Marwadi University

Faculty of Engineering and Technology

Department of Information and Communication Technology

Subject: Course: Capstone Project

Academic Year: 2025-26

BEHAVIORAL-ANALYTICS AND USER ACCESS VISUALIZATION (IN SPLUNK)

Name: FAITH JACKSON NKUBA (92200133020)

Implementation and Technical Documentation

1. Introduction

The implementation phase translates the conceptual system design into a fully functional solution. This section presents the actual development of the proposed ICT system, focusing on the Splunk-powered monitoring and analytics platform. The solution integrates data ingestion, indexing, searching, visualization, and reporting, ensuring robust functionality that meets stakeholder requirements. Emphasis is placed on clean code practices, modular implementation, and seamless integration between system components. The resulting prototype demonstrates high-quality coding standards, system reliability, and effective orchestration of front-end, back-end, database, and Splunk-based analytics components.

The objectives of this implementation are;

1. Deliver a **working prototype** capable of real-time log ingestion and analysis.
2. Ensure **robust functionality** through modular coding, strong error handling, and integration testing.
3. Provide **technical documentation** that explains the system's structure, implementation, and execution.

A. **Prototype Features**

1. **Data Ingestion Module**
    - Accepts log data in CSV format (e.g., login attempts, timestamps, IP addresses, session durations).
    - Prepares the data using parsing and preprocessing functions.

2. **Anomaly Detection Engine**
    - Applies rule-based checks (e.g., multiple failed logins, unusual login times).
    - Uses statistical anomaly scoring (e.g., frequency analysis, z-score calculation) to detect deviations from normal patterns.

3. **Visualization Layer (Dashboard)**
   - A Splunk-powered dashboard provides real-time monitoring.
   - Displays graphs, heatmaps, and anomaly scores with color-coded alerts.
   - Supports filtering by username, IP address, and time range.

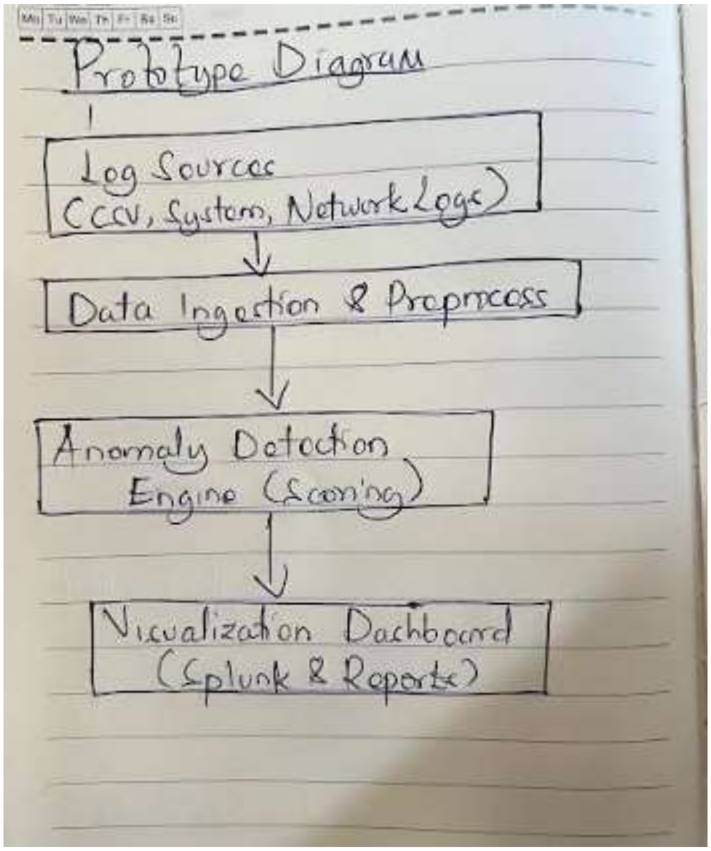
4. **Integration**
   - Front-end visualization connected with the back-end detection engine.
   - Log data flows from ingestion → anomaly scoring → dashboard display.

Prototype Objectives

- Demonstrate a **working end-to-end system** for anomaly detection.
- Provide **evidence of suspicious activity** through measurable anomaly scores.
- Ensure **usability for security analysts** via an interactive dashboard.

