

Building software systems to support interactively exploring time series energy data

Jason Mair, Michael Jack and David Evers

Working with time series energy data

- Characteristics of energy data:
 - Variety of sources
 - e.g., household, individual devices or meters
 - Varying frequency
 - e.g., daily, hourly, per minute
- Data is often split across multiple files, e.g. per household
- Can have complex relationships
 - Region \Leftrightarrow neighbourhood \Leftrightarrow household \Leftrightarrow devices
- Involves manually restructuring data to suit specific research questions, such as grouping by device type or by household

Challenge: working at scale (lots of data)

- Challenging to restructure data sets as they keep growing
 - Results in duplication of files
 - Files are often organised in a way only known to the creator
- Software such as Microsoft Excel cannot open large data files
 - Can get around these limitations by using many smaller files
 - ... it then becomes a problem to manage all of the separate files
- This impacts how analyses and visualisation are performed

Challenge: reproducibility of data processing

- Everyone has a different configuration of their computers
- Software may not work on all operating systems
 - Only runs on Microsoft Windows and not Apple's macOS
- Incompatible software versions
 - People don't keep their software up to date
 - ... this may be due to concerns over changes in functionality across versions
- Can be hard to get the same results as someone else
 - Not all of the required steps are provided
 - "It worked for me ..."

Software systems can have low reproducibility

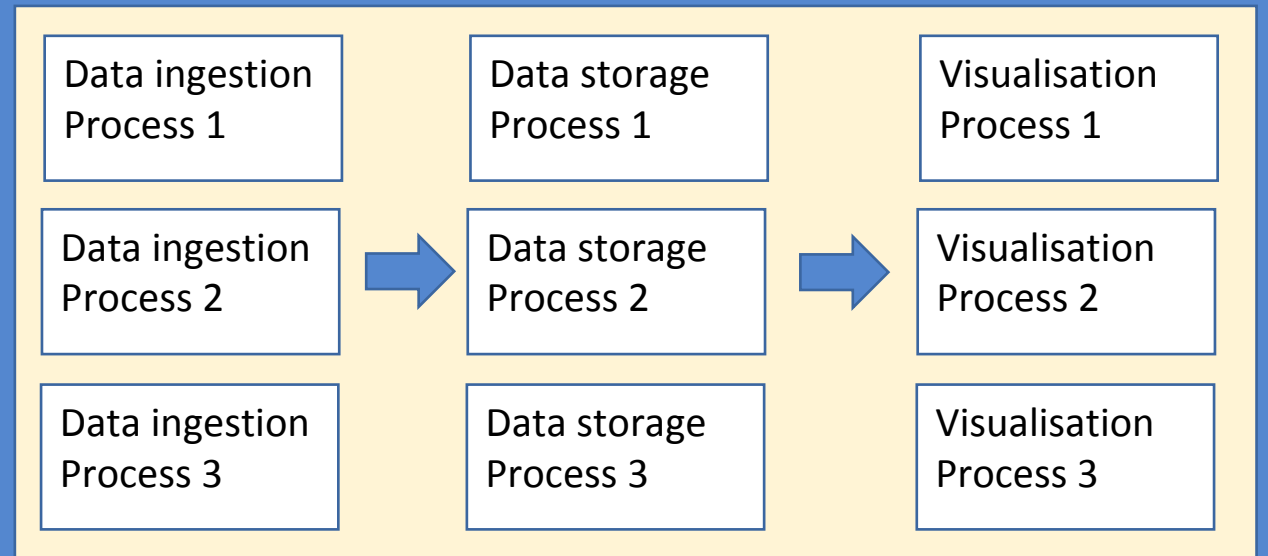
- The software you interact with is just the top of the stack:
 - There are many underlying dependencies, e.g.,
 - Independent, supporting software libraries
 - Functionality of the underlying operating system
- Different versions will have different support and dependencies
 - Leads to statements, "version X.Y.Y was used"
 - Can be difficult to get specific older version (X.Y.Y) to work when starting from scratch
- Such dependencies mean the complete software environment matters
 - There can be **lots** of configuration knobs and dials
- This is without even mentioning the requirement of tracking the state of the data set used within the computational pipeline

