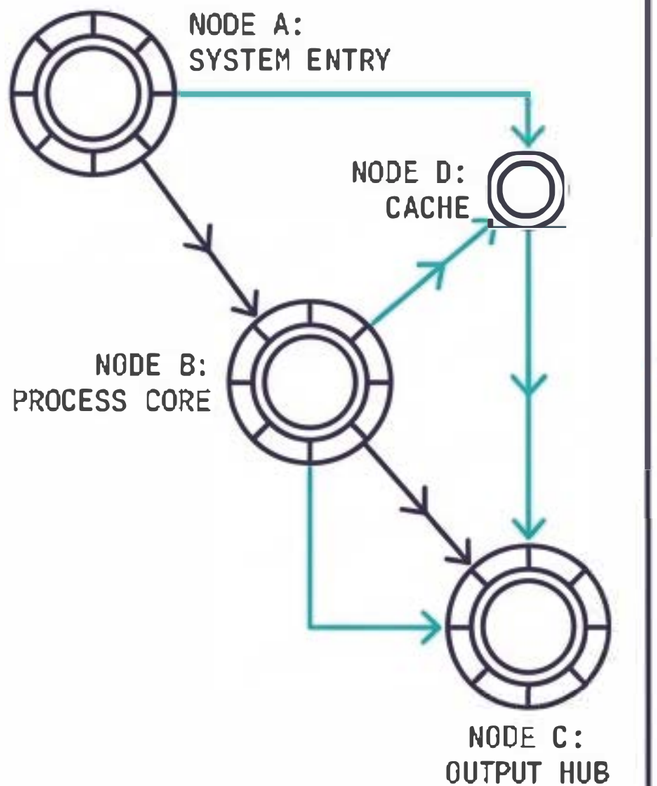
BLOCK 3: FLOW DEFINITION

BLOCK 4: PROGRESSION STEP

BLOCK 5: CONCLUSION

DUAL-CHANNEL PROCESSING
INTEGRATED UNDERSTANDING

Spatial / Relational



The "How"

Color Semantics

Intelligence Layer (Agents)

Data Layer (Ingestion, Storage)

Compute Layer (Processing)

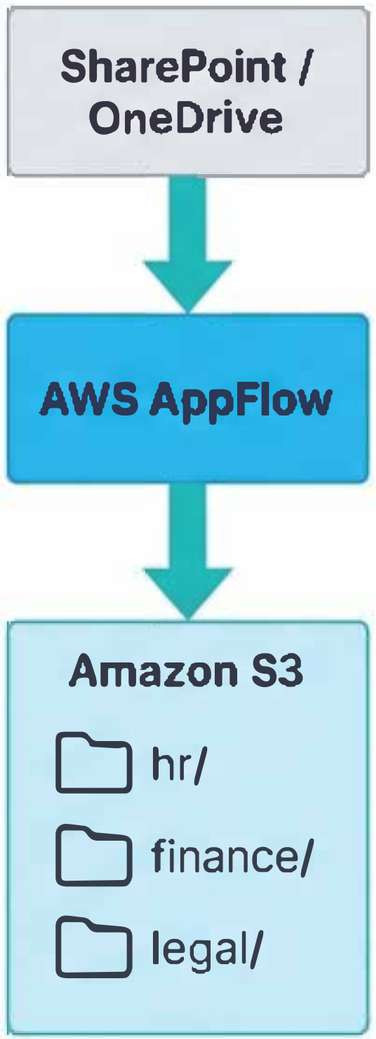
Phase 1: Automated Data Ingestion

1

Data Ingestion

AWS AppFlow syncs OneDrive and SharePoint documents to Amazon S3 every 3-5 hours.

Critical to this phase is hierarchy preservation: Folder structures (e.g., hr/policies/) are mapped directly to S3 prefixes to maintain context.



