

## Big Fortune Hidden in Big Data Files



*“Why, anybody can have a brain. That’s a very mediocre commodity. Every pusillanimous creature that crawls on the Earth or slinks through slimy seas has a brain. Back where I come from, we have universities, seats of great learning, where men go to become great thinkers. And when they come out, they think deep thoughts and with no more brains than you have. But they have one thing you haven’t got: a diploma.” – Wizard, **The Wizard of Oz, MGM (1939)***

Maybe the computer never delivered on the promise of the paperless office, but it did open the floodgate on one thing ... data.

The result has been a data explosion:

- 1,203 Exabytes of digital info was created, replicated around the globe last year
- 1 Petabyte of new info has been produced every 15 seconds this year
- The annual growth in information is 59 percent

All of this is because people want to do it themselves (O.K., maybe companies “encouraged/pushed us a little) with ATMs, self check-in/check-out, email/video/tweets, online tell your life story, personalized TV/video/music entertainment, online travel/banking/paying/shopping.

### **DIY World**

Folks just can’t do enough themselves.



**Social Noise** – *Social media has made some significant “contributions” to big data but even more comes from the growing networks of sensors and monitors. The great thing about social media though is that so many people volunteer to put in so much information about themselves, their wants/needs, friends and thoughts that enterprises are able to collect/analyze all of the data and use predictive technologies to plan for tomorrow.*

As if that wasn't good enough, all of that DIY work, every click, creates an even bigger digital shadow about you ... 8-10 times more than the data you produce.

The data flood (some call it a swamp) is growing because:

- Storage is cheap, it costs only about \$600 for a drive to store all of the world's music
- There were 5B mobile phones in use last year and the number is growing
- 30B pieces of stuff are given to Facebook each month
- Global data will grow in excess of 40 percent per year
- Global IT investment will only increase 5 percent per year ... bummer

That's huge because it means a clear, steady growing market for storage – hard drives, solid-state, optical.

Yes, and it means a really healthy market for storage services like the cloud because everyone – individuals and companies – finds it easier, better to dump all that stuff somewhere.



**Keeps On Coming** – *The constant stream of data in any organization is almost overwhelming. Without a lot of assistance and high-performance tools, making sense out of the data would be completely impossible. The tools are there now; all we need is the folks to use them.*

Looking at it all, the cowardly Lion said, “*I may not come out alive, but I’m going in there.*”

Your DIY/shadow stuff (an estimated 75 percent of the content) is in the hands of companies that are liable for it while they’re making sense of it, using it to make their own data:

- Google processes about 24 PB of data per day
- AT&T transfers about 19 PB of data through their networks each day
- Four Large Hadron Collider experiments produce about 15 PB of data per year
- The German Climate Computing Center (DKRZ) has a storage capacity of 60 PB of climate data
- As of last June, Isohunt has about 10.8 PB of torrents indexed files
- The Internet Archive contains about 3 PB of data, growing at the rate of about 100 TB per month
- World of Warcraft has about 1.3 PB of game storage
- Avatar required over 1 PB of 3D CGI

No wonder Google, IBM, Facebook, Amazon, MS, Apple, Oracle--everyone is building huge cloud storage farms around the globe.

### Storage Move

We always thought Pat Geslinger jumped from Intel to EMC because Otellini was too young to retire so he could move up.

Nope. He saw the writing on the wall...become the operations dude at a storage management company, buy a bunch of the smaller players, tweak/push the software, rent it to just everyone.

Oracle’s Ellison and Geslinger saw the same rainbow.

It’s not just about bigger, more buckets for the data, it’s making it big and important/useful for the folks who hold it, know how to use it.



**Swirling Content** – *From virtually every corner of the globe information, data, content is being created, disbursed, shared, parsed, moved and stored. In the right hands, it can help healthcare, education, government and business organizations do a better job of planning and working toward tomorrow. Big data holds the keys—the answers to questions that haven’t been asked.*

**Source – IDC**

All of that data, content, stuff is growing exponentially and what the world needs now is a way to tap into, use that information to make societal, technological, scientific, economic changes.

Everyone – from the largest companies and governments to small business – needs to wring out every bit of value from their large, small, rented digital universe.

