Lifescale Analytics

Data Analytics Pipeline

Theo Laughner, PE

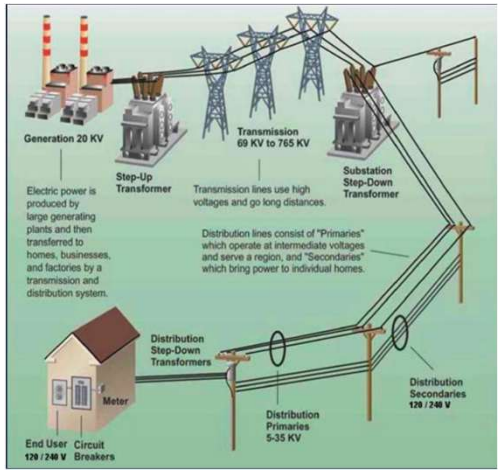
2025

Overview

- Drivers
- Vision
- Sensors
- Pipeline
- Resources
- Key Takeaways

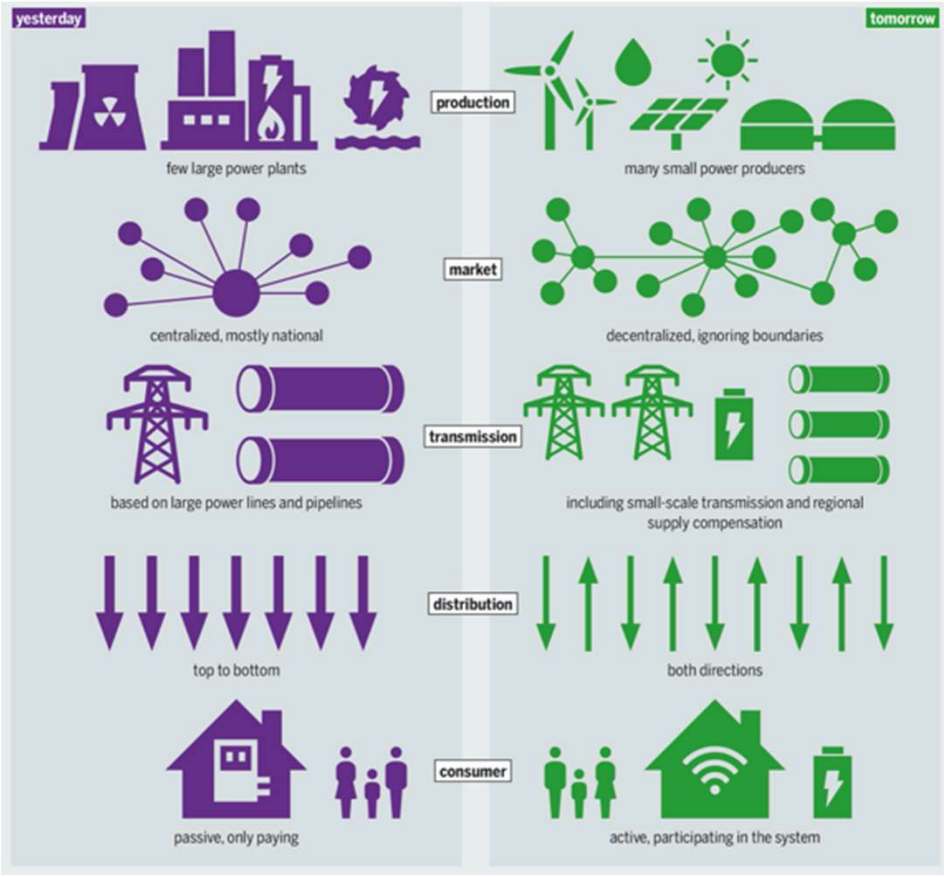
lifescale analytics

Drivers



Who is the consumer…

Generation?
Transmission?
Distribution?
Consumer?