Phase 2: Document Processing & Vectorization

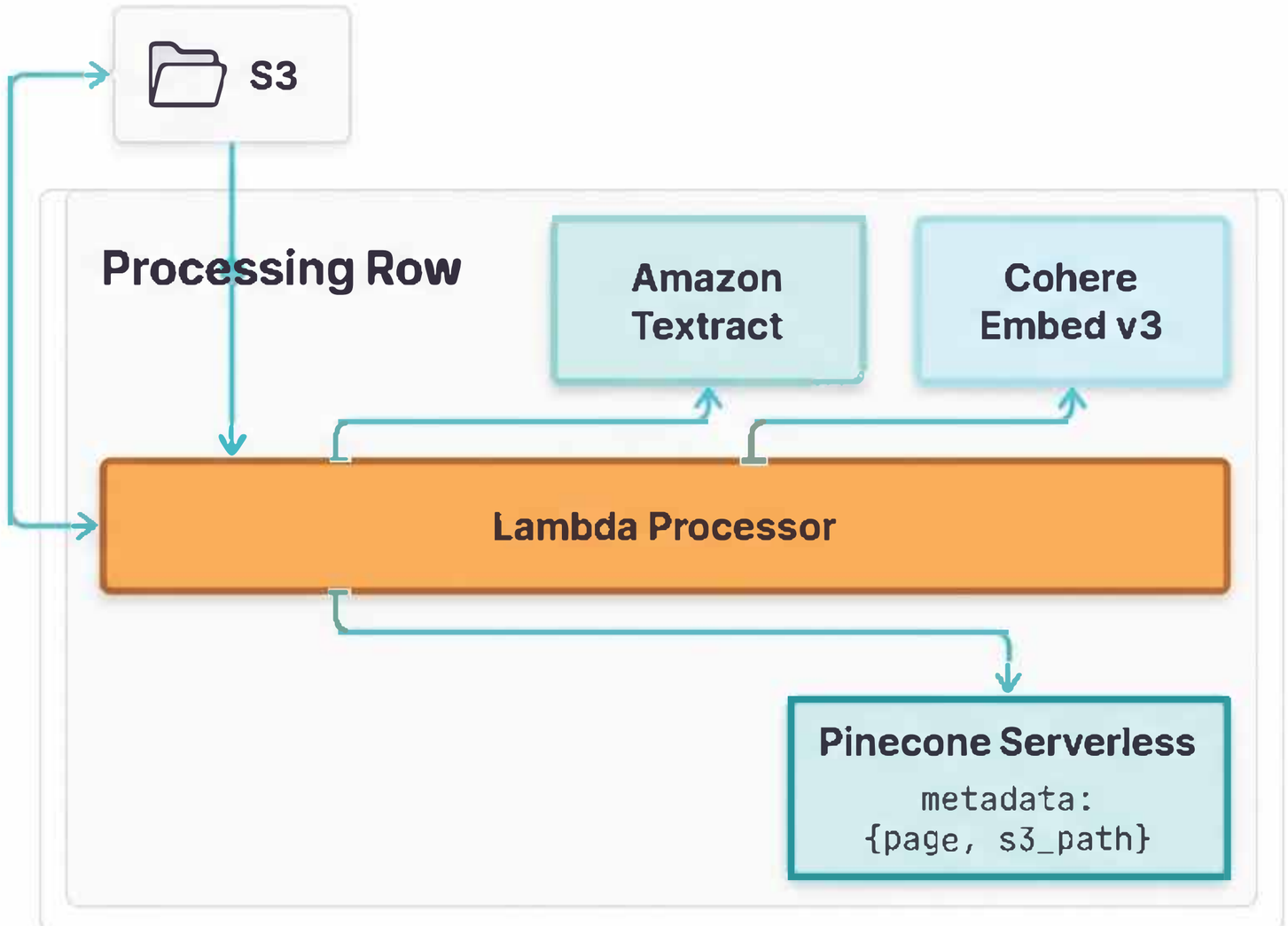
2

Document Processing

Event-driven architecture. S3 upload triggers Lambda.