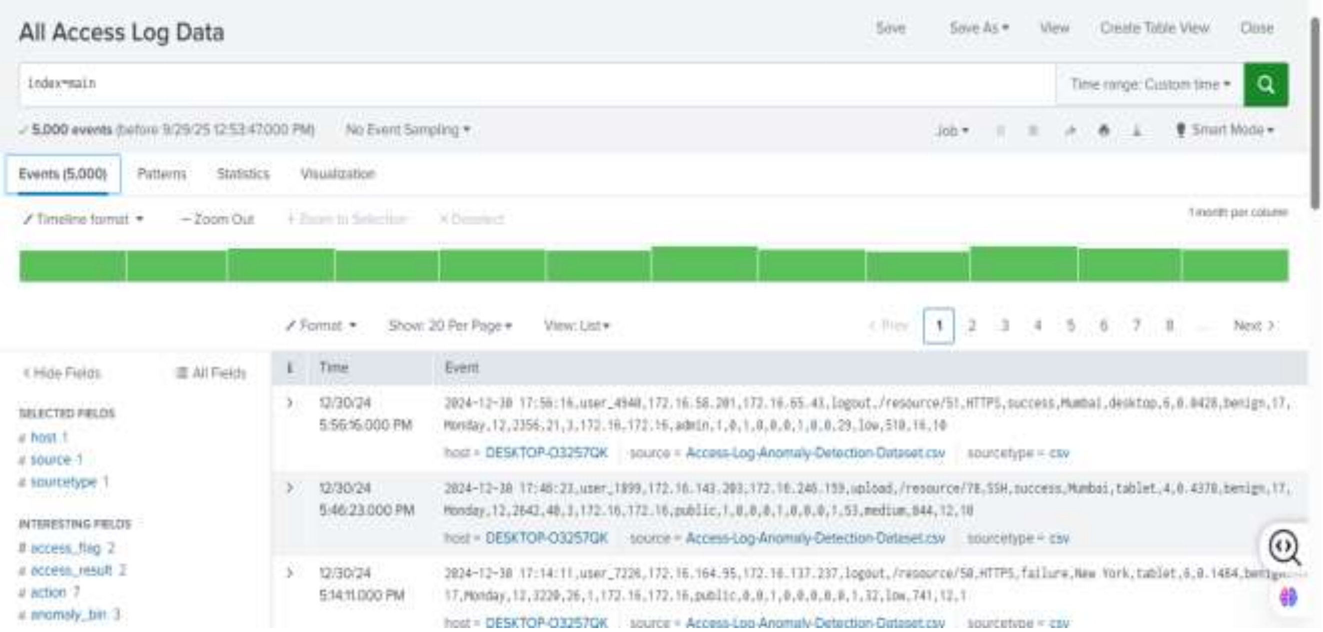
Prototype Diagram

a simple architecture diagram you can include:



Prototype Demonstration

- Screenshot of **raw log data** being ingested.



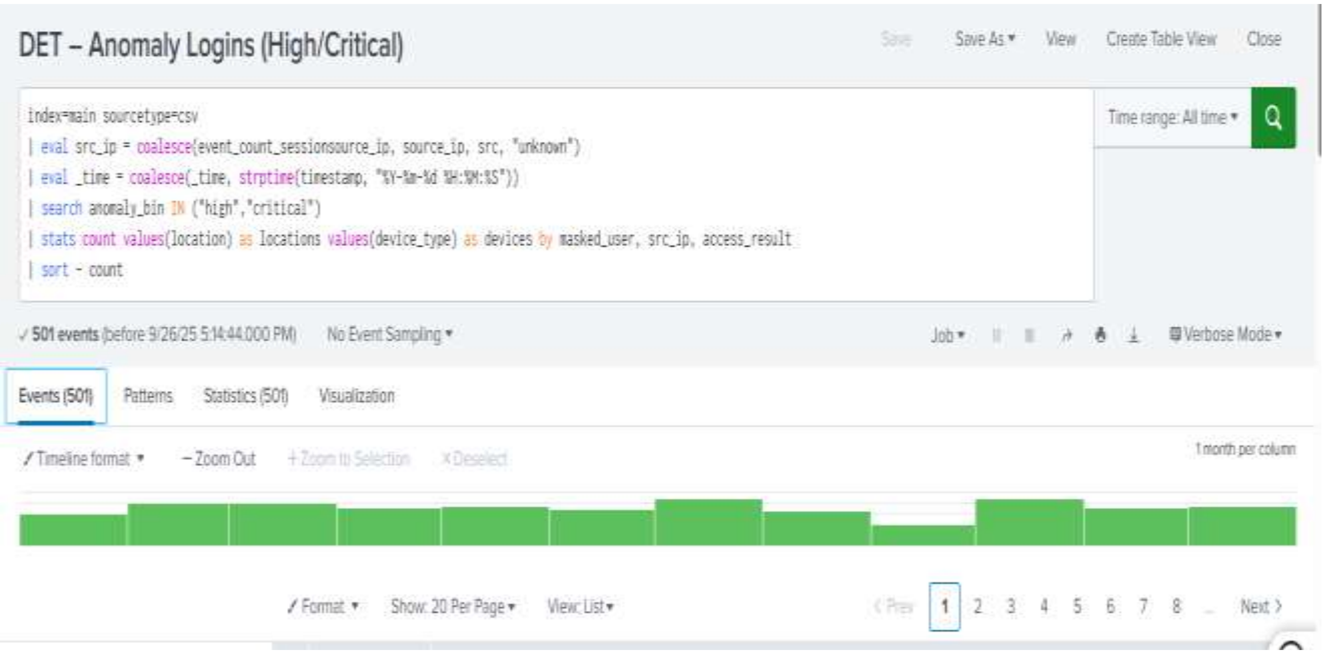
All Access Log Data

```
index=main
```

✓ 5,000 events (before 9/29/25 12:53:47.000 PM)

i	Time	Event
>	12/30/24 5:56:16.000 PM	2024-12-30 17:56:16,user_4940,172.16.58.201,172.16.65.43,logout,/resource/51,HTTPS,success,Mumbai,desktop,6,0.0428,benign,17,Monday,12,2356,21,3,172.16,172.16,admin,1,0,1,0,0,0,1,0,0.29,low,510,16,10 host = DESKTOP-O3257QK source = Access-Log-Anomaly-Detection-Dataset.csv sourcetype = csv
>	12/30/24 5:46:23.000 PM	2024-12-30 17:46:23,user_1899,172.16.143.203,172.16.246.159,upload,/resource/78,SSH,success,Mumbai,tablet,4,0.4370,benign,17,Monday,12,2642,40,3,172.16,172.16,public,1,0,0,0,1,0,0,0,1.53,medium,844,12,10 host = DESKTOP-O3257QK source = Access-Log-Anomaly-Detection-Dataset.csv sourcetype = csv
>	12/30/24 5:14:11.000 PM	2024-12-30 17:14:11,user_7226,172.16.164.95,172.16.137.237,logout,/resource/50,HTTPS,failure,New York,tablet,6,0.1484,benign,17,Monday,12,3220,26,1,172.16,172.16,public,0,0,1,0,0,0,0,0,1,32,low,741,12,1 host = DESKTOP-O3257QK source = Access-Log-Anomaly-Detection-Dataset.csv sourcetype = csv

