Increasing Business Insights and ROI with Microsoft’s Data Analytics Platform

Chris Fleming – Director Data Analytics

October 19, 2015
1. A few thoughts on becoming an “Analytic Competitor”.
2. High level overview of Microsoft Azure.
3. An overview of Data Services in Azure.
4. Spend some time on:
   • Data Science/Machine Learning – Azure ML
   • Insightful Analytics/Power BI - Demos
5. Q&A

Credit where credit is due: Many slides were sourced from Microsoft decks and I am grateful for the assistance.
Back to the future

What’s in a day?

January 1<sup>st</sup> 2006
Some companies have built their very businesses on their ability to collect, analyze, and act on data. Every company can learn from what these firms do.

Competing on Analytics

by Thomas H. Davenport
“It is nearly impossible for us to differentiate ourselves from our rivals by products alone”. To pull ahead we need to:

— Use sophisticated data-collection technology and analysis to wring every last drop of value from our business processes.

— Make analytics part of an overarching competitive strategy to out-think and out-execute your competitors.

— Push this capability down to decision makers at every level.
1. Champion Analytics from the top.
   - Understand our business and know how analytics can be applied to it.

2. Focus our Analytics efforts.
   - Channel our resources into analytics initiatives that most directly serve our overarching competitive strategy.

3. Establish an Analytics culture.
   - Instill an organizational wide respect for measuring, testing, and evaluating quantitative evidence.

4. Use the right technology.
   - Store information in a data warehouse, and make it easily accessible to everyone.

5. *Be nimble*…This is where Azure comes in!
WHAT IS AZURE?

A cloud offering from Microsoft that supports and integrates with an organization’s technology capabilities

Provides a platform for data center scale out, application development and analytics

Supports demanding and new scenarios where a world of devices and information is exponentially increasing

A trustworthy platform where your data remains your data

A network of services (~200) that build solutions
Huge infrastructure scale is the enabler

20 Regions ONLINE...huge datacenter capacity around the world...and we’re growing

- One of the top 3 networks in the world (coverage, speed, connections)
- 2 x AWS and 6x Google number of offered regions
- G Series – Largest VM available in the market – 32 cores, 448GB Ram, SSD...
Cloud vision: one cloud to host them all

- Acquiring new customers through new channels
- Redesigning business processes
- Increasing employee productivity with ubiquitous access
- Converting capital expenses to operating expenses
- Redeploying IT talent onto more strategic projects
- Boosting business agility with instant-on capacity
- Delighting customers with personalized experiences and service
- Entering new markets by innovating with digital products and services
- Instantly scaling to meet demand
- Speeding application development
What does Azure really do?

**Virtual Machines**
- Hypervisor
- Storage
- Gallery

**Cloud Services**
- Web Roles
- Worker Roles
- Queues and Storage

**Web Sites**
- Platform
- Deployment
- Gallery

**Mobile Services**
- Auth and Profiles
- Storage
- Notifications

**Data Services**
- SQL Databases
- Data Storage
- Hadoop/HDInsight
- Machine Learning

**Network Services**
- Traffic Manager
- Virtual Networks

**App Services**
- Service Bus
- Active Directory
- Media Services
- Scheduler
- Search
Azure Data Services

- Transform the datacenter
- Empower enterprise mobility
- Create the internet of your things
- Unlock insight on any data
- Enable application innovation

Unlock insight on any data
Create the internet of your things
Empower enterprise mobility
Transform the datacenter
Enable application innovation
SQL Capabilities in Azure
Azure Data Services SQL Server

- SQL DW
- Elastic, Petabyte Scale DW Optimized
- 99.99% uptime SLA, Geo-restore
- Azure Compliance (ISO, HIPAA, EU, etc.)
- True SQL Server Experience; Existing Tools Just Work
Symmetric Multiprocessing (SMP)

- CPU
- Cache
- Shared Memory
- I/O
- BUS

Massively Parallel Processing (MPP)

- Node 1
- Node 2
- Node 3
  - High-Speed Interconnect
  - A...F
  - G...R
  - S...Z
ADS Azure SDW:Data Warehouse Unit (DWU)

- **Measure of Power**: Simply buy the query performance you need, not just hardware
- **Transparency**: Quantified by workload objectives: how fast rows are scanned, loaded, copied
- **On Demand**: First DW service to offer compute power on demand, independent of storage

