



# Data Driven World means a Storage Drive Data Center

---

Hal Woods, VP Technology and Advanced Development  
November 7, 2015

[Hal.Woods@hgst.com](mailto:Hal.Woods@hgst.com)

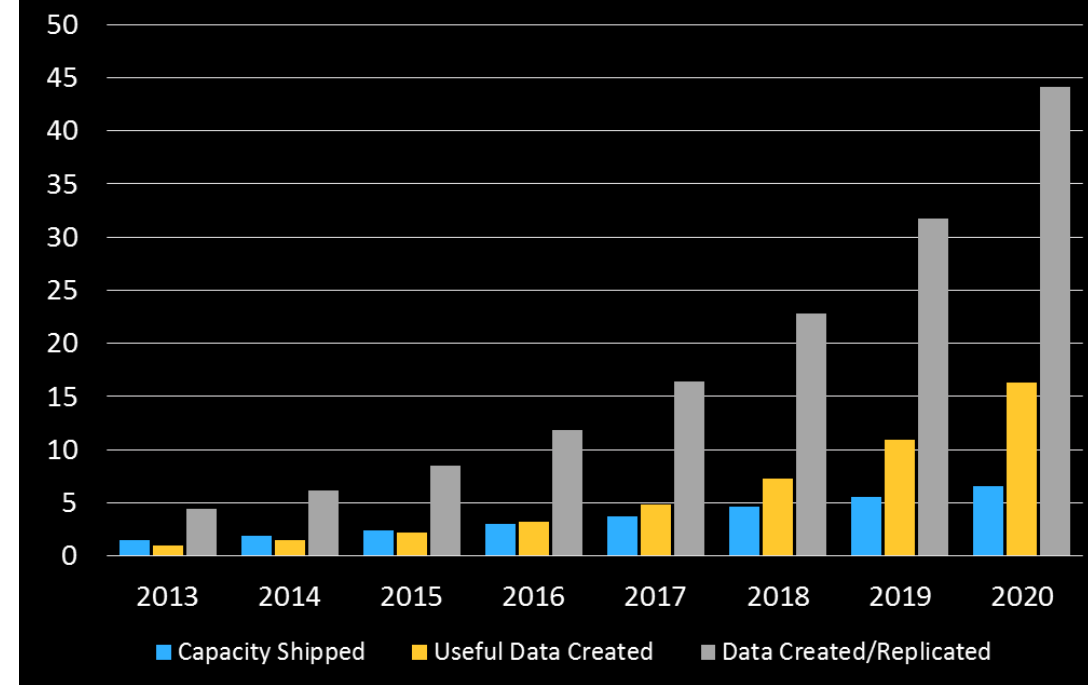


# I have some bad news for you... and good news for me...

- You are a data hoarder, an addict in fact
- You want to keep everything
- You admit that most of what you want to keep will never be useful
- You want to be even more, not less, of a data hoarder - if only you could afford it
- But you keep data because it might be useful, sometime, maybe, but probably not
- But when you need it, you really need it
- If only you can find the important pieces
- I am here to tell you, store it all, forever!



