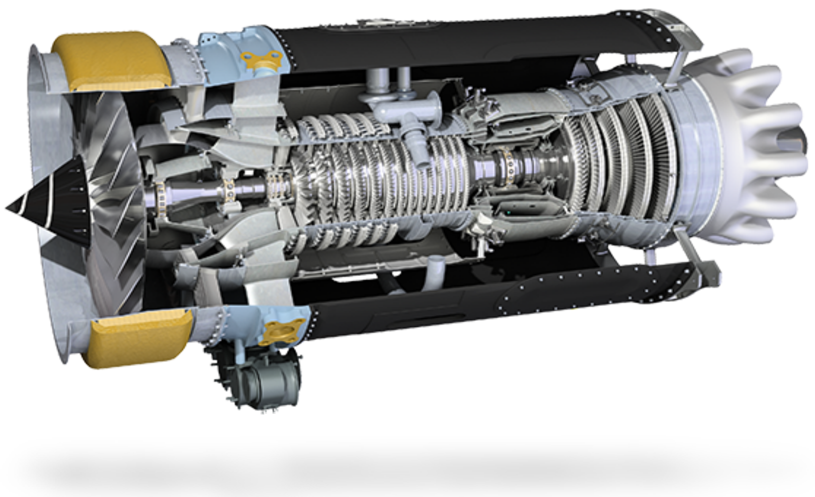
Scalable Architecture for Anomaly Detection and Visualization in Power Generating Assets

Paras Jain, **Chirag Tailor**, Sam Ford, Liexiao (Richard) Ding, Michael Phillips, Fang (Cherry) Liu, Nagi Gabraeel, Polo Chau

Background



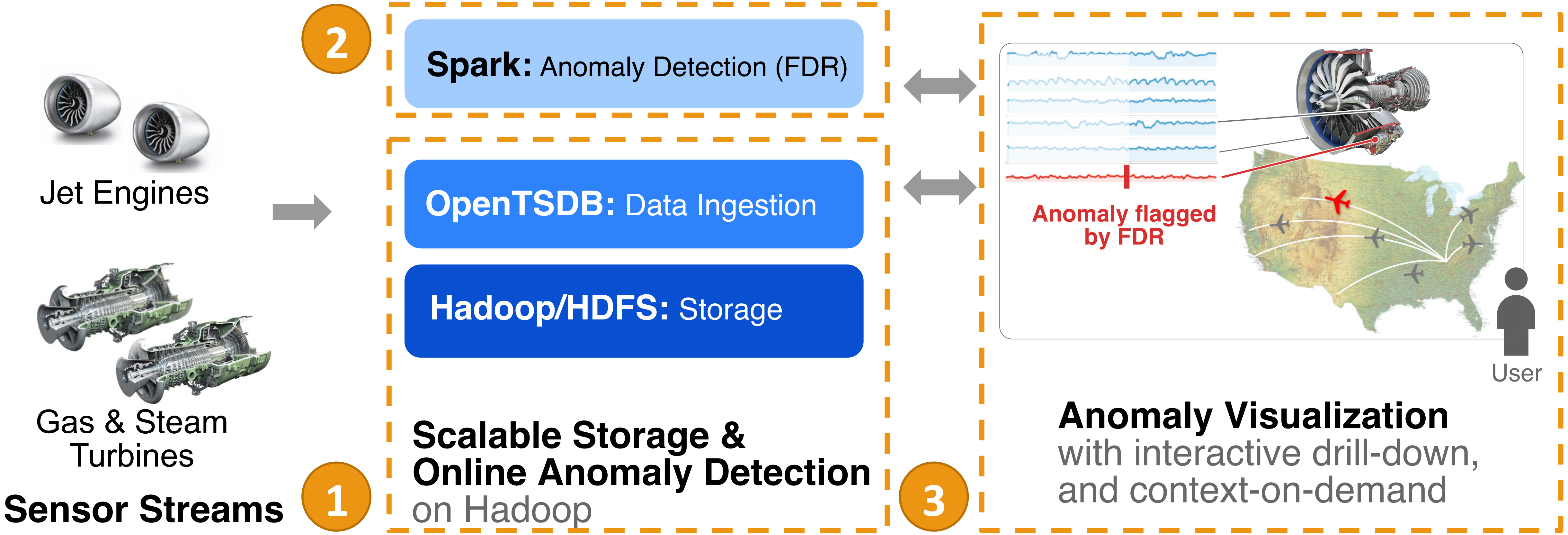
Power generating assets such as jet engines and gas turbines

