**Prevent, Mitigate, and Recover (PMR) Insight
Collective Knowledge System (PICK) Tool
Software Design Document**
Version 3.0
08 May 2020

Document Control

Approval

The Guidance Team and the customer shall approve this document.

Document Change Control

Initial Release:	0.1
Current Release:	3.0
Indicator of Last Page in Document:	**$**
Date of Last Review:	07 May 2020
Date of Next Review:	20 May 2020
Target Date for Next Update:	21 May 2020

Distribution List

This following list of people shall receive a copy of this document every time a new version of this document becomes available:

Guidance Team Members:
- Dr. Salamah Salamah
- Dr. Steven Roach
- Elsa Tai Ramirez
- Peter Hanson
- Jake Lasley

Customer:
- Dr. Oscar Perez
- Vincent Fonseca
- Herandy Denisse Vazquez
- Baltazar Santaella
- Florencia Larsen
- Erick De Nava

Software Team Members:
- Eduardo A Jiménez Todd
- Jacob N Torres
- Jorge I Felix
- Alejandro Zamora
- Matthew S Montoya

Change Summary

The following table details changes made between versions of this document

Version	Date	Modifier	Description
0.1	26 February 2020	Matt Montoya	Creation of Document
0.2	06 March 2020	Alex Zamora	Uploaded Database Schema based on ER diagrams by Jorge & Matt
0.3	07 March 2020	Jorge Felix	Finalized Collaboration Diagram with the help of Eddy & Jacob
0.4	08 March 2020	Eddy Todd	Finished Introduction with the help of Jacob & Matt
0.5	09 March 2020	Jacob Torres	Fixed Collaboration Diagram
1.0	09 March 2020	Matt Montoya	Fixed grammar errors, updated DB schema
1.1	22 March 2020	Matt Montoya	Updated *Maltego Class* Contract
1.2	22 March 2020	Matt Montoya	Updated *Log File Class* Contract
1.3	22 March 2020	Matt Montoya	Updated *Significant Log Entry Class* Contract
1.4	23 March 2020	Jacob Torres	Updated *Graph Class* Contract
1.5	24 March 2020	Alex Zamora	Updated *Vector Class* Contract
1.6	24 March 2020	Eddy Todd	Updated *OCR Class* Contract
1.7	24 March 2020	Eddy Todd	Updated *Event Configuration Class* Contract
1.8	24 March 2020	Eddy Todd	Fixed *Significant Log Entry Class* Contract
1.9	24 March 2020	Alex Zamora	Updated *Enforcement Action Report Class* Contract
1.10	24 March 2020	Alex Zamora	Updated Icon Class Contract
1.11	24 March 2020	Jorge Felix	Updated *Directory Class* Contract
1.12	24 March 2020	Jorge Felix	Updated *Search Class* Contract
1.13	25 March 2020	Matt Montoya	Added new introductions to Sections 2, 3, 4; Fixed grammar errors
2.0	30 March 2020	Alex Zamora, Matt Montoya, Jorge Felix, Jacob Torres	Fixed issues based on TA feedback
2.5	01 May 2020	Alex Zamora, Matt Montoya, Jorge Felix, Jacob Torres, Eddy Todd	Added notes based on further TA & Instructor Feedback
3.0	08 May 2020	Alex Zamora, Matt Montoya, Jorge Felix, Jacob Torres, Eddy Todd	Update and finalized SDD for PICK Tool

Table of Contents

1. Introduction

Section 1 shall introduce Team404's Software Design Document (SDD) document for the Spring 2020 Software II Project, *PICK Tool*, including the purpose and intended audience, overview, document references, as well as definitions, acronyms, and abbreviations.

1.1. Purpose and Intended Audience

The purpose of creating the software design document is to describe the Prevent, Mitigate Recover (PMR) Insight Collective Knowledge (PICK) Tool, providing the software development team guidance for the architecture of the project. This SDD details the specifications of the characteristics of the design components. The description of the design is necessary for Team404 to work under a single, guided vision of what PICK Tool shall be. Furthermore, the SDD shall act as a point of reference, outlining all parts of the software and the interactions between them.

