Big Data Storage Challenges for the Industrial Internet of Things

Shyam V Nath
Diwakar Kasibhotla

SDC

September, 2014

Agenda

- Introduction to IoT and Industrial Internet
- Industrial & Sensor Data
- Big Data Storage Challenges
  - Ingestion / Storage
  - Retrieval / Consumption
- Use Cases
- Wrap up

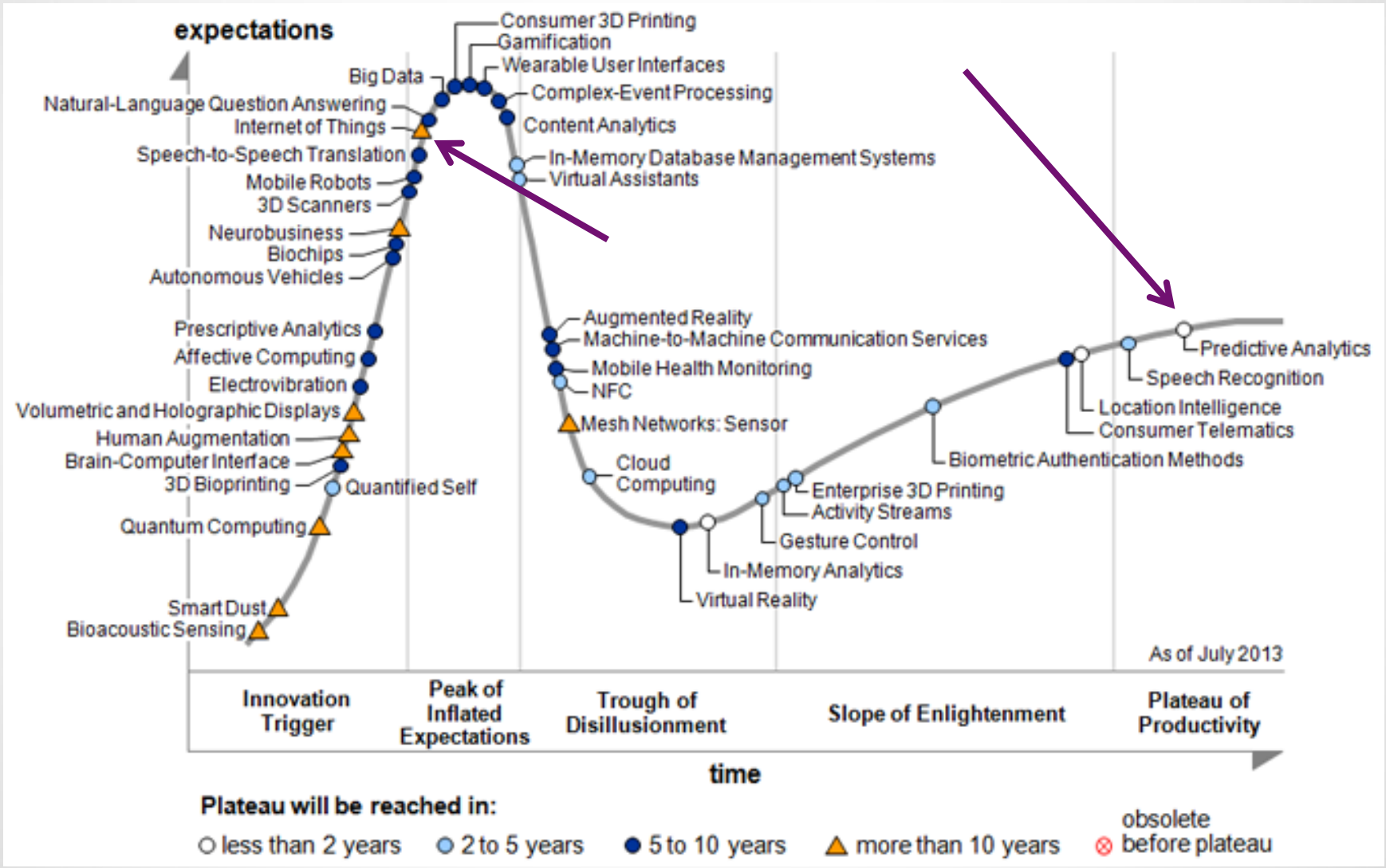
About Shyam

- Principal Architect – Analytics
- Board of Director (SIGs), 30K+ member User Group (IOUG)
- Started the IoT/ Industrial Internet Meetup in East Bay in June 2014, started other BI/Analytics related user groups
- Worked in IBM, Deloitte, Oracle and Halliburton, prior to GE
- Under grad from IIT (India), MS (Computer Science) and MBA (FAU)
- Regular speaker in large events like Oracle Openworld, Collaborate, BIWA Summit on IoT, Business Analytics and Data Warehousing / Engineered Systems related topics

About Diwakar

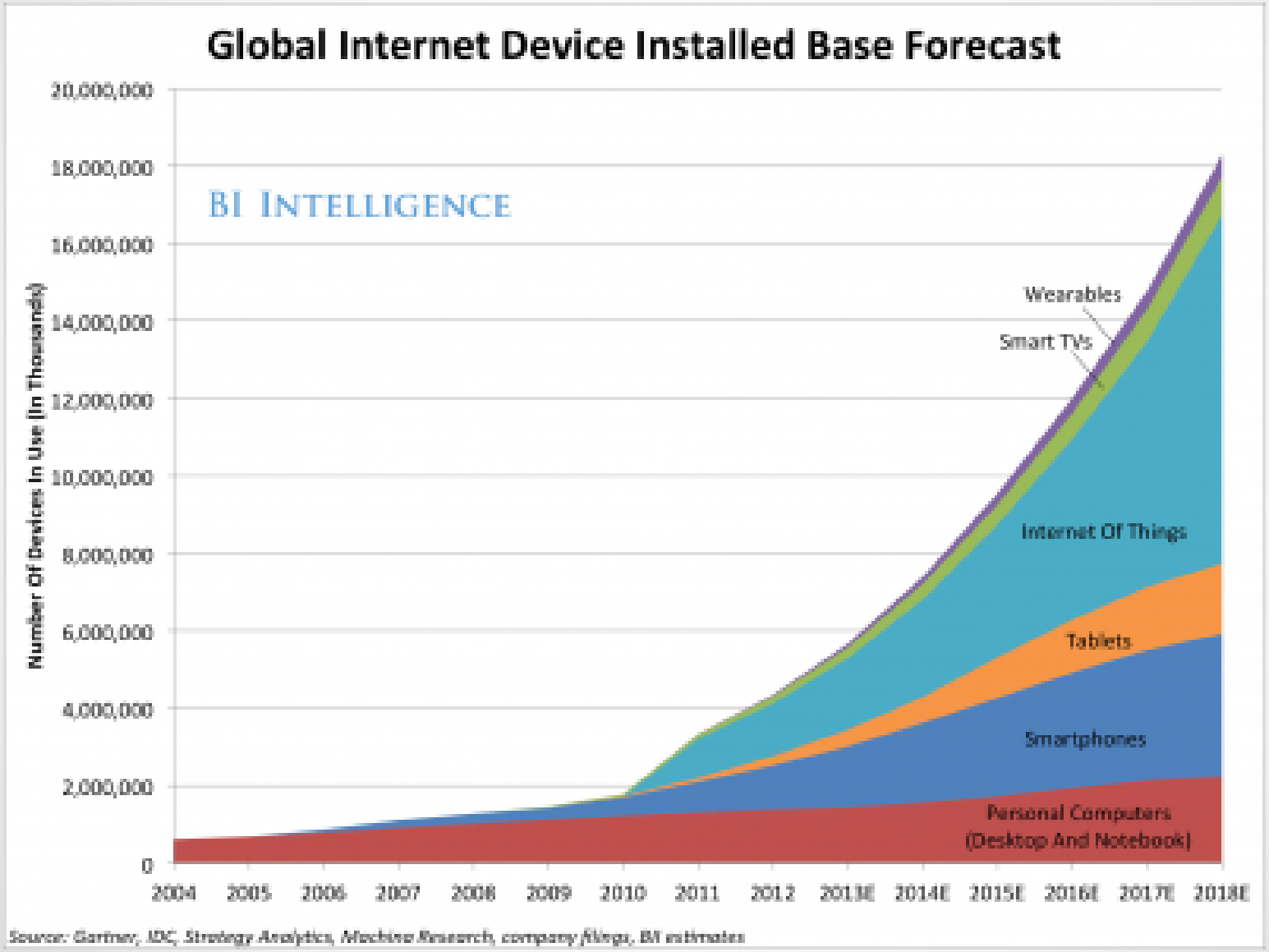
- Principal Architect – GE Aviation
- Worked at Oracle, EMC, Pivotal prior to GE
- Regular speaker in large events like Exadata SIG and Oracle Openworld
- Expertise: Data Integration, Database Appliances, Big Data, IOT