- Each unit instrumented with 1000s of sensors to signal incipient faults
- Difficult for humans to monitor
- Algorithms attempt to predict asset failure by detecting anomalies

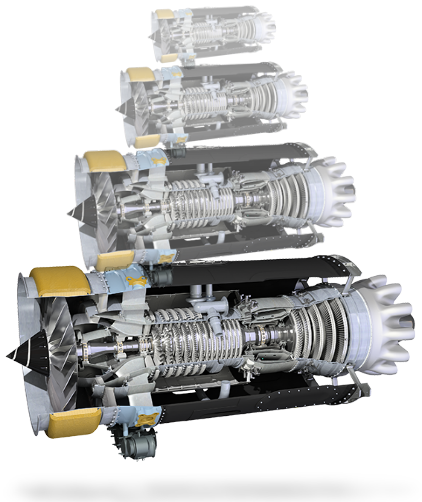
Key Challenges

- **Storage and Ingestion.** Huge volume of data from many machines in real-time.
- **Anomaly Detection.** Prevalence of false alarms leads to unnecessary downtime and maintenance.
- **Visualization.** Lack of an integrated visualization platform to understand and analyze flagged anomalies.

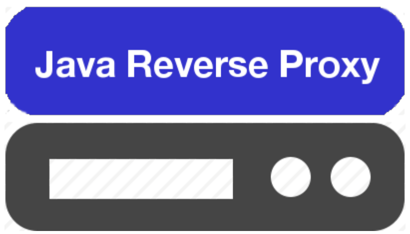
System Overview

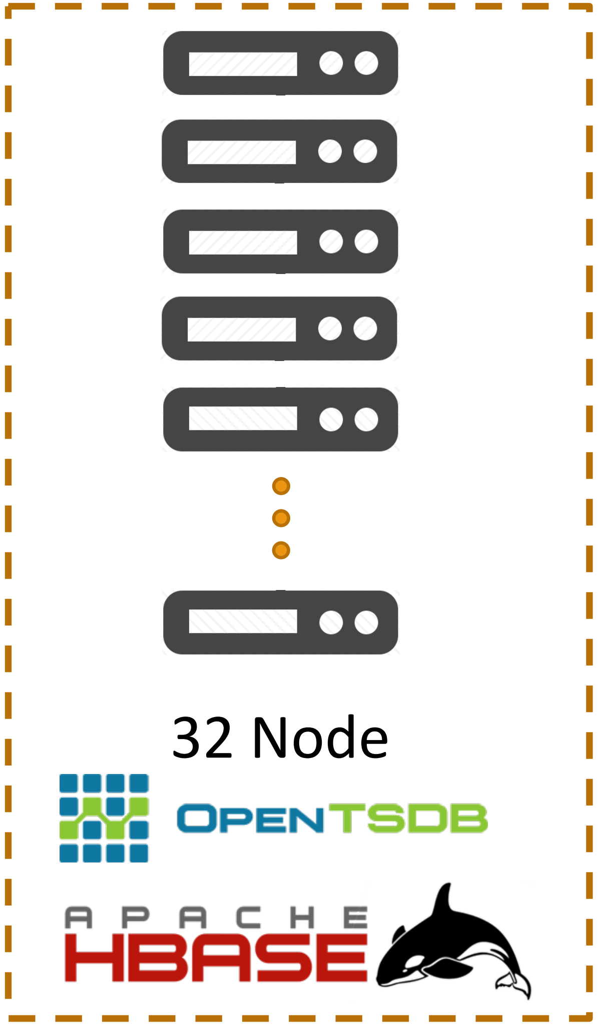


1 - Scalable Data Ingestion & Storage Architecture



100 Units
x 1000 Sensors
@ 1HZ





Goal Ingestion Rate: **100,000 sensor readings per second**

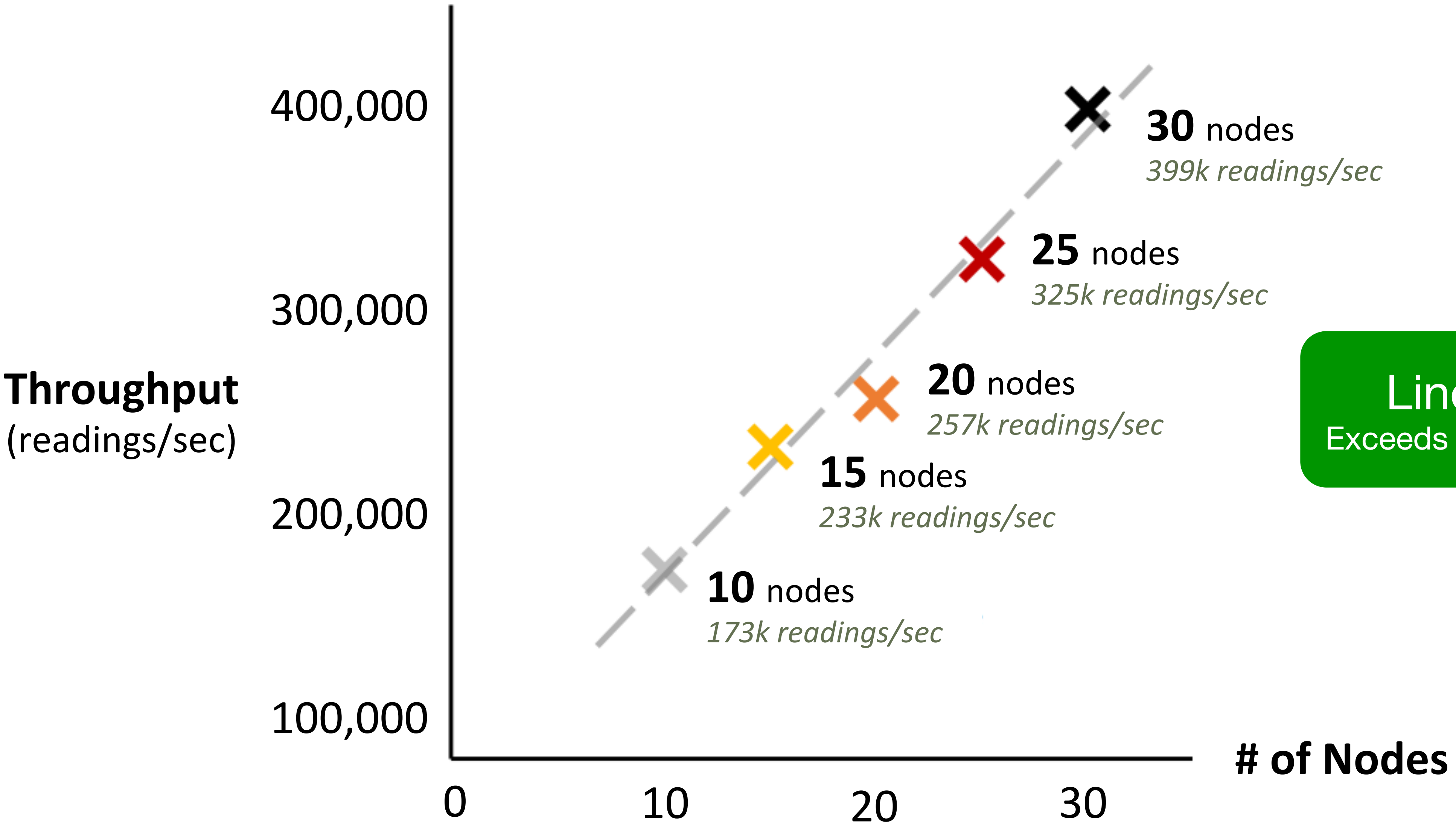
1 – Simulated Training Dataset

- **100 units with 1000 sensors producing readings at 1Hz.**