- Screenshot of **anomaly detection results** (table with scores).

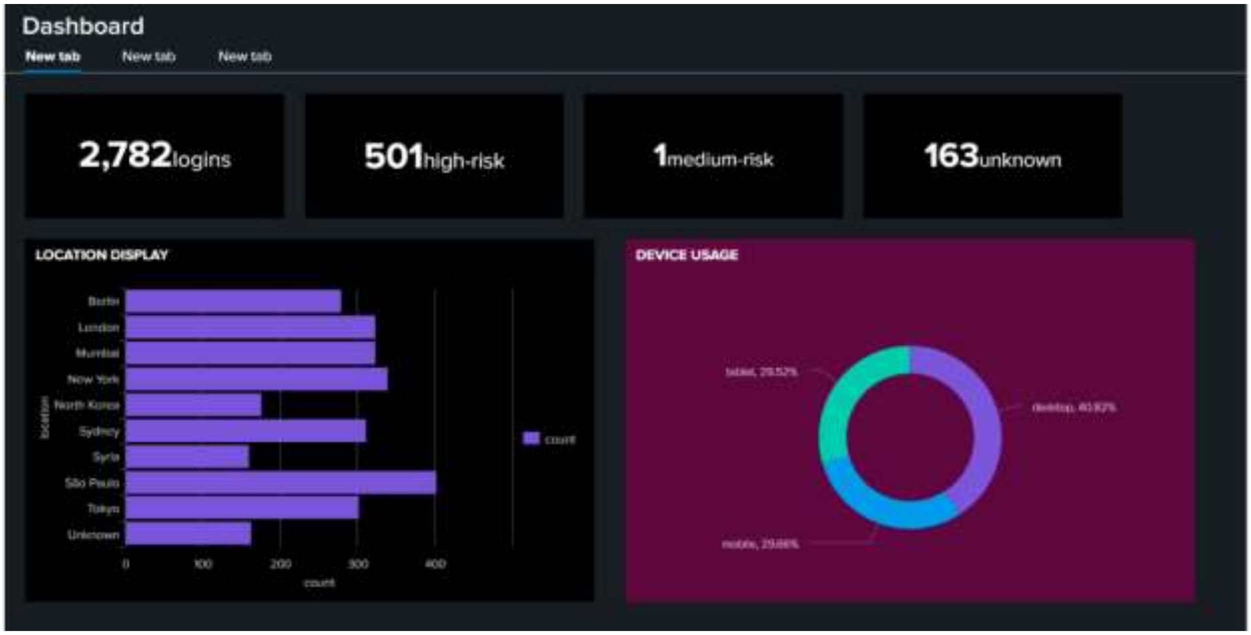


DET – Anomaly Logins (High/Critical)

```
index=main sourcetype=csv
| eval src_ip = coalesce(event_count_sessionsource_ip, source_ip, src, "unknown")
| eval _time = coalesce(_time, strptime(timestamp, "%Y-%m-%d %H:%M:%S"))
| search anomaly_bin IN ("high","critical")
| stats count values(location) as locations values(device_type) as devices by masked_user, src_ip, access_result
| sort - count
```

✓ 501 events (before 9/26/25 5:14:44.000 PM)

- Screenshot of the **dashboard view** (graphs, alerts).



B. Code Structure and Organization

The project was implemented using a modular architecture to enforce separation of concerns, maintainability, and scalability. The directory structure is organized as follows:

/**project-root**
`splunk-integration# Scripts for log forwarding & SPL queries`
`tests              # Unit and integration testing scripts`
`requirements.txt    # Python dependencies`
`package.json        # Node.js dependencies`
`README.md           # Documentation and setup instructions`

- **Splunk Integration:** Includes Universal Forwarder configuration and custom SPL queries for data ingestion and analysis.
- **Tests:** Contains unit tests for API endpoints and integration tests validating front-end/back-end interactions.