1.2. Scope of Product

PICK Tool shall facilitate the job of analysts during an Adversarial Assessment of a simulated cyber-attack, reducing the time it currently takes analysts to perform an assessment to about two weeks. In doing so, PICK Tool shall assist analysts in telling the true story pertaining to these simulated attacks. To help satisfy these needs, analysts will utilize PICK Tool to search through and filter through logs, or recorded notes from the systems of attackers and defenders, as they may pertain to a simulated attack. Though PICK Tool, analysts can use PICK Tool to construct a vector or visual graph of events that satisfy an objective. More information regarding logs can be found in Section 2 of this document.

1.3. References

[1] Wirfs-Brock, R., Wilkerson, B. and Wiener, L. (1990). Designing object-oriented software. Englewood Cliffs, N.J.: Prentice-Hall.

1.4. Definitions, Acronyms, and Abbreviations

1.4.1. Definitions

TERM	DEFINITION
Active Scene	The scene that is currently displayed.
Actor	A user or external system that interacts with the system in the use case diagram.
Adversarial Assessment	Analysis of a simulated cyberattack by the White Team.
Analyst	The Analyst is the primary user of PICK Tool. Multiple Analysts can access PICK Tool simultaneously. The analyst primarily uses PICK Tool to ingest log files, correlate logs, and create graphs within the system.
Associated Log	Logs with a cause and consequence log linked together.
Audio Transcription Tool	A software program that takes audio logs and transcribes them into text files.
Blue Team	Defenders during the simulated cyber-attack; the team that will defend their system from attack.
Client(s)/Customer(s)	The U.S. Army Combat Capabilities Development Command Data & Analysis Center: Lethality, Survivability & Human System Integration (LSH) Directorate; individuals from this directorate include Dr. Oscar Perez, Mr. Vincent Fonseca, Mr. Baltazar Santaella, Ms. Herandy Vazquez, Ms. Florencia Larsen, & Mr. Erick De Nava.
Client-Server Model	PICK Tool running on a server in a closed system.
Commit	A button that confirms the actions of the analyst to save the project.
Correlated Logs	Two log entries connected through cause and effect by the analyst.
Database	A structured set of data held in a computer, especially one that is accessible in various ways.
ETL Tool/Log Management Tool	Software that combines the Extract, Transform, Load database functions into one tool to take data from one or many sources into a destination system.
Event Log	A detailed record of system events stored in the System Event Viewer by the computer's operating system.
Filter	A way for the analyst to search for specific queries through specific conditions set by the analyst.
Filter Space	A certain condition set by the analyst.
Formatted Log Entry	A sanitized log entry that has been cleaned up based on a set configuration by the analyst.
Graph	A visual representation of the scenario between the BlueTeam

	and the Red Team shown through nodes and vectors. [1]
Graphing Tool	A software program that will construct the graph based on the vectors and correlations made by the analyst.
Graphical User Interface	The way that the system will display all components of the system to the user, the analyst.
Ingest/Ingestion	When files, specifically formatted log files, are put into our system to be correlated, edited and disregarded if need be.
Kali Linux	Operating System the clients will use.
Local Network	A data communications network within the system.
Log	An official record of events.
Log Entry	A log entry is the output of the log file. Log entries contain information from the log file and are recorded details of an ingested log.
Log File	A file is an input into the system that records either events that occur in an operating system or other software runs or messages between different users of communication software.
Natural Language Processor	An external software program that takes audio files and translates them into text files to be later validated.
Node	A node is a visual representation of a significant event that was marked for the current vector.
Normalize	To make rid of duplicate entries.
Objective	The goal of the attackers when they attack a system. It is the goal of the defenders to prevent attackers from achieving their goal(s).
Observer	A member of the white team that takes notes and assessments during the staged cyber-attack between the Blue Team and the Red Team.
Optical Character Reader	An external software program that takes image files and translates them into text files to later be validated.
PICK Tool	The software product which Team404 is tasked with constructing for the clients.
Red Team	The title given to the attackers in the simulated cyber-attack; The team that will attack the system the blue team will defend.
Sanitized Logs/Sanitized Log Entries	A log entry that has gone through validation.
Significant Event	Logs that are important to the overall cyber-attack scenario between the Blue Team and the Red Team.
Team404	CS4311 Team 6, the software development team; this includes Mr. Alejandro Zamora, Mr. Eduardo Jiménez Todd, Mr. Jacob Torres, Mr. Jorge Felix, and Mr. Matt Montoya.
Timeline	A set of events, logs, that help convey the overall story of the events that took place between the Blue Team and the Red Team.
Timestamp	The time that a log was recorded in the system during the

