# Reinventing ETL for Detection and Response Teams

BSides SF, May 2024, Josh Liburdi

## Who, Me?

- 10+ years of industry experience<sup>1</sup>
  - Staff Security Engineer at Brex<sup>2</sup>
  - Previously: Splunk, Target, CrowdStrike, GE
- Working on security data solutions<sup>3</sup> for several years

<sup>&</sup>lt;sup>1</sup>Detecting, hunting, responding, consulting, engineering, architecting, and much more that I'd rather forget about.

<sup>&</sup>lt;sup>2</sup> Warning: Opinions are mine and not representative of my employer or colleagues.

<sup>&</sup>lt;sup>3</sup> Most closed source, some open source. See my talk from BSides SF 2019 for an example.

## WTF is ETL?<sup>4</sup>

<sup>&</sup>lt;sup>4</sup>Don't worry, your SIEM vendor probably manages this for you. You should be worried if your SIEM vendor manages this for you.

#### ETL in Brief

#### **Extract**

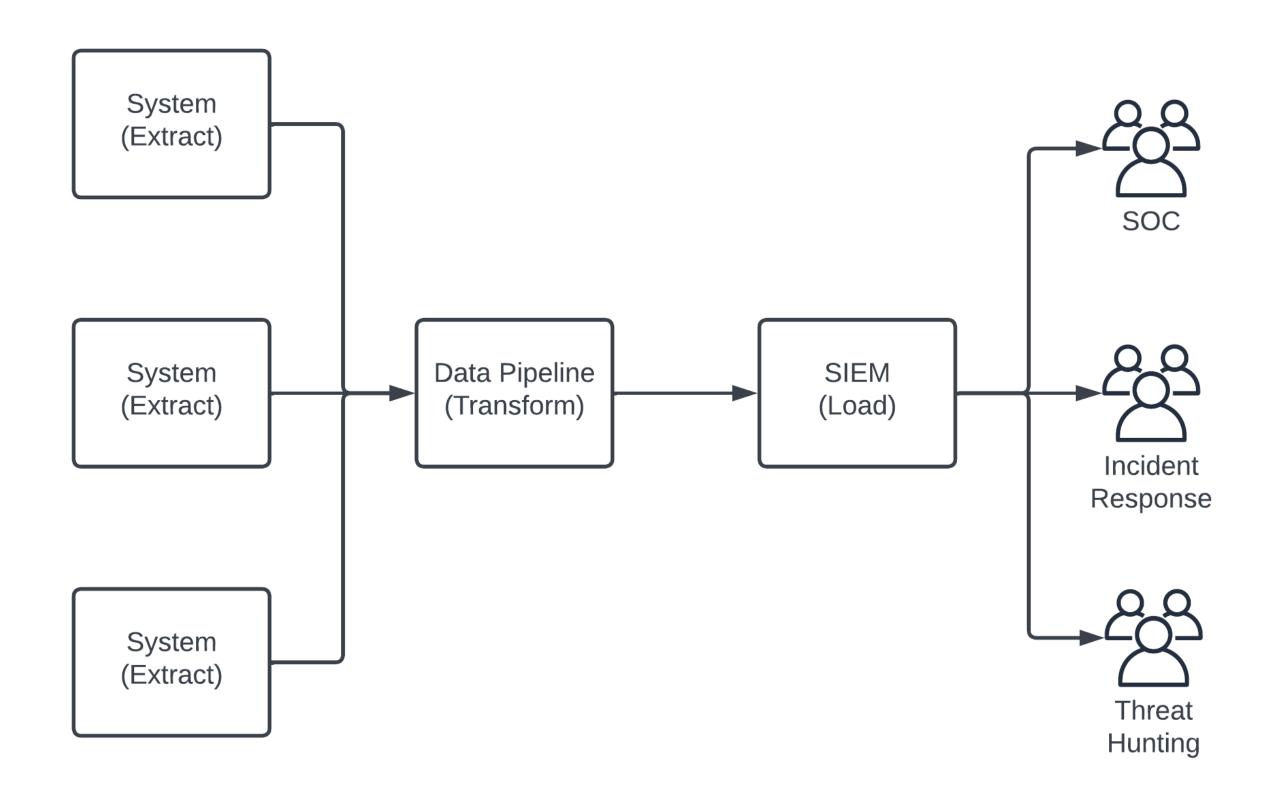
Pull, receive, sync, read, or capture data

