LUNCH & LEARN

March 26, 2021
1:00 - 2:00 pm EST

FEATURING:

Ameren Accelerator PingThings
Lunch & Learn Outline

1. Introductions
2. Project Overview
3. Platform Overview + Quick Demos
4. Technical Deep Dive
5. Conclusions
6. Questions
Introductions
Ameren Team

Full Time Team

Krushna Debata
Data Engineer

Scott Hixson
Grid of the Future & Analytics Solutions Manager

Patrick Doran
Scrum Master

Ankush Hore
Data Scientist

Cody Davis
Distribution/DERMS Engineer

James Kabbes
Data Scientist

Lynda Lin
M.S. Statistics
PhD Candidate Psychology

Scotlyn Belew
B.S Mech. Engineer

Ravi Donepudi
PhD Mathematics

August Gress
Computer Engineering

Dara Zirlin
PhD Mathematics

Steven Sohn
B.S Information Systems and Marketing
Speakers

Sean Murphy, CEO
sean@pingthings.io

Sean founded and quickly built a million-dollar consulting firm focused on data science and AI after completing his MBA at Oxford. Previously, he served as a senior scientist at the Johns Hopkins University Applied Physics Laboratory focused on machine learning, anomaly detection, and high performance and cloud computing.

Dr. Chris Ryan, Data Scientist
chris@pingthings.io

Chris is a data scientist broadly experienced in both data product development and analytics. He received his PhD in 2013 from the University of California, Berkeley, where he developed computer simulations to study shape transformations in biological cells. He has over 11 years of experience in data science & statistical physics research.
Overview

• Go to market in 2017
• 16 full time employees (tripled in size in 2020)
• Entire team is technical, no sales or marketing (yet)
• Fully remote team – MD, VA, DC, CA, WA, and NY
• Expected to hit 30 full time in 2021

Financial Background

• Raised ~$1.6M in seed funding since 2017
• Supported by multiple DOE, ARPA-E, and NSF grants totaling over $8M
• Large, multi-year contract with Dominion Energy
• Customers include transmission, distribution, telecom, and transportation
Open Innovation 2018
"A National Infrastructure for Artificial Intelligence on the Grid"
$6.25M over 3 years

Big Data Analysis of Synchrophasor Data
"Combinatorial Evaluation of Physical Feature Engineering and Deep Temporal Modeling for Synchrophasor Data at Scale"
$1.1M over 18 months
Project Overview
What: Data and Use Case

• 8TB of AMI data
• Solve uses cases
  • Secondary connectivity
  • Phase identification
• Strawman comparisons and synthetic benchmarks only go so far
How: Platform Comparison

Ameren AWS Services

1. Data collection from sensors.
2. Setting up AWS EMR Cluster.
3. Algorithm development in pyspark.

PingThings

4. Receive data from Ameren.
5. Set up the data in PT platform.
6. Refactor the same algorithm for PT platform.
<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PredictiveGrid platform setup</td>
</tr>
<tr>
<td>2</td>
<td>Ingestion of anonymized AMI data</td>
</tr>
<tr>
<td>3</td>
<td>Development of benchmark methodology</td>
</tr>
<tr>
<td>4</td>
<td>Training and support</td>
</tr>
<tr>
<td>5</td>
<td>Data quality assessment and reporting</td>
</tr>
<tr>
<td>6a</td>
<td>Prototype analytics development</td>
</tr>
<tr>
<td>6b</td>
<td>Scale prototype analytics</td>
</tr>
<tr>
<td>7</td>
<td>Testing and evaluation</td>
</tr>
<tr>
<td>8</td>
<td>Final reporting</td>
</tr>
</tbody>
</table>

**Week**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Data Acquisition**

**Actual Project**

3/30/21
The PredictiveGrid™ platform for time series data offers exponentially accelerated performance for analytics, ML, and AI development and deployment without massive costs.
The PingThings Platform

**Data Sources**
- T&D Electric Utility
- Sensor Fleets
- RC / ISO
- Data Concentrators
- Solar Fleet
- Historians
- Wind Farm
- Individual Sensors
- EV Charging Fleet
- Edge Devices
- Research Partners

**PredictiveGrid™**
- **BTrDB**
  High performance, extremely scalable time series database for real world sensor data at scale.
- **DISTIL**
  High performance stream and signal processing for time series measurements.
- **User Defined Functions**
- **Topology Store**
- **Metadata Store**
- **Geospatial Data Store**