	simulated cyber-attack.
User/Users	A person that interacts with the system; in PICK, the user will be an *Analyst*
Vector	The series of activities/steps an adversary executes or attempts to execute that is necessary to achieve an objective.
White Team	The title given to the observers during the simulated cyber-attack. They observe what happens between the Blue Team and the Red Team on the system.
Wildcard	Unsearchable characters, and any characters that can be searched through.
Zulu Time	The military and navigation parlance for the UTC time standard.

1.4.2. Acronyms

TERM	DEFINITION
AA	Adversarial Assessment
AKA	As Known As
CSV	Comma Separated Value.
DOD	Department of Defense.
ETL	Extract Transform Load.
GUI	Graphical User Interface
IEEE	Institute of Electrical and Electronics Engineers
JPG/JPEG	Joint Photographic Experts Group.
NLP	Natural Language Processor.
OCR	Optical Character Reader.
OS	Operating System
PDF	Portable Document Format.
PICK	PMR Insight Collective Knowledge.
PMR	Prevent, Mitigate Recover.
SDD	Software Design Document

1.4.3. Abbreviations

Admin	Administrator

1.5. Overview

1.5.1. Introduction

Section 1 shall introduce Team404's Software Design Document (SDD) document for the Spring 2020 Software II Project, *PICK Tool*, including the purpose and intended audience, overview, document references, as well as definitions, acronyms, and abbreviations.

1.5.2. Decomposition Description

Section 2 (Decomposition Description) introduces the System Collaboration Diagram (SCD), the Subsystem and Component Descriptions (CRC Cards), as well as the dependencies of the system, defining the responsibilities entities have for specific functions, and tracing system requirements to design entities.

1.5.3. Detailed Description of Components

Section 3 (Detailed Description of Components) provides complete, detailed descriptions of all components listed within the System Collaboration Diagram (Section 2.1 of the SDD). These components include external as internal components, as they relate to PICK Tool.

1.5.4. Database

Section 4 introduces Team404's database (DB) proposal for PICK Tool. During a previous demo session, the clients noted they were unsure if a database was needed. This section serves as Team404's proposed database implementation, using knowledge from UTEP CS 4342 (Database Management) course. This proposal includes the Entity-Relationship (ER) Diagram and database schema that represent the graphical representation of entities and logical configuration of the relational database, respectively.

2. Decomposition Description

Section 2 introduces the System Collaboration Diagram (SCD), the Subsystem and Component Descriptions (CRC Cards), as well as the dependencies of the system, defining the responsibilities entities have for specific functions, and tracing system requirements to design entities.