The Hype Cycle – Gartner July 2013

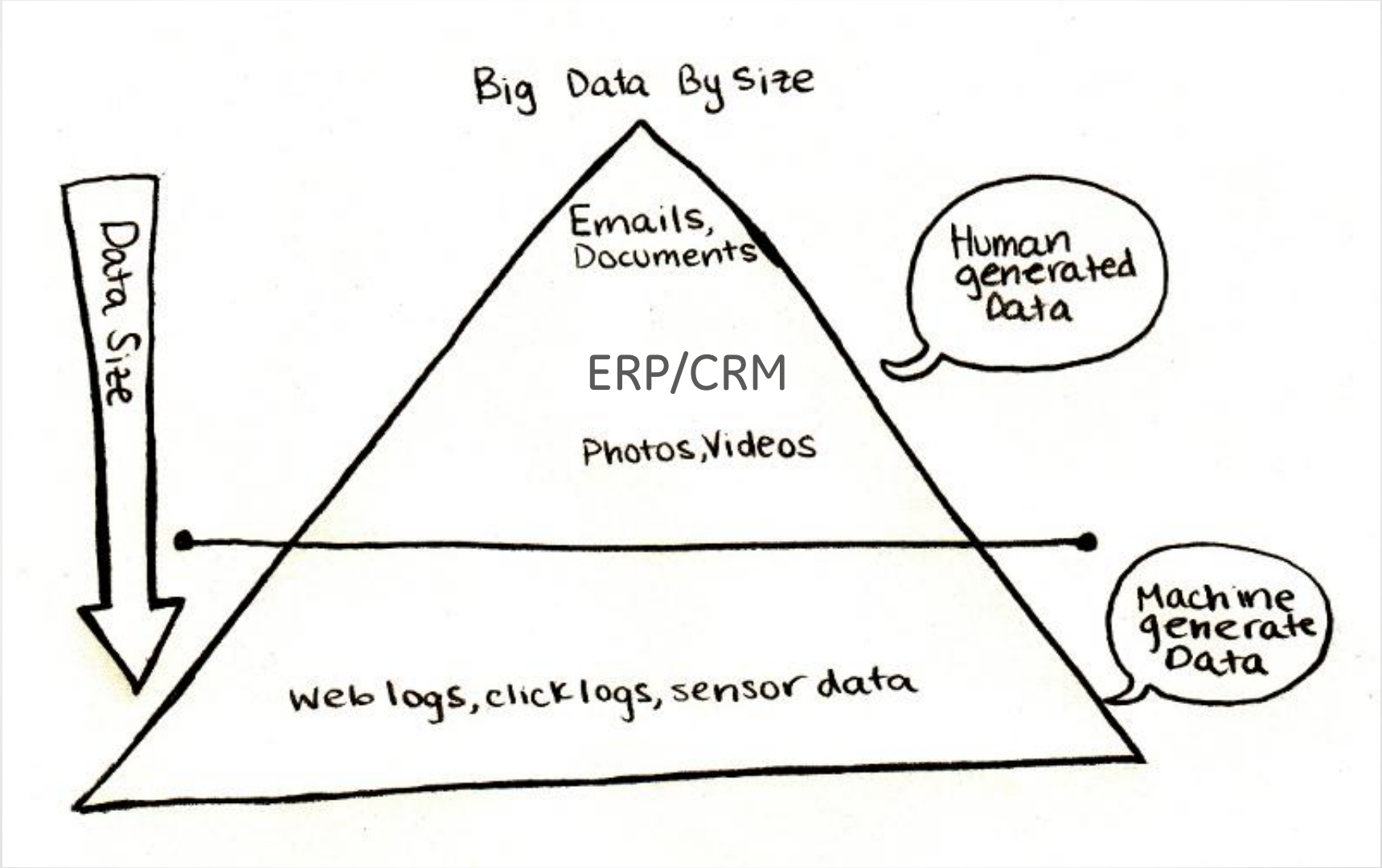


The 2013 Hype Cycle features Internet of Things, machine-to-machine communication services, mesh networks: sensor and activity streams.