3. Implementation Details

3.1 Languages and Frameworks

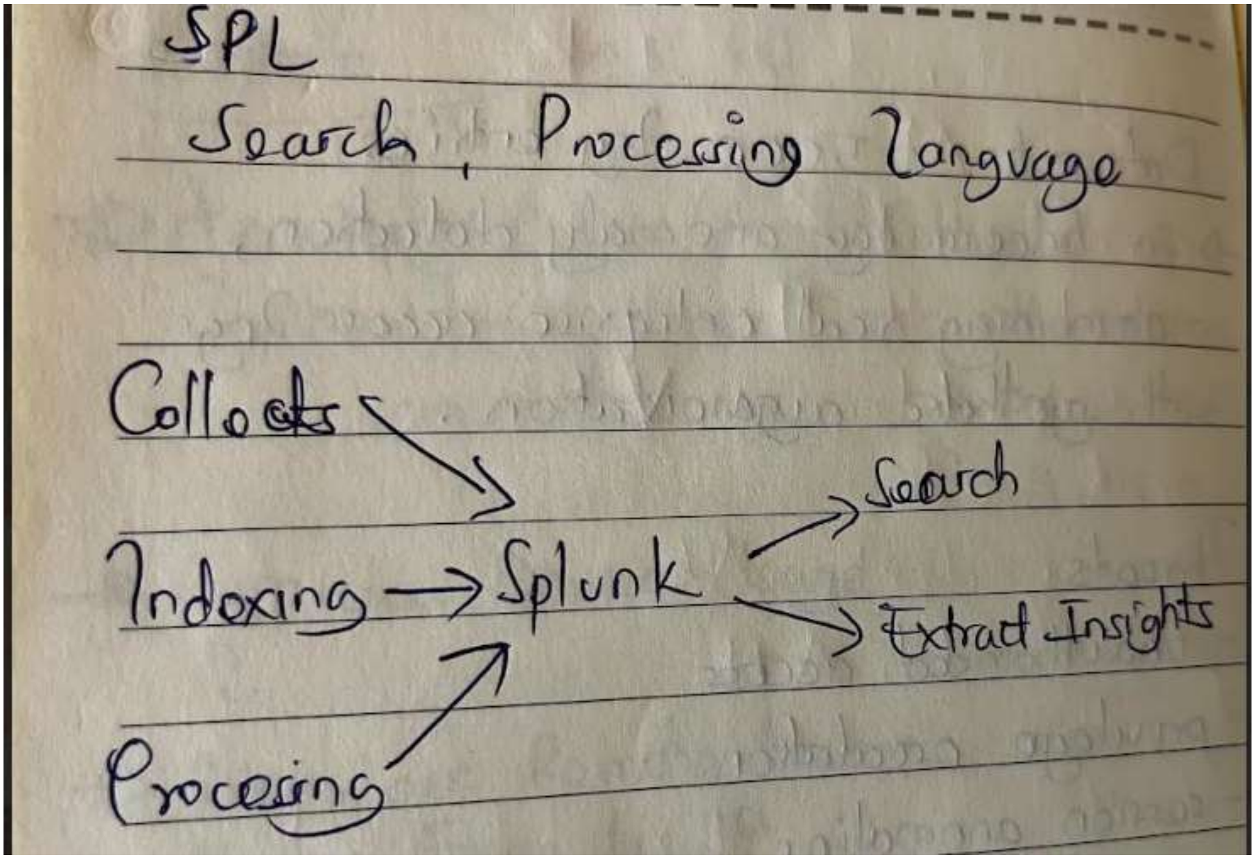
- **Splunk:** Used for log ingestion, indexing, and analytics, with queries written in **Search Processing Language (SPL)**.