2.1. System Collaboration Diagram
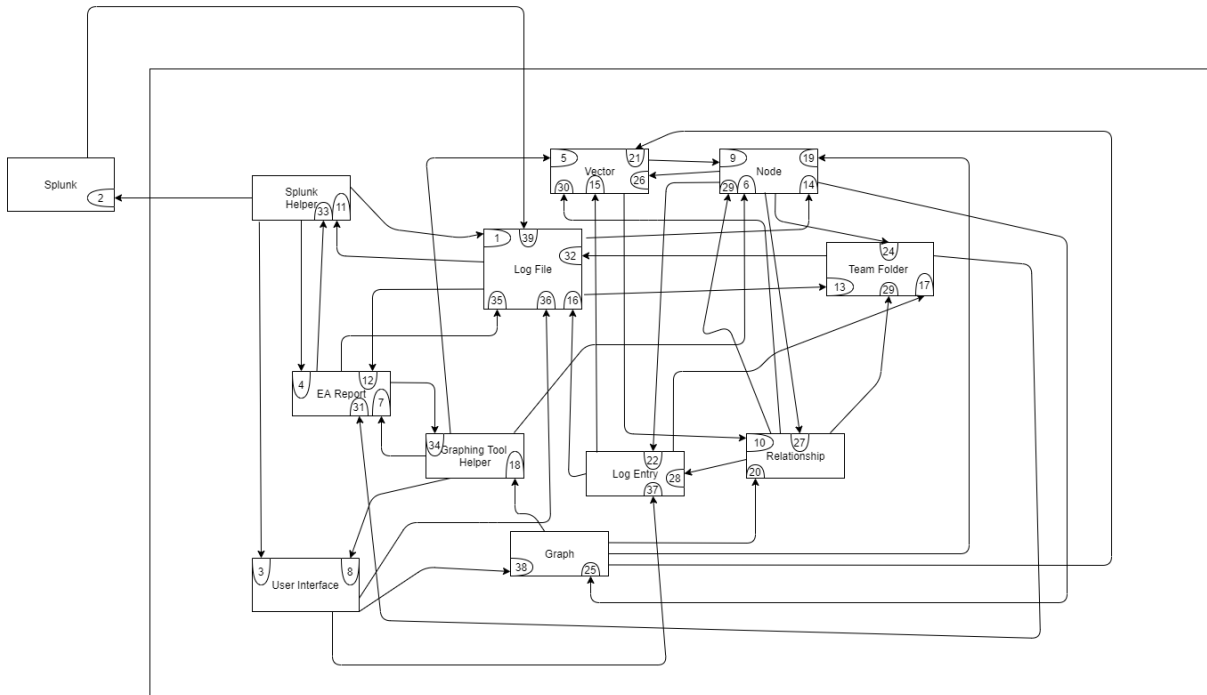
2.2. Subsystem and Component Descriptions

Splunk Interface - Communicates with Splunk

Graphing Tool Interface - third party software tool for graphing that communicates with graph

Vector - A description of a significant event communicates with nodes and relationships

Log File - Computer data object that stores logs

Log Entry - Logs that are important to the system to the overall scenario

Graph - Visual representation of a vector communicates with graphing tool interface

Node - Visual representation of a significant event communicates with log file

Relationship - stores details on relations between nodes

Filter - Allows the user to search for specific criteria

Change - Path and structure of the log files

Event Action (EA) Report - Handles errors

Splunk - Handles log ingestions and cleansing

OCR - recognize characters in images
User Interface - what the user interacts with.

2.3. Dependencies

Dependencies are external elements to the program. These dependencies are required by PICK Tool in order to work properly.

2.3.1. Dependency List

- Kali Linux OS
- Python 3 (version 3.7.5)
- Splunk Enterprise
- Splunk Python SDK (version 1.6.12)
- PyQt5 (version 5.14.2)
- QGraphViz (version 0.0.50)

2.3.2. Dependency Descriptions

Kali Linux is a Debian GNU / Linux based distribution designed primarily for general computer security and auditing.

Python is an interpreted programming language whose philosophy emphasizes the readability of its code. It is an interpreted, dynamic and cross platform language.

Splunk is a software to search, monitor and analyze big data generated through a web interface. *Splunk Enterprise* and the *Splunk Python SDK* enable for the development of Splunk-based applications.

PyQt is a binding of the Qt graphics library for the Python programming language.

QGraphViz is a PyQt5 widget that allows for the development of GraphViz graphs in python.

2.3.3. Dependency Installation Instructions

1. PICK Tool is built to run on **Kali Linux**. The Kali Linux ISO can be download directly from the Kali Linux website (https://www.kali.org/).