### Scan 1B rows

<table>
<thead>
<tr>
<th>DWU</th>
<th>Scan Rate</th>
<th>Loading Rate</th>
<th>Table Copy Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 DWU</td>
<td>= 297 sec</td>
<td>xx M row/sec</td>
<td>xx K row/sec</td>
</tr>
<tr>
<td>400 DWU</td>
<td>= 74 sec</td>
<td>xx K row/sec</td>
<td></td>
</tr>
<tr>
<td>800 DWU</td>
<td>= 37 sec</td>
<td></td>
<td>xx K row/sec</td>
</tr>
<tr>
<td>1,600 DWU</td>
<td>= 19 sec</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**ADS Azure SQL Data Warehouse : DW600**

**Azure SQL Data Warehouse**

- **Engine**
- **Worker1**
- **Worker2**
- **Worker3**
- **Worker4**
- **Worker5**
- **Worker6**

**Azure Storage Blob(s)**

- D1
- D2
- D3
- D4
- D5
- D6
- D7
- D8
- D9
- D10
- D11
- D12
- D13
- D14
- D15
- D16
- D17
- D18
- D19
- D20
- D21
- D22
- D23
- D24
- D25
- D26
- D27
- D28
- D29
- D30
- D31
- D32
- D33
- D34
- D35
- D36
- D37
- D38
- D39
- D40
- D41
- D42
- D43
- D44
- D45
- D46
- D47
- D48
- D49
- D50
- D51
- D52
- D53
- D54
- D55
- D56
- D57
- D58
- D59
- D60
Azure HDInsight
Azure HDInsight

Hadoop in Azure

- HBase as a columnar NoSQL transactional database running on Azure Blobs
- Storm as a streaming service for near real time processing
- Hadoop 2.4 support for 100x query gains on Hive queries
- Mahout support for machine learning + Hadoop
- Graphical User Interface for HIVE queries
Hadoop stores files (non-relational store)

Files could have a variety of semi-structured or unstructured data

Previously, these files may not have been seen as providing value or insights

Today, new business questions and insights are being uncovered through data science

**Sentiment**
Understand how your customers feel about your brand and products—right now

**Clickstream**
Capture and analyze website visitors’ data trails and optimize your website

**Sensors**
Discover patterns in data streaming automatically from remote sensors and machines

**Geographic**
Analyze location-based data to manage operations where they occur

**Server logs**
Research logs to diagnose process failures and prevent security breaches

**Unstructured**
Understand patterns in files across millions of web pages, emails, and documents
### ADS: A Hadoop distribution – a package of projects

Tested for consistency across entire package

<table>
<thead>
<tr>
<th>Hadoop and YARN</th>
<th>Data management</th>
<th>Data access</th>
<th>Governance and integration</th>
<th>Operations</th>
<th>Security</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDP 2.1 April 2014</td>
<td>2.4.0</td>
<td>0.40</td>
<td>0.12.1</td>
<td>0.13.0</td>
<td>0.98.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.00</td>
<td>1.5.1</td>
<td>0.9.1</td>
<td>0.9.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.7.2</td>
<td>0.5.0</td>
<td>1.4.4</td>
<td>1.4.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5.1</td>
<td>4.0.0</td>
<td>3.4.5</td>
<td>.04.0</td>
</tr>
<tr>
<td>HDP 2.0 October 2013</td>
<td>2.2.0</td>
<td>0.12.0</td>
<td>0.12.0</td>
<td>0.96.1</td>
<td>0.8.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.4.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.4.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.3.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.4.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.04.0</td>
</tr>
<tr>
<td>HDP 1.3 May 2013</td>
<td>1.1.2</td>
<td>0.11.0</td>
<td>0.11.0</td>
<td>0.94.6</td>
<td>0.7.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.4.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.3.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.3.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.4.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.04.0</td>
</tr>
</tbody>
</table>
ADS HDInsight supports

- Hive: Provides a SQL-like experience and queries on data in HDInsight
- Hbase: Column-oriented DBMS that runs on top of HDFS. Suited for sparse data sets, which are common in many big data use cases.
- Mahout: A library of scalable machine-learning algorithms.
- Storm: A distributed real time computation system for near-real time processing
- Spark: In memory computing. Contrast to Hadoop's disk-based MapReduce paradigm, Spark's in-memory capabilities can provide performance up to 100x faster for certain applications, by loading data into a cluster's memory and query it repeatedly
Azure Stream Analytics

Process real-time data in Azure

Consumes millions of real-time events from Event Hub collected from devices, sensors, infrastructure, and applications

Performs time-sensitive analysis using SQL-like language against multiple real-time streams and reference data

Outputs to persistent stores, dashboards or back to devices
Azure Data Factory

Orchestrate trusted information production in Azure

Connect to relational or non-relational data that is on-premises or in the cloud

Orchestrate data movement & data processing

Publish to Power BI users as a searchable data view

Operationalize (schedule, manage, debug) workflows

Lifecycle management, monitoring
Azure Machine Learning (ML)
What is Machine Learning?