#### Transform

· Filter, normalize, or enrich data

#### Load

Store, visualize, or alert on data



## ETL Turns Data into Information

## The Current State of Security ETL

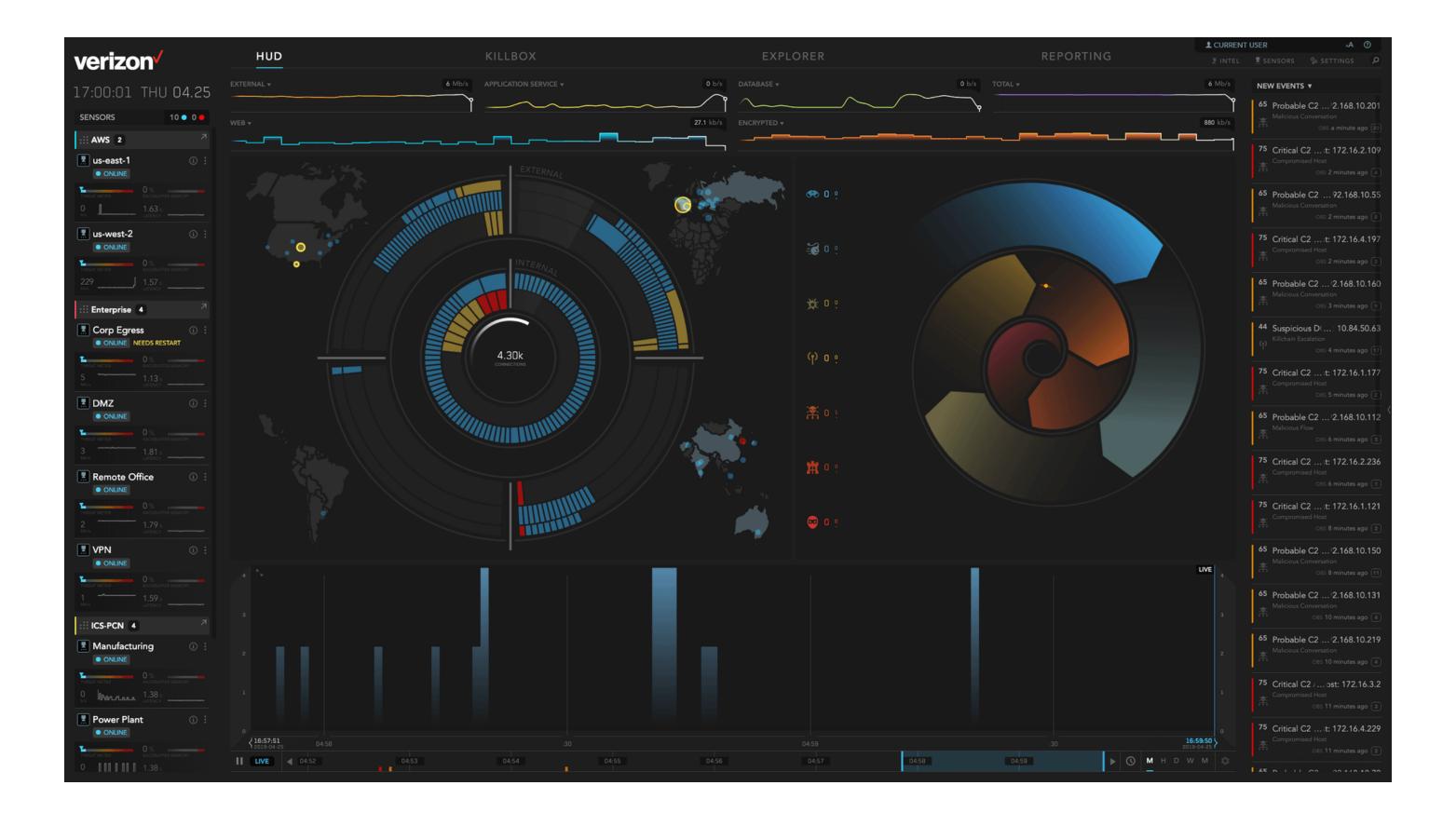


```
"LocalAddressIP4": "0.0.0.0",
"ContextBaseFileName": "Spotify",
"event_simpleName": "NetworkConnectIP4",
"ContextTimeStamp": "1714267028.786",
"ConfigStateHash": "2248159763",
"ConnectionFlags": "0",
"ContextProcessId": "339011099229862299",
"RemotePort": "443",
"aip": "[REDACTED]",
"ConfigBuild": "1007.4.0018305.1",
"event_platform": "Mac",
"LocalPort": "0",
"Entitlements": "15",
"name": "NetworkConnectIP4MacV13",
"EventOrigin": "1",
"id": "fb542841-cbea-4323-b62e-1e0b2ac90d07",
"Protocol": "6",
"EffectiveTransmissionClass": "3",
"aid": "[REDACTED]",
"RemoteAddressIP4": "35.186.224.39",
"ConnectionDirection": "0",
"InContext": "0",
"timestamp": "1714268077295",
"cid": "[REDACTED]"
```

```
"event": { "category": "network", "type": "connection" },
"host": {
 "id": "[REDACTED]",
 "name": "[REDACTED]",
  "public_ip": "[REDACTED]"
},
"process": {
  "command_line": "/Applications/Spotify.app/Contents/MacOS/Spotify",
 "name": "Spotify",
  "pid": "339011099229862299",
 "start": "2024-03-06T18:46:21.000000Z"
},
"network": { "direction": "outbound", "transport": "tcp" },
"server": { "ip": "35.186.224.39", "port": 443 },
"user": {
 "email": "[REDACTED]",
 "roles": ["Sr. Analyst, [REDACTED]"],
 "status": ["[REDACTED]_active"]
"@timestamp": "2024-04-28T01:47:50.039000Z"
```

```
WITH proc AS (
    SELECT TargetProcessId, CommandLine FROM edr WHERE type = 'process_start'
),
host AS (
    SELECT Id, ComputerName FROM edr WHERE type = 'host_online'
),
dvc AS (
    SELECT HostName, UserEmail FROM dvc WHERE type = 'device_checkin'
),
idp AS (
    SELECT UserEmail, Department, Title FROM auth WHERE type = 'user_login'
SELECT
    CONCAT(idp.Title, ', ', idp.Department) AS UserRoles,
    COUNT(*) AS NumberOfConnections
FROM
    edr net
JOIN
    proc ON net.ContextProcessId = proc.TargetProcessId
JOIN
    host ON net.Id = host.Id
JOIN
    dvc ON host.ComputerName = dvc.HostName
JOIN
    idp ON dvc.UserEmail = idp.UserEmail
WHERE
    proc.CommandLine LIKE '%/Spotify.app/%'
    AND net.ConnectionDirection = 0
    AND NOT RLIKE(net.RemoteAddressIP4, '^(10\\.|172\\.(1[6-9]|2[0-9]|3[0-1])\\.|192\\.168\\.).*')
    AND net.RemoteAddressIP4 != '0.0.0.0'
    AND dvc.UserEmail IS NOT NULL
    AND idp.Department IS NOT NULL
    AND idp.Title IS NOT NULL
GROUP BY
   idp.Title, idp.Department;
```

```
SELECT
    COUNT(*) AS NumberOfConnections
FROM
    events udm
WHERE
    udm.ProcessCommandLine LIKE '%/Spotify.app/%'
    AND udm. NetworkDirection = 'outbound'
GROUP BY
    udm.UserRoles;
```