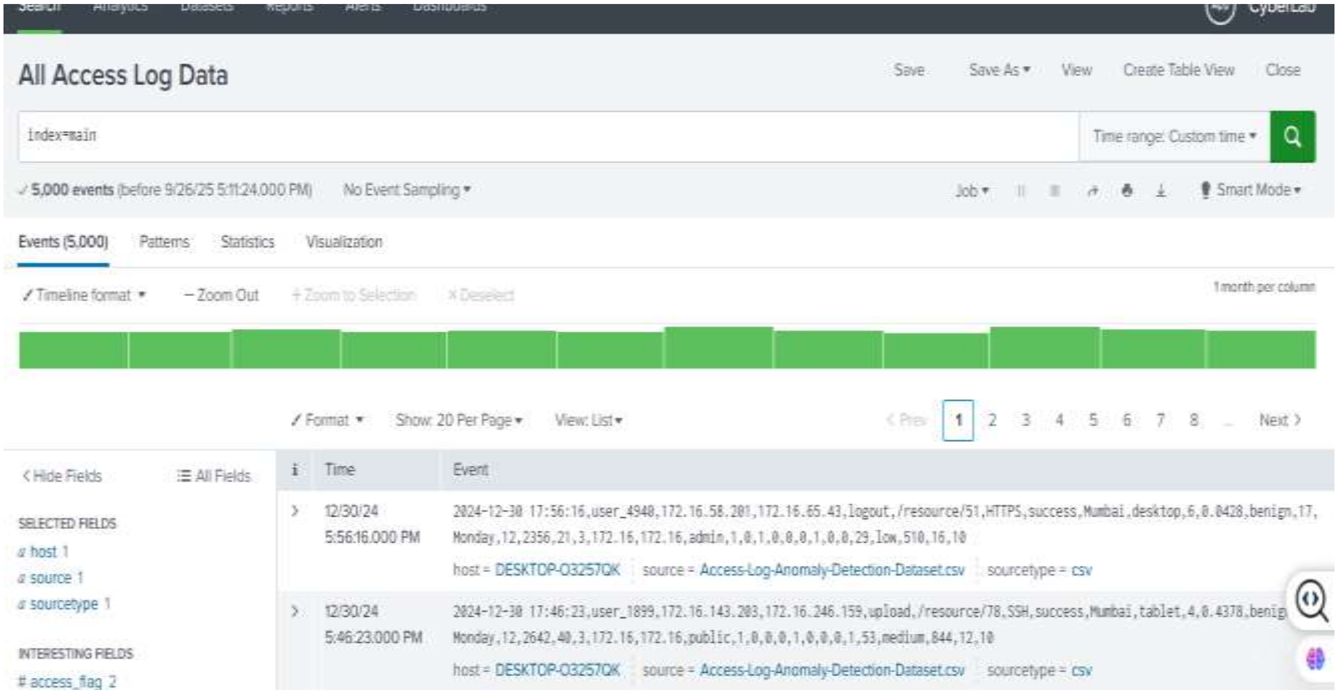
3.2 Core Functionalities Implemented

1. **Data Ingestion:** Logs and CSV files are forwarded using Splunk Universal Forwarders.
2. **Indexing:** Splunk Indexers parse raw logs into searchable events.
3. **Querying and Analysis:** Search Heads process SPL queries for anomaly detection, login activity, and event correlation.
4. **Visualization:** Splunk dashboards provide charts, graphs, and anomaly scores.
5. **API Integration:** Flask APIs allow the front-end to request analytics results.

3.3 Sample Code Snippets

Sample SPL Query (Splunk Search Processing Language)