  - Similar number of units owned by a regional energy provider.
  - On the order with 3000 sensors in Siemens SGT5-8000H gas turbine*.

- **Anomalous behavior modeled in dataset:**
  - Pure random noise.
  - Pure random noise plus gradual degradation.
  - Pure random noise plus sharp shift.

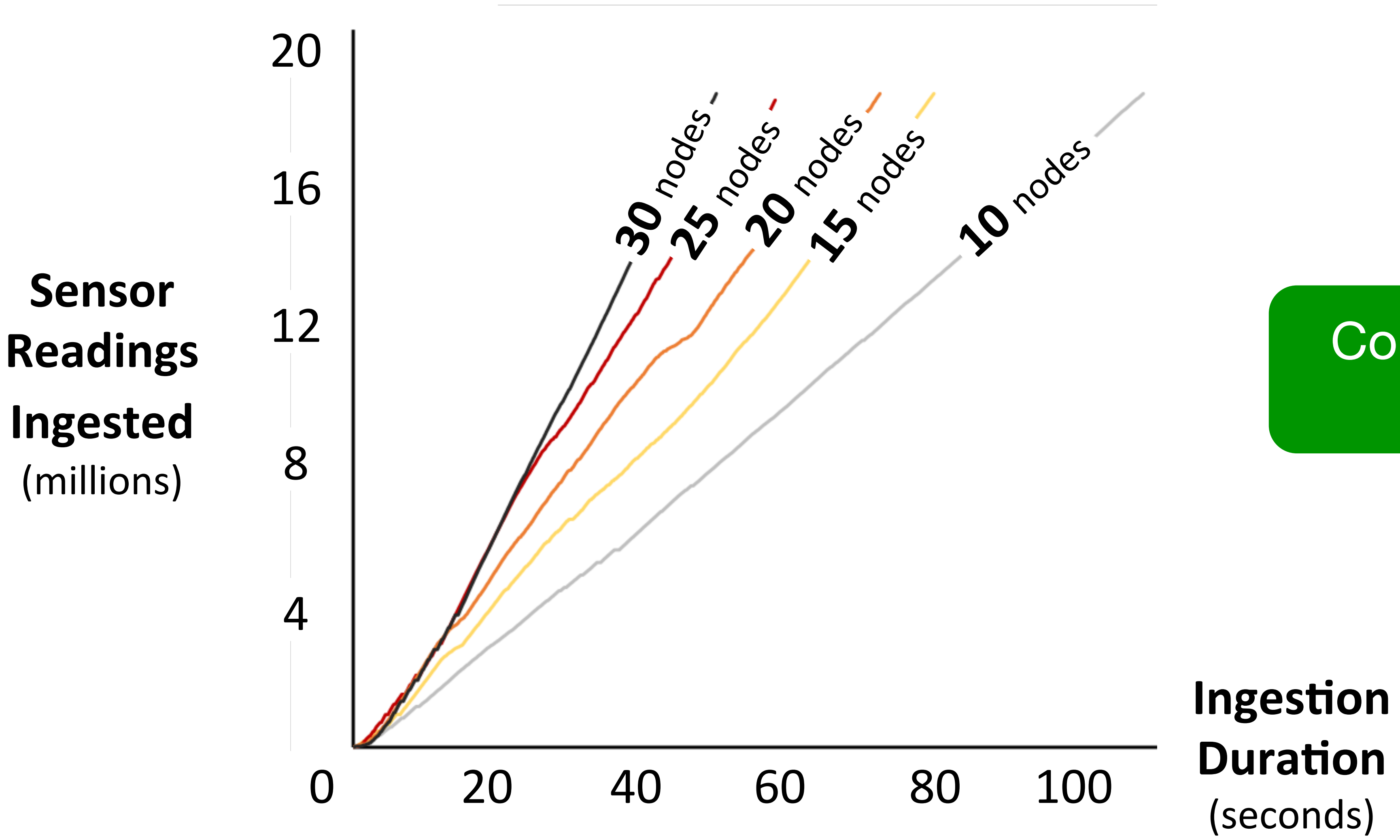
*P. Ratliff, P. Garbett, and W. Fischer, “The new siemens gas turbine sgt5-8000h for more customer benefit,”