## The Problem with Security ETL

- Security and audit logs are diverse and have no standards
- SIEM (et al) are building blocks, not solutions
  - Good for search, OK<sup>5</sup> for storage, bad for analysis
- Increases cognitive load on practitioners
  - Experience and fatigue can lead to inaccurate conclusions

<sup>&</sup>lt;sup>5</sup> Ish. Depends on how deep your CFO's pockets are.

# Inventing ETL for Detection and Response Teams

BSides SF, May 2024, Josh Liburdi

## Tenets of Security ETL

- 1. Data is available when it's needed
- 2. Data is easy to understand at a glance
- 3. Data is contextualized and actionable

## Tenets of Security ETL (Opinion<sup>6</sup> Ed.)

- 1. Data is available when it's needed
  - Federated SIEM is a Band-Aid<sup>™</sup> on a third-degree burn
- 2. Data is easy to understand at a glance
  - Unified data models are the standard, not nice to have
- 3. Data is contextualized and actionable
  - Deriving insights from data should be effortless

<sup>&</sup>lt;sup>6</sup> Again: Opinions are mine and not representative of my employer or colleagues.

# It would be nice if my data was better, but is this really a problem?

You, right now (probably)

## Meet Your New Friend, Data Decay!

...when the data in your database becomes outdated or incorrect due to the timesensitive nature of the data.

- 6sense<sup>7</sup>

...the rate of data decay amplifies as already degraded data is being input through disparate processes without governance or attention to detail.

Leadspace<sup>8</sup>

<sup>&</sup>lt;sup>7</sup>https://6sense.com/blog/data-decay/

<sup>&</sup>lt;sup>8</sup> https://www.leadspace.com/blog/data-decay-what-why-and-how/

## Decay? In My Data?

- Geolocation: 11% of IPs change their city in a week<sup>9</sup>
- Tor: 33% of routers up for less than a week<sup>10</sup>
- Proxies: ~10% daily churn in the largest provider networks<sup>11</sup>
  - Luminati Res. Proxy: 8,100,000+ active IPs, 11% daily churn
  - OxyLabs Proxy: 5,100,000+ active IPs, 9% daily churn

<sup>&</sup>lt;sup>9</sup> https://ipinfo.io/blog/how-many-ips-change-geolocation-over-a-year/

<sup>&</sup>lt;sup>10</sup> https://torstatus.rueckgr.at/index.php?SR=Uptime&SO=Desc

