Cloud Storage:

What is it and how will it enable your business?

Collecting data has rarely been a problem for business, and the falling cost of storage makes it relatively cheap to retain information indefinitely. But the low headline cost of storage masks a larger problem - the cost of managing ever increasing volumes of data.

79% of IT managers named managing storage growth as their most pressing business pain, well ahead of backup and recovery in second place.

79%
of IT managers named **managing storage growth** as their most pressing **business pain**

BACKUP & RECOVERY

Introduction

Collecting data has rarely been a problem for business, and the falling cost of storage makes it relatively cheap to retain information indefinitely. But the low headline cost of storage masks a larger problem - the cost of managing ever increasing volumes of data. **79%** of IT managers named managing storage growth as their most pressing business pain, well ahead of backup and recovery in second place.

Collecting and storing information is crucial to modern Big Data programs, but so too is the ability to retrieve and manage data effectively. Which is where Cloud Storage can help.

Contents

What is Cloud Storage?

Traditional storage techniques split files into evenly-sized 'blocks' of data, which are then assigned a physical address to assist with retrieval. These blocks are written to disk as-is, with nothing but the address to identify them.

Cloud Storage technologies have been designed with a view to assisting with the management of files. Instead of being split, files are written to disk as a single object (hence the name), with a physical address to aid with retrieval. More importantly still, objects can be assigned metadata that describes the content of each file, making it easier to describe and manage the information contained therein.

The metadata attached to each file can be as long or as complex as you need to properly describe and manage the data.

Why choose Cloud Storage?

Cloud Storage has some specific benefits that will make it an essential part of your data strategies in future. The technology is not new however more than half of the world's top 500 super computers are reliant on an Object-based file system. More visible are consumer services built on Cloud Storage platforms like Facebook, Spotify and Dropbox.

Every one of these applications has a single factor in common - the need to store huge amounts of data - typically in the form of files that are saved or synchronised to a Cloud-type platform.

But what can Cloud Storage do for your business?

Meeting the Big Data challenge

The name says it all - Big Data is big, requiring huge amounts of storage. Block-based storage allows businesses to kick-start Big Data programmes, but they will quickly run into problems with durability as capacity demands grow beyond a hundred terabytes.

According to Gartner, annual growth rates are expected to be at least 35%. And 52% of businesses already have over 100 terabytes of data to manage, which means they are wrestling with block storage, or have begun to deploy Cloud Storage alternatives to overcome the challenges of hyperscale data.

> *"Legacy infrastructure was designed in scale-up architecture, and has many shortcomings in terms of capacity, concurrency and data mining. As such, scale-out architecture is the answer to the data growth challenge. As well as this, legacy infrastructure is not designed for variety data and data intensive compute."*
>
> **Yanhua Xiao** - big data solution CMO, Huawei.

Cloud Storage is also perfect for Cloud deployments, allowing capacity and resilience to be increased simply by adding new nodes. This makes it particularly easy to deploy objects across a distributed system, simplifying remote access and high availability. Replicating data in Cloud

Storage also helps to reduce the number of points of possible failure in the platform, preventing catastrophic data loss. And the self-healing nature of a proper Cloud Storage deployment helps to prevent data loss down to the individual drive level.

Building platforms using Block Storage also helps to reduce running costs by simplifying management. The administrative overheads of managing multiple nodes and LUNs is exponentially higher in a file storage environment, ruling out options like multi-site replication for resilience and redundancy. In fact, in the age of the Cloud, file storage systems are completely at odds with the high volume, low costs model required.

The ability to tag objects with extensive metadata only serves to further increase the value of information in a Big Data deployment. Object metadata can be decentralised as required too, helping to further improve performance of queries and analytics.