2. PICK Tool is built using **Python 3.7.5** and requires this programming language to run. Python 3.7.5 can be downloaded from the python website (https://www.python.org/).

3. PICK Tool depends on **Splunk Enterprise** and requires dependency, which can be downloaded from the Splunk website (https://www.splunk.com/en_us/download/splunk-enterprise.html).

   a. Users will need the following information to login to and Download Splunk Enterprise (note: This information expires July 1, 2020):

      i. Username: jntorres217

      ii. Password: Jaconian@1520

4. PICK Tool depends on the **Splunk Python SDK** and requires this dependency

   a. Most Linux users can install the SDK using one of the following commands:

      i. pip install splunk-sdk==1.6.12

      ii. pip3 install splunk-sdk==1.6.12

   b. Debian Users may need to clone the Github repo (https://github.com/splunk/splunk-sdk-python) and follow the instructions listed therein.

5. PICK Tool depends on **PyQt5** and requires this dependency

a. Most Linux users can install PyQt5 using the following command:

   i. pip3 install PyQt5==5.14.1

b. Debian Users may need to use direct download links from PyPi (https://pypi.org/project/PyQt5/).

6. PICK Tool depends on the **QGraphViz** and requires this dependency

   a. Most Linux users can install PyQt5 using the following command:

      i. pip3 install QGraphViz==0.0.50

   b. Debian Users may need to use other download links from PyPi (https://pypi.org/project/QGraphViz/).

3. Detailed Description of Components

Section 3 provides complete, detailed descriptions of all components listed within the System Collaboration Diagram (Section 2.1 of the SDD). These components include external components such as the Database and Splunk, as well as internal components including Splunk.

3.1. Splunk Helper

Class: Splunk Helper
Superclass: None
Subclasses: None
Description: A third party software tool for transforming log files into normalized log entries.
(1) Ingestion Contract

Responsibilities:	Collaborations:
1. Can ingest log files.	1Log Files (1)Splunk (2)
2. Can transform a log file into normalized log entries.	2Log Files (1)Splunk (2)
3. Can store log entries in the normalized data files.	3Log Files (1)Splunk (2)
4. Can process incoming data into individual activities according to the nature of the data.	4Log Files (1)Splunk (2)
5. Can export log entries	5Log Files (1)Splunk (2)
6. Displays known information to the user	6 User Interface(3)Log Files(1)Splunk(2)EA Report(4)

3.2. Graphing Tool Helper

Class: Graphing Tool Helper	
Superclass: None	
Subclasses: None	
Description: A tool for creating graphs.	
(3) Create Graph	
Responsibilities: 1. Knows export format. 2. Can create graphs with nodes and relationships between nodes. 3. Can export graph. 4. Can display Known information to the user	**Collaborations:** 1 2vector(5) Node(6) 3EA Report(7) 4User Interface(8)vector(5) Node(6)EA Report(7)

3.3. Vector

Class: Vector	
Superclass: None	
Subclasses: None	
Description: A description of a significant event.	
(4) Vector Elements	
Responsibilities: 1. Knows Vector Name 2. Knows Vector Description 3. Knows Vector Entries	**Collaborations:** 1Nodes (9) 2Nodes (9)Relationship (10) 3Nodes (9)

3.4. Log File

Class: Log File

Superclass: None

Subclasses: None

Description: A computer data object that stores logs.

(5) Splunk Integration

Responsibilities:	Collaborations:
1Stores Log File Name 2Stores Cleansing Status 3Stores Validation Status 4Stores Ingestion Status 5Stores Acknowledgment, as well Status 6Stores Can export log files	1 2Splunk Helper(11) 3Splunk Helper(11) 4Splunk Helper(11) 5EA Report(12) 6Team Folder(13)

(7) Node Integration

Responsibilities:	Collaborations:
1. Can contain a timestamp per line, bounded by the start date, end date, start time, and end time specified in the event configuration.	1Node(14)