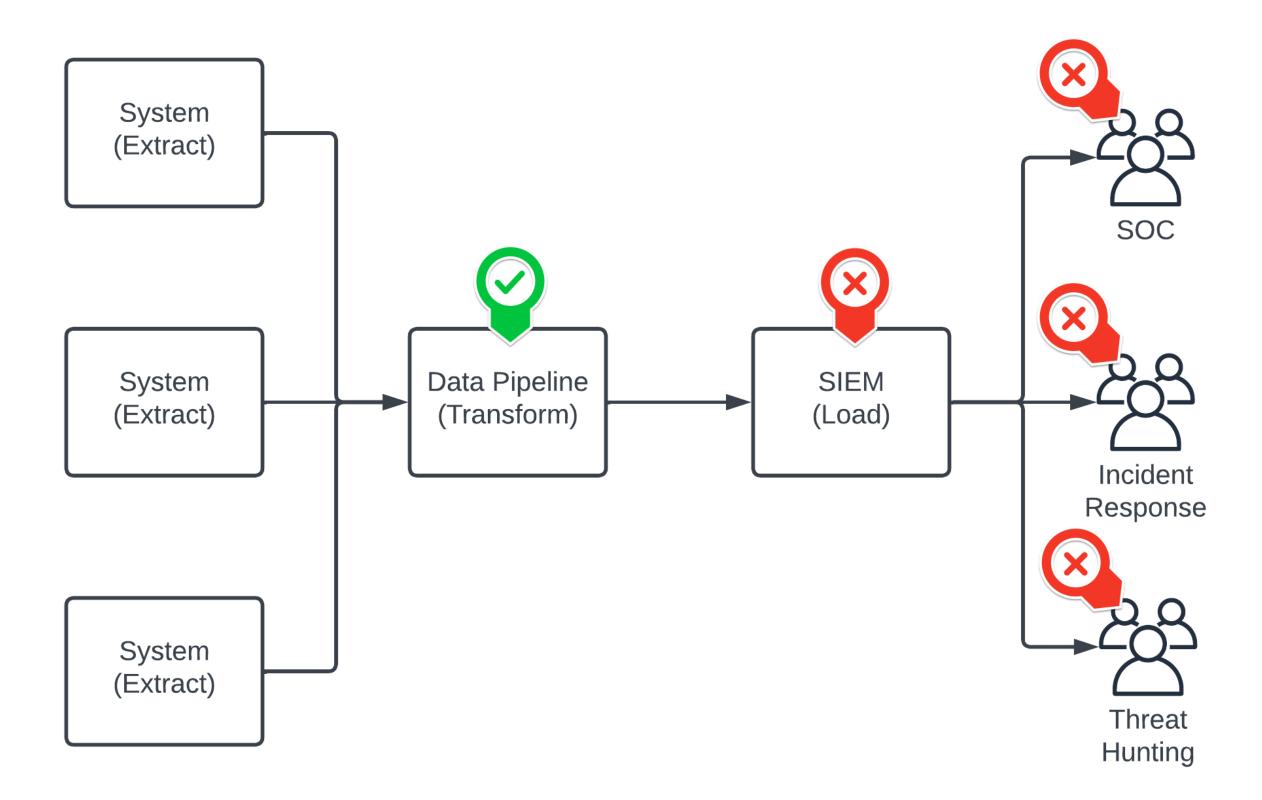
<sup>11</sup> https://spur.us/

## But Wait, There's More<sup>12</sup>...

- BGP Routing / ASN
- DNS Records
- WHOIS
- Domain Rank
- URL Reputation
- Open Services / Ports

- Cloud Resources
- User Groups / Roles
- Anti-Virus Results
- File Integrity
- Vulnerabilities
- Threat Intelligence

<sup>&</sup>lt;sup>12</sup> And more, but I ran out of space on this slide.



### Substation<sup>13</sup> from Brex

- Security analytics and data pipeline toolkit for the cloud
- Open source for 2+ years, used in production (AWS) for 3+ years
  - Billions of events and terabytes of data processed each day
  - Less than 1 hour of maintenance each week<sup>14</sup>
  - Costs a few cents per GB<sup>15</sup> of data processed

<sup>&</sup>lt;sup>13</sup> https://github.com/brexhq/substation

<sup>&</sup>lt;sup>14</sup> Usually it's zero, but YMMV.

<sup>&</sup>lt;sup>15</sup>This is *all AWS spend*.

## Substation Use Cases<sup>16</sup> and Examples

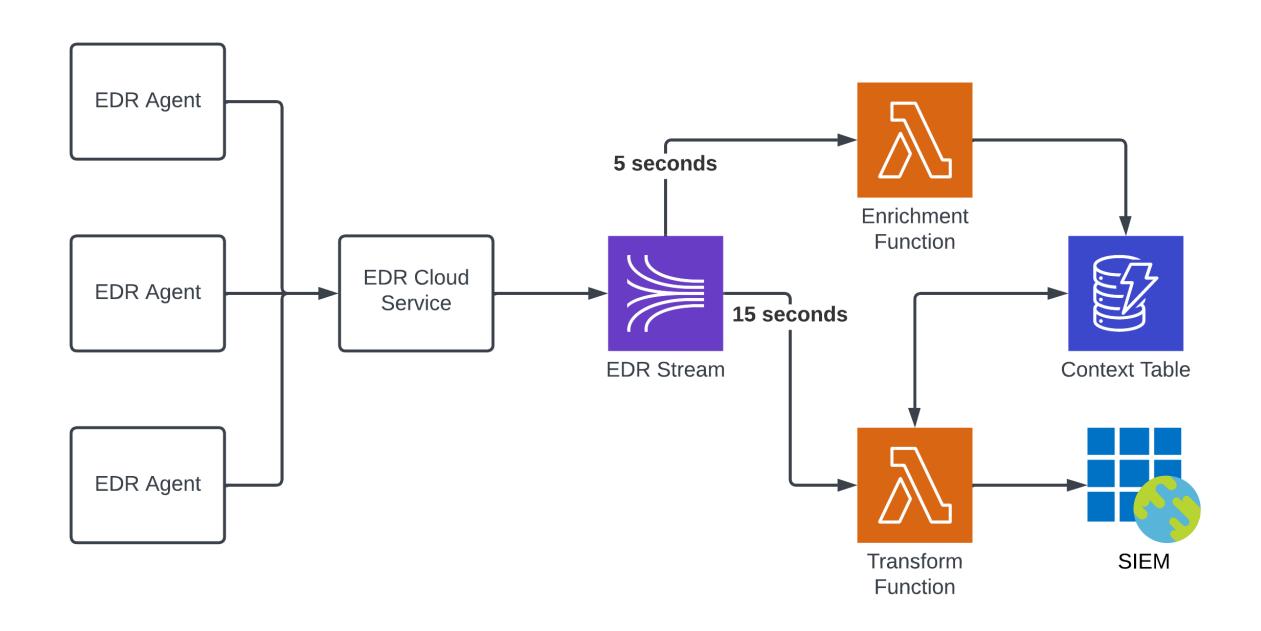
- Route data to / from almost anywhere (cloud & on-prem)
- Normalize data to any schema, open or proprietary
- Enrich data with asset, identity, and threat context