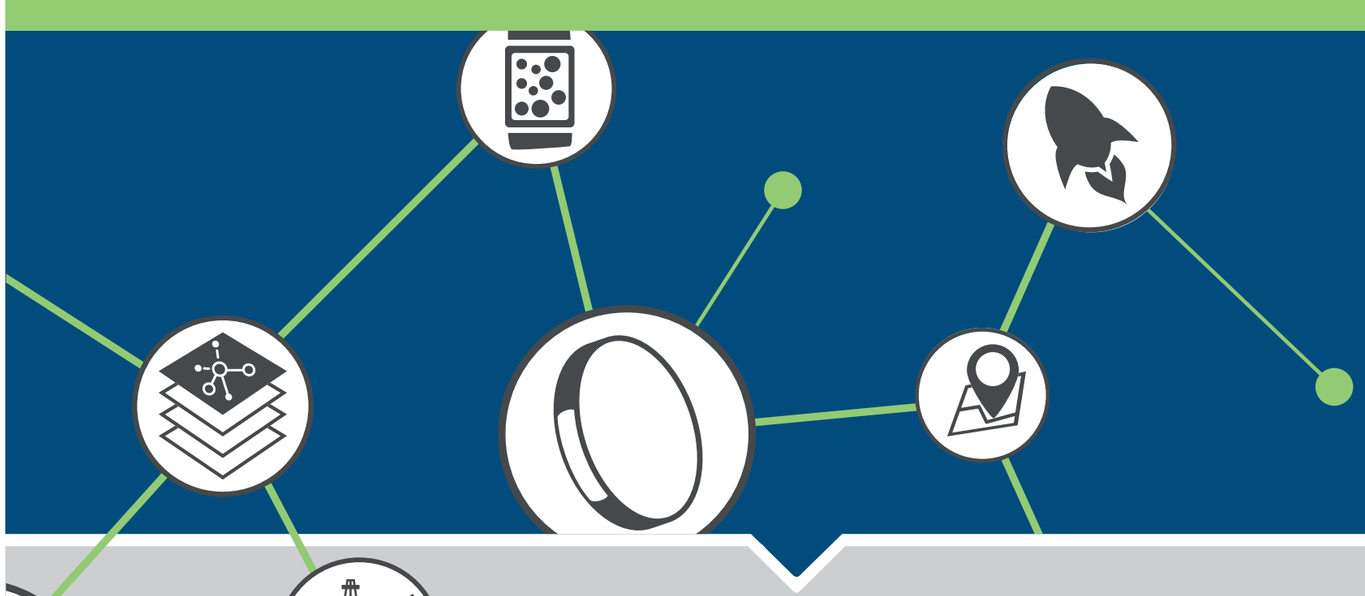
Enabling the Internet of Things

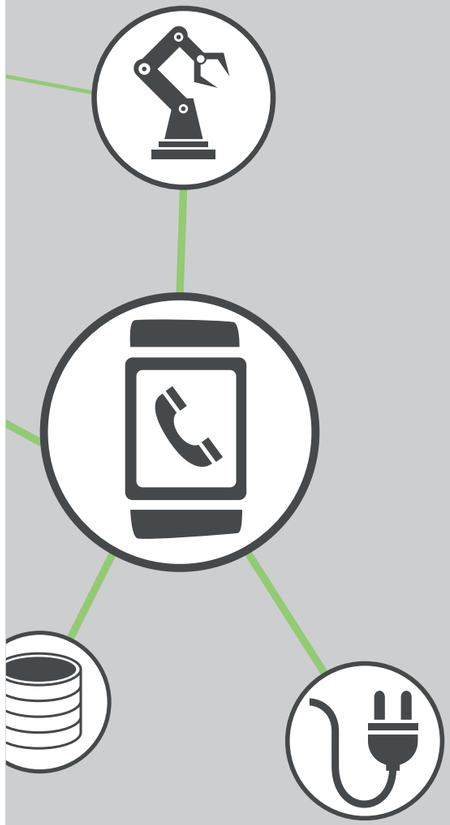
Current data growth rates are impressive, but as projects built around Internet of Things (IoT) technologies come online, that growth will accelerate further still. IBM report that 90% of the world's data has been produced in the last two years, and much of that from sensors and intelligent hardware.

> *"The issue is not only data growth. Users have always had a lot of data, much of which they might not have accessed. Now they are being forced to mine it for competitive insights. In terms of data growth, data is now of a granularity that was not dreamt of five years ago. IoT devices are a good example. Legacy infrastructure is struggling because the systems were not set up for dynamic, near real-time or real-time analytics and visualisation."*
>
> **Simon Garland** - chief strategist, Kx Systems.