*The data storage industry thanks you...*



**Better Decisions - 400 Gigabytes / in<sup>3</sup>**

**Faster Decisions - 150K IOPs / in<sup>3</sup>**

**How much capacity is consumed to store a gigabyte of useful data?**

■ .5 Gigabyte

■ 5 Gigabytes

■ 100 Gigabytes

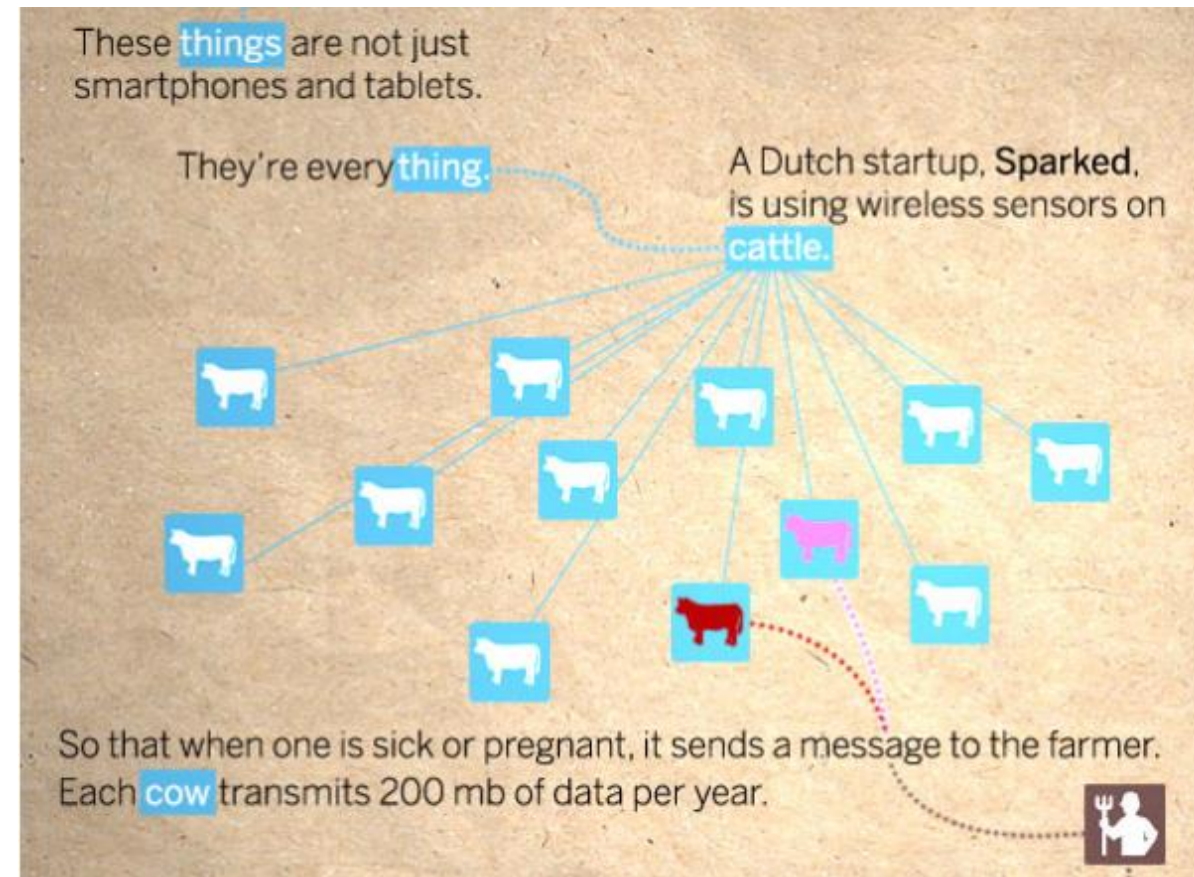
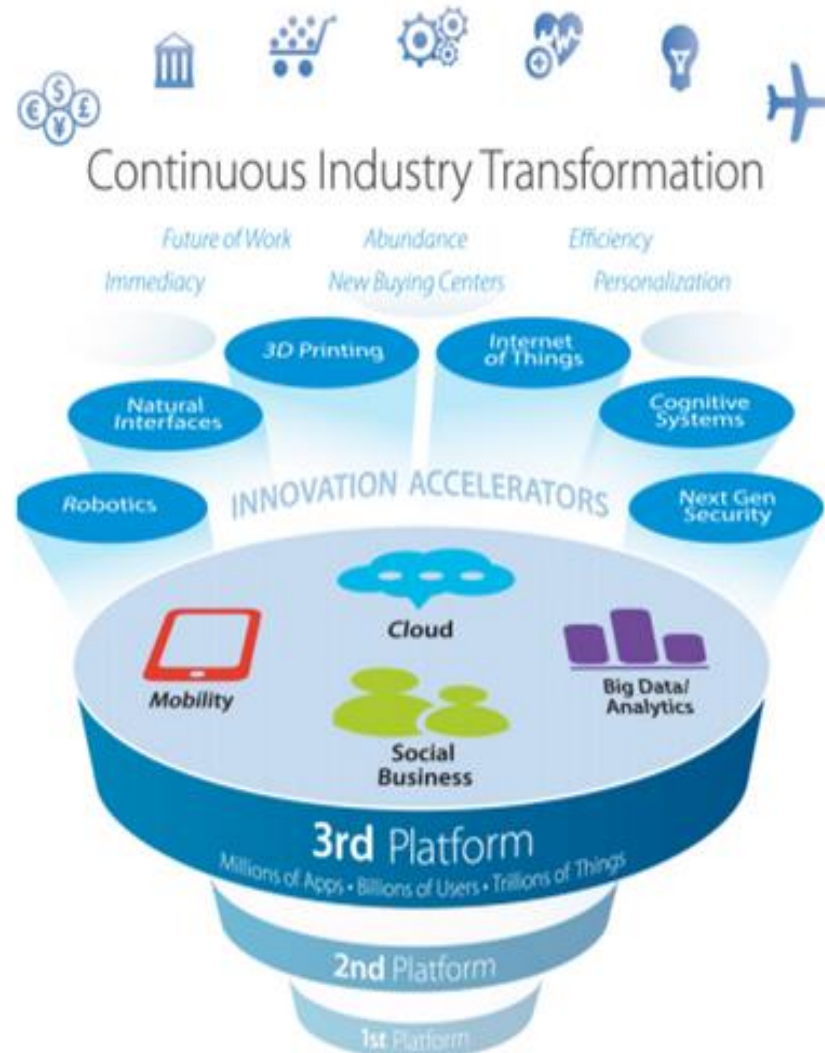
# Agenda