(9) EA Report Integration

Responsibilities:	Collaborations:
1. When a timestamp property of a previously saved event is changed, the impact of the change shall be restricted to the "not-validated" log files.	1EA Report(12)

3.5. Log Entry

Class: Log Entry	
Superclass: Log Entry	
Subclasses: None	
Description: Logs that are important to the overall cyber-attack scenario between the Blue Team and the Red Team.	
(10) Relational	
Responsibilities: 1Knows relevant information for a vector.	**Collaborations:** 1Vector(15)
(11) Filter Log Entry	
Responsibilities: 1Knows Log Entry Number 2Knows Log Entry Timestamp 3Knows Log Entry Content 4Knows Host 5Knows Source 6Knows Source Type	**Collaborations:** 1Log File(16) 2Log File(16) 3Log File(16) 4Team Folder(17) 5Team Folder(17) 6Team Folder(17)

3.6. Graph

Class: Graph
Superclass: None.
Subclasses: None.
Description: A visual representation of a vector.
(10) Graph Action Report

Responsibilities:	Collaborations:
1. A graph shall comprise at least one node.	1graphing tool Helper(18)
2. Can allow the analyst to add nodes, edit nodes, delete nodes, add relationships, edit relationships, and delete relationships.	2graphing tool Helper(18)Nodes(19) Relationship(20)
3. Can create a PNG of the graph.	3graphing tool Helper(18)
4. Knows how to export format:	4
5. Knows orientation	5graphing tool Helper(18)
6. Knows Vector Name	6Vector(21)
7. internal units	7
8. Knows interval	8
9. Knows position of nodes	9graphing tool Helper(18)
10. Knows position of relationships	10graphing tool Helper(18)

3.7. Node

Class: Node	
Superclass: None	
Subclasses: None	
Description: Visual representation of a significant event that was marked for the current vector.	
(11) Node Elements	

Responsibilities:	Collaborations:
1 Can represent a log entry as part of the graph	1 Log entry(22)
2 Knows Node ID	2
3 Knows Node Name	3
4 Knows Node Timestamp	4
5 Knows Node Description	5
6 Knows Log Entry Reference	6 log entry(22)
7 Knows Log Creator	7 log entry(22)
8 Knows Event Type	8
9 Knows Icon Type	9
10 Knows Source	10 Team Folder(24)
11 Knows Node Visibility	11

3.8. Relationship

Class: Relationship	
Superclass: None	
Subclasses: None	
Description: Represent the relationship between nodes.	
(18) Vector Relationships	
Responsibilities: 1. Knows the relationship ID 2. Knows the parent ID 3. Knows the child ID 4. Knows the label 5. Knows source of logs	**Collaborations:** 1 2 3 4 5log entry(28)Team Folder(29)
(19) Event Parameters	
Responsibilities: 1. Can correlate nodes to one another.	**Collaborations:** 1Node(30)

3.9. Team Folder

Class: Team Folder	
Superclass: None	
Subclasses: None	
Description: Path and structure of the log files.	
(23) Keeps Track of	
Responsibilities: 1Can perform structure check when data ingestion starts . 2Can generate root directory structure error if it fails to contain three folders or if the folder names specified in the event configuration don't match. 3Can store the log files in their corresponding folder. 4Knows Red Team Folder 5Knows Blue Team Folder 6Knows White Team Folder	**Collaborations:** 1EA Report(31) 2EA Report(31) 3Log Files(32)EA Report(31) 4 5 6

3.10. Event Action (EA) Report

Class: Enforcement Action Report	
Superclass: None	
Subclasses: None	
Description: Handles errors when anomalous events happen.	
(24) EA Elements	
Responsibilities: 1Can track error codes 2Can display error code 3Can track description of error codes 4Can track the state of the system 5Knows the Line Number 6Knows the Error Message	**Collaborations:** 1 2Splunk Helper(33) Graphing Tool Helper(34) 3 4Log file(35) 5 6