Like every other metric, your business is expected to store data captured from IoT sensors. Typically just 1% of your devices are Network-connected , but falling costs and the need for more accurate data and increased automation will see the number of deployed sensors grow. Gartner estimate a 30% increase in the number of IoT devices in use during 2016, rising to 6.4 billion in total.

Businesses will need to have a scalable storage platform in place before their IoT programs come on line, or they may encounter problems managing that new data efficiently. Approximately 5.5 million new devices are connected every day, so implementing a Cloud Storage back-end will become essential sooner rather than later.





When is Cloud Storage inappropriate?

Because files are stored as whole objects, it is impossible to incrementally edit a single part of a file - as you can with block storage file systems. Instead the whole object needs to be loaded, edited, and re-saved, potentially creating performance overheads which become even more significant when dealing with massive data sets.

The majority of Big Data and IoT programs collect and report on data though – there is rarely any call to update the files held in Cloud Storage. The overheads associated with updating objects are unlikely to have a major effect on Big Data programs. Which is another reason for using Cloud Storage-based platforms for bringing data from tape archives, back into the Hadoop data store.

Cloud Storage & the future

The platform of the future will need to deliver exceptional speed for line-of-business applications, and vastly scalable storage for Big Data analytics. Because these two requirements are at odds with each other, there is a case for deploying traditional block storage for applications requiring speed, and Cloud Storage for everything else.

As well as facilitating the convergence of Big Data and IoT platforms, Cloud Storage is set to further revolutionise the enterprise data environment.

NoSQL will dominate

The schema-less needs of unstructured Big Data need a similarly schema-less database to help index content. NoSQL databases provide a way to store, index and retrieve files, helping to unlock value from Big Data.

RDBMS vendors like Oracle and Microsoft will be the big losers as demand for NoSQL systems surge. Expect to see products from MongoDB, DataStax, Redis Labs, MarkLogic and DynamoDB become increasingly important in enterprise database enviornments.

Apache Spark will become the lingua franca of Big Data analytics

Cloud Storage may be slower performing than block storage, which is why any tools that can increase performance for Big Data projects will become vital. Originally part of the Hadoop ecosystem, Apache Spark has gone on to become a vital tool in its own right.

Spark Streaming makes it possible to batch process data to deliver near real-time analytics. The ability to generate analytics on the fly is crucial to realising the maximum value from Big Data - particularly when those insights may be time sensitive.

Hadoop will come of age

Business can see the value of Big Data, and early testing using Hadoop have proven the value of these proof of concept programmes.

The next step is to move these projects into production, widening their scope to encompass other business units.

76 percent of those who already use Hadoop plan on doing more within the next 3 months and almost half of the companies that haven't deployed Hadoop say they will within the next 12 months .

Cloud Storage also offers businesses the opportunity to put their tape archives back online permanently. With the problem of scale resolved, archive data can be stored on inexpensive hard drives, and plugged into the Hadoop engine, providing additional context to Big Data analytics programs.

Data discovery for end users will become simpler

The ongoing skills shortage means that data scientists will remain in short supply. Big Data programmes cannot wait indefinitely, so new tools are being developed to simplify the process of discovering data and insight for end users.

These data visualisation tools (from the likes of Alteryx, Trifacta, Paxata and Lavastorm) greatly reduce the time and expertise required to make generate actionable insights from Big Data. The ability to access objects and their accompanying metadata will further simplify the process of finding and using the most relevant information.

MPP data warehouse growth is moving to the Cloud

Cloud Storage makes use of commodity hardware, helping to drive down Data Centre acquisition costs. However, running costs continue to grow too; Gartner calculate that it costs $3351 to store 1 terabyte of data each year, plus up to $4000 to backup that same terabyte. A full time expert to manage data storage is a further $1000 per terabyte per year .
If the annual growth rate of unstructured data really is running at 35%, storage costs will quickly outstrip IT budgets.