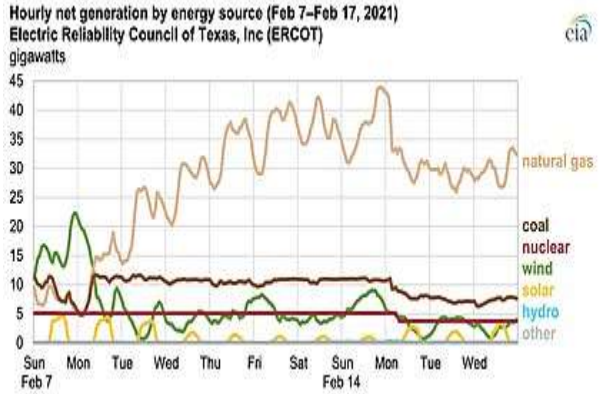


Operating the Old System In New Ways

Unintended Consequences

More Measurement

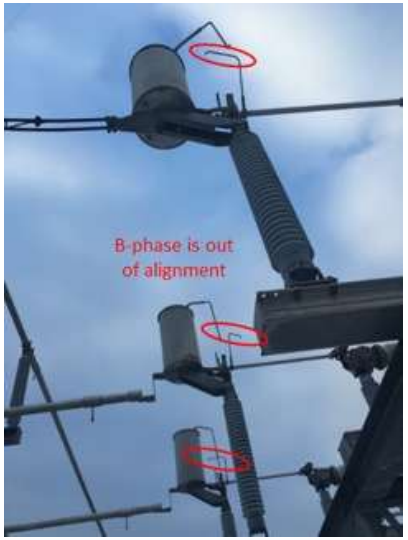
Regulation or Standards

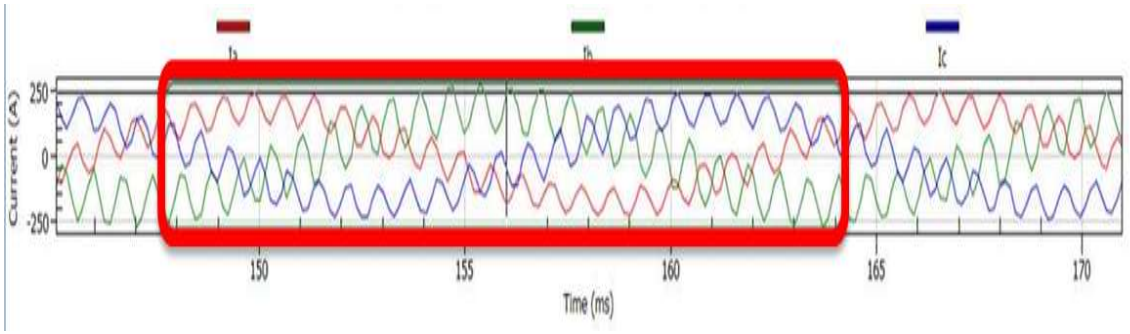


Resignations of key officials in ERCOT.
Bankruptcy of several retailers.

lifescale analytics

Increasing Complexity Results In Unforeseen Circumstances





Challenge

Prevent catastrophic **failure** of assets and **unplanned interruptions** to customers.

lifescale analytics

A Part of The Solution – Better Data Management

Data management is the **development** and **execution** of **architectures**, **policies**, **practices** and **procedures** in order to manage the information lifecycle needs of an enterprise in an effective manner.

*Source: http://searchdatamanagement.techtarget.com/definition/data-management

Policy

A *policy* is a statement of intent, and is implemented as a procedure or protocol.

Examples – laws, regulations

Procedure

What you are supposed to do.

Example

“Make a cake that says Best Wishes Suzanne,
Underneat that We will miss you”

Practice

What you actually do!

This seems like a lot of rigor!

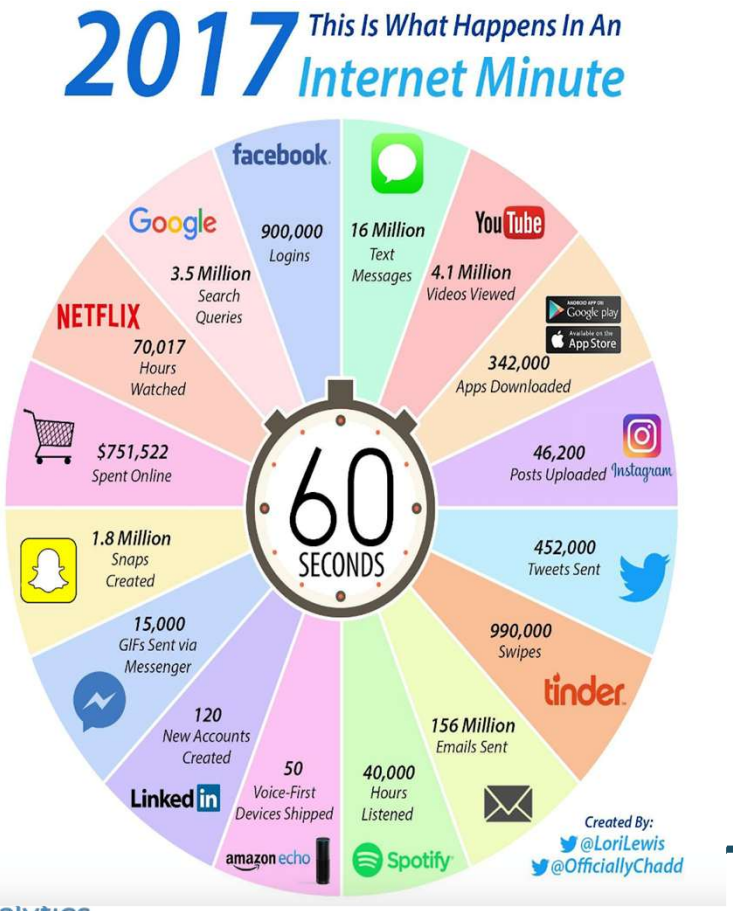
Is this really necessary?

Without due consideration of these, it is unlikely that you will get what you intended or desired.

