The Next Wave of Big Data Analytics: Internet of Things and Sensor Data

November 6, 2014
Hannah Smalltree, Director
There’s big data, then there’s really big data from the Internet of Things. IoT is evolving to include many data possibilities like new types of event, log and network data. The volumes are enormous, generating tens of billions of logs per day, which raise data challenges. Early IoT deployments are relying heavily on both the cloud and managed service providers to navigate these challenges. Learn about IoT, big data and deployments processing massive data volumes from wearables, utilities and other machines.
Defining “The Internet of Things”

- Gartner: The Internet of Things (IoT) is the network of dedicated physical objects ("things") that contain embedded technology to sense or interact with their internal state or external environment. The IoT comprises an ecosystem that includes things, communications, applications and data analysis.
- Varying degrees of connectivity and data involved with different types of "things"; Connected devices
- IDC: Approximately 212 billion "things" globally by the end of 2020 -- including 30.1 billion installed "connected (autonomous) things."
Connecting Big Data and IOT

• IOT: the network of dedicated physical objects (things) that contain embedded technology to sense or interact with their internal state or external environment.

• Big Data: high volume, velocity and/or variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.

• IOT is a HUGE source of Big Data.
Why IoT Data is Different

**Traditional Data**
- People, Transactions, Things
- Structured
- Hundreds, Thousands per Day
- Variable (often inside firewall)
- Customer Records, Transaction Records, Product Information, Business Data

**BIG Data**
- Actions, Observations, Events
- Multi-Structured
- Millions, Billions per Day
- Cloud-connected
- App Logs, Web Logs, Clickstream, Event Logs, Timeseries Data, Mobile Data

**IOT Data**
- Actions, Observations, Events, Interactions
- Multi-Structured
- Billions per Day
- Remote, Mobile
- Sensor Logs, Mobile Data, Event Logs, Operational Data
IoT Data Adds Up Quickly

• 20 sensors producing data every 5 min
  – 240 records/hour
  – 5760 records/day
  – 172,800 records/month
  – 2,102,400 records/year

• Multiply by 10,000 devices (not that many!)
  – 2,400,000 records/hour
  – 57,600,000 records/day
  – 1,728,000,000 records/month
  – 21,024,000,000 records/year
IoT Data Opportunity

Lots of Data:
- User Actions
- Sensor Logs
- Device Data
- Observations
- Errors
- Events
- Actions

Lots of Questions:
- How exactly are my products used?
- How can I make them better?
- How do I acquire more customers or raise avg. revenue per customer?
- How do I reduce operational costs?
Challenges of IOT Data

• Massive complexity, many components
• Lack of standards, rapidly evolving
• Network, device management often the focus
• Remote data collection
• Time to market, competitive landscape
• Core competency, focus
Why Cloud for IoT?

- Remote data
- Massive volumes
- Need for aggregation
- Minimize data movement/latency
- Bandwidth costs
- Simplicity
- Efficiency
- Access
How Treasure Data’s Cloud Service Works

Treasure Data Cloud Service

Collect
Real-time
Store
Scalable
Analyze
Use SQL or BI

Analytics Tools
SQL
Excel
Pioneer, #1 Auto Electronics Provider
Agile Wearables Innovation

Challenge

• Collecting data for R&D, improving devices and understanding customer interactions with device
• Existing tools (packaged analytic app) do not provide the scale, performance and flexibility required for analytics – it can be easy to collect data, hard to analyze it

Benefits

• Fast <2 week deployment
• Seamless scalability, analytic flexibility and ad hoc capabilities
• Use standard SQL on the data; easy to roll out to users
• Helps focus R&D for product improvements to maintain competitive edge
• Helps focus on analytics around understanding customers to reduce product returns and churn

Simplifying the Wearable/Mobile Data Pipeline

Excel

SQL
Utilities Provider
Key Considerations for IoT data processing

What should people think about when deciding how to handle IoT Data Processing?