- Why talk about data storage
- Storing your Data
- Protecting your Data
- Monetizing your Data
- Storage as Memory





# Why Talk About Storage: Transformation Demands New Approach



# Big Data: The Internet of Cow

1.5 Billion Cows

200MB / yr / cow

= 300PB / Year

= 100 racks of storage



Thursday, September 4<sup>th</sup>, 2014

IoT & Security: presented by Amit Rohatgi at CIE-SF

23



**10,000,000  
cow years  
per rack!!!**



# Storing your Data

## BIG

- Scale-out S3 Store
- 4.7 PB / 3.5GB/S per rack
- 15 9s of data durability
- Distributed erasure coding
- Site-level resiliency
- Co-located compute
- \$100 / TB

## FUTURE (2018)

- 10PB / 10GB/S
- Tiering
- Machine Learning



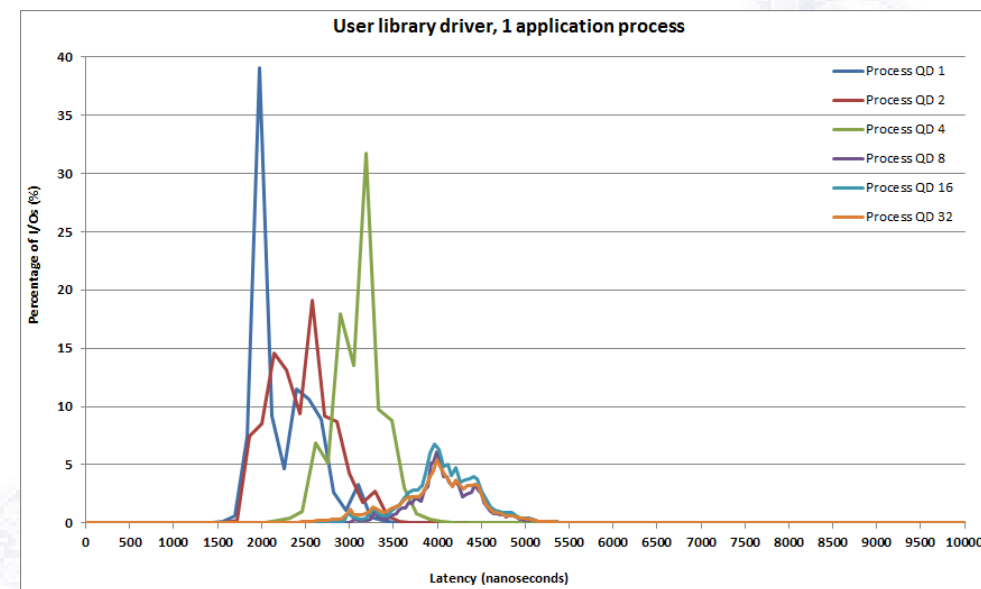
## FAST

- 3.2TB 750K IOPS per device
- 20-2000us latency
- 2M Hours MTBF 3DW/D
- PCIe Card and Drive Versions
- \$1000 / TB



## FUTURE

- 1M IOPS
- 10TB
- 2-5us
- eNVM



# Protecting Your Data

- What we used to worry about
  - Unrecoverable Errors
  - Simultaneous Failures
  - Natural Disasters
- What we (should) worry about now
  - Service and Maintenance
  - Upgrades
  - Bugs
  - Lost Metadata (keys, hashes,...)
  - Unnatural Disasters







# Protecting your data at scale – Looking for Answers

- Configuration Management
  - Was it broken from the beginning
- If it ain't broke don't fix it
  - Is the upgrade really needed
- Even if it is broke, don't fix it
  - Fail in place for predictable failures
- Partially broke may not be broken at all
  - Scale-out shared-nothing architectures designed for continuous availability using COTS
- Don't break it trying to make it better
  - Are you really sure you know better than the best practices
- Don't let others break it
  - Physical security still matters – public vs private cloud
- If it isn't broke don't let it be changed
  - Immutability of inactive data, create new versions, beware of obsolescence





# Monetizing your Data

<b>Big Data</b> Data is growing faster and bigger because of number of sensors	<b>10B+ rows 5TB+</b>		Wind turbine: 100 turbines x 100M rows per year Race car: 400M records / day x 365 days test drive Telco: 1.000 cells x 1.000 rows / sec x 1 days - wow Traffic analysis: 60M cars x 1 read / min x 365 days Oil rig: 1 rig = 8 billion records / day (not verified)
<b>Fast Data</b> Data streamed from sensors requires fast ingestion	<b>1M+ rows per sec</b>		Network monitoring: 1M rows per sec per cell Asset monitoring: 60M cars x 1 reading per minute Airplane monitoring: 4 turbines x 3k sensors x 100Hz Oil exploration: 10.000 wells x 100 sensors x 1Hz Oil rig: 1 drilling rig x 10.000 sensors x avg 100Hz
<b>Edge Analytics</b> IoT data is mostly generated at the 'Edges' of the network	<b>100+ Locations</b>		Manufacturing: 300.000 plants in US (2012) Cars / ships / airplanes: >1 billion world wide Telco: 190.000 cell towers in US (2013) Oil: 950.000 wells worldwide; 500.000 in US Mobile advertising: de-central ad serving / monitoring
<b>Real-Time Insights</b> Use cases require near real time analytics	<b>&lt;1 sec query response time</b>		Dashboarding: real-time visualization, many queries Network monitoring: root cause analysis, optimization Asset monitoring: conditional monitoring, safety Security: anomaly detection, building safety Traffic: location aware recommendations