Avoiding software sustainability challenges

- We are working to develop a software platform to make it easier to interact with and work across multiple time-series data sets
- Objectives for the software platform
 - Interactive
 - Easy to use
 - Provide meaningful insights
 - Scalable
 - Can work with large and small data sets
 - Reproducible
 - Works on Microsoft Windows, Apple macOS, Linux, etc.
 - Want it to 'just work' out of the box

The data workflow

- Working with the data has a basic workflow:
 - Pre-processing
 - Data cleaning
 - Transform data formats e.g. timestamps
 - Storage
 - Define a data structure
 - Continue to store new values
 - Visualisation and analysis
 - Load the required data
 - Perform analysis and visualisation



Building the underlying software platform

- Use Docker to work across different operating systems
 - Specifies exact versions of the software to be used
 - Uses lightweight virtualisation to run on different operating systems
- Use Kubernetes to orchestrate each component
 - Control the communication
 - e.g. can ensure storage is ready before starting data ingestion or visualisation
- These technologies allow everything to be packaged together
 - Ingestion process
 - Stored data
 - Visualisation/analysis process

Step 1 - Data ingestion

- Data pre-processing
 - Clean data
 - replace missing values
 - Convert timestamps
 - Handle daylight savings
 - Convert to consistent time zone
 - Unit conversion
- Performance
 - We have written custom scripts for loading data
 - Time taken to load reduced from 120 to 6 minutes
- Expandable
 - Update to handle live data feeds

Step 2 - Time series database

- TimescaleDB
 - Open source
- Efficiently stores time-based events
 - Sorted by time
- Fast data ingestion
- Scalable design
- Time series and normal database tables
 - Allows us to separate events from device information
- Performant queries to retrieve data

Time	Meter ID	Value
1420072200	4	0.934
1420072200	4	0.938
1420072200	4	9388
...

Meter ID	Location	Type
4	Building 1	UMG9

Step 3 - Visualisation

- Grafana
 - Open source, cross platform
- Provides a web-based user interface
- Analytics and interactive visualisation
 - Charts, graphs and alerts
- Can build complex monitoring dashboards
- Extendable: allows custom plugins to be created for visualisation and analysis

Plugins

- Grafana plugins for:
 - Data sources
 - Visualisation
- Public
 - Can download from Grafana website
- Private
 - Create custom plugins

Grafana Labs Grafana Products Open Source Learn Downloads Login [Contact us](#)

Features Contribute Dashboards **Plugins** Download

Plugins

Official & community built plugins

Product updates
Sign up to receive occasional product news and updates:

Name / Description

Panel plugins allow new data visualizations to be added to Grafana, for both time series and non-time series data.

Worldmap Panel
by Grafana Labs
World Map panel for Grafana. Displays time series data or geohash data from Elasticsearch...