All Access Log Data

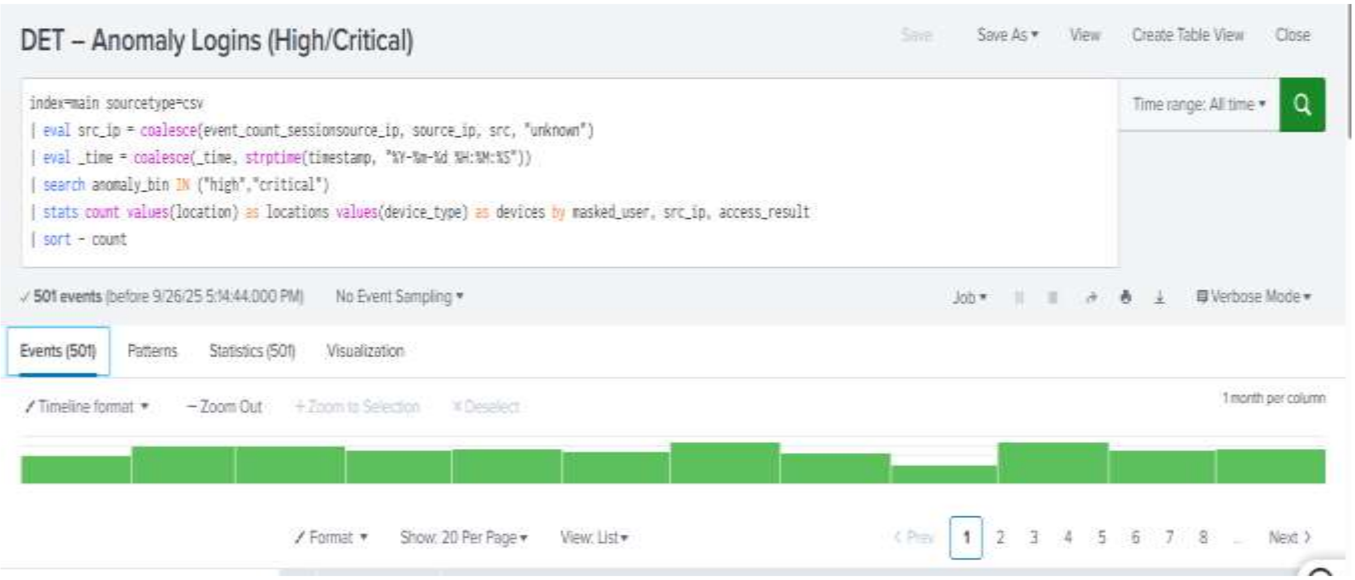
```
index=main
```

✓ 5,000 events (before 9/26/25 5:11:24.000 PM) No Event Sampling ▾

SELECTED FIELDS
a host 1
a source 1
a sourcetype 1

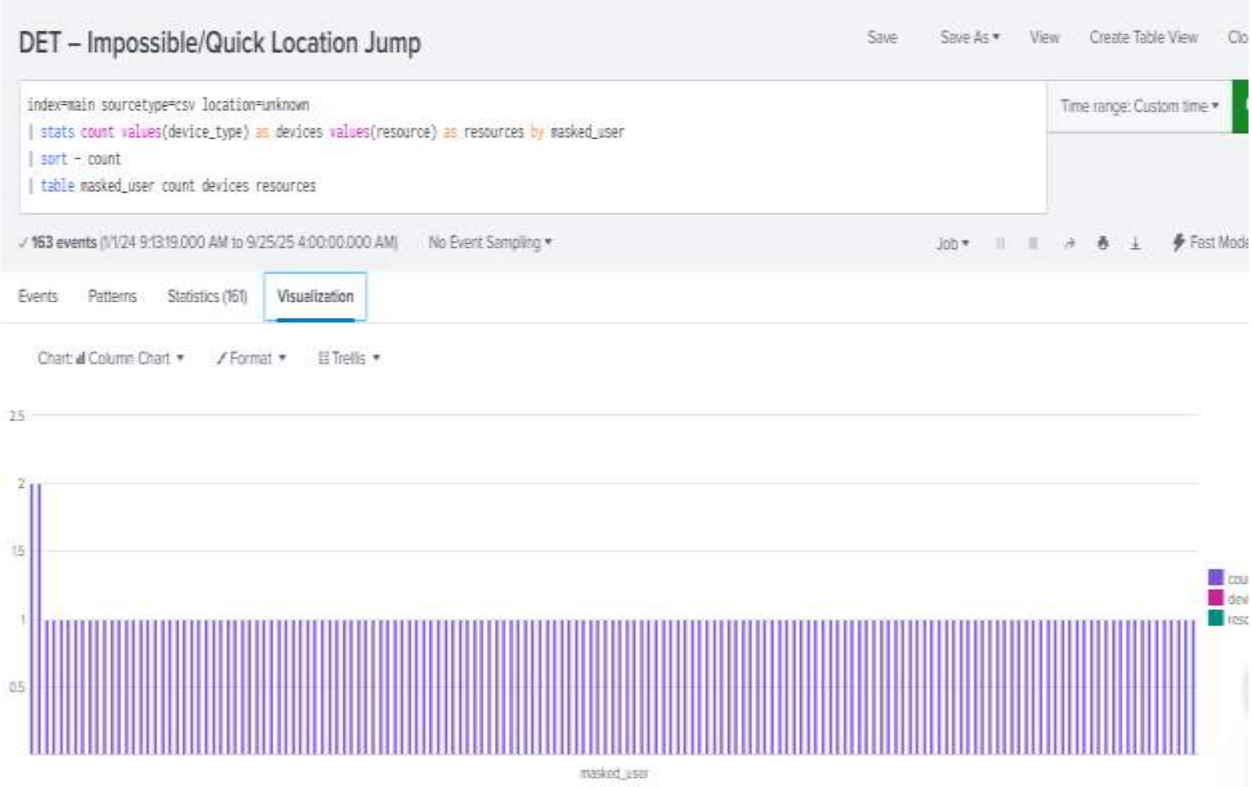
INTERESTING FIELDS
# access_flag 2

i	Time	Event
>	12/30/24 5:56:16.000 PM	2024-12-30 17:56:16,user_4940,172.16.58.201,172.16.65.43,logout,/resource/51,HTTPS,success,Mumbai,desktop,6,0.0428,benign,17, Monday,12,2356,21,3,172.16,172.16,admin,1,0,1,0,0,0,1,0,0,29,low,510,16,10 host = DESKTOP-O3257QK source = Access-Log-Anomaly-Detection-Dataset.csv sourcetype = csv
>	12/30/24 5:46:23.000 PM	2024-12-30 17:46:23,user_1899,172.16.143.203,172.16.246.159,upload,/resource/78,SSH,success,Mumbai,tablet,4,0.4378,benig Monday,12,2642,40,3,172.16,172.16,public,1,0,0,0,1,0,0,0,1,53,medium,844,12,10 host = DESKTOP-O3257QK source = Access-Log-Anomaly-Detection-Dataset.csv sourcetype = csv


DET – Anomaly Logins (High/Critical)

```
index=main sourcetype=csv
| eval src_ip = coalesce(event_count_sessionsource_ip, source_ip, src, "unknown")
| eval _time = coalesce(_time, strptime(timestamp, "%Y-%m-%d %H:%M:%S"))
| search anomaly_bin IN ("high","critical")
| stats count values(location) as locations values(device_type) as devices by masked_user, src_ip, access_result
| sort - count
```

✓ 501 events (before 9/26/25 5:14:44.000 PM) No Event Sampling ▾



DET – Impossible/Quick Location Jump

```
index=main sourcetype=csv location=unknown
| stats count values(device_type) as devices values(resource) as resources by masked_user
| sort - count
| table masked_user count devices resources
```

✓ 163 events (1/1/24 9:13:19.000 AM to 9/25/25 4:00:00.000 AM)