1. **OCR** via **Amazon Textract** (detects page geometry).
2. **Embedding** via **Cohere Embed v3**.
3. **Indexing** in **Pinecone Serverless**.

Metadata tagging ({page_number}, {s3_path}) is injected here.



Phase 3: The Router Agent (Claude Haiku)

3

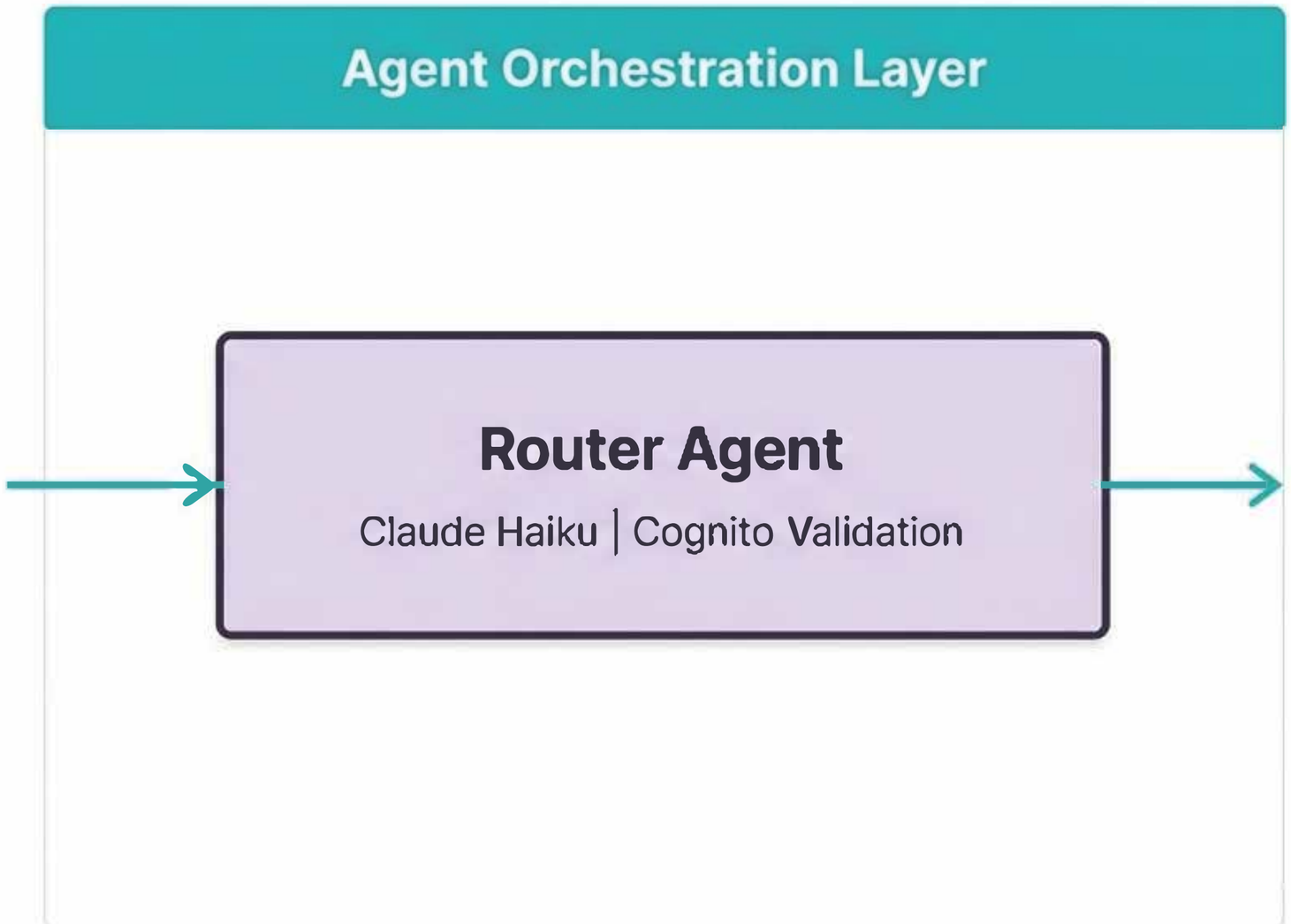
User Query

User asks: "Compare leave policies across HR and Finance divisions".