Or, as the Scarecrow said, *“Of course, some people do go both ways.”*



**Enterprises + Clouds** – *The volume of data that is developed and must be managed/used annually by organizations large and small is so massive and rich that IDC estimates that within a couple of years at least 20 percent of it will be stored in the clouds – private and public cloud storage services. An estimated 80 percent of the content will touch the cloud at some point in its journey.*

O.K., all the stuff is collected and stored pretty well.

The big problem for companies of all sizes is that the stuff is usually located in traditional organizational silos – marketing, engineering, manufacturing, finance, you name it.

You know... *“I want to leverage all their info but I need to keep mine separate because it’s different, better, really super important to us.”*

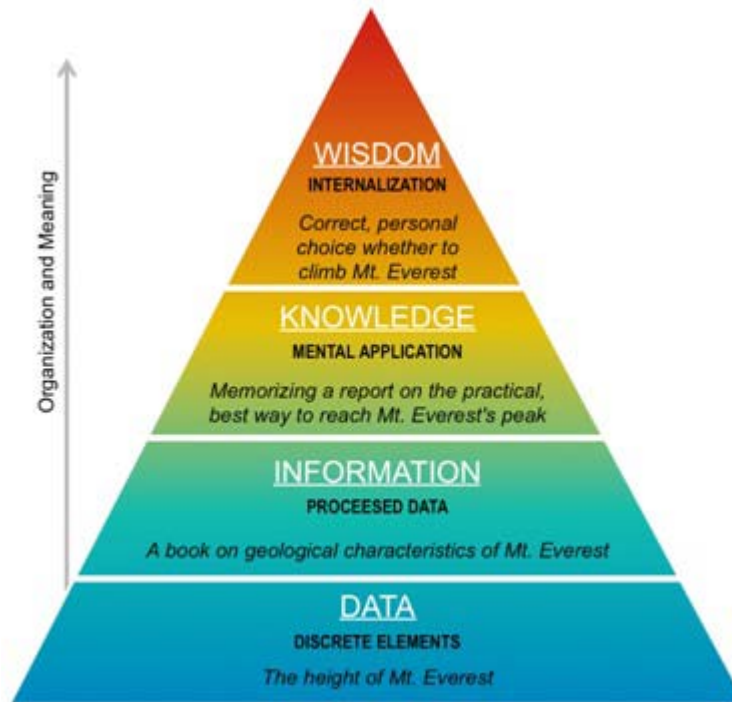
The marketplace and smart management are forcing the issue to answer the mission-critical question, *“How do we derive business value and insights from the data that can provide a competitive or market advantage?”*

### **Big Data Management**

No wonder Gelsinger, Ellison and the rest of the pack are buying up everything that touches or looks like it will touch big data.

It's big business. It's all about making exactly the right decision as quickly as possible.

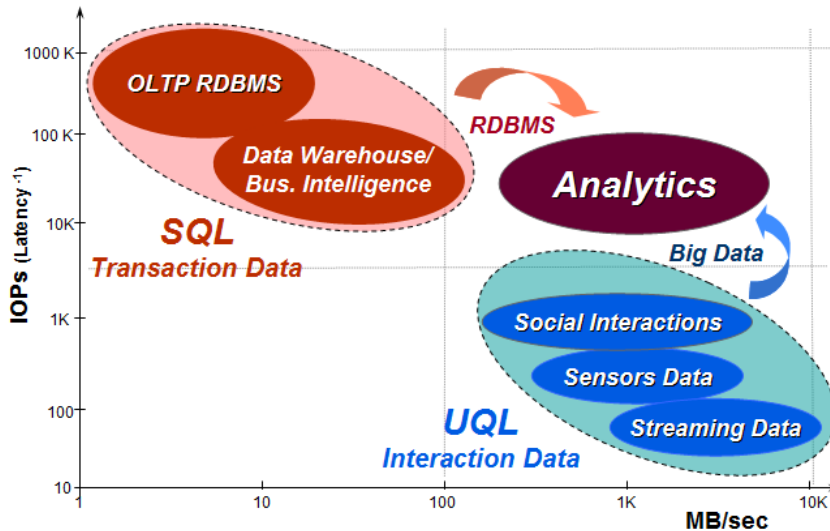
Looking at the swamp of data the Wizard said, *"You, my friend, are a victim of disorganized thinking."*