Considerations –

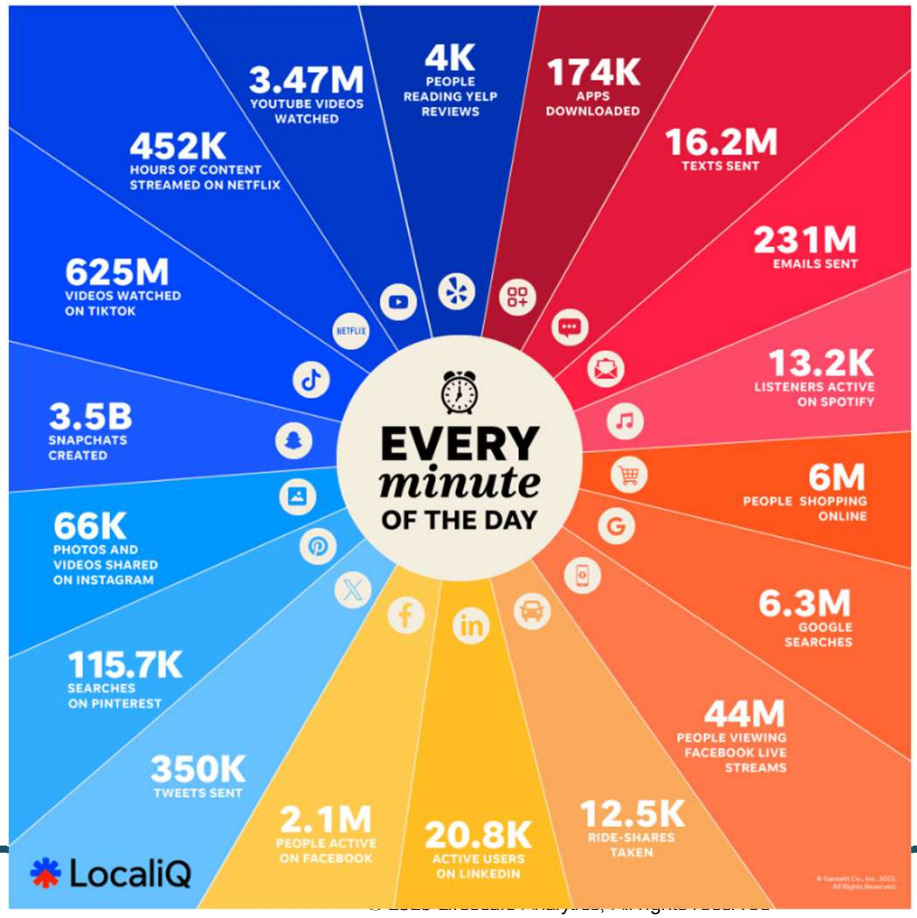
Volume, Velocity, Veracity, Variety

An Internet Minute…

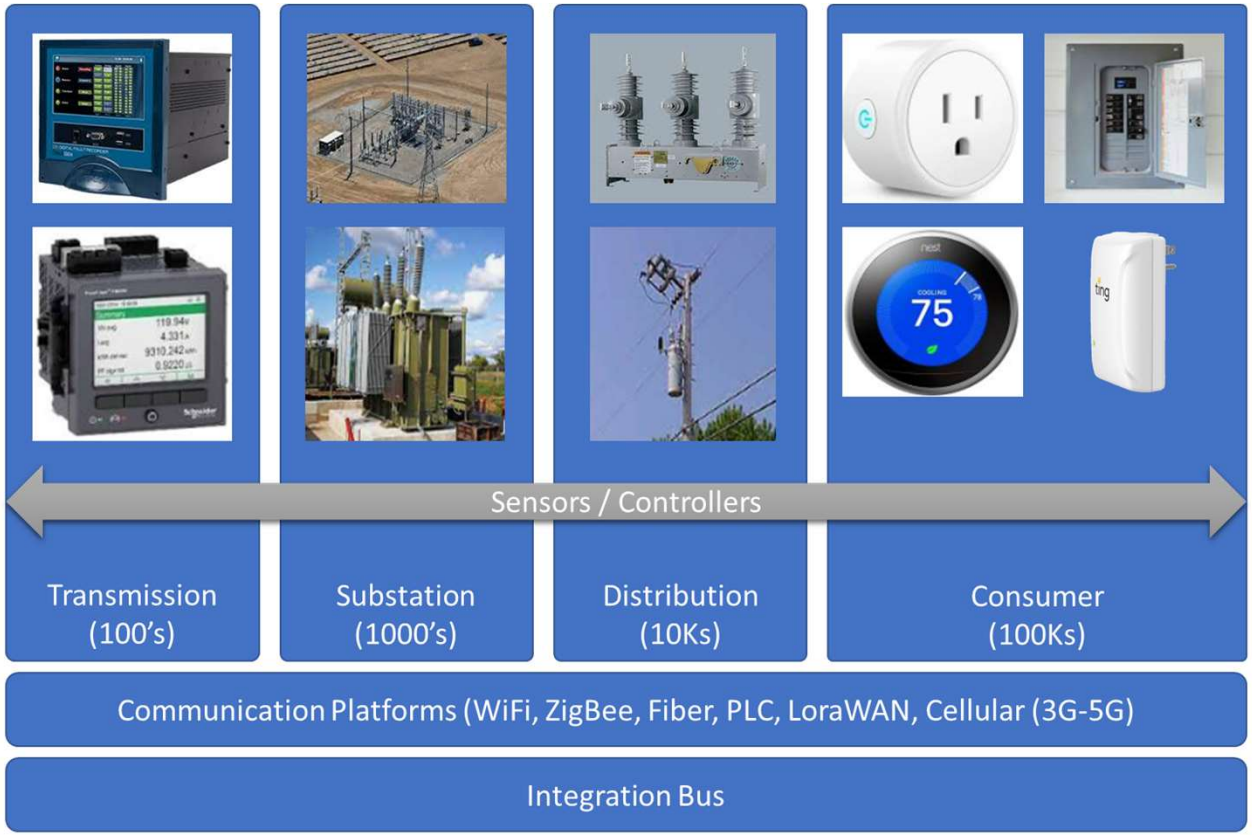


analytics

An Internet Minute…



Sensors



Digital Twin

Made Possible By:

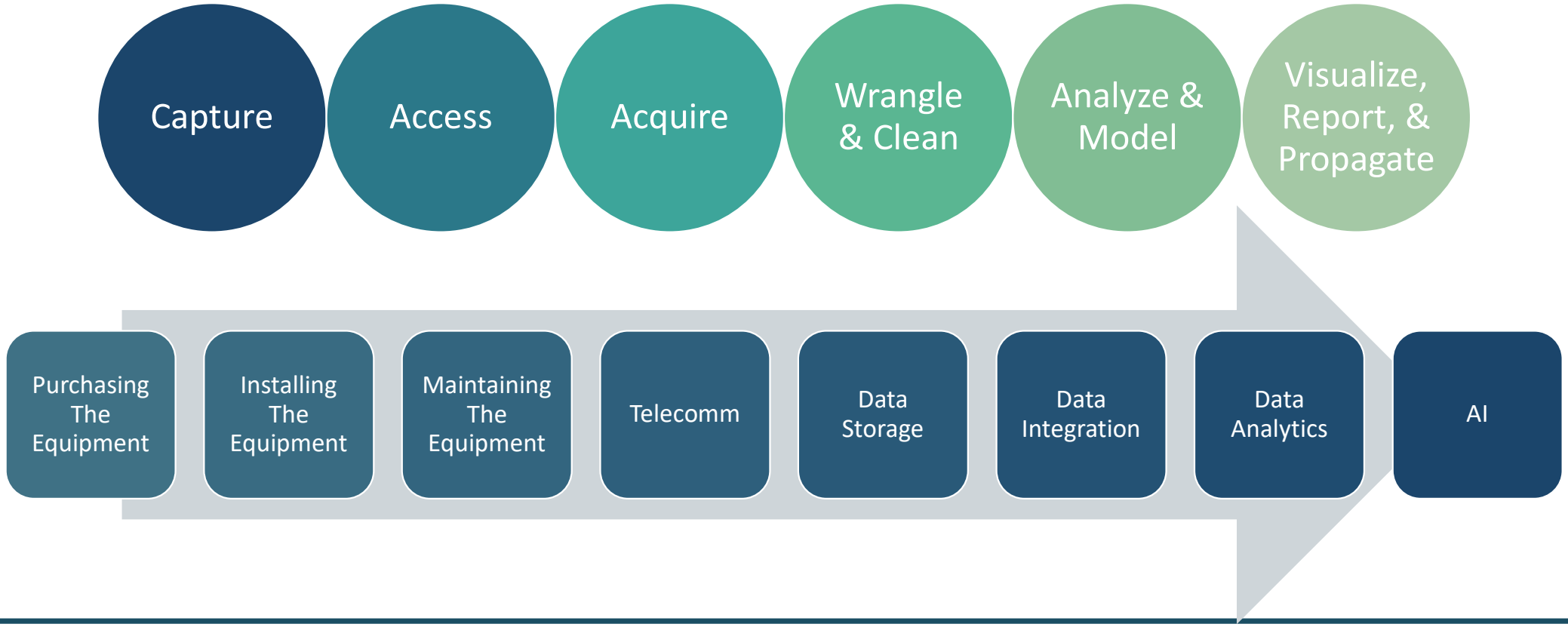
- **Ubiquitous Telecommunications**

Challenged By:

- **Variety of Sensors / IOT**
- **Cybersecurity**

lifescale analytics

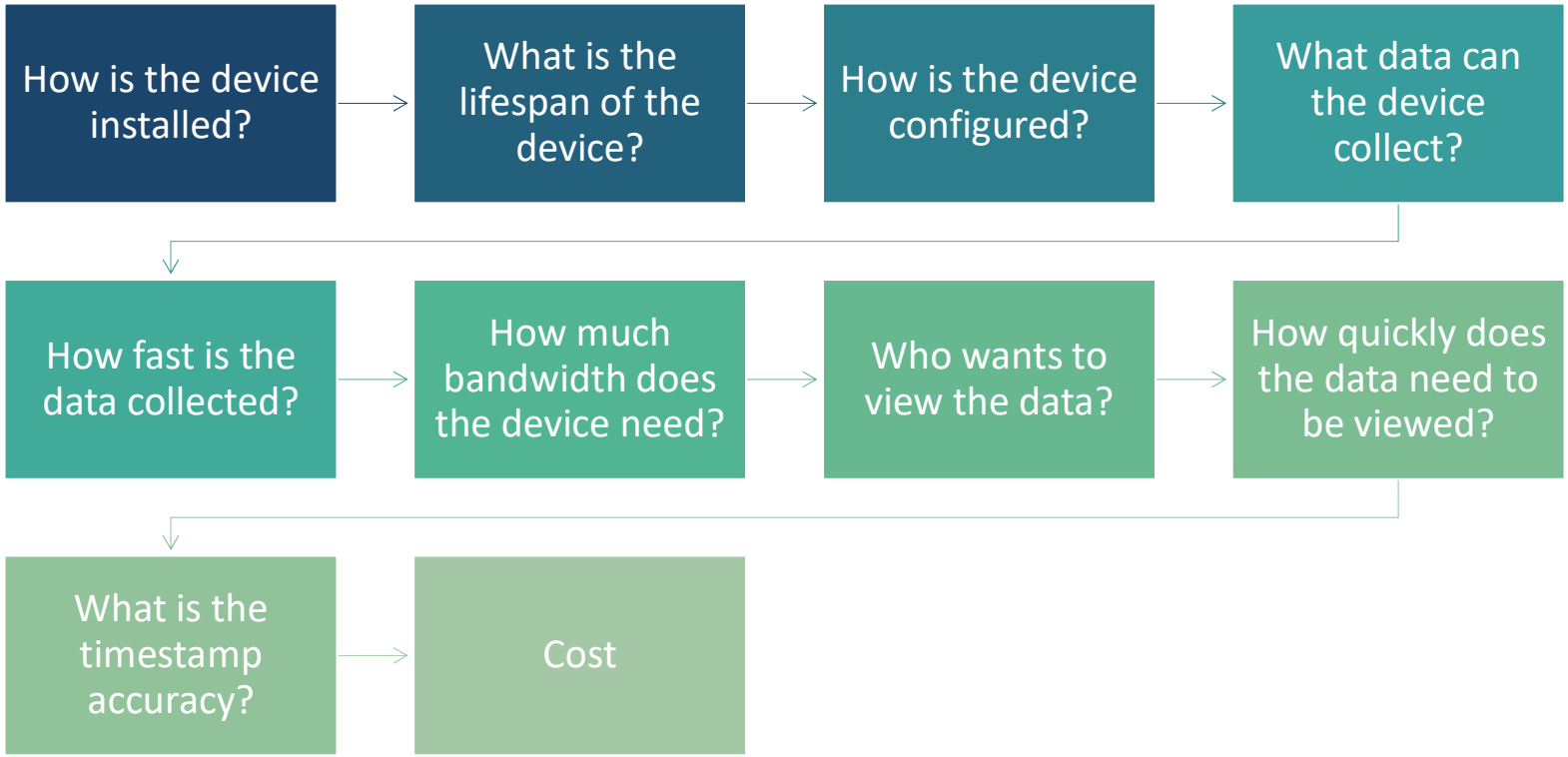
Pipeline



lifescale analytics

Pipeline

Capture

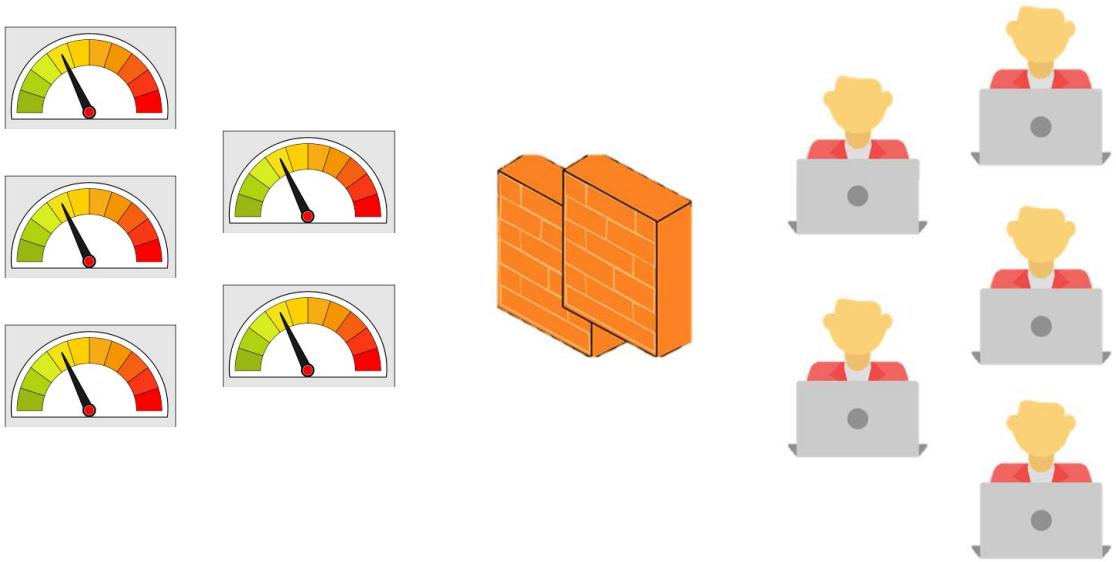


- Data is not always captured.
- Large numbers of devices are being deployed.
- Deactivated sensors.

lifescale analytics

Pipeline

Access



- Limited to few people.
- Limited bandwidth.
- Security.

Pipeline

Acquire

- Who will collect the data?
- What tools will be needed to collect the data?
- What protocols?
- What file formats / databases?
- How long will it be kept?
- Where will the data be stored?
- What are the security constraints?