- S3, Kinesis, SQS, SIEM,
   HTTP, local files, and more
- Model event data to OCSF with optional validation
- Static lookups, dynamic lookups, real-time lookups

<sup>&</sup>lt;sup>16</sup> These capabilities are table stakes for modern detection and response teams.

## Let's Look at Some Foundational ETL Solutions You Won't See at RSA

## Introducing... the Time Travel Pattern!



```
"event": { "category": "network", "type": "connection" },
"process": {
    "name": "Spotify",
    "pid": "339011099229862299"
},
"server": { "ip": "35.186.224.39", "port": 443 },
"@timestamp":"2024-04-28T01:47:50.039000Z"
}
```

```
"event": { "category": "network", "type": "connection" },
"process": {
  "command_line": "/Applications/Spotify.app/Contents/MacOS/Spotify",
 "name": "Spotify",
 "pid": "339011099229862299",
 "start": "2024-03-06T18:46:21.000000Z",
 "parent": {
   "command_line": "/usr/libexec/runningboardd",
   "name": "runningboardd",
   "pid": "338971324198273501",
   "start": "2024-03-06T18:17:49.000000Z",
   "parent": {
     "command_line": "/sbin/launchd",
     "name": "launchd",
      "pid": "338053280202314993",
      "start": "2024-03-06T18:17:45.000000Z"
"server": { "ip": "35.186.224.39", "port": 443 },
"@timestamp": "2024-04-28T01:47:50.039000Z"
```

```
{
   "event": { "category": "network", "type": "connection" },
   "process": {
        "name": "Spotify",
        "pid": "339011099229862299"
   },
   "server": { "ip": "35.186.224.39", "port": 443 },
   "@timestamp":"2024-04-28T01:47:50.039000Z"
}
```

```
"event": { "category": "network", "type": "connection" },
"process": {
 "command_line": "/Applications/Spotify.app/Contents/MacOS/Spotify",
 "name": "Spotify",
 "pid": "339011099229862299",
 "start": "2024-03-06T18:46:21.000000Z",
  "parent": {
   "command_line": "/usr/libexec/runningboardd",
   "name": "runningboardd",
   "pid": "338971324198273501",
    "start": "2024-03-06T18:17:49.000000Z",
    "parent": {
     "command_line": "/sbin/launchd",
     "name": "launchd",
     "pid": "338053280202314993",
      "start": "2024-03-06T18:17:45.000000Z"
"server": { "ip": "35.186.224.39", "port": 443 },
"@timestamp": "2024-04-28T01:47:50.039000Z"
```

```
"event": { "category": "network", "type": "connection" },
"process": {
    "name": "Spotify",
    "pid": "339011099229862299"
},
"server": { "ip": "35.186.224.39", "port": 443 },
"@timestamp":"2024-04-28T01:47:50.039000Z"
}
```

```
"event": { "category": "network", "type": "connection" },
"process": {
  "command_line": "/Applications/Spotify.app/Contents/MacOS/Spotify",
 "name": "Spotify",
 "pid": "339011099229862299",
 "start": "2024-03-06T18:46:21.000000Z",
  "parent": {
   "command_line": "/usr/libexec/runningboardd",
   "name": "runningboardd",
   "pid": "338971324198273501",
   "start": "2024-03-06T18:17:49.000000Z",
    "parent": {
     "command_line": "/sbin/launchd",
     "name": "launchd",
     "pid": "338053280202314993",
      "start": "2024-03-06T18:17:45.000000Z"
"server": { "ip": "35.186.224.39", "port": 443 },
"@timestamp": "2024-04-28T01:47:50.039000Z"
```