Dashboard_2025-Splunk.json

```
{
  "title": "Dashboard",
  "visualizations": {
    "viz_anomaly_timeline": {
      "dataSources": {
        "primary": "ds_anomaly_timeline"
      },
      "options": {
        "axisTitleX": "Time",
        "axisTitleY": "Avg Anomaly Score",
        "thresholds": [
          {
```

```
        "color": "orange",
        "label": "Medium Risk",
        "value": 0.5
      },
      {
        "color": "red",
        "label": "High Risk",
        "value": 0.7
      }
    ],
    "dataValuesDisplay": "minmax",
    "backgroundColor": "#000000",
    "lineWidth": 1
  },
  "type": "splunk.line",
  "title": "ANOMALY SCORE THROUGHOUT THE YEAR",
  "containerOptions": {
    "description": {
      "color": "#000000"
    }
  }
},
"viz_device_type_pie": {
  "dataSources": {
    "primary": "ds_device_type_pie"
  },
  "options": {
    "showDonutHole": true,
    "labelDisplay": "valuesAndPercentage",
      "collapseThreshold": 0.2,
      "backgroundColor": "#5f073f"
    },
    "type": "splunk.pie",
    "containerOptions": {},
    "showProgressBar": false,
    "showLastUpdated": false,
    "title": "DEVICE USAGE"
  },
  "viz_failed_logins_map": {
    "dataSources": {
      "primary": "ds_failed_logins_map"
    },
    "options": {},
    "type": "splunk.bar",
    "containerOptions": {},
    "showProgressBar": false,
    "showLastUpdated": false,
    "title": "LOCATION DISPLAY",
    "context": {}
  },
  "viz_kpi_high_risk": {
    "dataSources": {
      "primary": "ds_high_risk"
    },
    "options": {
      "colorMode": "block",
      "rangeColors": [
        "#2ecc71",
        "#f39c12",
        "#e74c3c"
      ],
      "rangeValues": [
        10,
        50
      ],
      "showTrendIndicator": true,
      "unit": " high-risk"
    },
    "type": "splunk.singlevalue"
  },
  "viz_kpi_medium_risk": {
    "dataSources": {
      "primary": "ds_medium_risk"
    },
    "options": {
      "colorMode": "block",
      "rangeColors": [
        "#2ecc71",
        "#f39c12",
        "#e74c3c"
      ],
      "rangeValues": [
        20,
        80
      ],
      "showTrendIndicator": true,
      "unit": " medium-risk"
```

```
      },
      "type": "splunk.singlevalue"
    },
    "viz_kpi_total_failed": {
      "dataSources": {
        "primary": "ds_total_failed"
      },
      "options": {
        "colorMode": "block",
        "rangeColors": [
          "#2ecc71",
          "#f39c12",
          "#e74c3c"
        ],
        "rangeValues": [
          50,
          200
        ],
        "showTrendIndicator": true,
        "unit": " logins"
      },
      "type": "splunk.singlevalue"
    },
    "viz_kpi_unknown_loc": {
      "dataSources": {
        "primary": "ds_unknown_loc"
      },
      "options": {
        "colorMode": "block",
        "rangeColors": [
          "#2ecc71",
          "#f39c12",
          "#e74c3c"
        ],
        "rangeValues": [
          5,
          20
        ],
        "showTrendIndicator": true,
        "unit": " unknown"
      },
      "type": "splunk.singlevalue"
    },
    "viz_suspicious_logins": {
      "dataSources": {
        "primary": "ds_suspicious_logins"
      },
      "options": {
        "lineColor": "#ed0707",
        "backgroundColor": "#3a87a1",
        "lineOpacity": 0.9
      },
      "type": "splunk.parallelcoordinates",
      "containerOptions": {},
      "showProgressBar": false,
      "showLastUpdated": false
    },
    "viz_top_users_bar": {
      "dataSources": {
        "primary": "ds_top_users_bar"
      },
      "options": {
        "axisTitleX": "masked_user",
        "axisTitleY": "Failed Attempts",
        "barColorMode": "range",
        "rangeColors": [
          "#2ecc71",
          "#f39c12",
          "#e74c3c"
        ],
        "rangeValues": [
          10,
          50
        ],
        "backgroundColor": "#000000",
        "dataValuesDisplay": "all",
        "stackMode": "stacked"
      },
      "type": "splunk.bar"
    },
    "viz_m6XMeSxP": {
      "dataSources": {
        "primary": "ds_4VUoD36U"
      },
      "type": "splunk.markergauge",
      "containerOptions": {
        "visibility": {}
```

