**Splunk Architecture**

**What is Splunk**: It is a proprietary software used by companies to collect and analyze data they produce. Also, to create Dashboards and visualizations to analyze metrics

Uses:

1. Monitoring System performance
2. Data informed Decisions
3. Security Cognizance
4. Monitor System health
5. Improve product quality

Graphical user interface, diagram, text, application

Description automatically generated

**Different stages in Data Pipeline:**

1. **Data Input Stage**:
   1. Splunk consumes raw data stream from its sources.
   2. Breaks it in to **64K blocks and annotates with metadata keys**( hostname, source, source type)
   3. **Forwarder**:
      1. Component used to collect logs from other remote machines
      2. We can distribute forwarders into multiple machines independent of main splunk instance.
      3. Splunk forwarders consume very less CPU ~1-2% and we can scale them up to thousands of systems easily and collect terabytes of data with minimal impact on performance.
      4. Two types:
         1. Universal Forwarder
         2. Heavy Forwarder
2. **Data Storage Stage:** It has two Phases:
   1. **Parsing**:
      1. Splunk software examines, analyzes, and transforms the data to extract only the relevant information.
      2. **Event processing**: during this phase splunk breaks the data stream into individual events.
         1. Breaking the stream of data into individual lines
         2. Identifying, parsing (eliminate unwanted data) , and setting **timestamps**
         3. Annotating individual events with metadata
         4. Transform event data according to **Regex** transform rules
   2. **Indexing**: (Benefit of Indexing is to easily be accessed during searching)
      1. Splunk writes parsed events to the index on disk.
      2. It writes both compressed raw data and corresponding index file.
      3. Splunk indexer creates below files
         1. Raw data in compressed form
         2. Indexes that point to raw data and meta data files
         3. These files reside in directories (buckets).
         4. Data replication
3. **Data Searching Stage: (Search Head)**
   1. Component used for interacting with splunk. GUI to perform various operations.
   2. Can Install Search heads on separate servers.
   3. **Search Peer:** Sits on top of indexer to perform actual searches.
   4. Search Head sends query to search peer.
   5. As part of search function, splunk software stores User-created Knowledge objects, such as reports event types, dashboards, alerts, and field extractions

Diagram

Description automatically generated

Graphical user interface, website

Description automatically generated