Tame your data and boost your confidence in pressing delete!

IRMMS Conference, 21st May 2018

Richard Keeble
Client Principal, Professional Services
richard.keeble@microfocus.com
Who exactly, are you?
Micro Focus acquired HPE Software
A long history of stability, innovation and acquisitions

### Micro Focus

**History and Background**
A global software company with 40 years' experience empowering organizations to innovate safely and meet complex business demands. Micro Focus unifies emerging technologies and the IT investments that customers rely on every day

- Headquarters: Newbury, UK
- 20,000+ Customers
- 91 of the Fortune Global 100 companies
- 4,500+ employees

**Key Acquisitions**
- SERENA
- NetIQ
- Borland
- Novell
- SUSE
- Attachmate

**Key Innovations**
- Visual COBOL
- Multi-Factor Authentication in Host Connectivity
- Platespin Workload Management

**Annual Revenue**
$1.4 billion

### HPE Software

**History and Background**
A global enterprise IT provider born directly from the rich heritage of HP innovation – originally formed as part of HP's Networking group more than 23 years ago. Since then, the business has evolved through continued innovation and acquisition to become a market leader

- Headquarters: California, USA
- 30,000+ customers
- 98 of the Fortune Global 100 companies
- 13,500+ employees

**Key Acquisitions**
- ArcSight
- Voltage
- Peregrine
- Vertica
- Opsware
- Iron Mountain

**Key Innovations**
- Application Defender
- Cloud Service Automation
- Mobile Center
- Service Virtualization

**Annual Revenue**
$3.1 billion
Combined Micro Focus Portfolio Overview
Strong, broad and deep portfolio

Testing and lifecycle tools for traditional and DevOps

Automate, orchestrate and transform IT operations

Manage cloud, from private to public

Protect users, apps and data across locations and devices

Help customers understand, organize, control and protect information

Linux and open source for the enterprise

Extract value through advanced analytics

ALM Unified Functional Testing

AppPulse Service Management, Operations Bridge, Data Center Automation, Network Management

Cloud Service Automation, Hybrid Cloud Management

Digital Safe, Data Protector, Control Point, Structured Data Manager, Storage Optimizer

Vertica

IDOL

Cobol Development and Mainframe Solutions, Application Development and Testing

Development and IT Operations Management, Host Connectivity Collaboration

Identity-based Access Governance and Security

Identity-based Access Governance and Security

Enterprise Linux, OpenStack Private Cloud, Software-defined Storage
What’s the problem?
I remember when...
Then came computers...
...and storage
…and more storage
...and more storage
...and we keep on filling it up
The problem gets worse

Traditional
- Spreadsheets
- Word processing
- Presentations
- Databases

New media
- Video
- Digital Audio
- Pictures
- Music
- CCTV
Where does all the data come from?
Big data

Given an infinite length of time, a chimpanzee punching at random on a typewriter would almost surely type out all of Shakespeare's plays.

(Wikipedia)
Connected people, apps and things generating massive data in many forms

10x faster growth than traditional business data

Human data

Machine data

Business data
Exactly how much data?

- 90% of all digital data was created in the past 2 years
- 2.5 quintillion bytes added every day
- That’s 2,500,000,000,000,000 bytes
- 2,500,000,000,000 KB
- 2,500,000,000 MB
- 2,500,000 GB
- 2,500 Terabytes
- 2.5 Petabytes
- DAILY

(Source – betanews.com)
But we stored it all. So what’s the problem?
In a changing world
And let’s not forget the regulations change too

Data Protection Act 1998

General Data Protection Regulation

Freedom of Information Act 2000
The consequences of not handling data properly?
Financial penalties
Reputational damage
A quick recap

- We have too much information
- It’s stored in too many places
- We don’t know where it all is
- We don’t know what it contains
- There is a huge amount of regulation
- We have to protect the information
- There’s a massive cost for failing
I know all this. What’s the answer?
Artificial Intelligence and Machine Learning
Human Information Processing – an example

Key Words are small amounts of very strong information without “context”

“Mercury”

Is it a planet?
Is it an element?
Is it a car?

With high certainty; it’s an element!