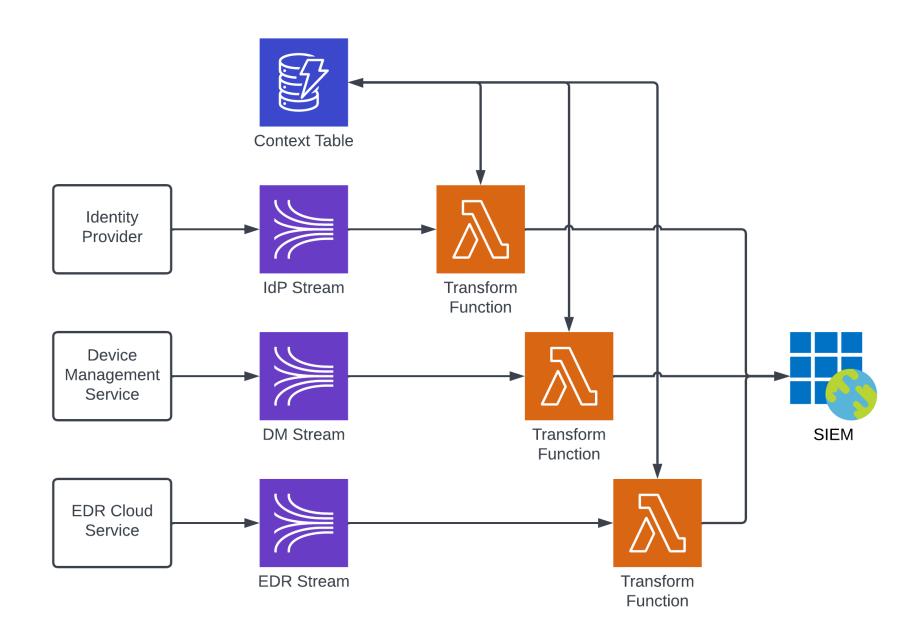
```
"event": { "category": "network", "type": "connection" },
"process": {
    "name": "Spotify",
    "pid": "339011099229862299"
},
    "server": { "ip": "35.186.224.39", "port": 443 },
    "@timestamp":"2024-04-28T01:47:50.039000Z"
}
```

```
"event": { "category": "network", "type": "connection" },
"process": {
  "command_line": "/Applications/Spotify.app/Contents/MacOS/Spotify",
 "name": "Spotify",
 "pid": "339011099229862299",
 "start": "2024-03-06T18:46:21.000000Z",
  "parent": {
   "command_line": "/usr/libexec/runningboardd",
   "name": "runningboardd",
    "pid": "338971324198273501",
    "start": "2024-03-06T18:17:49.000000Z",
    "parent": {
     "command_line": "/sbin/launchd",
     "name": "launchd",
     "pid": "338053280202314993",
     "start": "2024-03-06T18:17:45.000000Z"
"server": { "ip": "35.186.224.39", "port": 443 },
"@timestamp": "2024-04-28T01:47:50.039000Z"
```

```
"event": { "category": "network", "type": "connection" },
"process": {
    "name": "Spotify",
    "pid": "339011099229862299"
},
"server": { "ip": "35.186.224.39", "port": 443 },
"@timestamp":"2024-04-28T01:47:50.039000Z"
}
```

```
"event": { "category": "network", "type": "connection" },
"process": {
  "command_line": "/Applications/Spotify.app/Contents/MacOS/Spotify",
 "name": "Spotify",
 "pid": "339011099229862299",
 "start": "2024-03-06T18:46:21.000000Z",
 "parent": {
   "command_line": "/usr/libexec/runningboardd",
   "name": "runningboardd",
   "pid": "338971324198273501",
   "start": "2024-03-06T18:17:49.000000Z",
   "parent": {
     "command_line": "/sbin/launchd",
     "name": "launchd",
      "pid": "338053280202314993",
      "start": "2024-03-06T18:17:45.000000Z"
"server": { "ip": "35.186.224.39", "port": 443 },
"@timestamp": "2024-04-28T01:47:50.039000Z"
```

## Introducing... the Telephone Pattern!



```
{
  "event": { "category": "network", "type": "connection" },
  "host": {
      "id": "eb67b0b6a1d04086b75ee38d02018a10",
    },
  "process": {
      "name": "Spotify",
      "pid": "339011099229862299"
    },
      "server": { "ip": "35.186.224.39", "port": 443 },
}
```

```
"event": { "category": "network", "type": "connection" },
"host": {
 "id": "eb67b0b6a1d04086b75ee38d02018a10",
  "name": "C02TG3H6JGH1"
"process": {
 "name": "Spotify",
  "pid": "339011099229862299"
"server": { "ip": "35.186.224.39", "port": 443 },
"user": {
  "email": "alice@brex.com",
  "roles": ["Manager", "Security", "Engineering"],
  "status": ["idp_active"]
```

```
{
  "event": { "category": "network", "type": "connection" },
  "host": {
      "id": "eb67b0b6a1d04086b75ee38d02018a10",
    },
  "process": {
      "name": "Spotify",
      "pid": "339011099229862299"
    },
  "server": { "ip": "35.186.224.39", "port": 443 },
}
```

```
"event": { "category": "network", "type": "connection" },
"host": {
 "id": "eb67b0b6a1d04086b75ee38d02018a10",
 "name": "C02TG3H6JGH1"
"process": {
 "name": "Spotify",
 "pid": "339011099229862299"
"server": { "ip": "35.186.224.39", "port": 443 },
"user": {
 "email": "alice@brex.com",
 "roles": ["Manager", "Security", "Engineering"],
 "status": ["idp_active"]
```

```
"event": { "category": "network", "type": "connection" },
"host": {
    "id": "eb67b0b6a1d04086b75ee38d02018a10",
},
"process": {
    "name": "Spotify",
    "pid": "339011099229862299"
},
"server": { "ip": "35.186.224.39", "port": 443 },
}
```

```
"event": { "category": "network", "type": "connection" },
"host": {
 "id": "eb67b0b6a1d04086b75ee38d02018a10",
 "name": "C02TG3H6JGH1"
"process": {
 "name": "Spotify",
 "pid": "339011099229862299"
"server": { "ip": "35.186.224.39", "port": 443 },
"user": {
 "email": "alice@brex.com",
 "roles": ["Manager", "Security", "Engineering"],
 "status": ["idp_active"]
```

