Internet of Things
ANALYZING INTERNET OF THINGS USING BIG DATA ECOSYSTEM

Internet of Things matter for...
- Industrial Manufacturers
- Transportation
- Healthcare, Life Sciences
- Financial Services
- Retail
- Telecom and Media
IOT MARKET OPPORTUNITY

Global Product Data Interoperability Summit | 2015

• “IoT will grow to a $19 trillion industry by 2022.” (Cisco, revised)

• “Incremental revenue will exceed $300 billion by 2020.” (Gartner)

• “The Industrial Internet of Things will transform companies and countries, opening up a new era of economic growth and competitiveness.” (Accenture)

• “This will inevitably create entirely new markets to buy and sell algorithms, generating significant incremental revenue for existing companies and spawning a whole new generation of specialist technology start-ups.” (Gartner)
The Power of 1

Global Product Data Interoperability Summit | 2015

Driving Outcomes That Matter

- Increasing Freight Utilization Rail
- Predictive Maintenance Healthcare
- Predictive Diagnostics Power

One Percent Improvement Equals

- $27B Industry Value by Reducing System Inefficiency
- $63B Industry Value by Reducing Process Inefficiency
- $66B Industry Value with Efficiency Improvements In Gas-fired Power Plant Fleets

Source: General Electric
THE INTERNET OF THINGS JOURNEY

Global Product Data Interoperability Summit | 2015

STORE
• Structured
• Unstructured
• High Volume
• High Velocity

ANALYZE
• Predictive Analytics
• Machine Learning
• Advance Data Science
• Real-time Analytics

DEVELOP
• Advanced Analytic Pipelines
• Real-time Analytical Applications
• Global Scale Data-Driven Applications
• Enterprise, Consumer, IoT, and Mobile

INNOVATE
• Agile Dev Expertise
• DevOps
• Hybrid Cloud
• Continuous Delivery
• Closed Loop Applications

ENTERPRISE PAAS

AGILE DEVELOPMENT

PREDICTIVE ANALYTICS

BIG DATA
LARGE ENTERPRISE BIG DATA TROUBLE

80% of CEOs thinking data mining and analysis are strategically important (1)

0% of CIOs think their IT infrastructure is fully prepared for big data (3)

44% of new applications failed to meet performance expectations (5)

But…

4% of companies use analytics effectively (2)

30% of companies have deployed advanced analytics, 11% big data analysis (4)

90% of companies allocate at least 2X more cloud capacity than needed to ensure performance (6)

(1) 2015 PWC CEO Survey  (2) 2013 Bain and Company - The Value of Big Data  (3) 2014 IT Infrastructure Conversation - IBM  (4) Ernest and Young - 2014 Enterprise IT Trends and Investments  (5) 2014 Riverbed Technologies - The Transformers  (6) 2014 ElasticHosts CIO Study
The Data Divide

**BIG DATA CHASM**

- Digital universe is estimated to grow from 4.4 ZB in 2013 to 44 ZB in 2020
- 35% of the digital universe will contain valuable information
- >5% of potential useful data is analyzed
- <0.5% of operationalized data is being operationalized
Ease-of-Use more pressing issue than cost *

Fragmentation

Complexity

Constraints

* Dimensional Research, March 2015, Internet of Things Meets Big Data and Analytics: A Survey of IoT Stakeholders
Nearly 90 percent of IoT project stakeholders believe that more flexible analytics would significantly increase ROI. *

* Dimensional Research, March 2015, Internet of Things Meets Big Data and Analytics: A Survey of IoT Stakeholders
JOURNEY TO AN AGILE DATA-DRIVEN ENTERPRISE

Perform advanced analytics
Discover insights

Modernize data infrastructure

Deploy analytic apps and automate at scale
MODERNIZE DATA INFRASTRUCTURE

Global Product Data Interoperability Summit | 2015

**REQUIREMENTS**

- Elastic, Scale-out storage and processing
- Flexible data types and pipelining
- Cloud friendly and open-source based

**BENEFITS**

- Higher quality analytics
- Lowered storage/processing cost
- ETL on demand: low operational cost
- Expanded use cases
- Less fragmented ecosystem
- Reduced vendor lock-in

Cloud friendly and open-source based
ADVANCED ANALYTICS

Global Product Data Interoperability Summit | 2015

REQUIREMENTS

- Massive stream processing
- SQL-compliant batch and interactive queries
- Machine learning and advanced analytics

BENEFITS

- Internet of Things use cases
- Rapid time to insights
- Leverage existing skills and tools
- Rapid time to insights
- Solve business problems
- Predictive insights: proactive execution
ANALYTIC APPS AND AUTOMATION AT SCALE

**Requirements**

- Resilient, scale-out messaging and object storage
- Agile analytic app-dev with enterprise PaaS
- Low-latency, distributed in-memory transactions

**Benefits**

- Reduced time to insights
- Flexible ingestion: low operating cost
- Reduced time to action
- Low ‘analytics ↔ app-dev’ integration cost
- High performance: low operating cost
- Transactional safety: business critical ops
AGILE – What is it?