**Maslow's Pyramid** – *The more and better the data, the more quickly and more accurately organizations and management can make accurate decisions based on quality information and knowledge. The better the foundation, the more accurate the solutions.*

All that data in all that storage looks like serious money to mathematicians, computer scientists, statisticians and engineers because they seem to know how to make sense out of it and make it valuable by:

- Making it transparent, usable faster
- Creating, storing transactional data digitally so they can collect more accurate, detailed info on everything that goes on in the company, including sick days
- Collecting data, analyzing it, doing controlled experiments so better management decisions can be made
- Testing, tuning for JIT (just in time) everything
- Collecting, analyzing social media in realtime for proactive support, maintenance, product/service adjustments/refinements
- Using sophisticated analytics to improve decision making
- Doing a better job of developing next-generation products, services



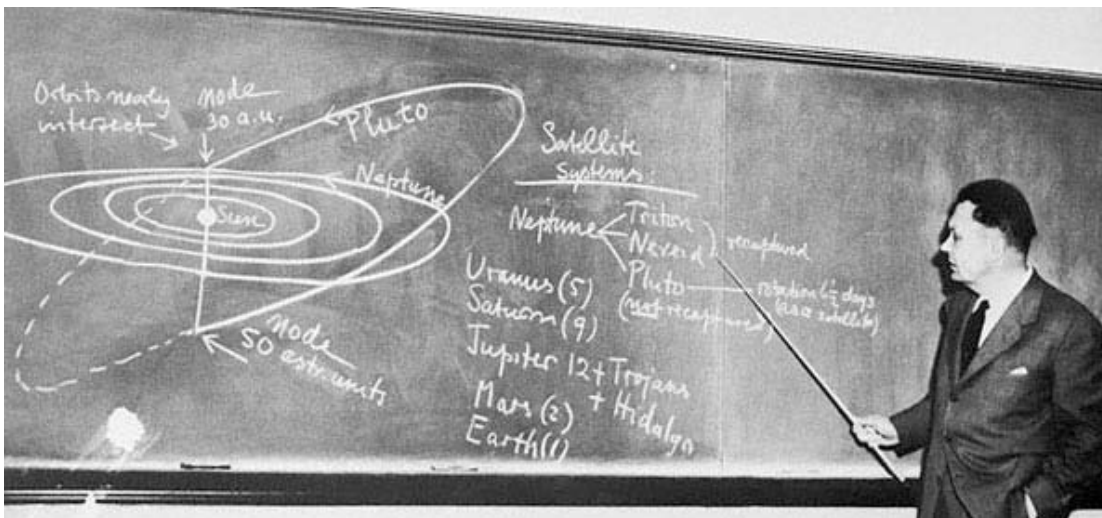
Source – IMEX

**Powerful Tools** – Robust databases and a growing volume of interactive data are brought together and massaged using sophisticated big data analytics tools to harvest valuable information that can be used and even sold. **Source – IMEX**

That’s right, there’s extremely valuable information in the patterns, behaviors identified in the data ... even your tweets.

### Fast, Accurate

The trouble is, the answer doesn’t just jump out and say, “here I am,” even with the best big data software. Interpreting, explaining, using the universe of data takes really smart people.



**Explanation, Interpretation** – While big data, big content, big stuff is growing faster than Moore’s Law, the biggest issue facing the industry is that not enough engineers, scientists, computer experts are being “produced” by our educational institutions.

It’s like Hal Varian, Google’s chief economist, said, “The sexy job in the next ten years will be statisticians... The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it.”

The way our systems, mobile devices, sensors are cranking out data; it's no wonder files will increase by 75X for the foreseeable future.

The challenge is, the number of new rock star techies will only increase by 1.5X.

The upside is, our son can intelligently carry on about things like string theory ... we have to work at making a ball outta' string.



Like the Scarecrow, our son looks at us and says, *“I don't know ... But some people without brains do an awful lot of talking... don't they?”*

###