1 - Scalable Data Ingestion & Storage Results



Linear Scale Up
Exceeds 100k samples/sec goal

1 - Scalable Data Ingestion & Storage Results



Constant & Stable Ingestion

1 - Interesting Findings

1. **Salting.** HBase keys generated by OpenTSDB must be salted since continuous value timestamps all map to the same HBase node.
2. **Backpressure.** HBase does not provide backpressure to OpenTSDB.

OPENTSDB APACHE HBASE

2 - Flagging Anomalies with Low False Alarm Rates

We use the **False Discovery Rate** (FDR) algorithm.

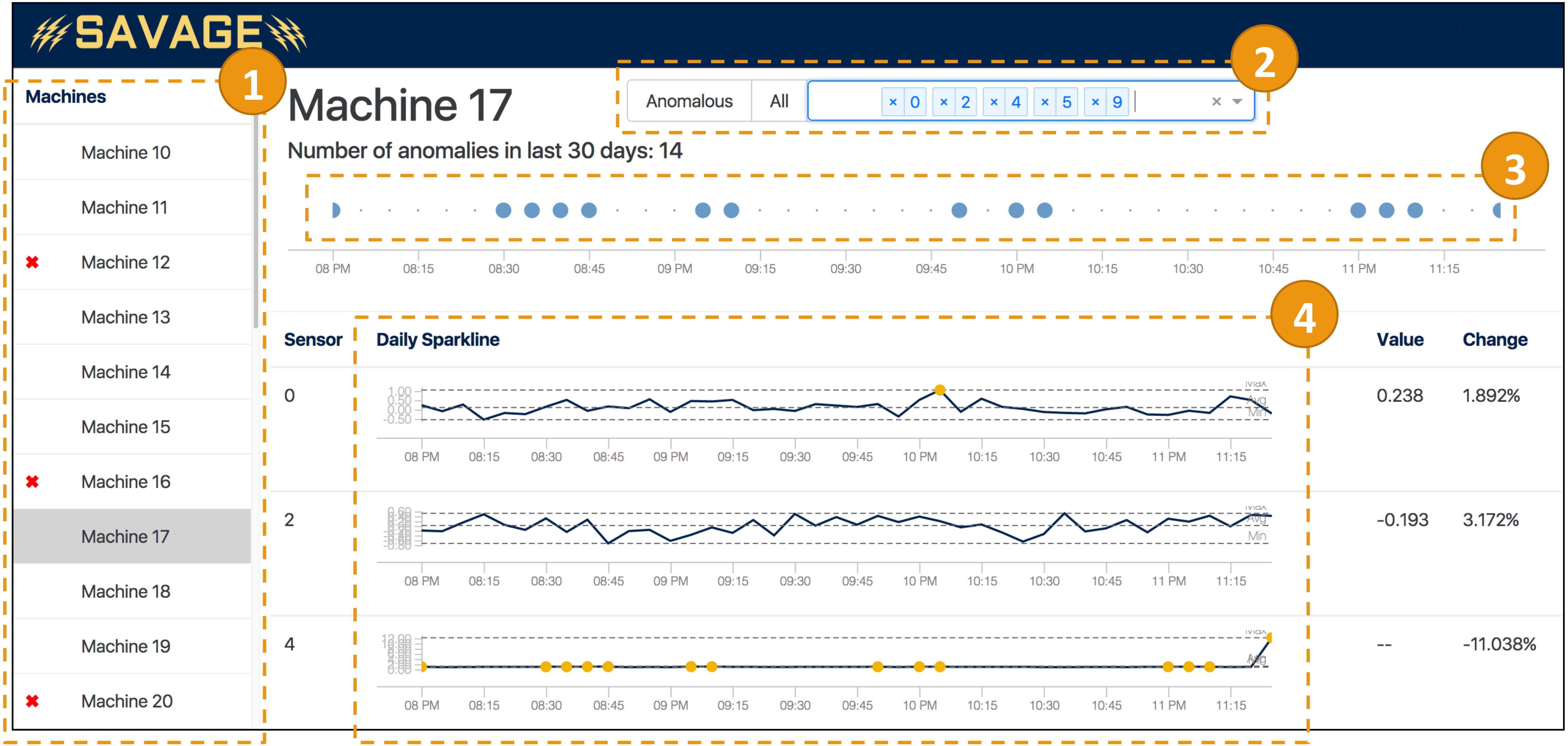
1. First introduced by **Benjamini and Hochberg in 1995** and used in multiple inference clinical trials*.