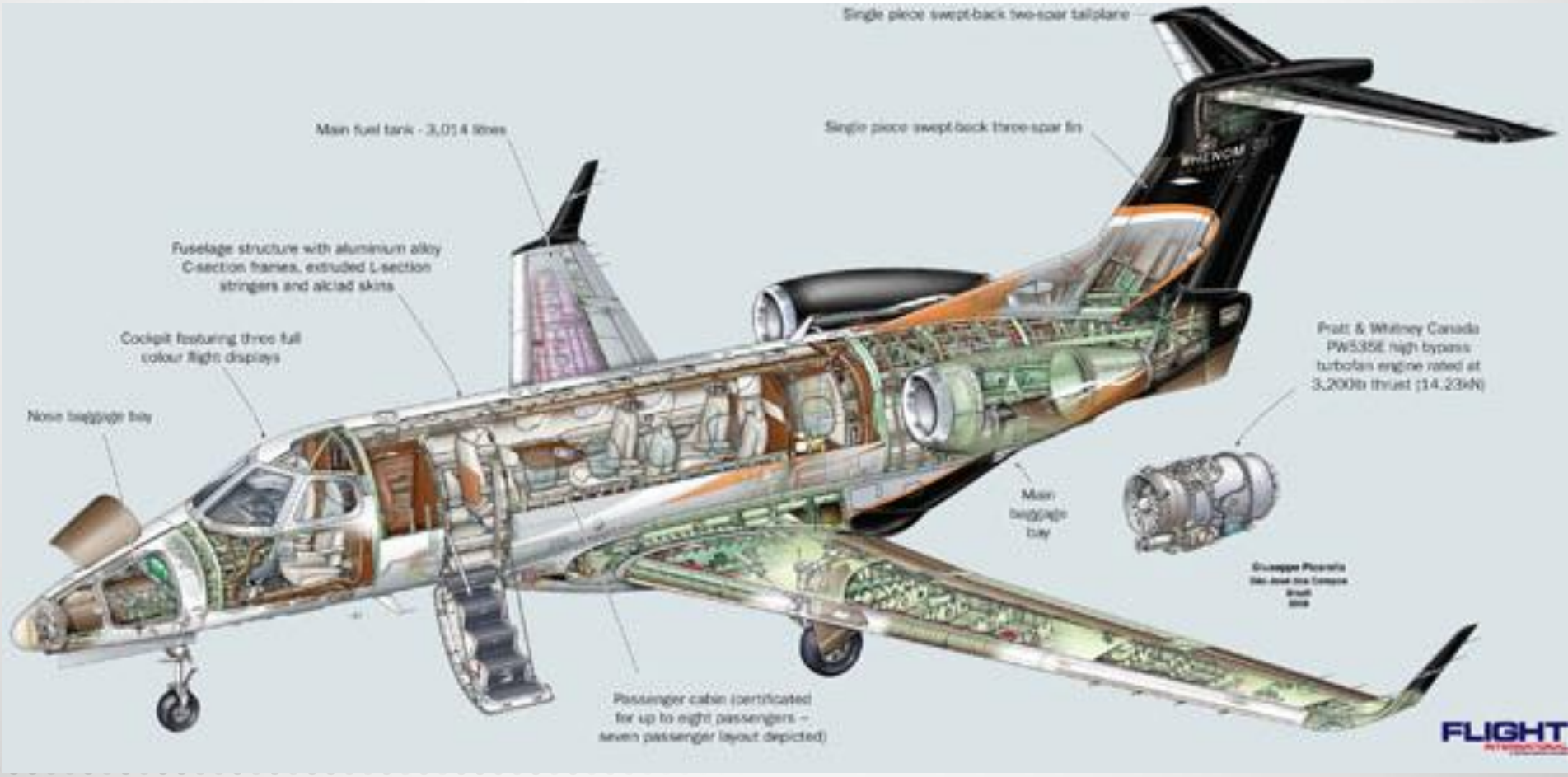
What are the "Things?"



Big Data and IoT



Different "Views" of Aircraft - as collection of sensors



http://www.flightglobal.com/cutaways/civil/phenom-300/

Data from Jet Engine

We Used to Get…

Takeoff
Diagnostics Data
(Averaged)

Cruise
Diagnostics Data
(Averaged)

Landing
Diagnostics Data
(Averaged)

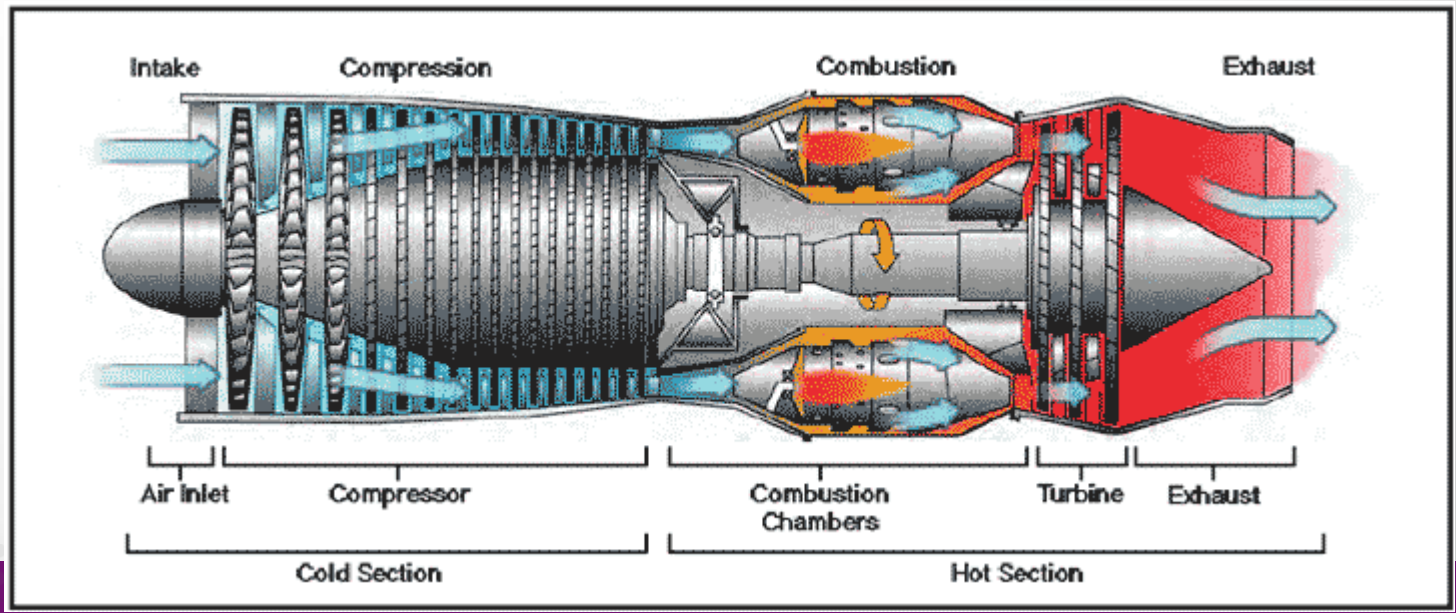
Making "Sense" of the "Sensors"

EGT = Exhaust Gas Temperature

The temperature of the exhaust gases as they enter the tail pipe, after passing through the turbine

