IoTAbench: an Internet of Things Analytics benchmark

Martin Arlitt, Manish Marwah, Gowtham Bellala, Amip Shah, Jeff Healey, Ben Vandiver

Problem Statement

- Internet of Things (IoT) is expected to be a large market
  - IDC (2013) estimates 212 billion connected “things”, $8.9 trillion market by 2020

- Challenge for organizations – IoT is “Big Data”
  - How best to store and analyze sensor data?

- Our challenge: How to quantitatively evaluate Big Data platforms for IoT
  - No industry standard IoT benchmark
  - Only small empirical datasets available
  - Need to systematically control Big Data experiments

Solution

Create (and use) a benchmark

IoTAbench

- IoT can have many different applications
- We could build benchmarks for each of them under the IoTAbench umbrella
- Our initial focus is on a benchmark for smart metering in Utilities vertical

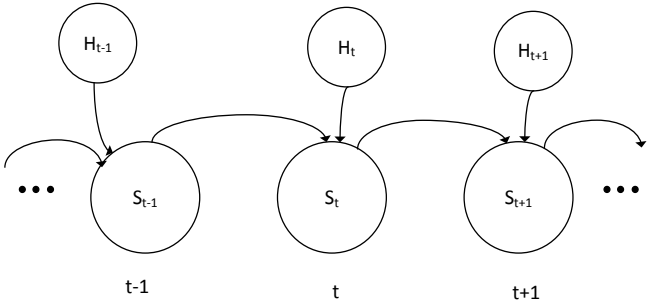
- A general solution involves:
  - Generation of synthetic data with realistic properties
  - Creation of a set of relevant “business” queries that stress the database system
  - A test harness to systematically control “Big Data” experiments

Synthetic Data Generator

- An augmented Markov chain model simulates consumption process

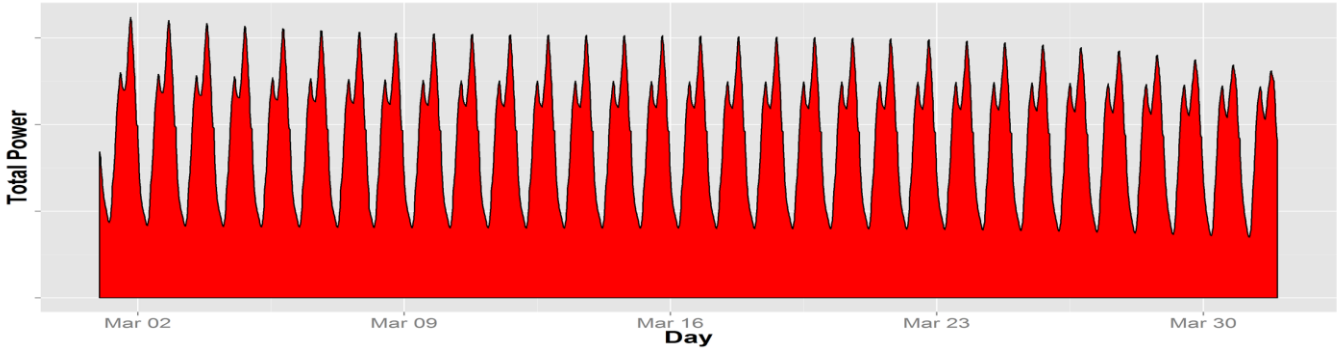
Training

- Train the model on a small empirical dataset
- We develop a model for each month of the year