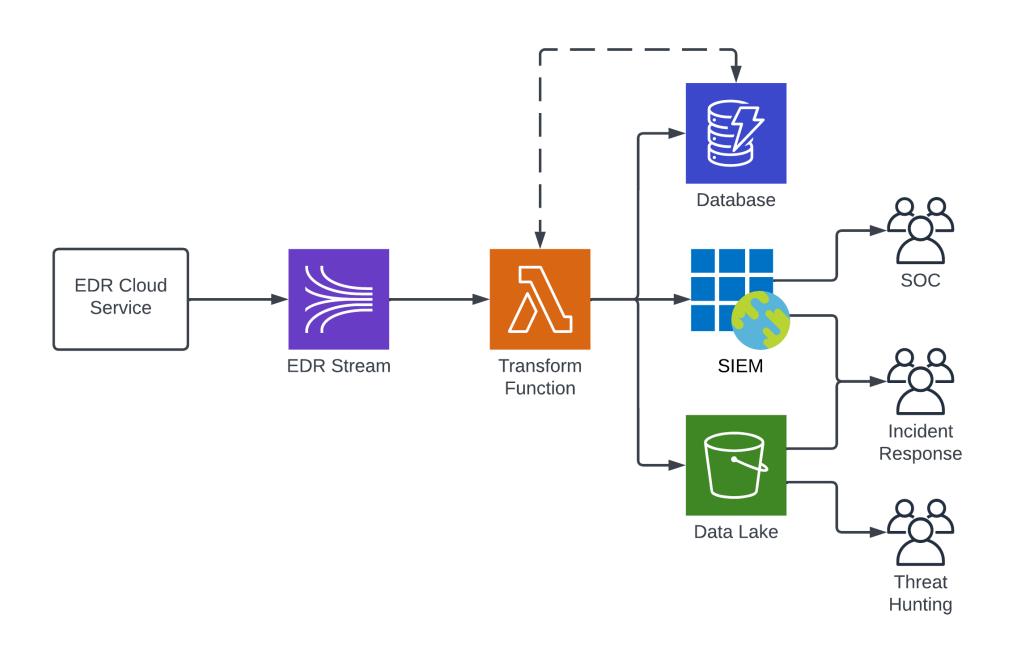
```
"event": { "category": "network", "type": "connection" },
"host": {
    "id": "eb67b0b6a1d04086b75ee38d02018a10",
},
"process": {
    "name": "Spotify",
    "pid": "339011099229862299"
},
"server": { "ip": "35.186.224.39", "port": 443 },
}
```

```
"event": { "category": "network", "type": "connection" },
"host": {
 "id": "eb67b0b6a1d04086b75ee38d02018a10",
 "name": "C02TG3H6JGH1"
"process": {
 "name": "Spotify",
 "pid": "339011099229862299"
"server": { "ip": "35.186.224.39", "port": 443 },
"user": {
 "email": "alice@brex.com",
 "roles": ["Manager", "Security", "Engineering"],
 "status": ["idp_active"]
```

```
{
  "event": { "category": "network", "type": "connection" },
  "host": {
      "id": "eb67b0b6a1d04086b75ee38d02018a10",
    },
  "process": {
      "name": "Spotify",
      "pid": "339011099229862299"
    },
      "server": { "ip": "35.186.224.39", "port": 443 },
}
```

```
"event": { "category": "network", "type": "connection" },
"host": {
 "id": "eb67b0b6a1d04086b75ee38d02018a10",
  "name": "C02TG3H6JGH1"
"process": {
 "name": "Spotify",
  "pid": "339011099229862299"
"server": { "ip": "35.186.224.39", "port": 443 },
"user": {
  "email": "alice@brex.com",
  "roles": ["Manager", "Security", "Engineering"],
  "status": ["idp_active"]
```