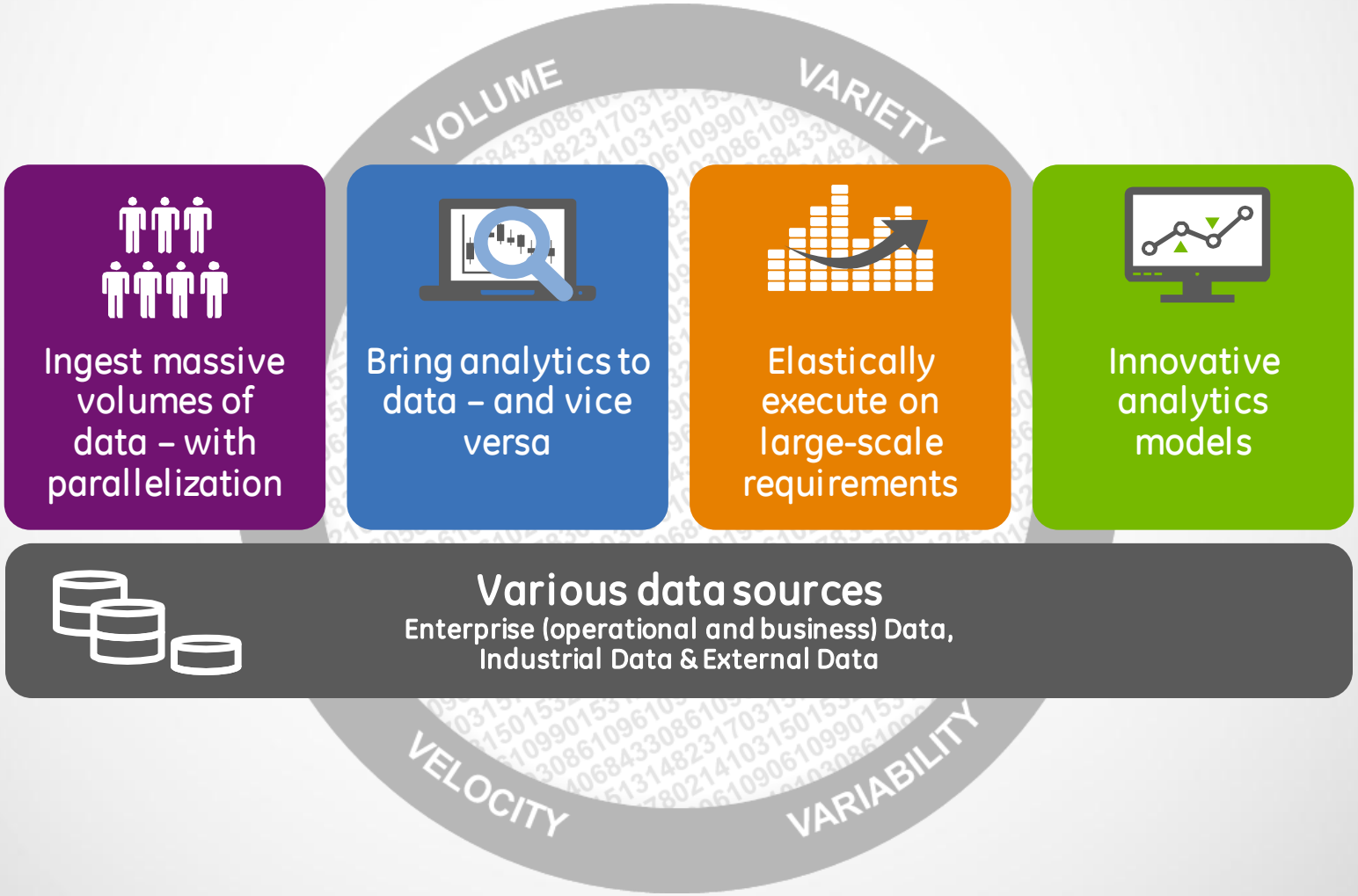
A good indicator of the health of engine (just like human body temperature)

Recording and interpreting the EGT can help to detect several jet engine problems.



Industrial Internet: Big Data Analytics

Delivering sharper insights to users

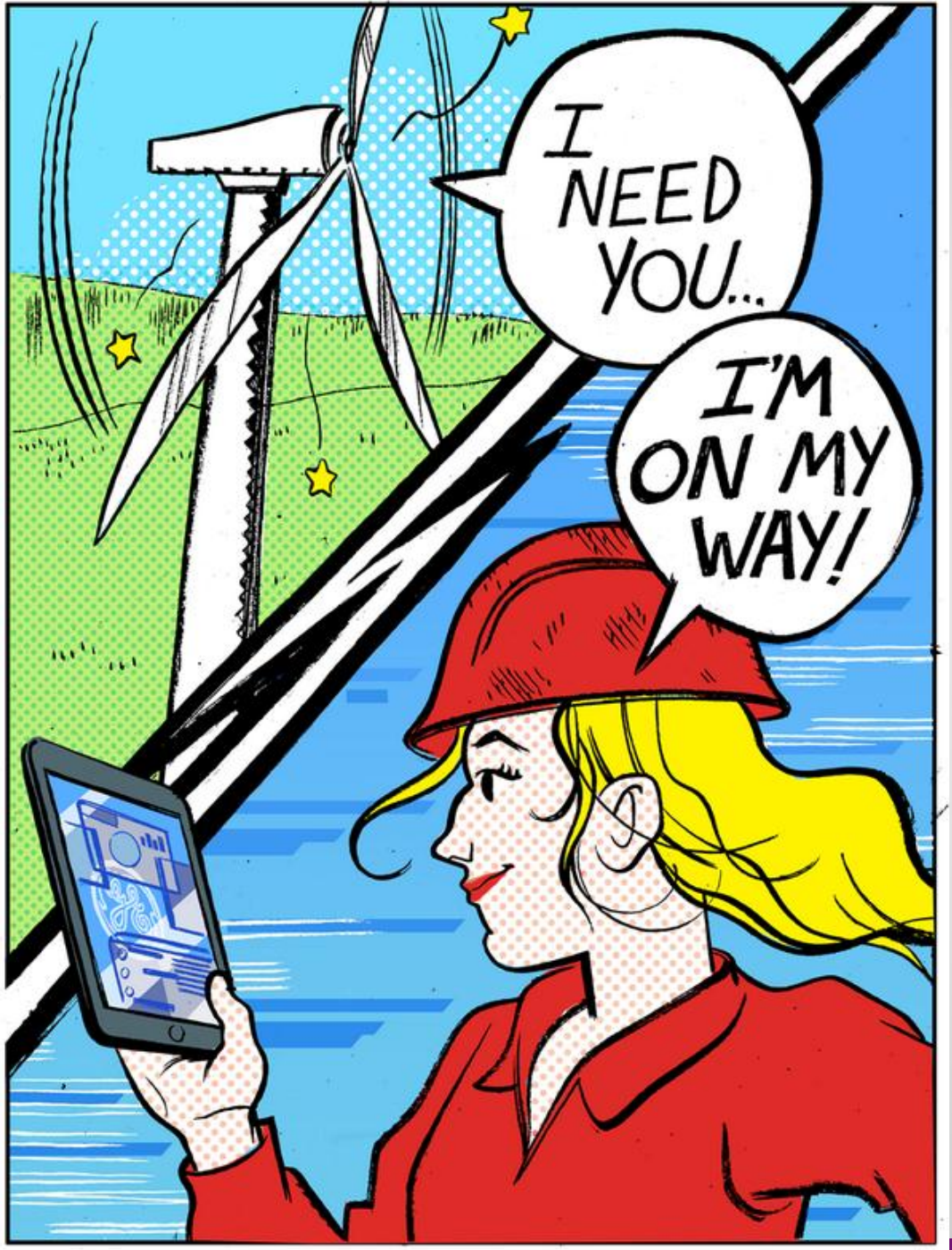


Wind Farms Explained Via Visuals!

