Data-Driven Education, Research, & Trade Using the Cloud

Dilum Bandara, PhD
Dept. of Computer Science & Engineering
University of Moratuwa
Dilum.Bandara@uom.lk

Seychelles and Sri Lanka: Investment and Education – 2016
Big Data, Cloud, & Analytics

<table>
<thead>
<tr>
<th>Cloud Computing</th>
<th>Big Data</th>
<th>Data Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Computing as a Utility</td>
<td>• Volume</td>
<td>• Science of examining data</td>
</tr>
<tr>
<td>• Pay as You Go</td>
<td>• Velocity</td>
<td>• Data driven decisions</td>
</tr>
<tr>
<td>• XaaS</td>
<td>• Variety</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Veracity</td>
<td></td>
</tr>
</tbody>
</table>

Science of examining data
Data Analytics

#1 Descriptive
Analytics that helps you understand how things are going.

#2 Predictive
Analytics that help you forecast future performance and results.

#3 Prescriptive
Analytics that suggest a prescribed next step or action.

Source: https://moz.com/blog/when-it-comes-to-analytics-are-you-doing-enough
Figure 2. Gartner Analytic Ascendancy Model

Tools can help reduce difficulty
Case Studies – Data Analytics
Your Options

Consolidation

Colocation

Public Cloud
Virtualization

- Act of creating a virtual version of something
  - Hardware platforms, operating systems, storage devices, & computer network

Source: VMWare
Public/Hybrid Cloud

- Pay-as-you go
- High agility & scalability
- Low cost
- Better Security
- Low control & privacy concerns
- Hybrid clouds
  - Simultaneous scalability, cost, performance, & privacy
  - Private & public cloud should be interoperable
Private Cloud

- Same as a public cloud, but within organization
- Better agility, performance, control, & security
- Still limited by available resource pool

Source: Microsoft
Virtualization in Cloud Computing

Cloud Computing = Software as a Service + Platform as a Service + Infrastructure as a Service + Data as a Service
Choosing among IaaS, PaaS, vs. SaaS

- **IaaS**
  - Flexibility, finer control, & performance
  - Still need some level of infrastructure maintenance

- **PaaS**
  - Speedy development, better integration, automated scaling, no maintenance needs
  - Relatively low-customization, Vendor lock-in

- **SaaS**
  - Fastest for common applications
  - Little customization

**SaaS First, PaaS Second, & IaaS Last**
Selection of a Cloud Service

- Know what you want first
- What services are available?
- What is your pricing model?
- What are your scaling options?
- What are your security measures?
- Where are your datacenters located?
- What are SLA terms?
- Customer support
- Reputation
When Not to Use Cloud

- Performance critical applications
  - Unless virtualizing for scale & reliability
- I/O intensive applications
- Colocation of conflicting workloads
- When restricted by licensing terms
  - Applications/OSs with hardware keys
- Not every OS works on latest hypervisors
  - Hypervisors are known to have driver issues
  - Old virtualized OSs may still have driver issues on newer hardware
- When restricted by regulation
  - Be creative
  - Check possibility of changing regulation
Case Studies – Consolidation, Colocation, Public Cloud

- Revenue Administration Management Information System (RAMIS)
- Integrated Treasury Management Information System (ITMIS)
- Online University Admissions
- National Payment Platform (NPP)
- Dept. of Computer Science & Engineering
  - Google Web Apps for Education
  - Office 365
- Touch Pass
  - NFC based bus ticketing
- Microsoft Azure Cloud in Germany
- Mexico IRS
Case Studies – Cloud Research

Lecturer
Need 15 identical machines for a b for 15 students from 8:00 am to 10:00 am

Scientist/Researcher
Need an 8 core machine with 50 GB storage and 16 GB memory for Seismic Wave analysis Application

Student
Need a dual core machine with 8 GB memory for a Data Mining application

Rule Based Engine

Auto Scaler

Graph showing cost over time with different labels and metrics.
Data Science & Cloud Computing in CSE Curriculum

**Undergraduate**
- Distributed Systems
- Mobile Computing
- High Performance Networking
- Machine Learning
- Data Mining & Information Retrieval
- Bioinformatics

**Postgraduate**
- MSc in Data Science
- MBA in IT in Data Analytics
- MPhil & PhD
- Distributed Computing
- Cloud Technologies and Systems
- Software Development on Cloud Platforms
- Data Science
- Business Intelligence
- Descriptive and Predictive Analytics
- Enterprise Information Management
- Data Mining
- Pattern Recognition
- Natural Language Processing