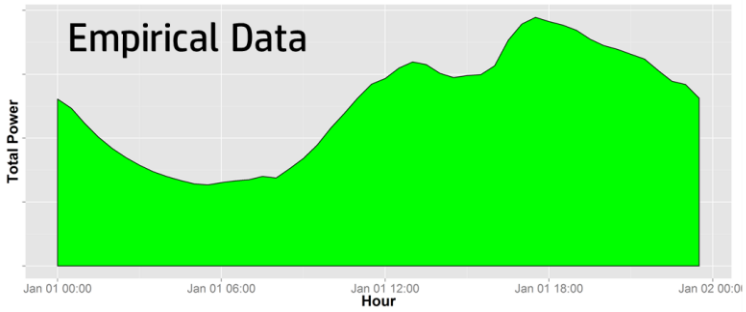
Generation

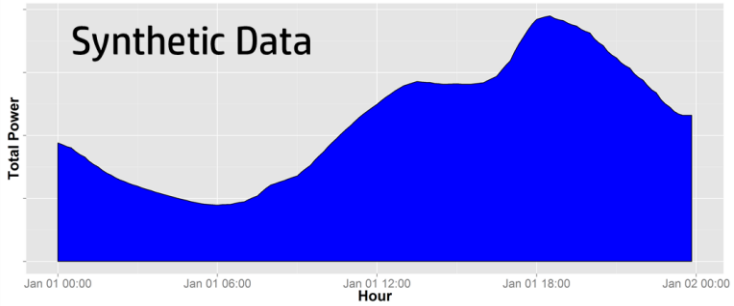
- A random walk on the chain generates synthetic data
- We can scale to arbitrarily large numbers of sensors
- A mixture model smooths the transition between months

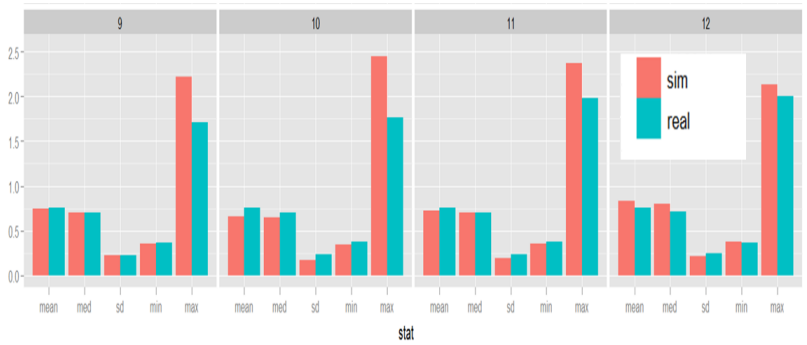


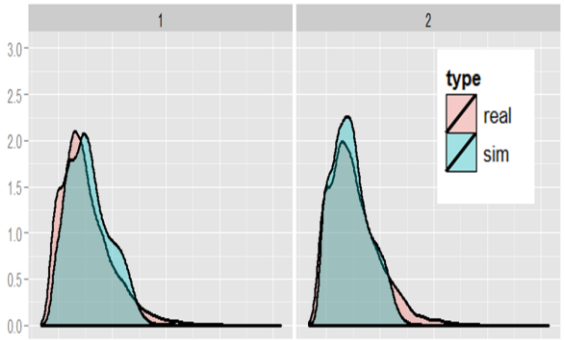
Synthetic Data Generator

Validation









Benchmark Queries

Smart Metering use case

	SELECT			WHERE			ORDER BY		
QUERY	TS KEY	METER KEY	Power Watts	TS KEY	METER KEY	Power Watts	TS KEY	METER KEY	Power Watts
Total Readings		<any>		✓					
Total Consumption			✓	✓					
Peak Consumption	✓		✓	✓					✓
Consumption Timeseries	✓		✓	✓			✓		
Top Consumers		✓	✓	✓	✓				✓
Time of Usage (ToU) Billing		✓	✓	✓	✓			✓	

System Under Study

Hardware: eight HP ProLiant DL380p g8 servers

- Each node configured with:
  - Two 8-core 2.6 GHz CPUs (Intel Xeon E5-2670)
  - 128 GB memory
  - 2@ 300 GB mirrored drives for OS and database catalog
  - 22@ 900 GB drives (RAID 10) for data
  - Two dual-port 10 Gb Ethernet NICs
  - RHEL 6.4

  Bottom node is benchmark controller

Software: HP Vertica Analytics Platform

Experimental Design

Use Case:

A large electric utility with:

- 40 million meters
- 10 minute reading frequency
- 1% of readings missing

	Experiment 1 (Vertica 6.1.2)	Experiment 2 (Vertica 7.0.0)
Total Readings	15.1 trillion	22.8 trillion
Data Duration	7 years 2 months	10 years 10 months
Raw Data Size	0.478 PB	0.726 PB