**Applications**
- **General Purpose**
  - Data Explorer
  - Dashboards
  - Notebooks
  - API and Access Control Layer
  - Alerting
  - Admin
  - Platform Mgmt

- **BI Tool Integrations**
- **Industry-Specific**
  - Solar Integration
  - Model Validation
  - Small Signal Analysis
  - Wildfire Detection
  - GMD/GIC
  - Capital Optimization

- **Customer and 3rd Party**
- **Applications**
  - App 1
  - App 2
  - App 3
  - App 6
  - App 9

**Secure, Scalable, and Reliable Infrastructure**

This document is proprietary. No part of this document may be disclosed in any manner without the prior written consent of Ping Things, Inc.
A New Data Structure for Time Series

Solved Hard Problems Others Haven’t Considered

Natively supports:

- nanosecond time precision
- dynamic sampling rates and multiresolution time series
- out of order insertion and data dropouts
- data versioning
- data quality assessment

State of the art data compression at every level of the tree.

We have a 5-year head start.
More Scalable

**Horizontal Scalability in Data and Analytics**

Scales to
- petabytes of data
- millions of streams
- MHz sampling rates
- 100’s of millions of points/s

**Faster**

**Blazing Fast Performance**
- >10-1000x faster
- Proven 10,000x faster for common operations
- Log(n) aggregations
- Read 15M pts/s/node
- Write 30M pts/s/node

**Built for Analytics**

**Analytics and AI at Scale**
- Custom distributed signal processing framework
- Rapidly develop and operationalize use cases
- Integrates with leading big data frameworks

**Cost Effective**

**Designed for ROI**
- Scale compute and storage independently
- Leading data compression
- Intelligent storage tiering to optimize cost and performance
- Multiple levels of redundancy
- Fully containerized
- Horizontally scalable
- Intrinsically distributed
- Cloud agnostic
Centralized

Distributed/Edge

Cloud

On-Prem

Fractal Architecture

- Consistency across levels
- No “black box” algorithms
- Identical APIs
- Same algorithms run centralized or at the edge
A Spectrum of Solutions

Generic Cloud Providers

- All assembly required – you build your own platform
- You write your own analytics
- Only pay for what you use
- Everything is billed
- Complex pricing
- Not industry specific
- Reasonable performance

Universal Sensor Analytics Platform

PingThings

- Built for performance and cost effectiveness
- No maintenance required
- Quickly build your own analytics
- Visualization built-in
- Optimized workflows for time series data
- Sensor agnostic – supports sensor data rates up to 1GHz

Traditional Industry-Specific Solutions

- Fixed functionality
- Expansion options limited
- Designed for data lock-in
- Just enough performance for base case
- Does not play well with other sensor types

“Any customer can have a car painted any colour that he wants, so long as it is black.”
Technical Deep Dive
Experiences

• Stood up components of AWS to assemble “platform”
• Extracted Data from Teradata and loaded to S3
• Developed prototype algos on laptops using small sample data
• Tried to refactor code using various AWS services
• Implemented utilizing S3 and EMR/Spark (only available option)

PingThings

• Created ingestors to load AMI data into platform (one-time step)
• Visually explored data to identify and understand data quality issues
• Rapidly prototyped algos in Jupyter Notebook
• Iterated and optimized algorithms to yield scale-out versions
Sequence of Steps
- Loading .parquet data files from S3 into EMR
- EMR performing:
  - Data transformations
  - Calculations

Calculations - EMR Cluster
- Cost ~ $1,000 per month
- 8 x r5.2xlarge instances
- Each r5.2xlarge:
  - 8 vCPU, (8000 series circa 2017-2019)
  - Turbo CPU up to 3.1 GHz
  - 64 GB of RAM
  - Up to 10 Gbps network

Time Loading Data: ~1 minutes
Total Time to Compute: ~60 minutes
Conclusion

• PingThings Predictive Grid is a high-performance solution for all grid time series data

• Working with Smart Meter data has greatly informed engineering efforts moving forward for PingThings:
  • Next generation query engine
  • Enhancements to visualization system and APIs
  • New perspectives on large scale compute requirements

• Ameren to review findings in detail to help direct our strategy for sensor data + analytics with cloud team and architects