lifescale analytics

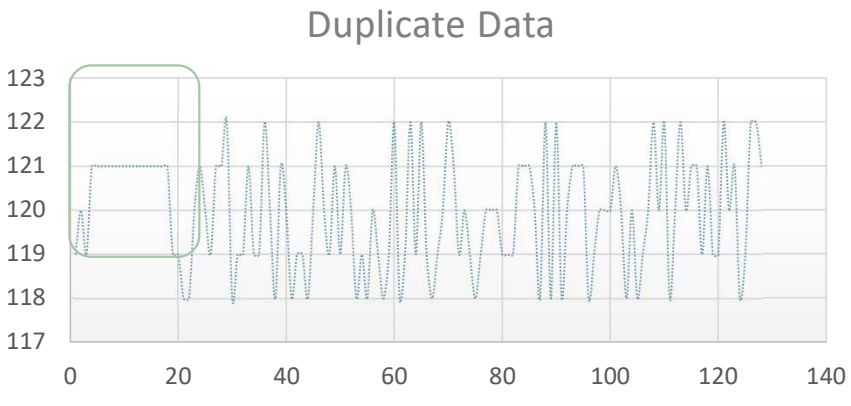
Pipeline

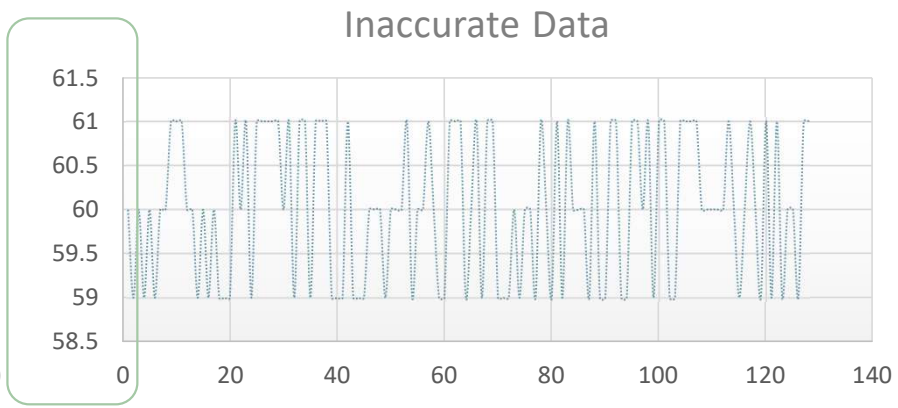
Acquire

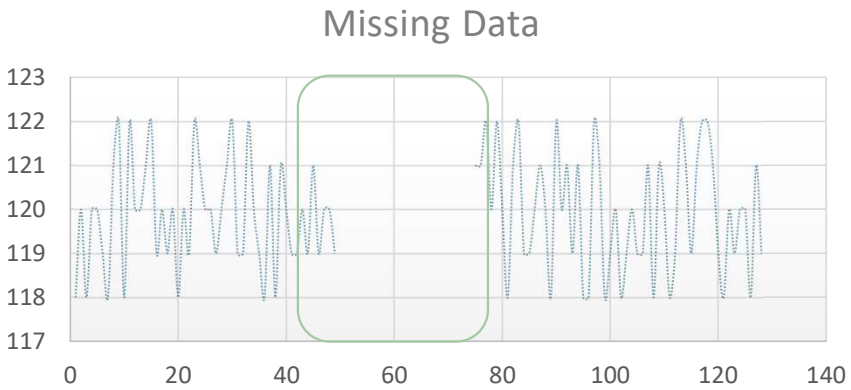
Read Speed (data points per sec)	Time to Read 1.89 Trillion Data Points	Channels (Samples/Cycle @ 166 Samples/Cycle)
10,000	6 years	1 (Voltage, Current, etc)
100,000	7.3 months	10 (1 Bus)
1,000,000	3.125 weeks	100 (1 Substation)
10,000,000	2.1875 days	1,000 (10 Substations)
100,000,000	5.25 hours	10,000 (100 Substations)
1,000,000,000	31.5 minutes	100,000 (1000 Substations)