Altamont Pass Wind Farm

Turbines near Livermore, California

Location	Altamont Pass, Alameda County, California
Coordinates	37°43'57"N 121°39'9"W
Commission date	1981
Power generation	
Primary fuel	Wind
Units operational	4930
Nameplate capacity	576 MW
Annual generation	1.1 TWh



David Gilford @dgilford · 20h

"A single power generating unit creates 1 TB of data each day" - @JeffImmelt on Industrial Internet #**IIoT** #BNEF2014 pic.twitter.com/OPt9Hend3e

Hide photo Reply Retweeted Favorite More

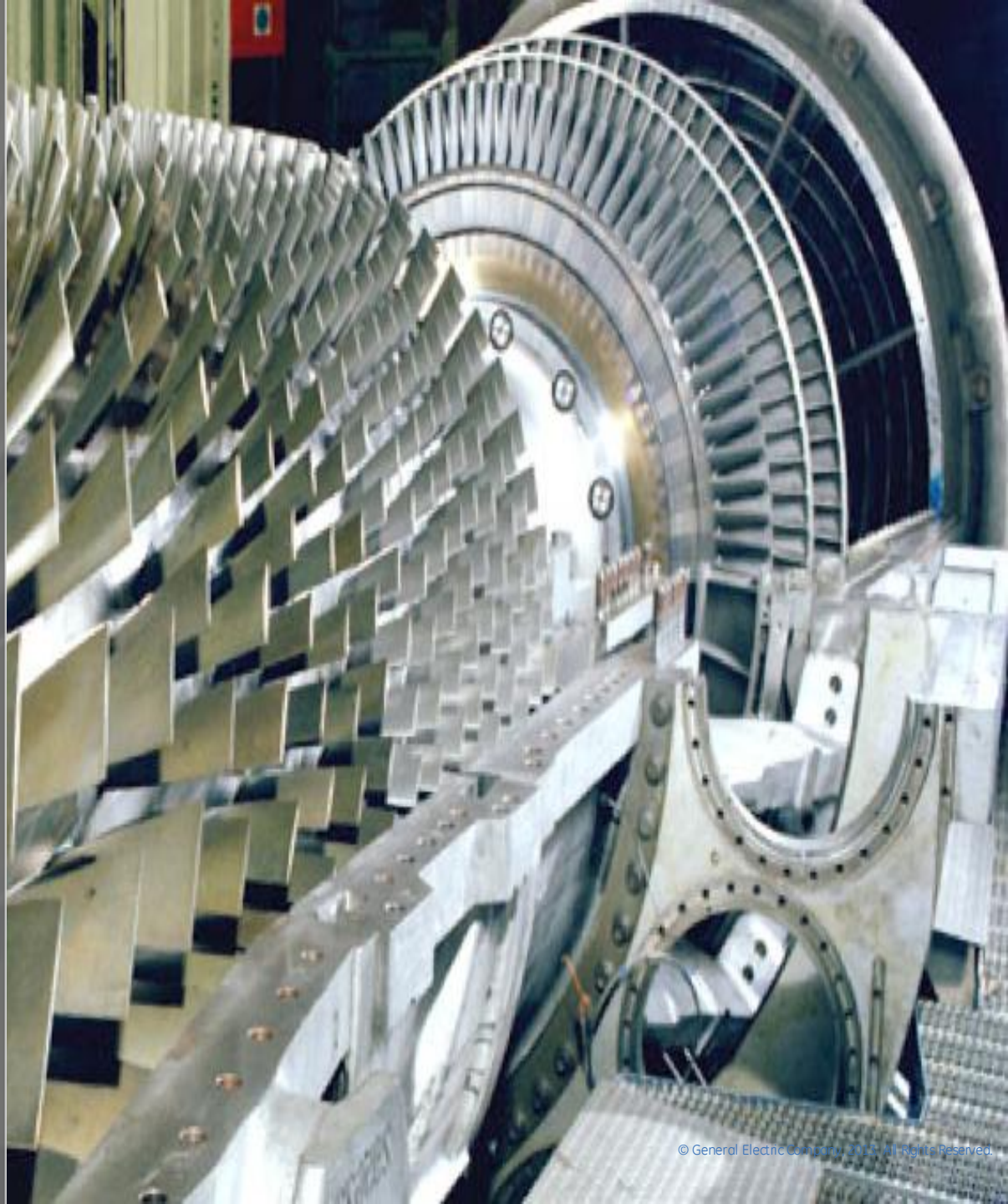
RETWEETS 3 FAVORITE 1

2:27 PM - 7 Apr 2014 · Details Flag media

Value Of Big Data Analytics

1 Gas Turbine Compressor Blade Monitoring Potential:

500 Gigabytes Per Day