Clock
by Grafana Labs
Clock panel for grafana

Pie Chart
by Grafana Labs
Pie chart panel for grafana

Alert List
by Grafana Labs
Shows list of alerts and their current status

Dashboard list
by Grafana Labs
List of dynamic links to other dashboards

Graph
by Grafana Labs
Graph Panel for Grafana

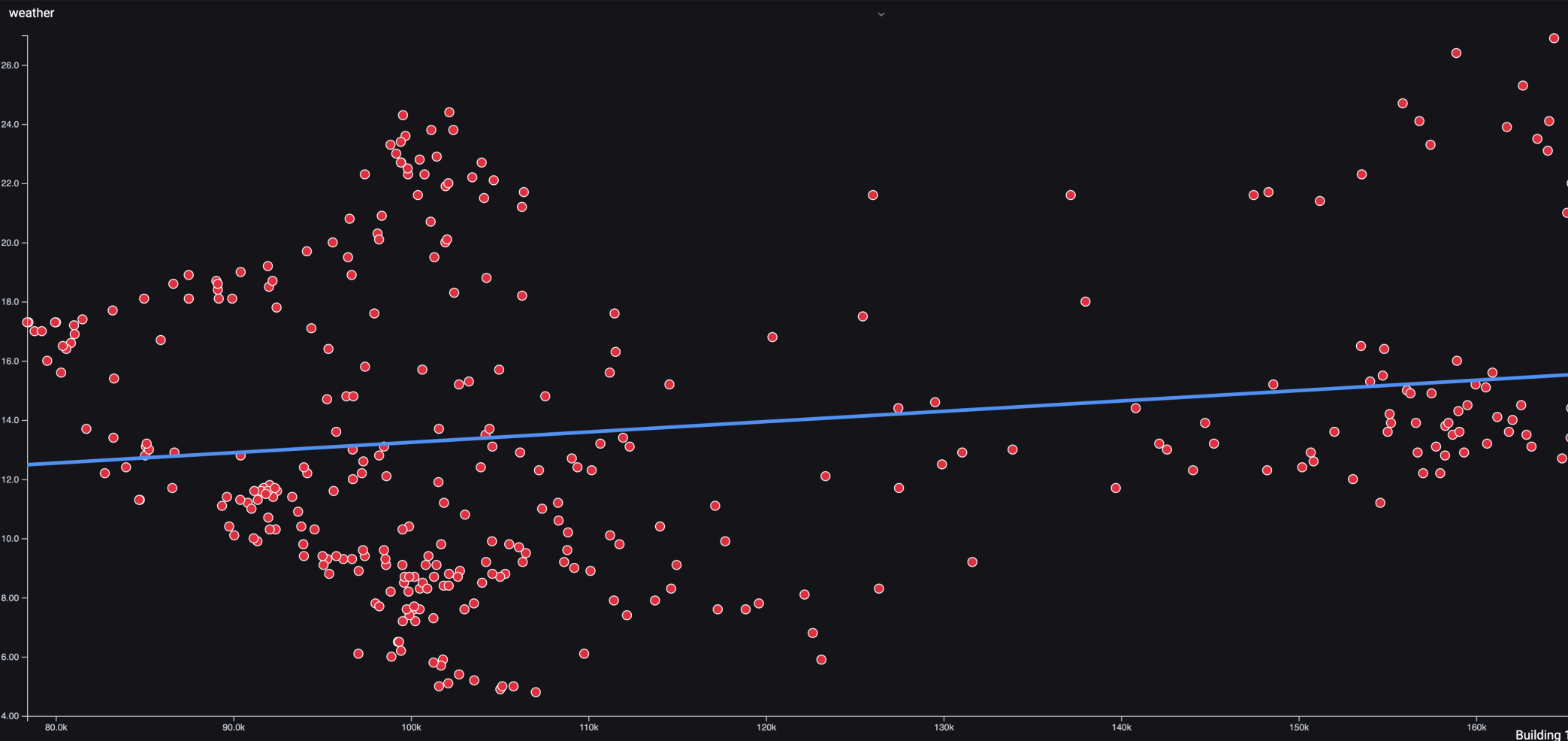
Plugin Type:
 All
 Panel
 Data Source
 App

Enterprise Plugins:
 All Plugins
 Enterprise Only

Sort By:



Building Building 1 measurement Active Power type Sum L1-L3 Weather temp



Correlations
 Pearson:
 0.174
 Spearman:
 0.0583
 Point:
 Building 1
 165379.72
 weather
 22