3.11. Splunk

Class: Splunk	
Superclass: None	
Subclasses: None	
Description: Handles log ingestion and cleansing.	
(25) log handling	
Responsibilities: 1. handles log cleansing 2. handles log ingestion	**Collaborations:** 1Log File(39) 2Log File(39)

### 3.12. User Interface

Class: User Interface	
Superclass: None	
Subclasses: None	
Description: The view that the user sees and interacts with.	
(27) display	
Responsibilities: 1. display current window user is interacting with 2. keep track of parallel tasks	**Collaborations:** 1log file(36)log entry(37)graph(38) 2

4.   Database

Section 4 introduces Team404's database (DB) proposal for PICK Tool. During a previous demo session, the clients noted they were unsure if a database was needed. This section serves as Team404's proposed database implementation, using knowledge from UTEP CS 4342 (Database Management) course. This proposal includes the Entity-Relationship (ER) Diagram and database schema that represent the graphical representation of entities and logical configuration of the relational database, respectively.
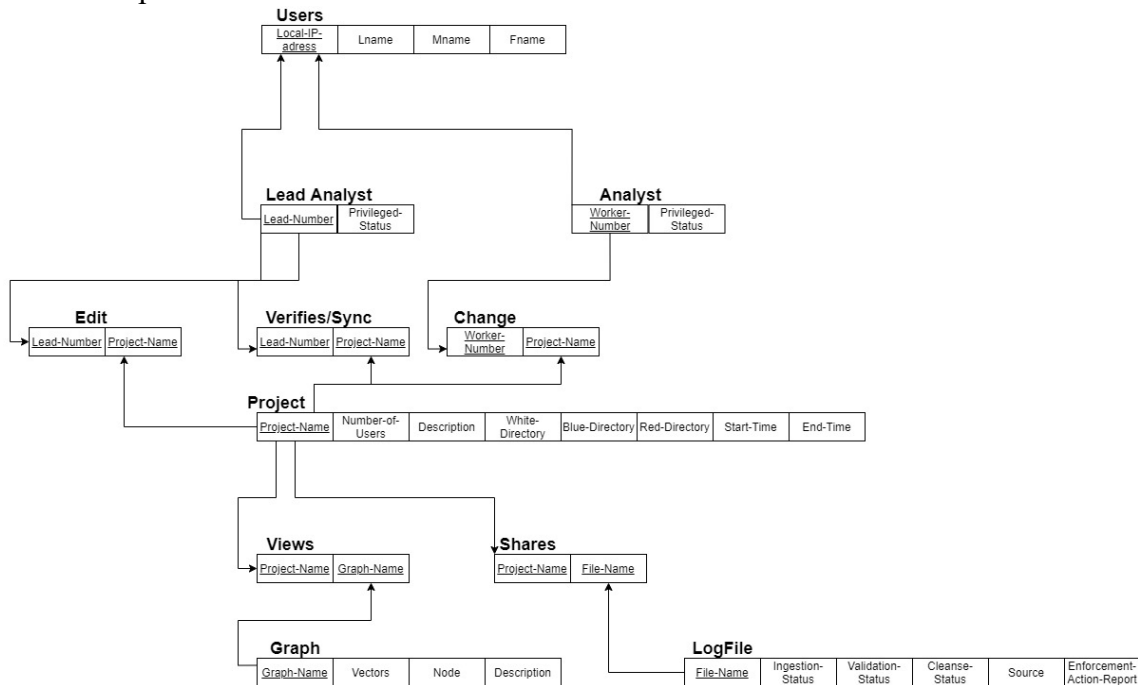
4.1. Entity-Relationship Diagram