Industrial Internet computing requirements

Cloud
for efficiency and agility

Going mobile: anytime/ anywhere
Access

End-to-end
Security

Predictive insights from
Big Data

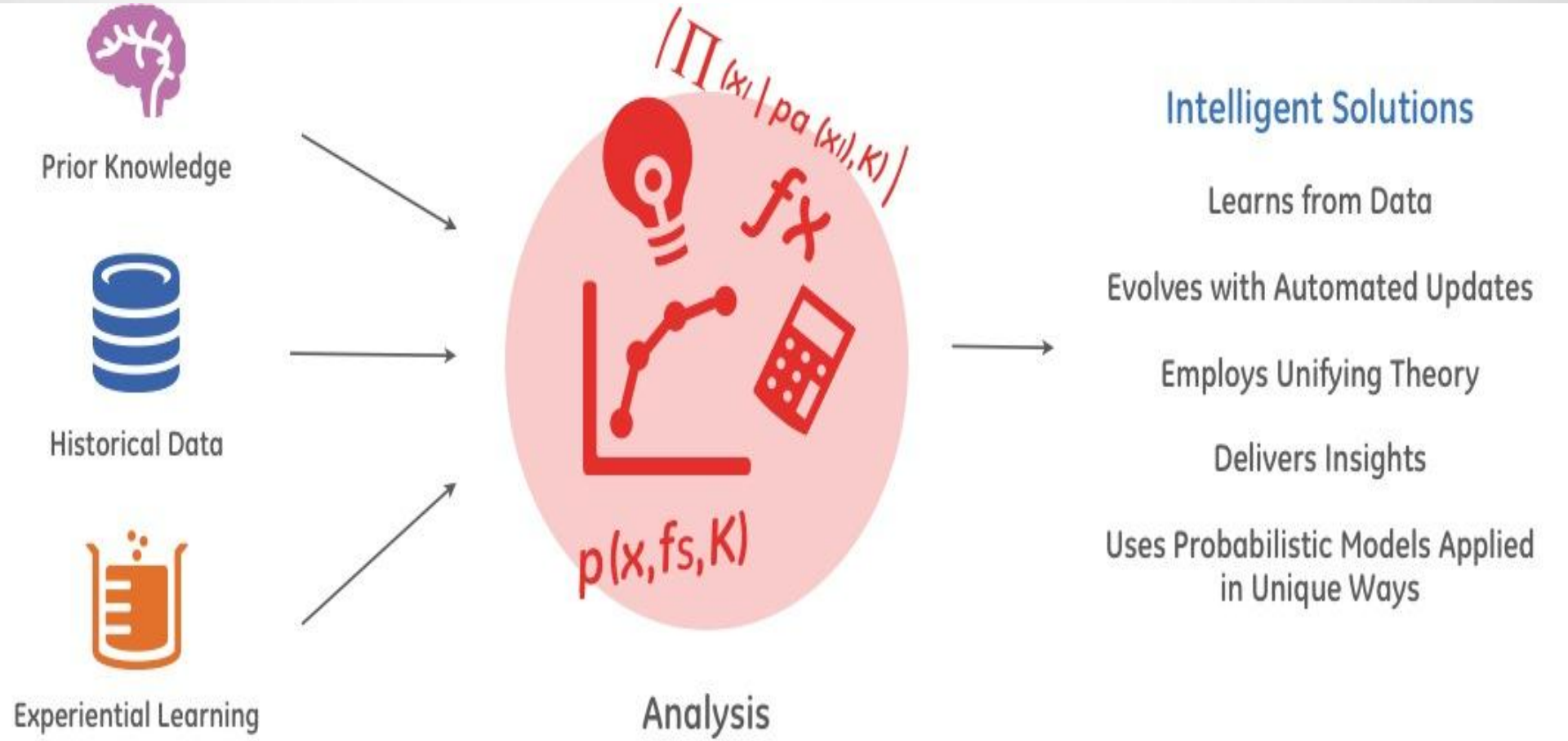
021010308
013161090
040109010
104078050

Consistent and meaningful
User experience

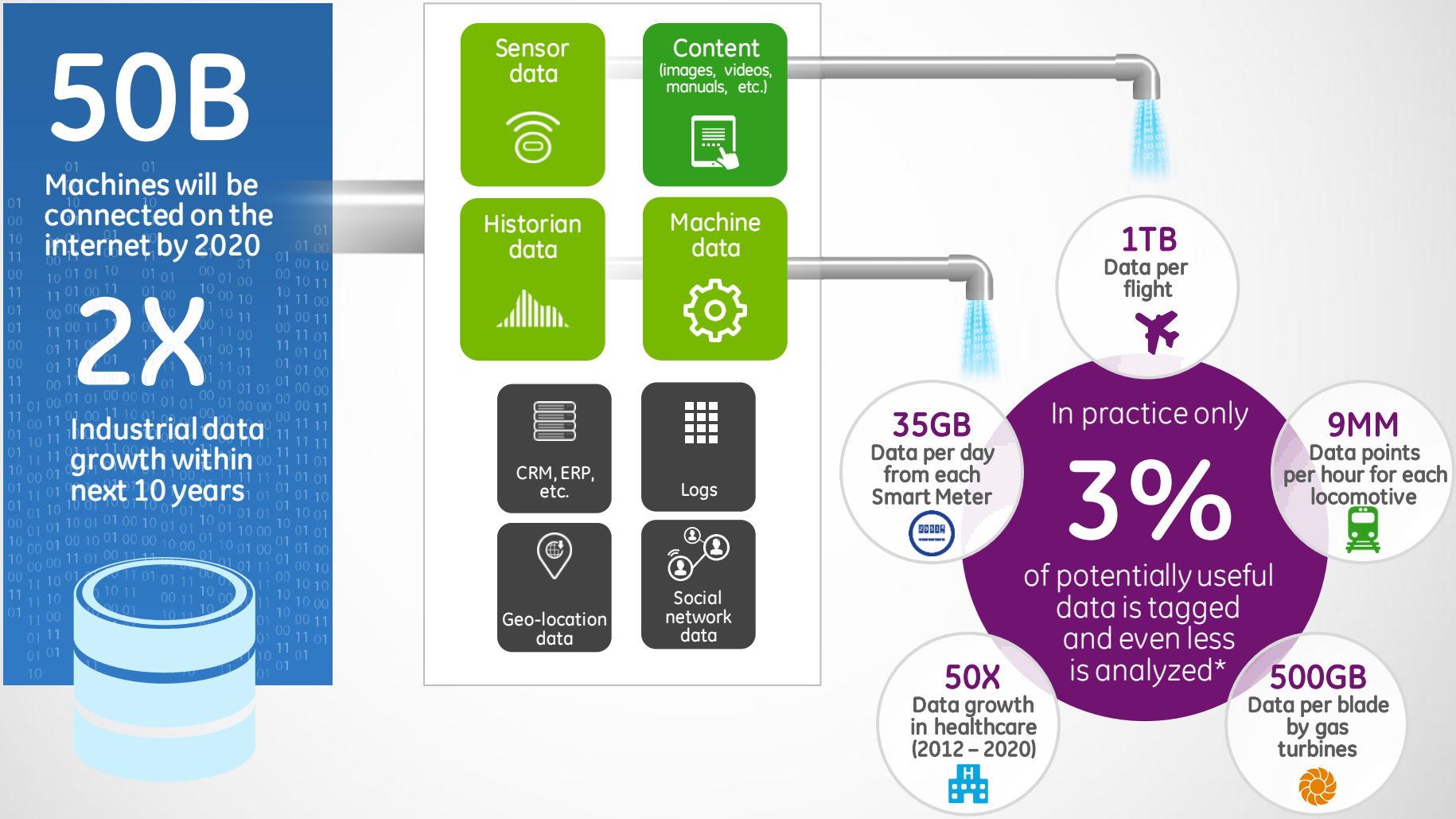
Cloud based Integrated
Asset Management

Transition to
"Brilliant machines"

Apply Batch or Real-Time Analytics to the Machine-Generated Data



Industrial Big Data – fast and vast



*Source: IDC

*Source: IDC

Today's approaches are not prepared for onslaught of Industrial Big Data

80% of an analytics project typically involves gathering and then preparing the data for analysis*