## Introducing... the nXDR Pattern!



```
"process": {
    "command_line": "/usr/libexec/security_authtrampoline
        /usr/sbin/installer auth 22 -verboseR -allowUntrusted
        -pkg /private/tmp/xp-6100/epsvcp.pkg -target /",
    "name": "security_authtrampoline",
    "parent": {
        "command_line": "/private/tmp/update_XP-6100
        Series/EPSON.app/Contents/MacOS/EpsonInstaller",
        "name": "EpsonInstaller",
    }
}
```

```
"process": {
  "command_line": "/usr/libexec/security_authtrampoline
   /usr/sbin/installer auth 22 -verboseR -allowUntrusted
   -pkg /private/tmp/xp-6100/epsvcp.pkg -target /",
  "name": "security_authtrampoline",
  "parent": {
    "command_line": "/private/tmp/update_XP-6100
     Series/EPSON.app/Contents/MacOS/EpsonInstaller",
    "name": "EpsonInstaller",
"threat": {
  "signals": [{
      "description": "Identifies when an authentication prompt
      is generated by the AuthorizationExecuteWithPrivileges API.",
      "name": "privilege_escalation_elevated_execution_with_prompt",
      "references": [
        "objective-see.com/blog/blog_0x2A.html"
      ],
      "risk_score": 73
```

```
"process": {
    "command_line": "/usr/libexec/security_authtrampoline
        /usr/sbin/installer auth 22 -verboseR -allowUntrusted
        -pkg /private/tmp/xp-6100/epsvcp.pkg -target /",
    "name": "security_authtrampoline",
    "parent": {
        "command_line": "/private/tmp/update_XP-6100
            Series/EPSON.app/Contents/MacOS/EpsonInstaller",
            "name": "EpsonInstaller",
        }
    }
}
```

```
"process": {
  "command_line": "/usr/libexec/security_authtrampoline
   /usr/sbin/installer auth 22 -verboseR -allowUntrusted
   -pkg /private/tmp/xp-6100/epsvcp.pkg -target /",
  "name": "security_authtrampoline",
  "parent": {
    "command_line": "/private/tmp/update_XP-6100
     Series/EPSON.app/Contents/MacOS/EpsonInstaller",
    "name": "EpsonInstaller",
"threat": {
 "signals": [{
      "description": "Identifies when an authentication prompt
      is generated by the AuthorizationExecuteWithPrivileges API.",
      "name": "privilege_escalation_elevated_execution_with_prompt",
      "references": [
        "objective-see.com/blog/blog_0x2A.html"
      "risk_score": 73
```

```
"process": {
  "command_line": "/usr/libexec/security_authtrampoline
   /usr/sbin/installer auth 22 -verboseR -allowUntrusted
   -pkg /private/tmp/xp-6100/epsvcp.pkg -target /",
 "name": "security_authtrampoline",
 "parent": {
   "command_line": "/private/tmp/update_XP-6100
     Series/EPSON.app/Contents/MacOS/EpsonInstaller",
    "name": "EpsonInstaller",
"threat": {
 "signals": [{
     "description": "Identifies when an authentication prompt
     is generated by the AuthorizationExecuteWithPrivileges API.",
     "name": "privilege_escalation_elevated_execution_with_prompt",
     "references": [
       "objective-see.com/blog/blog_0x2A.html"
     ],
     "risk_score": 73
```

```
"process": {
    "command_line": "/usr/libexec/security_authtrampoline
        /usr/sbin/installer auth 22 -verboseR -allowUntrusted
        -pkg /private/tmp/xp-6100/epsvcp.pkg -target /",
    "name": "security_authtrampoline",
    "parent": {
        "command_line": "/private/tmp/update_XP-6100
        Series/EPSON.app/Contents/MacOS/EpsonInstaller",
        "name": "EpsonInstaller",
    }
}
```

```
"process": {
  "command_line": "/usr/libexec/security_authtrampoline
   /usr/sbin/installer auth 22 -verboseR -allowUntrusted
   -pkg /private/tmp/xp-6100/epsvcp.pkg -target /",
 "name": "security_authtrampoline",
 "parent": {
   "command_line": "/private/tmp/update_XP-6100
     Series/EPSON.app/Contents/MacOS/EpsonInstaller",
    "name": "EpsonInstaller",
"threat": {
 "signals": [{
     "description": "Identifies when an authentication prompt
     is generated by the AuthorizationExecuteWithPrivileges API.",
     "name": "privilege_escalation_elevated_execution_with_prompt",
     "references": [
       "objective-see.com/blog/blog_0x2A.html"
     "risk_score": 73
```

```
"process": {
    "command_line": "/usr/libexec/security_authtrampoline
        /usr/sbin/installer auth 22 -verboseR -allowUntrusted
        -pkg /private/tmp/xp-6100/epsvcp.pkg -target /",
    "name": "security_authtrampoline",
    "parent": {
        "command_line": "/private/tmp/update_XP-6100
        Series/EPSON.app/Contents/MacOS/EpsonInstaller",
        "name": "EpsonInstaller",
    }
}
```

```
"process": {
  "command_line": "/usr/libexec/security_authtrampoline
   /usr/sbin/installer auth 22 -verboseR -allowUntrusted
   -pkg /private/tmp/xp-6100/epsvcp.pkg -target /",
  "name": "security_authtrampoline",
  "parent": {
    "command_line": "/private/tmp/update_XP-6100
     Series/EPSON.app/Contents/MacOS/EpsonInstaller",
    "name": "EpsonInstaller",
"threat": {
  "signals": [{
      "description": "Identifies when an authentication prompt
      is generated by the AuthorizationExecuteWithPrivileges API.",
      "name": "privilege_escalation_elevated_execution_with_prompt",
      "references": [
        "objective-see.com/blog/blog_0x2A.html"
      ],
      "risk_score": 73
```

## These Solutions, and Dozens More, Can Be Deployed *Right Now* with Substation!<sup>17</sup>

```
aws configure && \
make -s check && \
make -s build && \
make -s deploy EXAMPLE=terraform/aws/kinesis/time_travel
```

<sup>&</sup>lt;sup>17</sup> https://github.com/brexhq/substation#testing

## Open Source, Extra Reading, Contact Info

- github.com/brexhq/substation
- · medium.com/@jshlbrd
- linkedin.com/in/joshliburdi