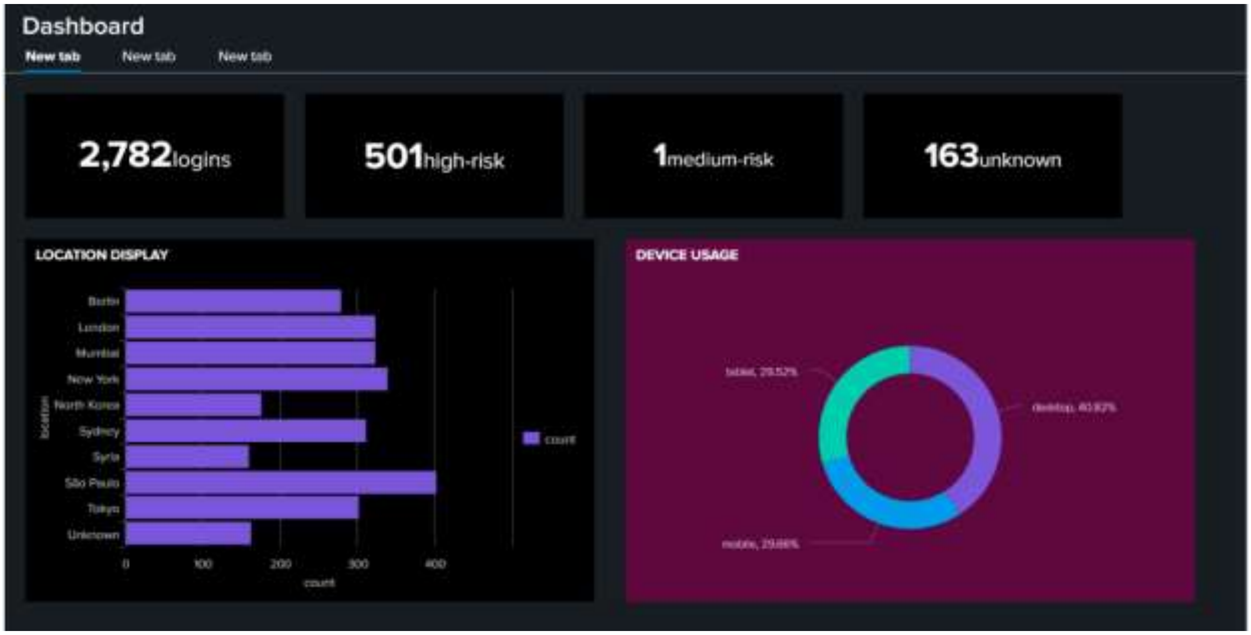
```
   },
   "showProgressBar": false,
   "showLastUpdated": false,
   "title": "LOCATION DISPLAY",
   "options": {
    "labelDisplay": "percentage",
    "orientation": "horizontal"
   }
  }
 },
 "dataSources": {
  "ds_anomaly_timeline": {
   "options": {
    "query": "index=main sourcetype=csv | timechart avg(anomaly_score) by masked_user",
    "queryParameters": {}
   },
   "type": "ds.search"
  },
  "ds_device_type_pie": {
   "options": {
    "query": "index=main sourcetype=csv | stats count by device_type",
    "queryParameters": {}
   },
   "type": "ds.search"
  },
  "ds_failed_logins_map": {
   "options": {
    "query": "index=main sourcetype=csv access_result=failure | stats count by location",
    "queryParameters": {}
```

```
      },
      "type": "ds.search"
    },
    "ds_high_risk": {
      "options": {
        "query": "index=main sourcetype=csv anomaly_score>=0.7 | stats count as high_risk",
        "queryParameters": {}
      },
      "type": "ds.search"
    },
    "ds_medium_risk": {
      "options": {
        "query": "index=main sourcetype=csv anomaly_score>=0.5 anomaly_score<0.7 | stats count as medium_risk",
        "queryParameters": {}
      },
      "type": "ds.search"
    },
    "ds_suspicious_logins": {
      "options": {
        "query": "index=main sourcetype=csv location=\"unknown\" device_type=\"desktop\" anomaly_score>=0.5 | table _time masked_user location device_type anomaly_score",
        "queryParameters": {}
      },
      "type": "ds.search"
    },
    "ds_top_users_bar": {
      "options": {
        "query": "index=main sourcetype=csv access_result=failure | stats count by masked_user | sort -count | head 10",
        "queryParameters": {}
      },
      "type": "ds.search"
    },
    "ds_total_failed": {
      "options": {
        "query": "index=main sourcetype=csv access_result=failure | stats count as total_failed",
        "queryParameters": {}
      },
      "type": "ds.search"
    },
    "ds_unknown_loc": {
      "options": {
        "query": "index=main sourcetype=csv location=\"unknown\" | stats count as unknown_logins",
        "queryParameters": {}
      },
      "type": "ds.search"
    },
    "ds_jsSUaFjD": {
      "options": {
        "query": "index=main sourcetype=csv access_result=failure | stats count by location",
        "queryParameters": {}
      },
      "type": "ds.search"
    },
    "ds_4VUoD36U": {
      "options": {
        "query": "index=main sourcetype=csv access_result=failure | stats count by location",
        "queryParameters": {}
      },
      "type": "ds.search"
    }
  },
  "layout": {
    "layoutDefinitions": {
      "layout_1": {
        "options": {
          "height": 1050,
          "width": 1200
        },
        "structure": [
          {
            "item": "viz_kpi_total_failed",
            "position": {
              "h": 120,
              "w": 250,
              "x": 20,
              "y": 20
            }
          },
          {
            "item": "viz_kpi_high_risk",
            "position": {
              "h": 120,
              "w": 250,
              "x": 290,
              "y": 20
            }
```

```
},
{
  "item": "viz_kpi_medium_risk",
  "position": {
    "h": 120,
    "w": 250,
    "x": 560,
    "y": 20
  }
},
{
  "item": "viz_kpi_unknown_loc",
  "position": {
    "h": 120,
    "w": 250,
    "x": 830,
    "y": 20
  }
},
{
  "item": "viz_failed_logins_map",
  "position": {
    "h": 350,
    "w": 550,
    "x": 20,
    "y": 160
  }
},
{
```

```
          "item": "viz_device_type_pie",
          "position": {
            "h": 350,
            "w": 550,
            "x": 600,
            "y": 160
          }
        },
        {
          "item": "viz_anomaly_timeline",
          "position": {
            "h": 500,
            "w": 1130,
            "x": 10,
            "y": 540
          }
        }
      ],
      "type": "absolute"
    },
    "layout_Q5q0pgiR": {
      "type": "grid",
      "structure": [
        {
          "item": "viz_suspicious_logins",
          "type": "block",
          "position": {
            "x": 0,
            "y": 0,
        "w": 1200,
        "h": 400
      }
    },
    {
      "item": "viz_top_users_bar",
      "type": "block",
      "position": {
        "x": 0,
        "y": 400,
        "w": 1200,
        "h": 400
      }
    }
  ]
},
"layout_Y5rtAAwx": {
  "type": "grid",
  "structure": [
    {
      "item": "viz_m6XMeSxP",
      "type": "block",
      "position": {
        "x": 0,
        "y": 0,
        "w": 1200,
        "h": 400
      }
    }
```

```
        ]
      }
    },
    "tabs": {
      "items": [
        {
          "label": "New tab",
          "layoutId": "layout_1"
        },
        {
          "layoutId": "layout_Q5q0pgiR",
          "label": "New tab"
        },
        {
          "layoutId": "layout_Y5rtAAwx",
          "label": "New tab"
        }
      ]
    }
  },
  "defaults": {
    "dataSources": {}
  }
}
```



ONE OF THE OUTPUT

Conclusion

The implementation of this project successfully produced a functional system that meets the defined objectives. All major components were developed, integrated, and tested to ensure proper performance. The system demonstrates good code quality, reliable functionality, and smooth interaction across modules. Overall, this implementation provides a strong foundation that fulfills the project requirements and can be further improved in the future.