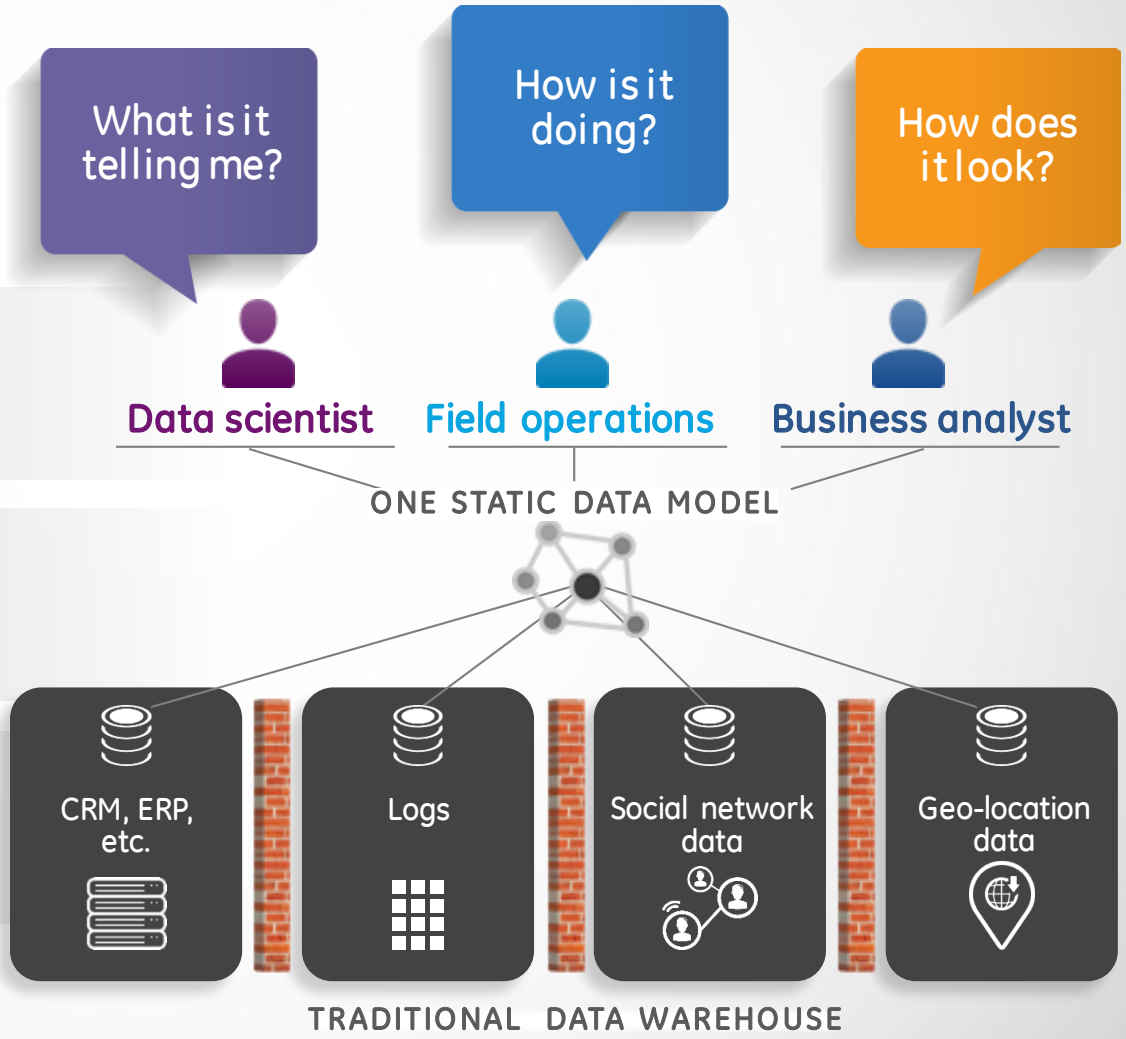
*Source: IDC

Yesterday's data warehouse architecture

1 All over the place
Data across multiple locations

2 Limited data types
Mostly structured and semi-structured data types

3 Snapshot
Limited to narrow snapshots and time

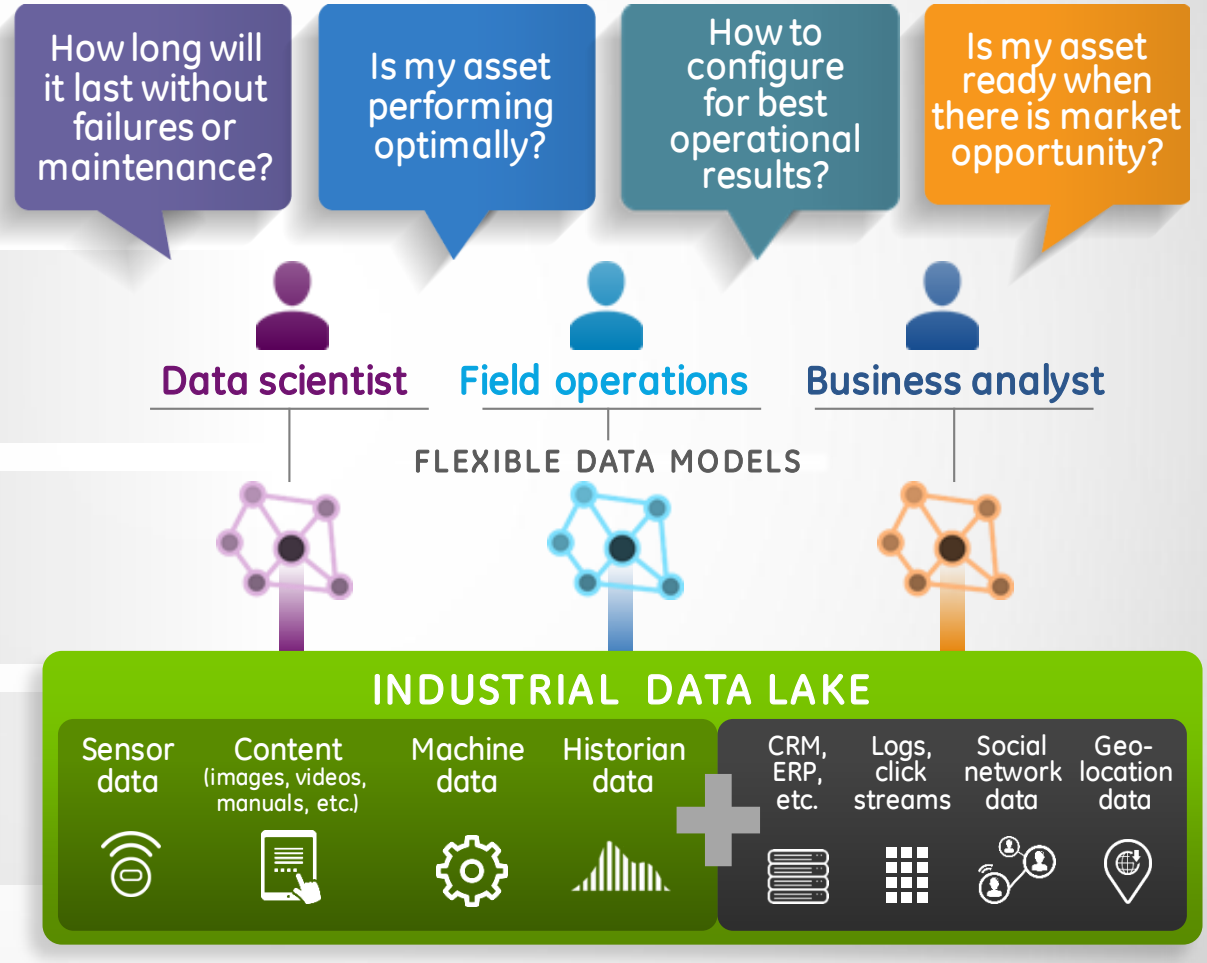


New approach – Industrial Data Lake architecture

1 One place
Access to all data in one place to quickly respond to the speed of business change

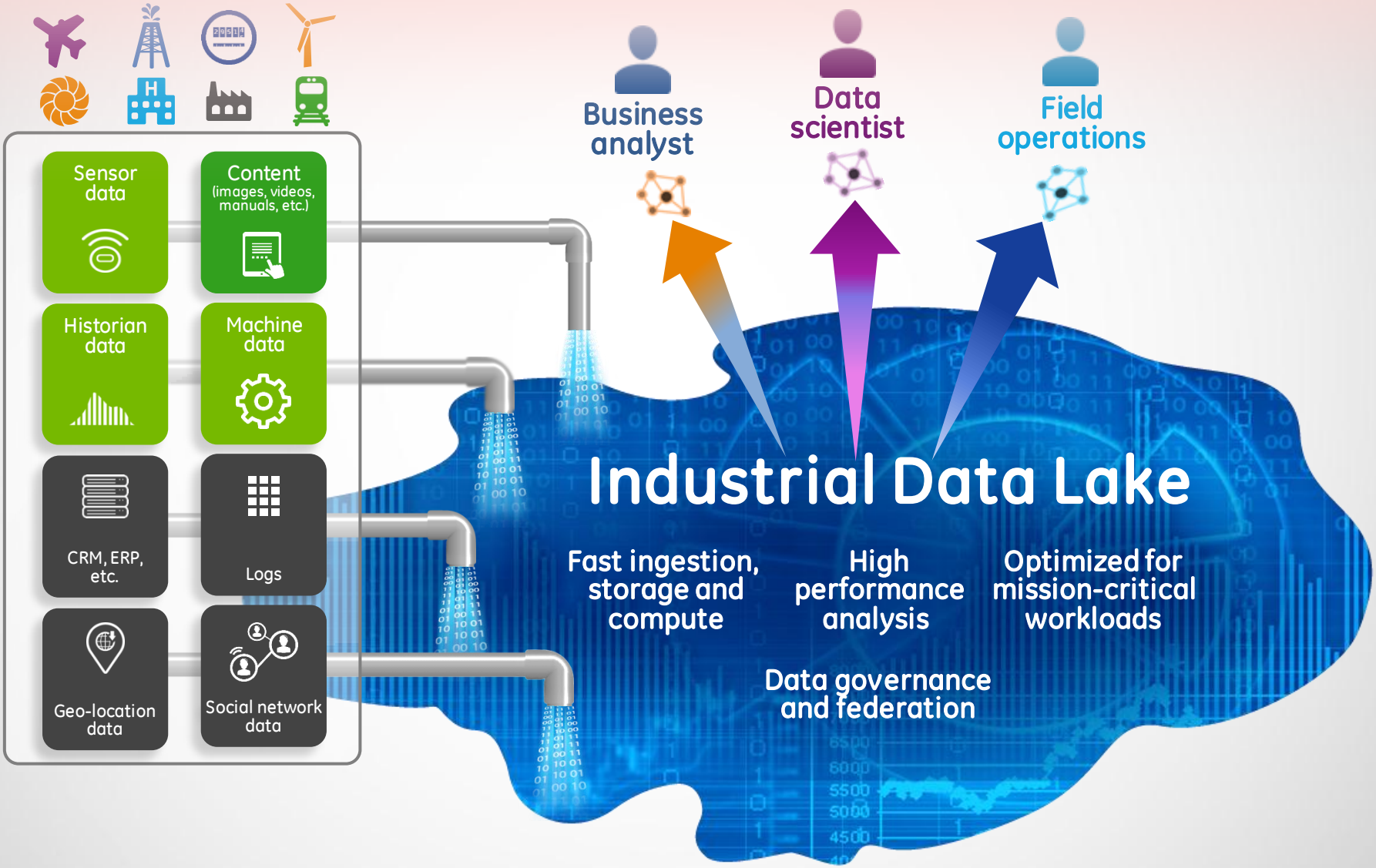
2 Any data
Handing of all data types including documents, images machine data, sensor data

3 All data
Access to real-time data and historical data and not limited to snapshot of data