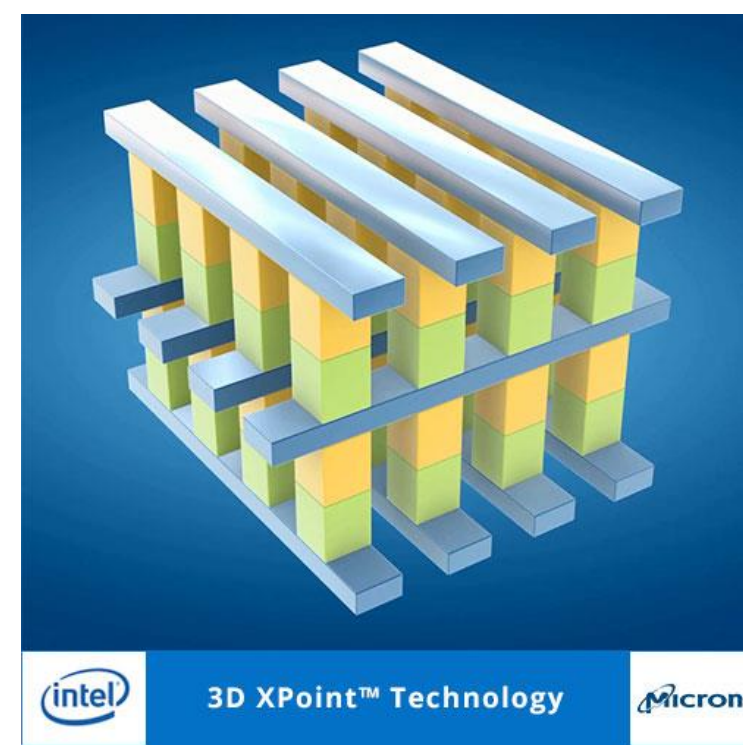
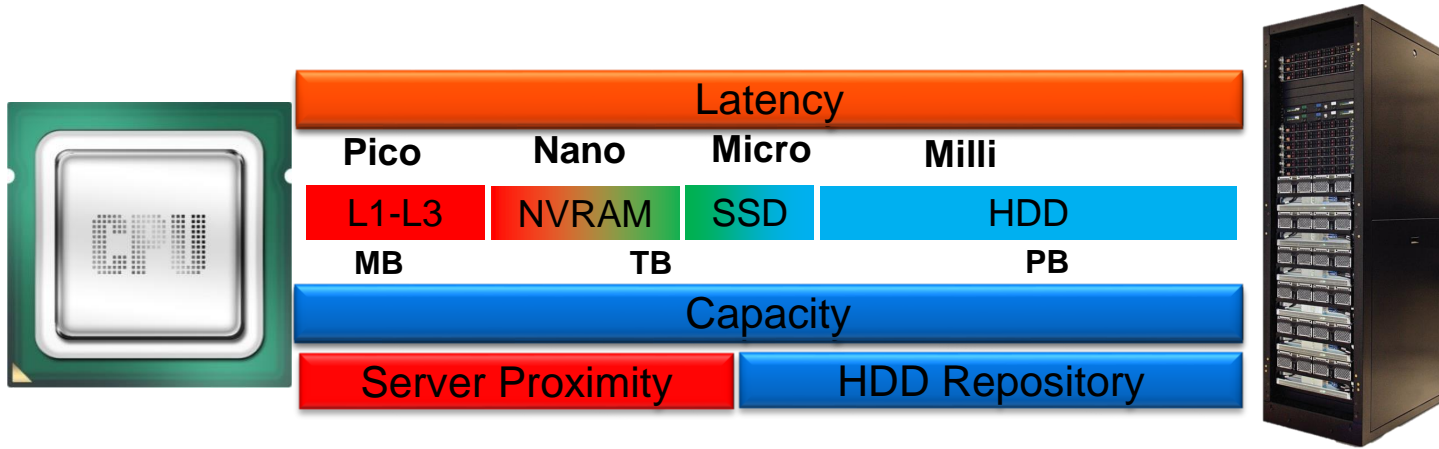
**Volume**

**Speed**

**Concurrency**

**Actionable  
Intelligence**

# Storage as Memory



- What if I could keep my entire working data set close to the CPU, shared, available
- What if the rest of my data, including copies, were affordably accessible in milliseconds
- New memory Hierarchy enablers:
  - Software defined storage for tiering and high availability
  - Solid state technology memory alternative
  - Low latency commodity interconnects

## Non Volatile Memory Enables a New Style of Application – starting in 2017

Store it All, Store it Forever...

...We'll make more

---

Signed: *Your Friends in the Data Storage Industry*