- Semi-Structured Log Files
  - Created by printf statement in server processes:
    - Web, database, network file servers, operating system components
  - Human-readable text format files
    - Very rarely actually read by a human
    - Can store/archive in binary or compressed format
  - Format published or “defined” by code
    - Can be very difficult to parse
- Apache Web Server Log Format
  - Components:
    - First component - Client IP address
    - Second component - User identity from remote machine (hyphen means not available)
    - Third component - User identity from local logon (hyphen means not available)
    - Fourth component - Request time
    - Fifth component - Client request
      - Request method (e.g., Get, Post, etc.)
      - Endpoint (a Uniform Resource Identifier)
      - Client protocol version
    - Sixth component - Status code the server sent back to the client
      - OK respones (2xx), others: 3xx, 4xx, 5xx
      - Size of the object returned to client
        - “-“ if no content returned, or sometimes 0
- Lab: Explore Web Server Access Log
- Some Log Analysis Questions
  - Overall
    - What are the statistics for content being returned?
  - Temporal
    - How man unique host per day?