4

Router Agent

Claude Haiku validates identity via Cognito. Classifies intent. Routes to specific Research Agent.

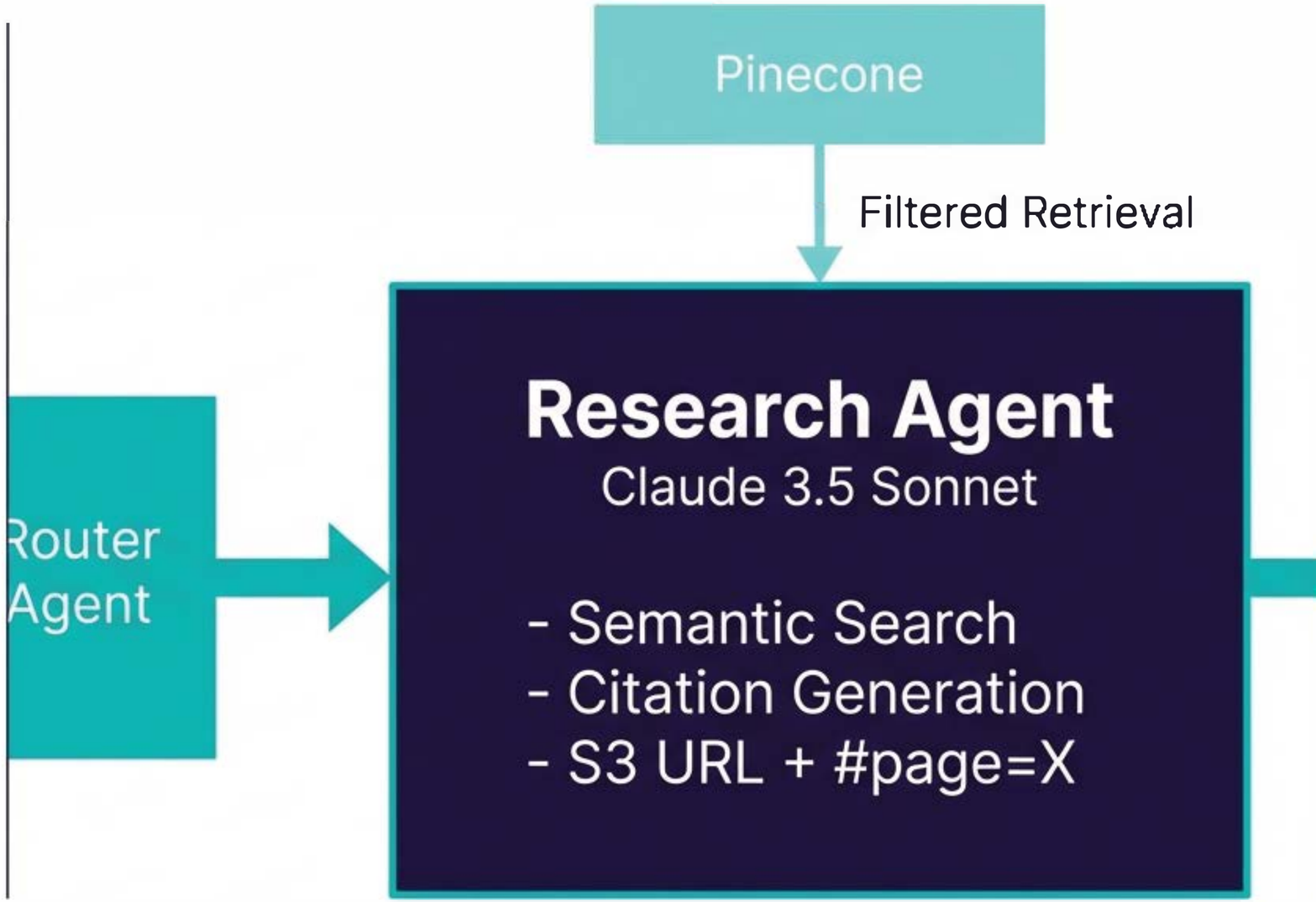


Phase 4: The Research Agent (Claude 3.5 Sonnet)

5

Research Agent

The heavy lifter. Claude 3.5 Sonnet performs:
- Filtered Retrieval (Pinecone query by division)
- Semantic Synthesis (Constructs answer from chunks)
- Citation Injection (Maintains references to source metadata)



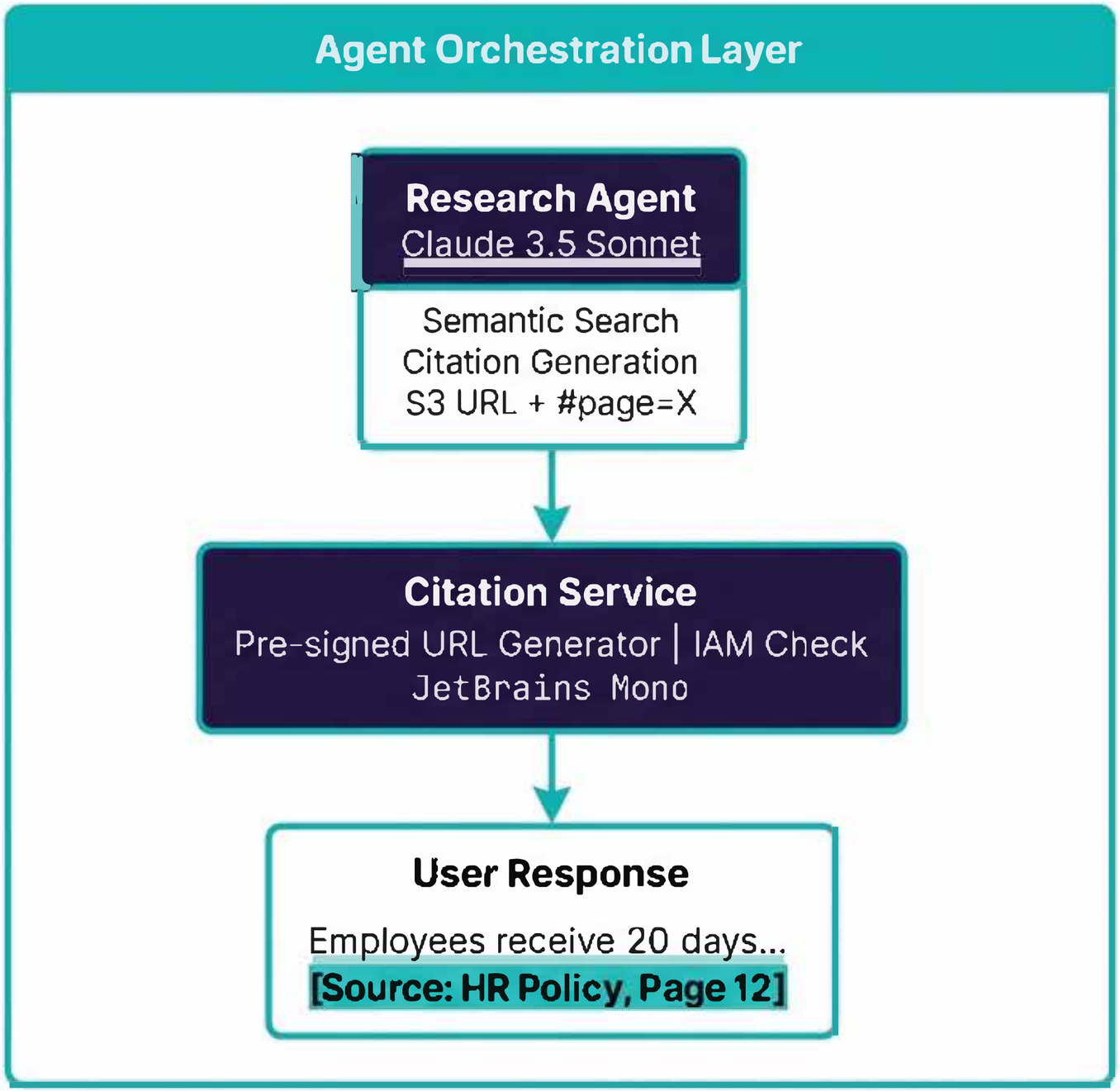
Feature Spotlight: Clickable Citation Generation