“A heavy element and the only metal that is liquid at standard conditions for temperature and pressure with the symbol Hg and atomic number 80, commonly known as quicksilver”

\[
\begin{array}{|c|c|c|c|c|c|c|c|}
\hline
\text{H} & \text{He} & \text{Li} & \text{Be} & \text{B} & \text{C} & \text{N} & \text{O} \\
\hline
\text{F} & \text{Ne} & \text{Na} & \text{Mg} & \text{Al} & \text{Si} & \text{P} & \text{S} \\
\hline
\text{Cl} & \text{Ar} & \text{K} & \text{Ca} & \text{Sc} & \text{Ti} & \text{V} & \text{Cr} \\
\hline
\text{Mn} & \text{Fe} & \text{Co} & \text{Ni} & \text{Cu} & \text{Zn} & \text{Ga} & \text{Ge} \\
\hline
\text{As} & \text{Se} & \text{Br} & \text{Kr} & \text{Rb} & \text{Sr} & \text{Y} & \text{Zr} \\
\hline
\text{Nb} & \text{Mo} & \text{Tc} & \text{Ru} & \text{Rh} & \text{Pd} & \text{Ag} & \text{Cd} \\
\hline
\text{In} & \text{Sn} & \text{Sb} & \text{Te} & \text{B} & \text{Be} & \text{Al} & \text{Si} \\
\hline
\text{Ga} & \text{Ge} & \text{As} & \text{Se} & \text{Te} & \text{Pb} & \text{Bi} & \text{Po} \\
\hline
\text{At} & \text{Rn} & \text{Fr} & \text{Ra} & \text{Ac} & \text{Th} & \text{Pa} & \text{U} \\
\hline
\end{array}
\]
Traditional information association

- Create keywords to search the data
  - Mispelingz?
  - Abbr.
  - Synonyms
- Train the algorithm with a known dataset
  - Prone to failure when structures change
Natural Language Processing
Using Cognitive Analysis to form a human-like understanding of content

- Fundamentally created to understand natural human language using probabilistic modeling and NLP algorithms
  - Allows incoming data to dictate the model, not pre-defined rules, dictionaries, or semantic webs
- Self-Learning / Machine Learning
  - Updates as more data is added or removed
  - Adapts to changing definitions or meaning
- Fundamentally language-independent
  - Treats words as symbols

\[ P(\theta \mid x) = \frac{P(x \mid \theta) \cdot P(\theta)}{\sum_{\theta' \in \Theta} P(x \mid \theta') \cdot P(\theta')} \]

Information Theory and Bayesian Inference

\[ H = -\sum p_i \log_2(p_i) \]
Eduction – augmenting NLP

Hundreds of conceptual entities

- Quickly narrow search results with auto-identified facets and conceptual entities such as employee names from documents

- Validate or customize entities
  - Is this a valid credit card number?
  - What are all docs that contain drivers licenses?
  - If area code is 01273, output as Brighton

Names
Places
IP addresses
Companies
Events
Relationships
Medicines
Airports
Cars
National Insurance numbers
Phone numbers
Credit cards
Dates
Holidays
Job titles
Currencies
... many more
Secure Content Management: pieces and process

Active system repositories - unstructured data

File Analysis
- Identify
- Analyse
- Categorize
- Curate

Structured Data Management
- App retirement
- Data extraction
- Data masking

Legacy databases – structured data

Enterprise Content Management
- Lifecycle management
- Intelligent archiving

Redundant, Obsolete and Trivial (ROT)

Apply policy to manage in place

Manage-in-place

Auto classify
- Policy
- Security
- Redaction
- Encryption
- Authorized access
- Reporting
- Audit logs

Defensibly dispose

Connect

Identify PII, PCI, PHI

Declare content and records

Identify

Validate data

Legacy data and ROT
Where is this heading?
(Sorry, this does mention our solution)
Chatbots driven by AI & Machine Learning

Have you ever heard this?

“The searching capabilities are amazing, with support for Boolean Logic search”

There is an alternative...
Introducing IDOL Natural Language Question Answering

FAQ support documents

IDOL Answer Bank

Factual insight from both structured & unstructured data sources

IDOL Fact Bank

Trusted document repositories and public data sources

Relevant passages extracted from documents

IDOL Passage Extraction

Factual answers from database or entities in documents

Question variations mapped to approved question/answer

IDOL Answer Server
Put it together & what do you get?

- Huge volumes of data, indexed and “understood” by AI
- New classification and clusters created by Machine Learning
  - some human direction, building information governance policies that AI applies to the data
  - *no need for user classification or declaring records*
  - Important information automatically copied/moved to managed and secure storage
- Interaction with all the data through Natural Language Processing – including verbal interaction
- Automated selection of data for disposal – AI & ML finding information based on previous disposition and suggesting it to you
- Reduced time, cost and risk
Use case – come see us on the Micro Focus / Oyster ims*
Stand upstairs.

Follow the smell of the best coffee at the conference!