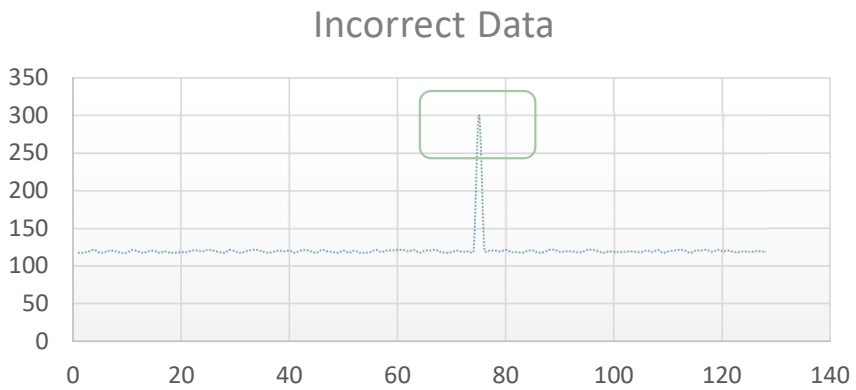
Pipeline

Wrangle & Clean





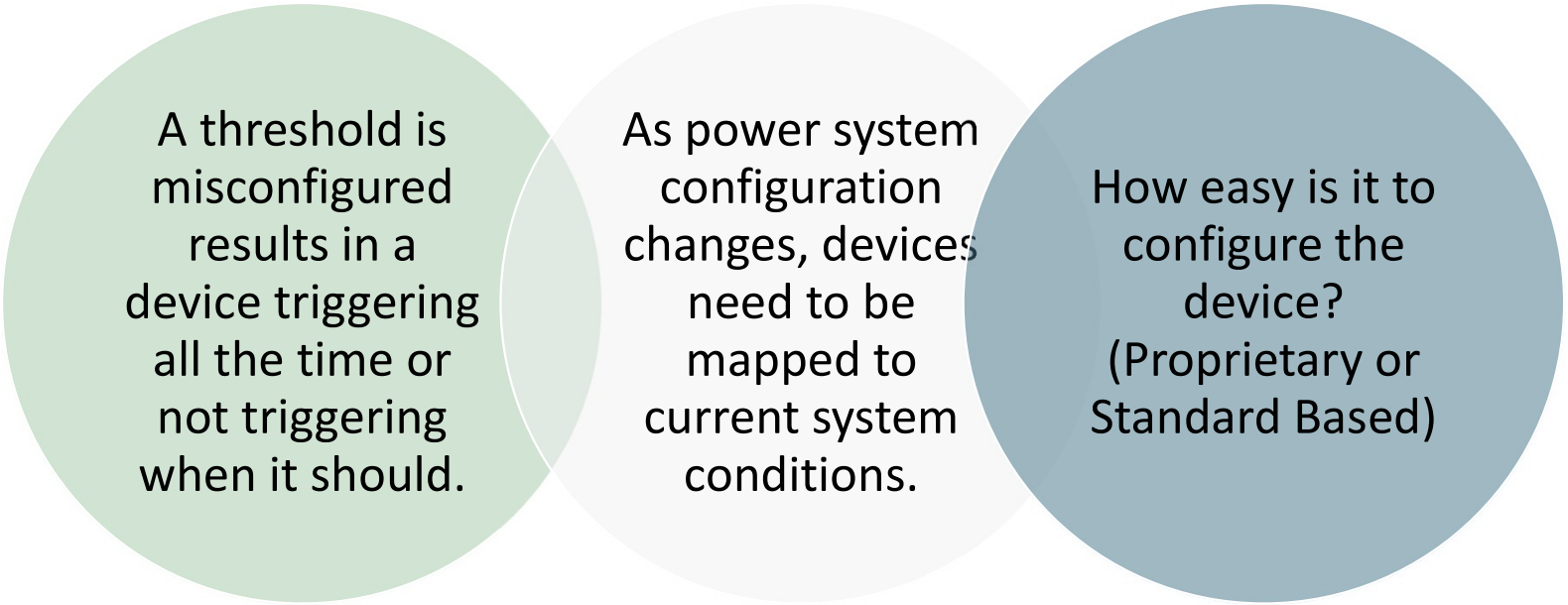




lifescale analytics

Pipeline

Wrangle & Clean



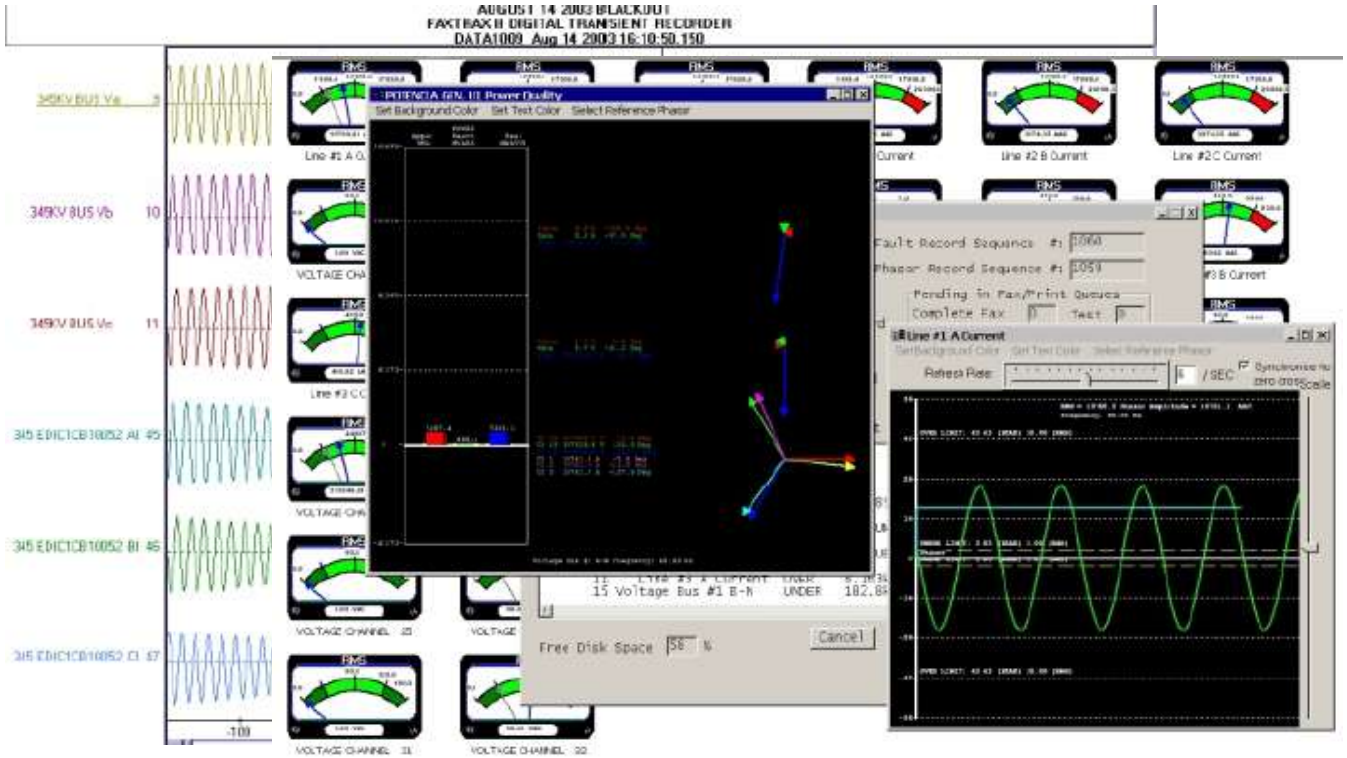
Poor configuration management leads to bad/misleading data.

lifescale analytics

Pipeline

Analyze & Model

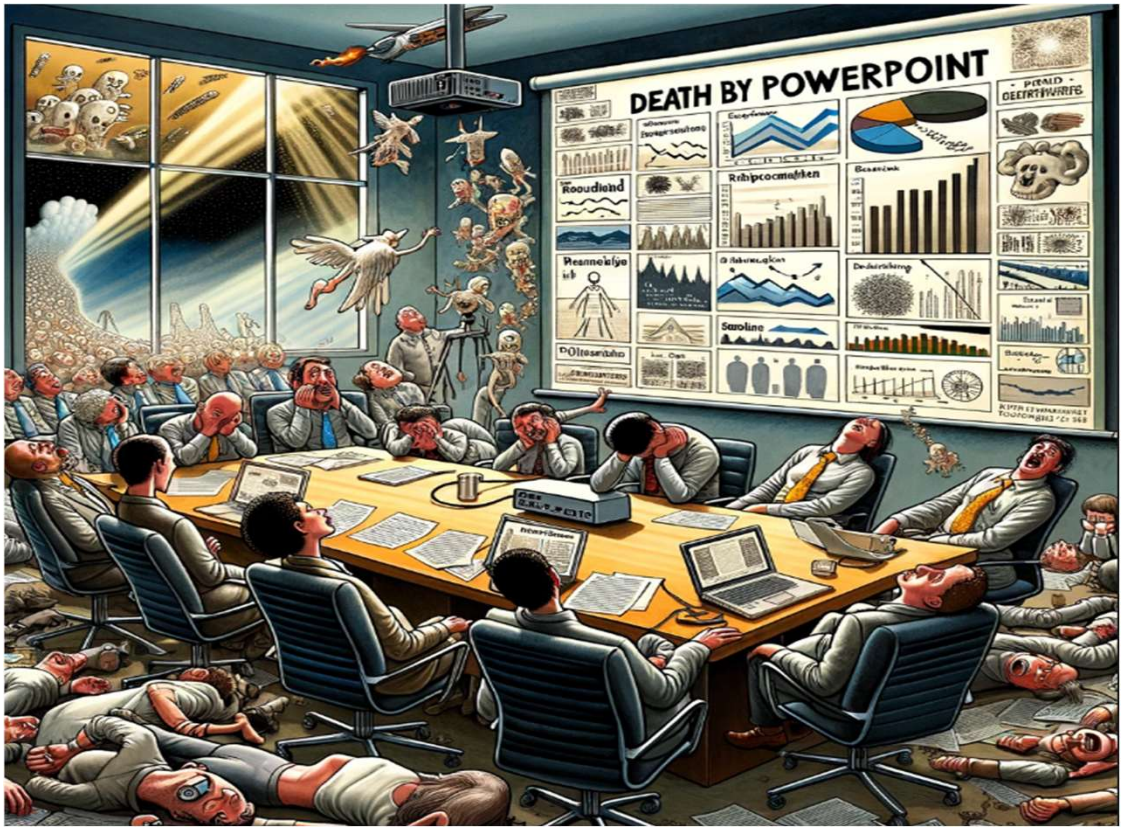
- Excel
- MATLAB
- Too much data to review manually!
- **Reminder 1 PMU Channel – 60GB of data per year.**