Technical
• Data collection, real-time, batch, reliability, flexibility
• Data movement, latency, reliability, bandwidth
• Data storage style, scaling, tuning, maintenance, upgrades
• Analytics functions, ad-hoc vs. canned, SQL access
• Export/integration with other systems & processes

Business
• Available skills for setup, mgmt.
• Available skills for analysis
• Time to implement
• Costs – startup, ongoing
• Flexibility to change
• Tool/process impact
• Support/upgrades/maintenance
• Monitoring/Security
About Treasure Data
How Treasure Data Works

Treasure Data Cloud Service

Collect
Store
Analyze

Real-time
Scalable
Use SQL or BI

Analytics Tools
SQL
Excel

Copyright ©2014 Treasure Data. All Rights Reserved.
Quick Facts

About Treasure Data

• Cloud service
• Collect, store and analyze massive BIG data volumes
• 100% managed, monitored; SaaS model
• Named Gartner Cool Vendor in Big Data – 2014
• 100+ Corporate Customers (thousands free)
• Service launched Sept 2012
• Based in Mountain View, CA
Key Differentiators

• Cloud Service Model, 100% managed
• Fast setup time (<2 weeks)
• End to End, integrated, “all in one” service
• Pricing Model (monthly subscription, scale as you grow)

Technology Stack Differentiators
• Real-time data collection technology; Semi-structured
• Export and integration capabilities
• SQL, Ad hoc, Multiple Engines, Optimized Performance
• Schema-flexible, scalable data storage
Why are customers choosing Treasure Data for IOT Data Analytics?

Technical
• Unique technology for data collection
• End to end “all in one” system - collect, store and analyze
• SQL access to raw product data
• Ask any questions of your data, flexible ad hoc analyses are now possible

Business
• Managed Service, fully supported and monitored, no need to staff
• 2 week setup, get up and running fast
• Monthly subscription free, flat rate, predictable
• Flexible, easy integration, developer-friendly
Treasure Data Customer Focus Areas

• **BIG data from web applications, large sites and web engines:**
  – Large websites
  – Software, web & ad tech
  – Media, entertainment, gaming
  – Retail, e-comm

• **BIG data from mobile apps and devices:**
  – Mobile apps
  – Media, entertainment, gaming

• **BIG data from products or devices:**
  – Automotive telematics
  – Wearables
  – Manufacturing
  – Energy/utilities

---

Behavioral Analytics  
Data Refinery  
Operational Analytics  
Research & Dev.  
Customer Reporting  
New Services
<table>
<thead>
<tr>
<th>Trusted By Customers Across Industries</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internet/social</strong></td>
</tr>
<tr>
<td>wish, ContextLogic, GETJAR, YAHOO! JAPAN, COOKPAD (<a href="http://cookpad.com">http://cookpad.com</a>)</td>
</tr>
<tr>
<td><strong>Ad-tech/media</strong></td>
</tr>
<tr>
<td>MobFox, DashBid, Apper, dentsu, ADREPUBLIC, viki, HAKUHODO</td>
</tr>
<tr>
<td><strong>Gaming</strong></td>
</tr>
<tr>
<td>GREE, SPRYFOX, GameSalad</td>
</tr>
<tr>
<td><strong>IoT</strong></td>
</tr>
<tr>
<td>Pioneer, pebble, NTT docomo</td>
</tr>
<tr>
<td><strong>SaaS</strong></td>
</tr>
<tr>
<td>heroku, Cloud9 IDE, Yahoo! Japan</td>
</tr>
<tr>
<td><strong>Other Enterprise</strong></td>
</tr>
<tr>
<td>EQUIFAX, MUJI, NTT docomo, CLASS DOJO, MONEX</td>
</tr>
</tbody>
</table>
Q&A & Resources

TreasureData.com

• Read Whitepapers and Info
  – Untangling the Internet of Things whitepaper
• Request a Demo
• Try the Treasure Data Cloud Service for Free

Thank You!