Experimental Methodology

Month1: generate → load → repair → analyze → design → Re-analyze

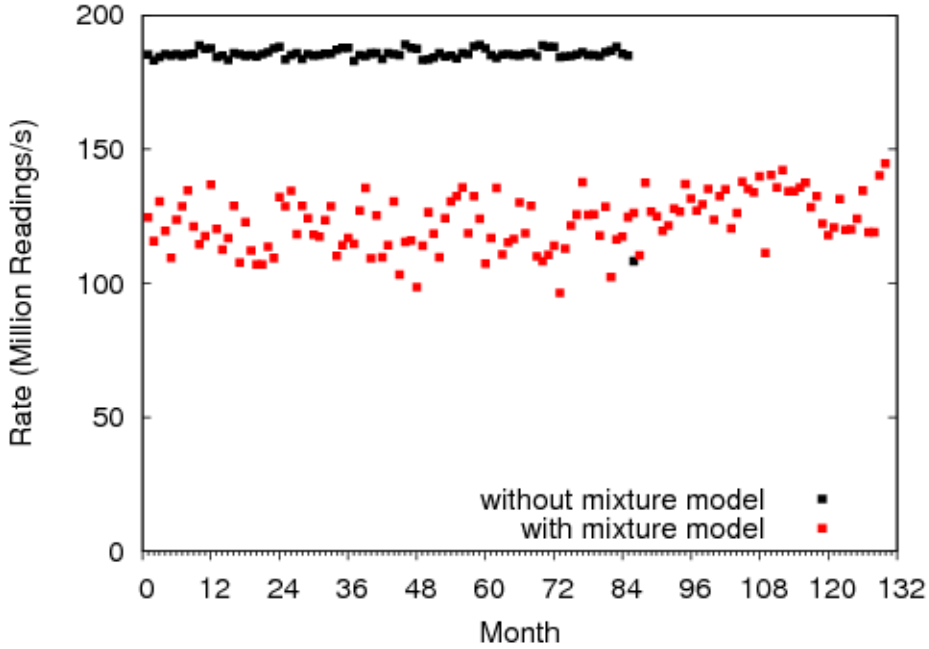
Month 2-N: generate → load → repair → analyze

Once N is determined:

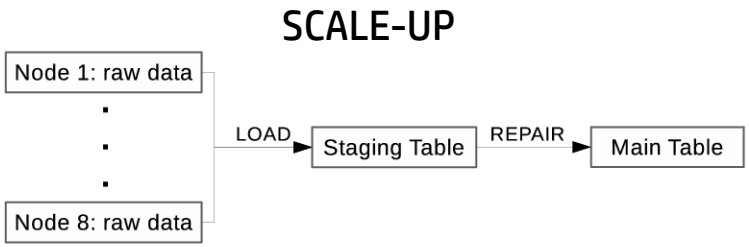
Month 1-N: Re-analyze

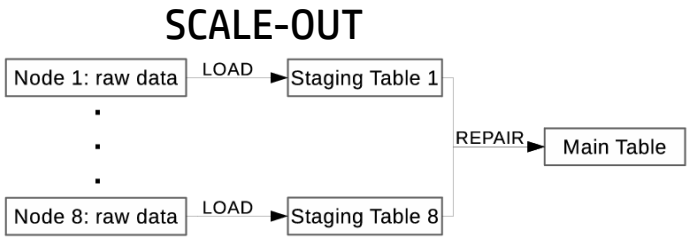
Once N is reached, exit

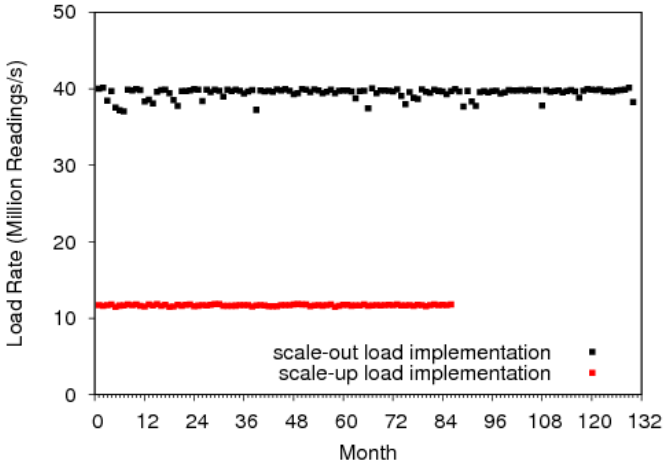
Generator Performance

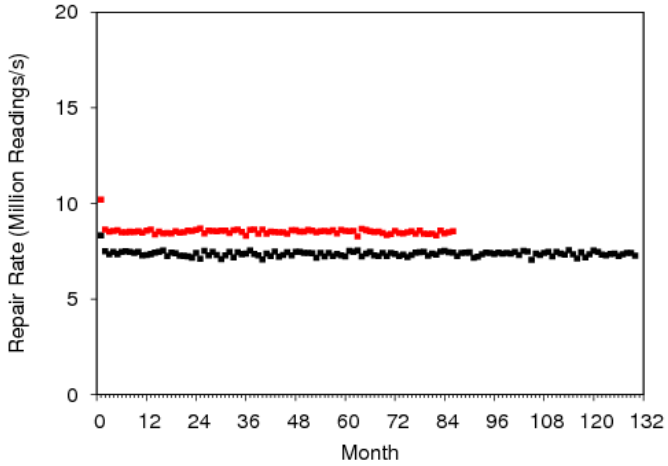


Load and Repair Performance





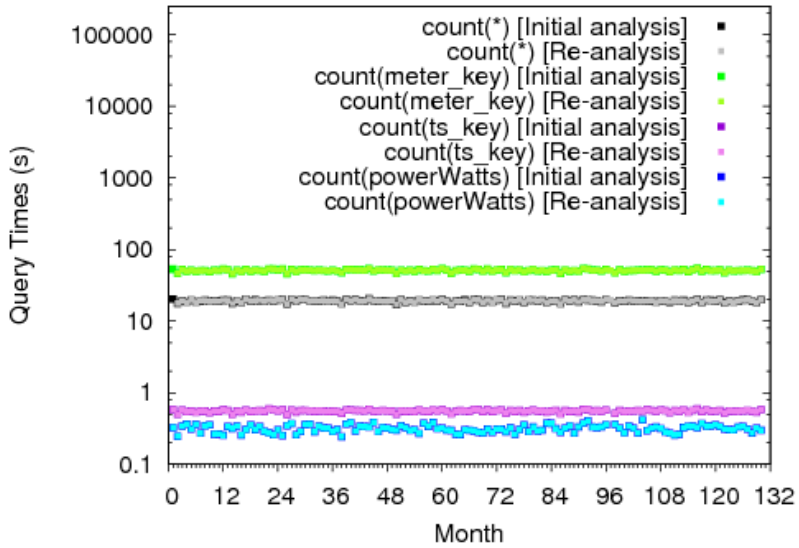




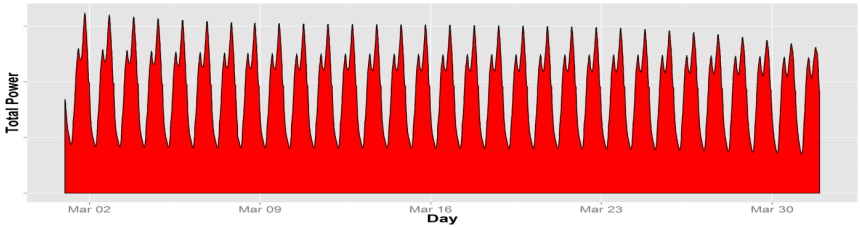
Total Readings

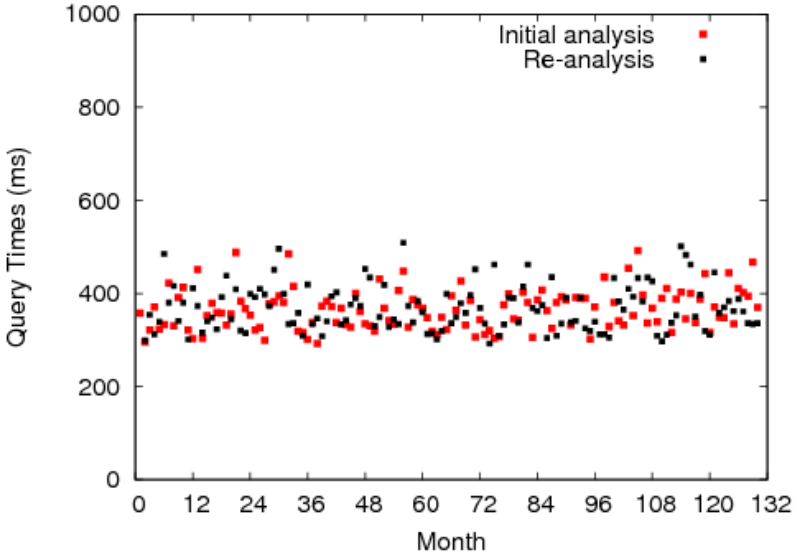
Comparison of:

- Different encoding methods
- Initial analysis and re-analysis

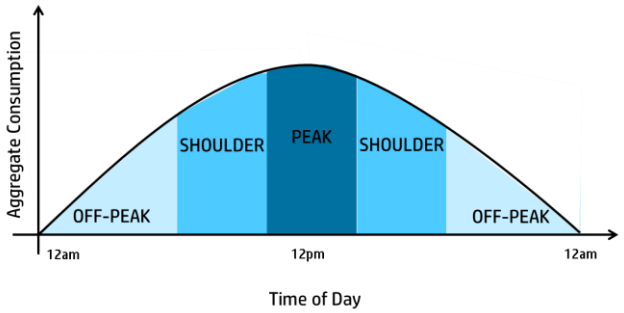


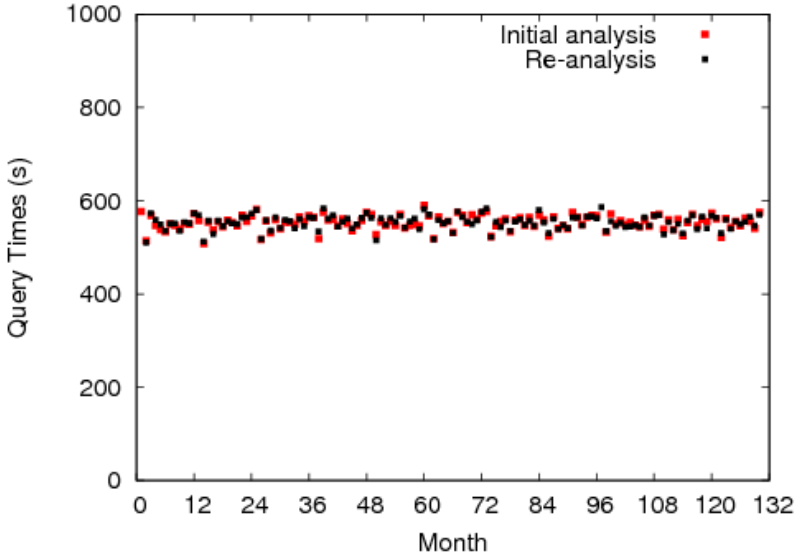
Consumption Timeseries





Time of Usage Billing





Analyzing the Entire Dataset

Query	Time (s)
Total Readings	10.859
Total Readings (ts_key)	10.131
Total Readings (meter_key)	6,673.220
Total Readings (powerWatts)	11.549
Total Consumption	10.636
Peak Consumption	149.034
Consumption Timeseries	155.412
Top Consumers	51,135.932
Time of Usage Billing	54,695.257

Summary

- Internet of Things is an important emerging area
- IoTAbench can quantitatively evaluate the performance of IoT solutions
  - Enable “apples-to-apples” comparisons
- We conducted two extensive IoT/Big Data experiments

Next Steps

- Assist in creating an industry standard IoT benchmark leveraging IoTAbench
- Extend IoTAbench to evaluate more advanced analytics features
- Develop additional benchmark examples with the IoTAbench portfolio
  - E.g., benchmarks tailored to other verticals

Thank you

Contact: martin.arlitt@hp.com