> *"As companies recognize the advantage of side-stepping Hadoop hardware requirements, which becomes outdated every 18 months, Cloud adoption will surge."*
>
> **Stefan Groschupf** - CEO of Datameer.

Most businesses have already begun to address this challenge using Cloud platforms, that allow them to offset most of the running costs in return for a utility billing model. Second generation services like Snowflake, BigQuery and Teradata have begun to eat into the market share of established players like Redshift from Amazon though

Hosted Cloud Storage-based data warehousing services will continue to grow in popularity in the near future.

How Exponential-e can help you?

Professional Services:

Our highly qualified and experienced Professional Services team at Exponential-e will examine your existing infrastructure, consider your business' future needs and mitigate any risks to ensure a successful transition from Windows Server 2003. These practitioners are multi-skilled and uniquely placed to drive your business to an IT environment that will enable growth, flexibility, mobility of data, added security and compliance.

www.**exponential-e**.com/professional-services

IT Audit:

Avoid the headache of manually reviewing your IT estate with our IT Audit solution. Monitor all activity across your infrastructure with our single reporting tool, allowing you to see minute detail on things such as software installations and the utilisation statistics for individual users. The performance of individual machines will also not be affected due to the non-intrusive nature of the install.

www.**exponential-e**.com/it-audit

Cloud Storage demands attention now

Any business can build a Big Data programme, but without the correct planning and implementation, they will never realise the benefits expected. It is absolutely vital that scalability is built in from the start, or businesses may have to rebuild platforms as they encounter the limitations of the underlying file system.

Cloud Storage is the perfect way for organisations to increase capacity indefinitely, and to increase future expansion options. Migration to a hybrid or hosted Cloud platform is greatly simplified by using the same underlying file system, further increasing available expansion options and keeping costs low.

For any organisation serious about mastering Big Data and IoT projects, Cloud Storage offers unrivalled flexibility and extensibility, unavailable through any other file system.

To learn more about Cloud Storage and how it can be used by your business to reach its data goals, please get in touch.

Works Cited: **Managing Storage: Trends, Challenges, and Options (2013-2014) – EMC - http://www.richardsoneyres.co.uk/wp-content/uploads/2015/11/EMC_Managing_Storage_Trends.pdf. Open-source Lustre gets supercomputing nod – CNET - http://www.cnet.com/news/open-source-lustre-gets-supercomputing-nod/**
How to Calculate the Total Cost of Cloud Storage – Gartner - http://docs.media.bitpipe.com/io_12x/io_121933/item_1123732/Gartner%20-%20How%20to%20Calculate%20the%20Total%20Cost%20of%20Cloud%20Storage.pdf
Managing Storage: Trends, Challenge and Options (2013-2014) – EMC - http://www.richardsoneyres.co.uk/wp-content/uploads/2015/11/EMC_Managing_Storage_Trends.pdf
What is Big Data? – IBM - https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html
Big Data and What it Means – US Chamber of Commerce Foundation - https://www.uschamberfoundation.org/bhq/big-data-and-what-it-means
Garter Says 6.4 Billion Connected "Things" Will Be in Use in 2016, Up 30 Percent From 2015 – Gartner - http://www.gartner.com/newsroom/id/3165317
Big Data Industry Predictions for 2016 – Inside Big Data - http://insidebigdata.com/2015/12/08/big-data-industry-predictions-2016/
IT Key Metrics Data 2015: Key Infrastructure Measures: Storage Analysis: Multiyear – Gartner - https://www.gartner.com/doc/2934817/it-key-metrics-data-
Big Data Industry Predictions for 2016 – Inside Big Data - http://insidebigdata.com/2015/12/08/big-data-industry-predictions-2016/

exponential-e
APPLIED INNOVATION

www.**exponential-e**.com

Telephone
+44 (0) 845 470 4001

Visit the website
www.exponential-e.com

bsi.
ISO 9001 Quality Management
ISO 14001 Environmental Management
ISO/IEC 20000-1 Information Technology Service Management
ISO/IEC 27001 Information Security Management
CSA STAR Cloud Security
ISO 22301 Business Continuity Management
ISO 50001 Energy Management

REF:1153