The ER diagram, shown below, is a conceptual design of the relationship between entities, or things in the real-world as they pertain to PICK Tool. These entities are denoted by the rectangular shapes in the model. These entities may contain attributes that describe the entities, as well as relationships between entities. Attributes are housed within the oval shapes, where relationships are housed within the diamond shapes; both of which are connected to the entities.
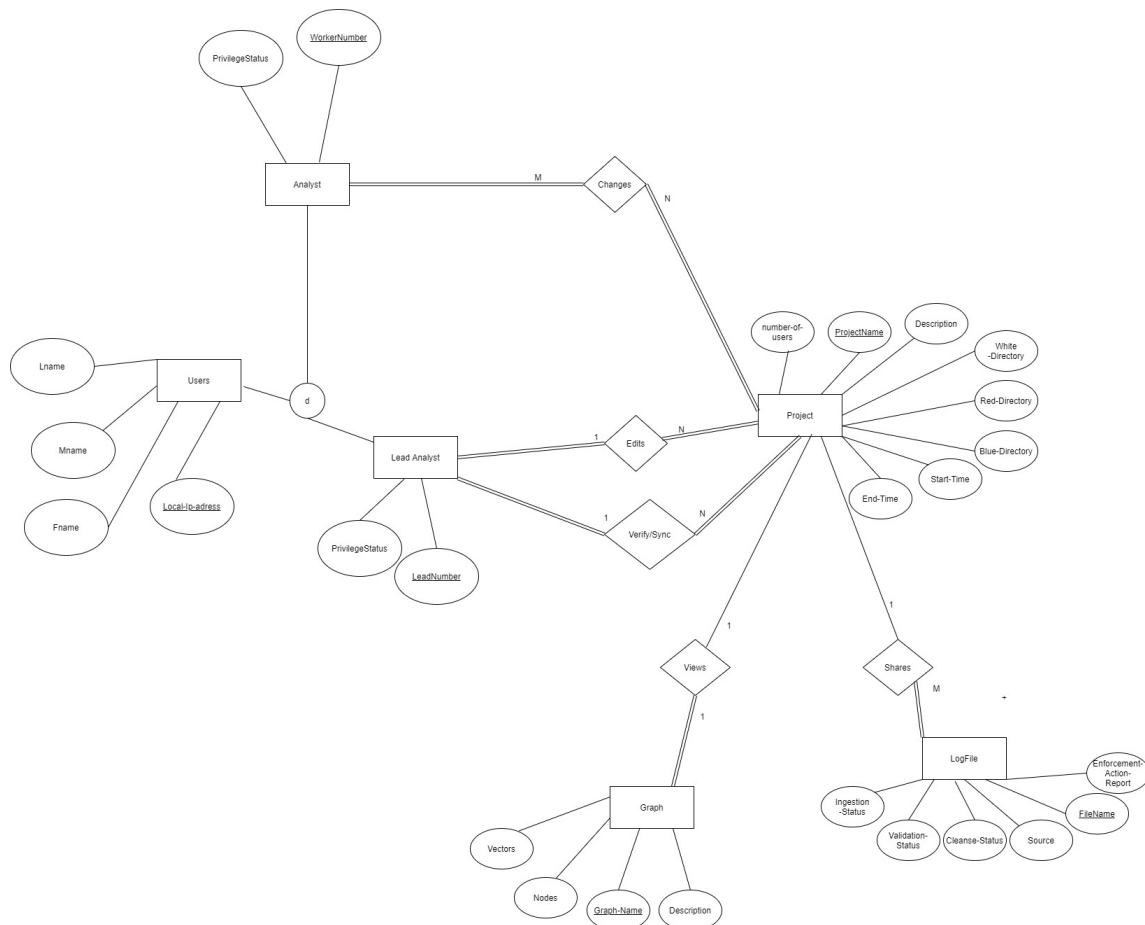
4.2. Database Schema

The Database Schema (or relational model) shown below, is another visual representation of the database, depicting the logical configuration of the database, including the set of rules that govern as they pertain to PICK Tool. This schema serves as a description of the database itself and is derived from the ER Diagram in Section 4.1. The conversion of high-level (ER diagram) to logical design (DB Schema) follows a seven-step algorithm: Mapping regular entities to relations, mapping weak entities to relations, identifying 1:1 binary relationship types, identifying binary 1:N relationship types, identifying M:N relationship types, separating multi-valued attributes, and creating new relations for N-ary relationships.

$