**DATA MODELING:**

Given the sample data in file, show what tables you would create in a new data warehouse off that source. Additionally, state what other data would be helpful to have, and any assumptions made.

**File: DataModelWideTakeHome.csv**

Based on the sample data provided, we would create **four** tables in a new data warehouse**:**

**TABLE 1: Events Table:**

**Columns:**

**EVENTUHUB\_ID** (Primary Key) - uniquely identifies each event.

**LOGGED** - timestamp when the event was logged. (date is in the format of YYYY/MM/DD)

**DEVICEID** (Foreign key)– linking to the devices table.

**DEVICETYPE** – type of the device generating the event.

**Table 2: Devices Table:**

**Columns:**

**DEVICEID** (Primary Key) - uniquely identifies each device.

**DEVICETYPE** - describes the type of device.

**COMPANY\_ID** (Foreign Key) - links devices to the companies they belong to.

**TABLE 3: Alarms Table:**

**Columns:**

**EVENTUHUB\_ID** (Foreign Key) - serves as a foreign key linking events to alarms.

**ALARMS** - contains information about the alarms associated with each event. (date is in the format of YYYY/MM/DD)

**ALARMID (**Primary key**)** – (assuming it would be useful if would generate or have alarmID)

**TABLE 4: Companies Table:**

**Columns:**

**COMPANY\_ID** (Primary Key) - uniquely identifies each company.

**EVENTUHUB\_ID** (Foreign Key) - serves as a foreign key linking events to alarms.

**COMPANY\_NAME** (assuming it would be useful to store the names of the companies)

**Table 5: Sensor data Table**

**Columns:**

**EVENTUHUB\_ID**, **DEVICEID** (Primary Key) (here both serves as a composite key to uniquely identify events)

**P1, P2, P3, P4, P5, P6, P7, P8, P9, P10** (assuming these are different parameters or attributes related to the event)

**Other helpful data and assumptions:**

1. **Device Attributes**: Including additional attributes related to devices such as location, model, manufacturer, and operational status can provide more context and facilitate deeper analysis of events and alarms.
2. **Event Metadata**: Storing metadata associated with events such as event severity, event type, and event category can help classify and prioritize events effectively.
3. **Alarm History**: Maintaining a historical record of alarms, including timestamps, severity levels, and actions taken, can support trend analysis and performance monitoring.
4. **Company data**: More data about the company like company name, domain name etc.
5. **Users Table**: If there are users associated with events, storing user information like user ID, username, email, etc., could be beneficial.
6. **Locations Table**: If events are associated with specific locations, storing location-related information such as latitude, longitude, address, etc., could be useful.

Please find the sample ER diagram below,

A diagram of a computer

Description automatically generated

**Notes: (Based on my assumptions from the data, but it may change based on the business logic)**

1. **EventHub Data to Sensor Data**: One-to-One Relationship (based on assumption – one set of sensor data)
2. **Device and EventHub**: One-to-Many Relationship (one device can be registered in different hub ids)
3. **Devices and Company Data:** One-to-Many (a device can be registered in multiple companies)
4. **EventHub and Alarm:** One-to-One (one event hub can have one alarm registered based on the data)
5. **Device and Sensor**: One-to-Many (a device can generate different set of sensor data)