The future of Unstructured Data Workloads

#OFADevWorkshop
Tasneem Maistry Pivotal

- Apps power businesses, and those apps generate data.
- Analytic insights from that data drive new app functionality, which in-turn drives new data.
- The faster you can move around that cycle, the faster you learn, innovate & pull away from the competition.
Pivotal’s Opportunity

- Uniquely positioned to help enterprises modernize each facet of this cycle today
- Comprehensive portfolio of products spanning Apps, Data & Analytics
- Converging these technologies into a coherent, next-gen Enterprise PaaS platform
The exponential increase in business expectations is unsustainable for IT.

Pivotal Can Help Bridge the Gap

IT Technology Eras

Mainframe  |  Client-Server  |  Mobile-Cloud

IT at the Speed of Business

Business Expectations

Ability of IT to Deliver

Steady IT Budgets*

IT is Constrained by Technology

Silos

72% of IT operating budget spent on maintaining old technologies

Modern Silos

Legacy Silos

[Images of various hardware and software logos, such as IBM, Linux, Windows Server, Oracle, and Sun.]
Liberating Resources Serves Customers

Data isn’t available.

Analysis takes too long.

App dev takes too long.

Provisioning takes weeks.

Production workloads are managed via email.

Industry Average: IT Spend on Innovation

Liberating Resources Serves Customers

Any user, any device, any app... automatically.

Apps get rolled out at “the speed of business.”

Provisioning a production environment takes minutes.

You can routinely deploy any workload...anywhere.

Goal: IT Spend on Innovation

Source: VMware Journey Benchmark Survey, 4th Wave 2013
## Interfaces and Usage Patterns vs. Hosting Solutions

<table>
<thead>
<tr>
<th>Service Offered</th>
<th>Users and usage</th>
<th>More dynamic and autonomous usage patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>High performance file system</td>
<td>Increased set-up time, but longer usage</td>
<td></td>
</tr>
<tr>
<td>Shared Scientific Analysis hosts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hosted Processing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hosted Infrastructure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application Hosting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual Storage for other organisations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloud Federation / Brokering</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PaaS - Hosted Analysis Environments</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bare metal

Increasing virtualisation
The Need for an Application Centric Layer

- Developers can focus on development and not infrastructure plumbing
- Separate the concerns of AppDev and Operations
- Eliminate the bottleneck of provisioning and deployment processes
Build your cloud for today and tomorrow

IT Operations

Manual

Automated

Storage, Compute, Network

Server Virtualization

App and Desktop Virtualization

Cloud Orchestration

Private & Public IaaS

Multi-Service Broker

IaaS

DaaS

SaaS

PaaS

STaaS

ITaaS

Infrastructure

Apps

Workloads

Services

Business Impact

IT Only

Departmental

All
Why is this relevant and important

- Sensitivity analysis of each platform component on the outcomes (QoS, analysis etc)
  - What does it need .. Logs, users, processing, and processes
  - What is missing -- processes to capture and collect these
  - What is missing -- people who can analyze this and come up with learnings
  - What can AWB provide -- lots, users, processing
  - NEED VOLUNTEERS TO do research on the cluster
Pivotal Analytics Workbench (AWB)

- 1000 node cluster
- Collaborative project with industry leaders (Mellanox, Intel, Seagate)
- Provides mixed mode environment (GPDB, GemXD, PCF, PHD)
- Contains entire Hadoop stack (HDFS, HBASE, PIG, HIVE)
Partners

- Intel
  - Contributed 2,000 6-core CPUs.

- Mellanox
  - Contributed >1,000 network cards and 72 switches.

- Micron
  - Contributed 6,000 8GB DRAM modules.

- Seagate
  - Contributed 12,000 2TB Drives
Cluster
Cluster Size

- Physical Hosts - More than 1,000 nodes
- Processors - Over 24,000 CPU’s
- RAM – Over 48TB for memory
- Disk capacity – More than 24PB of raw storage.
  - “Equivalent to nearly half of the entire written works of mankind from the beginning of recorded history”
Typical Target Use Cases

- Pivotal demonstration
- Partner engagements (Intel, Mellanox, Seagate)
- Industry and academia collaboration (Alpine, Informatica)
Submitting your Proposal

- Request invite @ www.analyticsworkbench.com
- Project determination made within 2 weeks of submission
Thank You