Pipeline

Visualize, Report, & Propagate

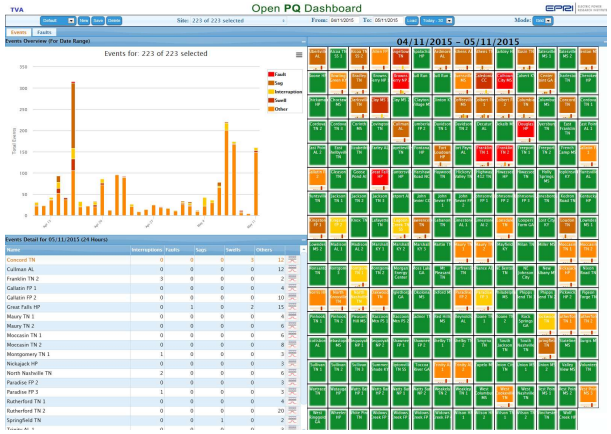
- Email
- Spreadsheets
- Powerpoint
- Journals
- Publications

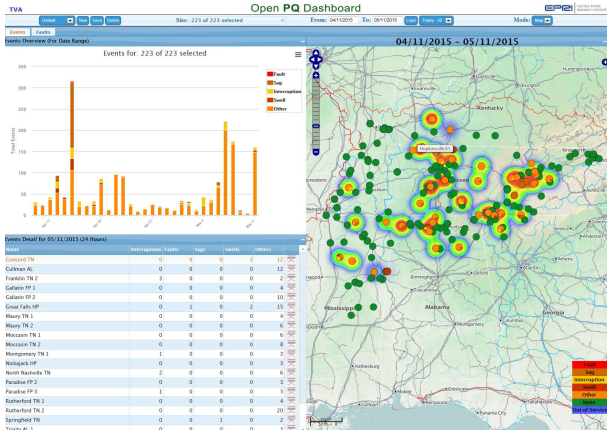


lifescale analytics

Pipeline

Visualize, Report, & Propagate





Find voltage sags over 3 months of data.

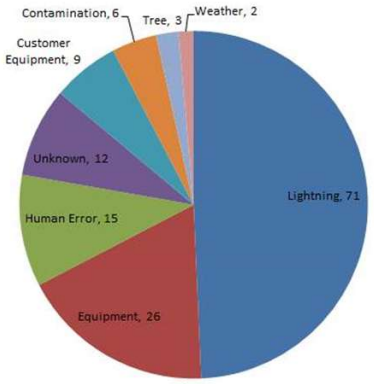
We look at voltage sags from November 19, 2016 to February 4, 2017.

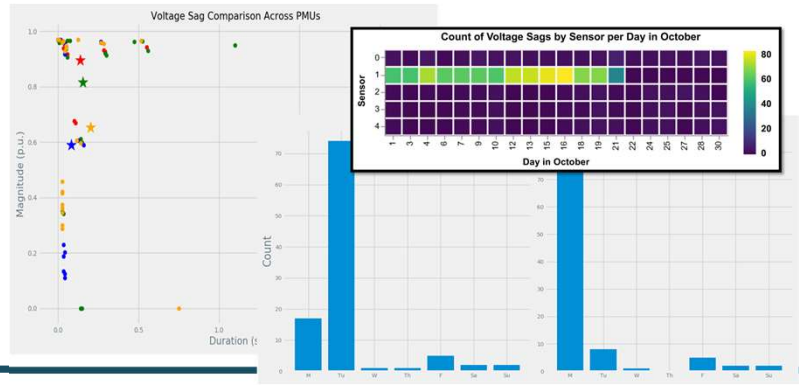
In [11]:
```
# Choose the stream
# PMU 1 from sunshine dataset
stream = streams["35bdb8dc-bf18-4523-85ca-8ebe384bd9b5"];
# Get nominal voltage of stream
vnom = get_mean_value(stream);

# Start and end times of period to study
start = "2016-11-19T00:00:00.000Z"
end = "2017-02-04T00:00:00.000Z"

# Threshold below which data is considered a voltage sag
thresh = 0.99 * vnom;

# Find voltage sag data points
sags = find_vsags_dfs(stream, thresh, start=start, end=end);
# Get features of voltage sags
starts, durs, mags = sag_survey(sags, verbose=False);
```





lifescale analytics

Resources for the Electric Power Industry

File Formats –	PQDIF COMTRADE PQds
Data –	EPRI Disturbance Library ORNL Disturbance Library
Tools –	Predictive Grid / NI4AI – Open API PQView – Commercial OpenXDA – Open Source
Benchmarks/Outage Info –	DOE – Outage Data Initiative Grid Metrics – Power Outages Whisker Labs – CPQI

lifescale analytics

Key Takeaways

- The grid is increasing in complexity.
- We don’t know what we don’t know.
- Measure everything let algorithms sort it out.
- There are tools and techniques for managing data.
- The data pipeline should lead to actionable information.

Additional Resources

- https://btrdb.readthedocs.io/en/latest/
- https://github.com/PingThingsIO/ni4ai-notebooks/
- https://blog.ni4ai.org/
- https://pingthings.io
- https://www.gridprotectionalliance.org
- https://www.python.org
- https://www.r-project.org
- https://gesl.ornl.gov/
- https://www.pqview.com/
- https://odin.ornl.gov/
- https://poweroutage.fyi/
- https://power-quality.tingfire.com