TRANSPARENCY  All aspects are visible and known

INSPECTION  Identify unacceptable variances

ADAPTABILITY  Adjust quickly and effectively
Migrating from a Reactive, Static and Constrained Model…

Ingest → Store → Analytics

**Coding based**
No real-time information
Based on expensive ETL

**Hard to change**
Labor intensive
Inefficient
To Pro-Active, Self-Improving, Machine Learning Systems

Global Product Data Interoperability Summit | 2015

Multiple Data Sources
Real-Time Processing
Store Everything

Data Stream Pipeline
HDFS
In-Memory Real-Time Data
Expert System / Machine Learning

Continuous Learning
Continuous Improvement
Continuous Adapting

To Pro-Active, Self-Improving, Machine Learning Systems

Global Product Data Interoperability Summit | 2015

Multiple Data Sources
Real-Time Processing
Store Everything

Data Stream Pipeline
HDFS
In-Memory Real-Time Data
Expert System / Machine Learning

Continuous Learning
Continuous Improvement
Continuous Adapting
"50-80% of the time on data science projects is spent on data wrangling."

Still…

Data Feeds

Stream Processing
Expert Systems
Machine Learning

Business Value
Smart Decisions

Historical Data

Data Lake

HDFS
Data Streaming Needs an Agile, Scalable, Automated and Fast Solution

Global Product Data Interoperability Summit | 2015

Ingest → Transform → Sink

SpringXD

Spark

GemFire

Data Lake

HAWQ

GPDB

Madlib

Python

R

BOEING is a trademark of Boeing Management Company
Copyright © 2015 Boeing. All rights reserved.
Copyright © 2014 Northrop Grumman Corporation. All rights reserved.
GPDIS_2015.ppt | 20
Data Streaming Reference Architecture

Data Feeds
- Distributed Computing
- Real-Time Data
- Expert Systems & Machine Learning
- Advanced Analytics

Transaction Apps
- HDFS

Analytic Apps
- Tableau
- SAS

Data Stream Pipeline

Data Lake
Data Streaming Reference Architecture

Data Feeds

Data Stream Pipeline

SpringXD

Data Lake

Spark

GemFire

HDFS

Transactional Apps

Analytic Apps

HAWQ

GPDB

MATlib python R

BOEING is a trademark of Boeing Management Company
Copyright © 2015 Boeing. All rights reserved.
Copyright © 2014 Northrop Grumman Corporation. All rights reserved.
GPDIS_2015.ppt | 22
## THE INTERNET OF THINGS JOURNEY WITH PIVOTAL

### Global Product Data Interoperability Summit | 2015

### STORE
- Structured
- Unstructured
- High Volume
- High Velocity

### ANALYZE
- Predictive Analytics
- Machine Learning
- Advance Data Science
- Realtime Analytics

### DEVELOP
- Advanced Analytic Pipelines
- Realtime Analytical Applications
- Global Scale Data-Driven Applications
- Enterprise, Consumer, IoT, and Mobile

### INNOVATE
- Agile Dev Expertise
- DevOps
- Hybrid Cloud
- Continuous Delivery
- Closed Loop Applications

### Data Engineering
- Spring XD
- Spark
- Pivotal HD & Open Data Platform

### Data Science
- Spring XD
- Pivotal Greenplum Database
- Pivotal HAWQ

### Pivotal Labs
- Spring XD
- Pivotal GemFire
- Redis
- RabbitMQ

### ENTERPRISE PAAS
- Spring IO
- Groovy
- Pivotal BDS on PCF
- Pivotal Cloud Foundry

### AGILE DEVELOPMENT

### BIG DATA

### PREDICTIVE ANALYTICS

---

**THE INTERNET OF THINGS JOURNEY WITH PIVOTAL**

- **STORE**
  - Structured
  - Unstructured
  - High Volume
  - High Velocity

- **ANALYZE**
  - Predictive Analytics
  - Machine Learning
  - Advance Data Science
  - Realtime Analytics

- **DEVELOP**
  - Advanced Analytic Pipelines
  - Realtime Analytical Applications
  - Global Scale Data-Driven Applications
  - Enterprise, Consumer, IoT, and Mobile

- **INNOVATE**
  - Agile Dev Expertise
  - DevOps
  - Hybrid Cloud
  - Continuous Delivery
  - Closed Loop Applications

**Data Engineering**
- Spring XD
- Spark
- Pivotal HD & Open Data Platform

**Data Science**
- Spring XD
- Pivotal Greenplum Database
- Pivotal HAWQ

**Pivotal Labs**
- Spring XD
- Pivotal GemFire
- Redis
- RabbitMQ

**ENTERPRISE PAAS**
- Spring IO
- Groovy
- Pivotal BDS on PCF
- Pivotal Cloud Foundry

**AGILE DEVELOPMENT**

**BIG DATA**

**PREDICTIVE ANALYTICS**
Energy, Maritime, Aviation

• Goal
  – Preventative maintenance vs. remediation
  – Increase profitability and reduce liability
  – Store and process fire-hose of data

• Solution
  – Ingest all data for data discovery and model development
  – Store 10 TB of HIGH VELOCITY turbine engine data in memory (GemFire)
  – VERY LOW LATENCY and HIGH SPEED data access
  – Currently realize 1 percent savings in fuel management alone
Pivotal
BUILT FOR THE SPEED OF BUSINESS