Rapid access to all data for analytics

Industrial Data Lake



Aviation and Big Data

Mon, Aug 11, 2014, 5:54 PM EDT - U.S. Markets closed

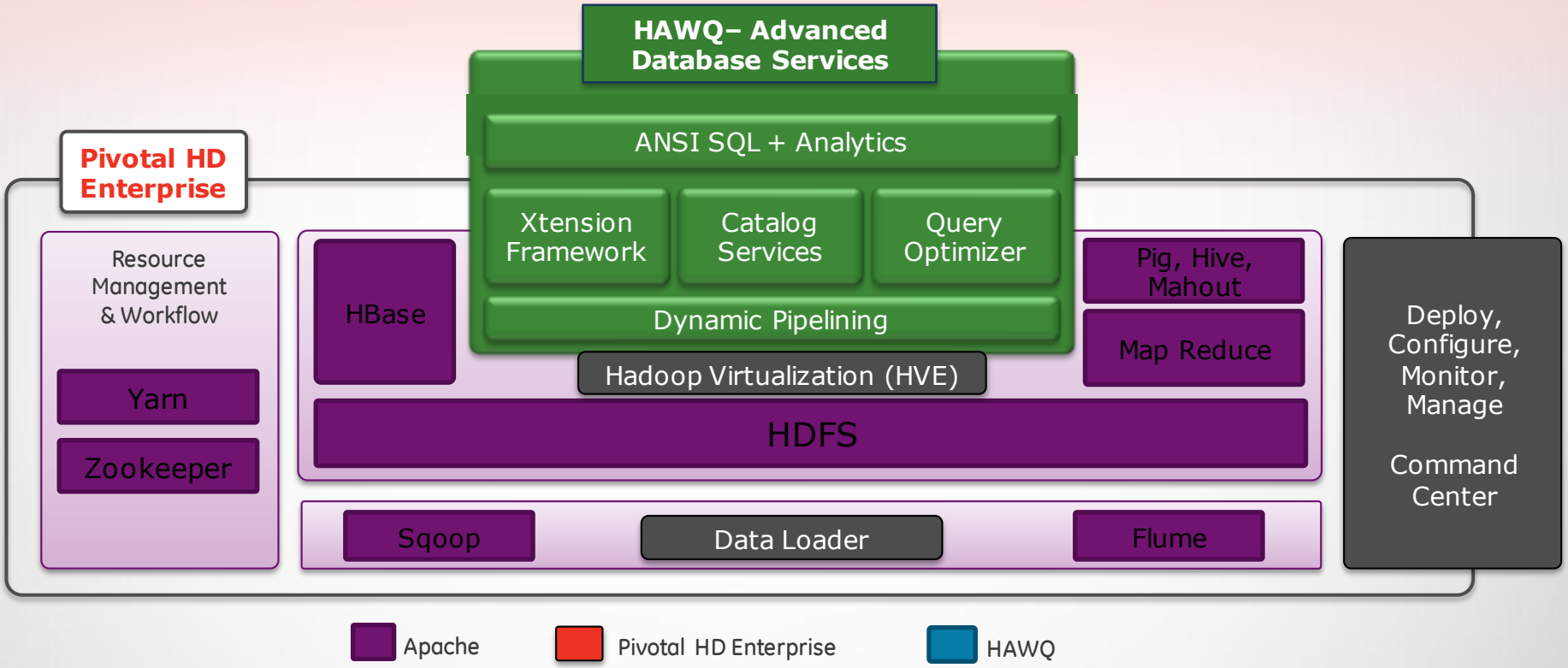
GE starts rolling out Pivotal's big data technology to its own customers

GIGAOM By Barb Darrow
16 hours ago

General Electric, which has touted the potential advantages of applied big data for a few years and last

"GE expects the data collection to grow to 10 million flights and 1,500 terabytes of full flight operational data by 2015."

Pivotal Architecture



Q&A

Shyam Nath
Shyam.Nath@GE.com
Diwakar Kasibhotla
Diwakar.kasibhotla@GE.com

Thank You!