2. **Suppresses false alarms:** Performs a test on an increasing ratio of the original significance level for each sensor's z-score.

3. **Scalable**: our implementation using Spark processes over **939,000 sensor samples per second**

*Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing,"

3 - Anomaly Visualization

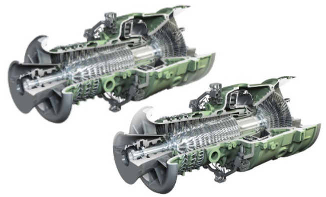


Ongoing Work

- Scaling up ingestion and analysis throughput with additional nodes.
- Migrate anomaly detection algorithm to Spark Streaming for online evaluation.
- Evaluate our system with domain users and industry partners like General Electric (GE).

Thanks!

Jet Engines



Gas & Steam Turbines

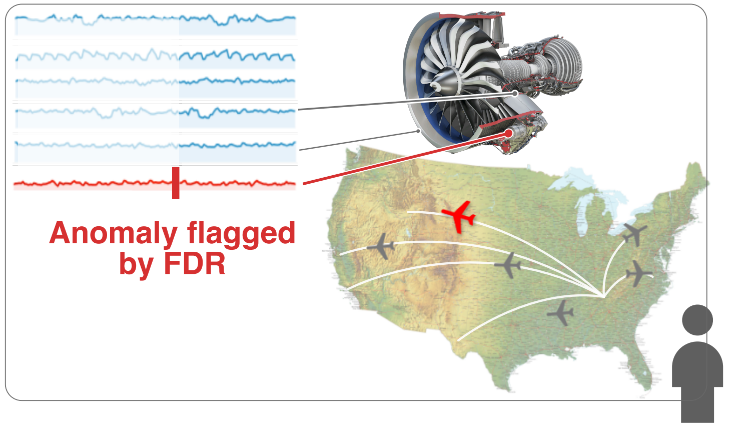
Sensor Streams

Spark: Anomaly Detection (FDR)

OpenTSDB: Data Ingestion

Hadoop/HDFS: Storage

Scalable Storage & Online Anomaly Detection on Hadoop



Anomaly Visualization with interactive drill-down, and context-on-demand

Scalable Architecture for Anomaly Detection and Visualization in Power Generating Assets

Georgia Tech

Paras Jain, **Chirag Tailor**, Sam Ford, Liexiao (Richard) Ding, Michael Phillips, Fang (Cherry) Liu, Nagi Gabraeel, Polo Chau