6 Citation Link Gen

Lambda generates S3 pre-signed URL (1hr expiry) + PDF anchor (#page=12).

7 Interactive Response

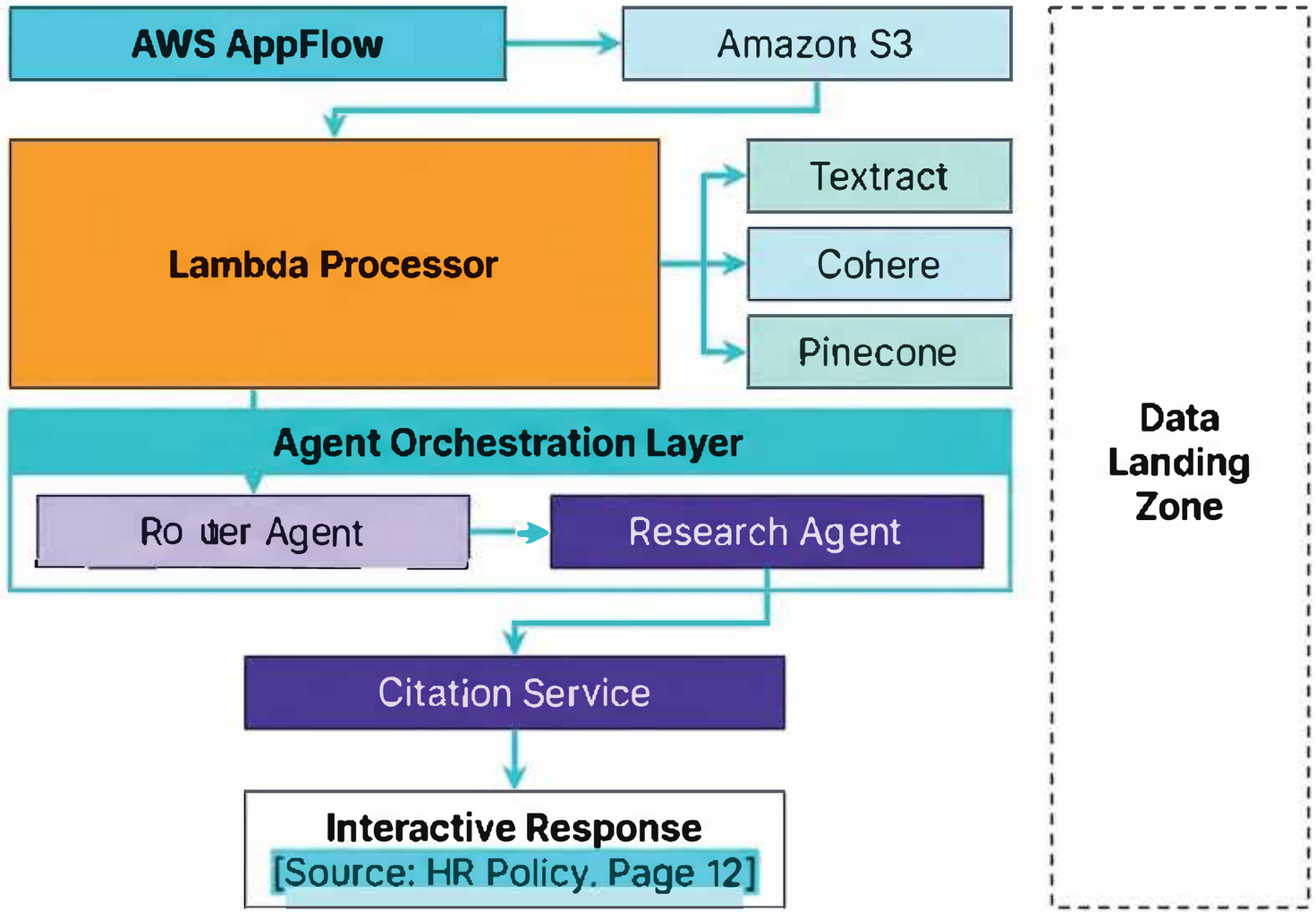
User sees clickable source. Browser opens PDF at exact page.