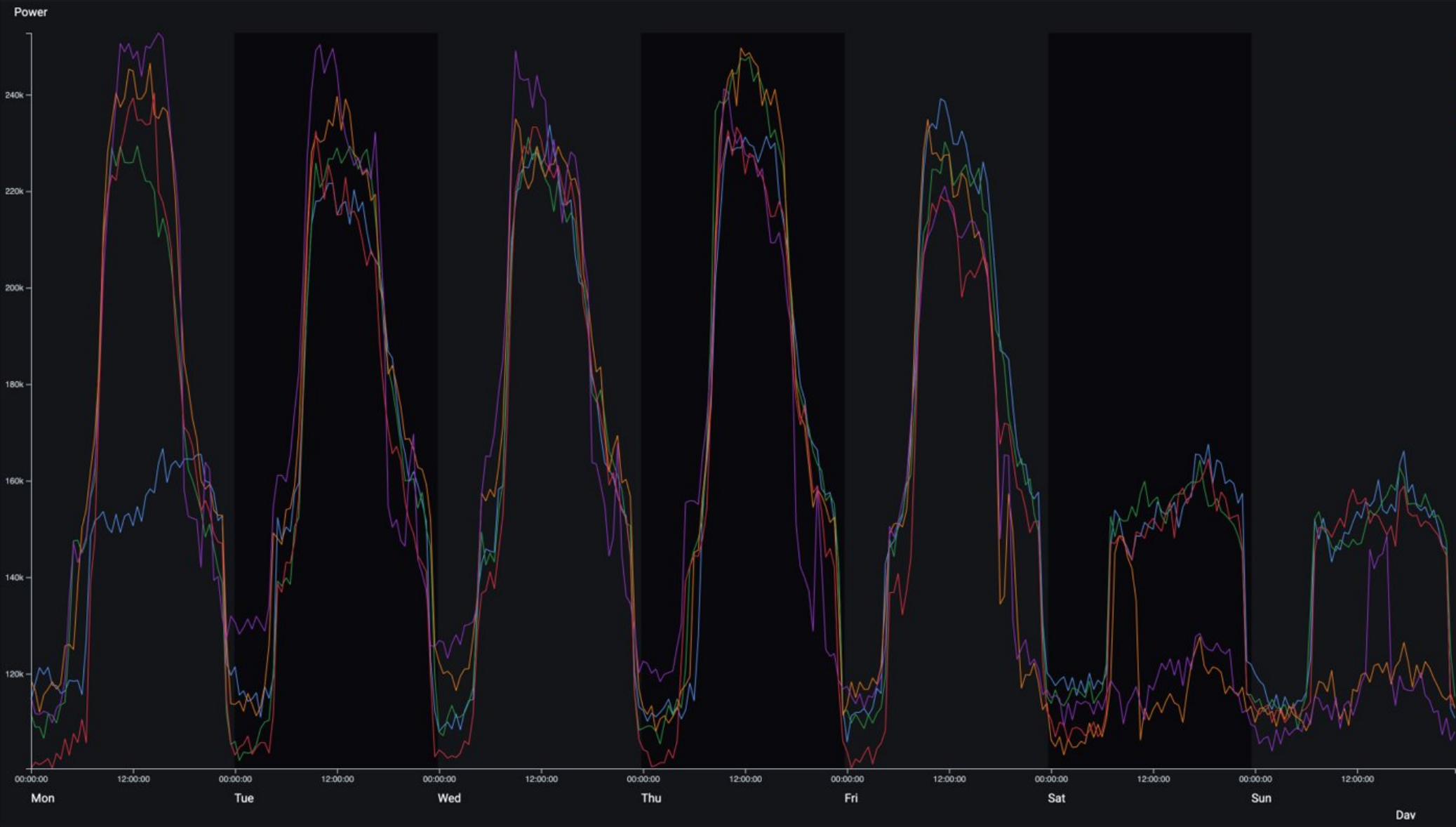
Some text





Building power

Building Building 2 ▾



Legend

- 1/6/2015 - 7/6/2015
- 8/6/2015 - 14/6/2015
- 15/6/2015 - 21/6/2015
- 22/6/2015 - 28/6/2015
- 29/6/2015 - 5/7/2015



Data from UoOtago used in our prototype

- 2015 historic values
- Some university buildings
 - Different levels of power metering
 - 4 buildings
 - Every 30 minutes
 - 4.1 million records
 - ~1 Gb file
- Weather from the Department of Physics weather station
 - Every 5 minutes
 - 148,000 records
 - 10.7 MB

Platform architecture—pipeline

Platform