Complete Architecture: Multi-Division Knowledge Management

Agentic RAG Architecture with Clickable Page-Level Citations

1. **Data Ingestion**
2. **Document Processing**
3. **User Query**
4. **Router Agent**
5. **Research Agent**
6. **Citation Link Gen**
7. **Interactive Response**

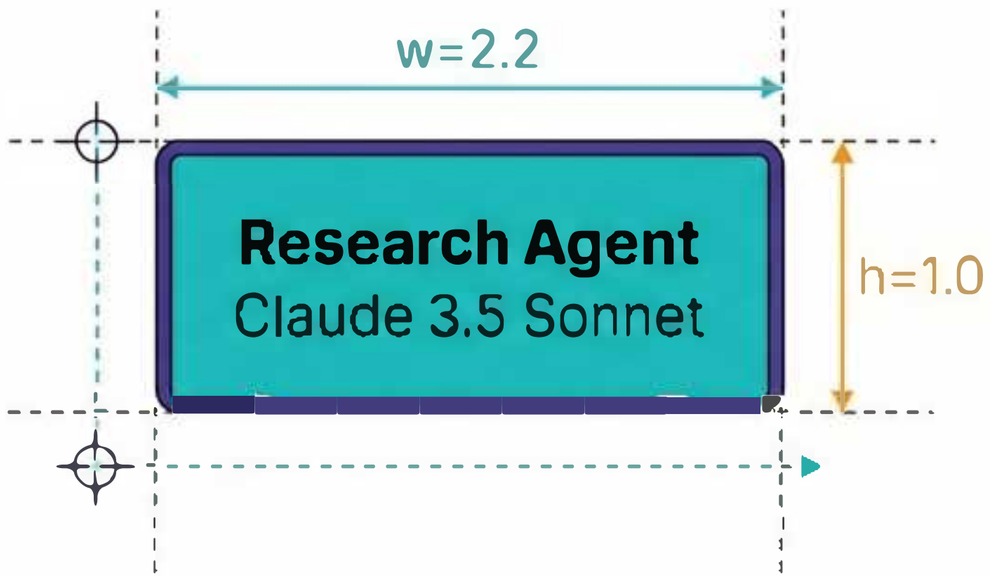


Automated Documentation Generation

Code

```python
def add_arch_box(x, y, w, h, text):
    shape = slide.shapes.add_shape(...)
    shape.fill.fore_color.rgb = COLORS['turquoise']
    return shape
```

Result



Infrastructure as Code extends to Documentation as Code. Using python-pptx to ensure architectural diagrams remain perfectly synchronized with implementation state.