Computing systems that become smarter with experience

“Experience” = past data + human input
Using past data to predict the future

Machine learning & predictive analytics are core capabilities that are needed throughout your business.
Azure Machine Learning

Get started with just a browser
Requires no provisioning; simply log on to your Azure subscription or try it for free off azure.com/ml

Experience the power of choice
Choose from hundreds of algorithms and packages from R and Python or drop in your own custom code

Take advantage of business-tested algorithms from Xbox and Bing

Deploy solutions in minutes
With the click of a button, deploy the finished model as a web service that can connect to any data, anywhere

Connect to the world

Microsoft Azure Machine Learning Studio Modeling environment (shown)

Microsoft Azure Machine Learning API service Model in production as a web service

Microsoft Azure Machine Learning Marketplace APIs and solutions for broad use
Azure ML Capabilities

- State of art ML Algorithms

Classification

Regression

Recommenders

Clustering
## Advanced Analytics scenarios

### EXAMPLE SOLUTIONS

<table>
<thead>
<tr>
<th>Sales and marketing</th>
<th>Finance and risk</th>
<th>Customer and channel</th>
<th>Operations and workforce</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales forecasting</td>
<td>Fraud detection</td>
<td>User segmentation</td>
<td>Agent allocation</td>
</tr>
<tr>
<td>Demand forecasting</td>
<td>Credit risk management</td>
<td>Personalized offers</td>
<td>Warehouse efficiency</td>
</tr>
<tr>
<td>Sales lead scoring</td>
<td></td>
<td>Product recommendation</td>
<td>Smart buildings</td>
</tr>
<tr>
<td>Marketing mix optimization</td>
<td></td>
<td></td>
<td>Predictive maintenance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Supply chain optimization</td>
</tr>
</tbody>
</table>
Advanced analytics architecture

**Data**
- Cloud
  - HDFS
  - RDBMS
  - NoSQL stores
  - Blobs and tables
- Local
  - Desktop files
  - Spreadsheets
  - Server stores
  - Sensors

**Data preparation**
- Data factory
- Stream analytics

**Modeling**
- Machine learning
- HDInsight

**Deployment**
- Marketplace
- Azure portal

**Apps, dashboards and processes**
- Microsoft Azure Marketplace
- Monetize this API

**Business value**
- Power BI
- Apps

**Data to model to web services in minutes**
- ML Studio
  - Integrated development environment for machine learning
- Web
  - Model is now a web svc
  - Monetize this API

**API**
- Apps
Excel/Power BI
See all your data in a single pane of glass
Live dashboards and interactive reports

- Monitor live dashboards for the data that matters most
- Track your data in real-time with support for streaming data
- Drill through to underlying reports to explore and discover new insight
- Pin new visualizations and KPIs to monitor performance
Stay connected from any device

- Native apps for iPad, iPhone, and Windows devices
- Receive alerts to important changes in your data
- Share and collaborate with colleagues and take action
Power BI overview

**Business analyst tools**
- Excel
- Power BI Designer

**On-premises data**
- SQL Server analysis services
- Databases and other data sources

**Microsoft cloud**
- Azure data services
- Office 365
- Dynamics
- Partner apps

**Web browser**

**Power BI**

**Non-Microsoft cloud software systems**
(e.g., SaaS applications, IoT scenarios, and data streams)

**Mobile apps**
Power BI Content Packs

Prebuilt Power BI experience for your service

TODAY

Microsoft Dynamics CRM
Visual Studio Online
GitHub
twilio
Dynamics Marketing
appfigures
Google Analytics
zendesk
MailChimp
SweetIQ
Acumatica
uservoice

COMING SOON

zuora
Office 365
sage
sumologic
SQLSentia
comScore
inkling
Adobe Analytics
SpaceCurve
Azure Mobile Engagement
Troux by planview
tyGraph

Circuit ID
PLANVIEW
Alteryx
insightly
Create Power BI content
Visual data exploration & interactive reporting

Desktop tool for data exploration and interactive reporting
Freeform drag and drop canvas
Effortless publishing to the Power BI service
Unlock your Excel data
Analyze data and publish your workbooks to Power BI

Prep and model data with familiar and flexible Excel

Publish to Power BI and view through Excel Online

Connect live to Excel workbooks in OneDrive
Ask questions of your data

Natural Language Query

Type a question and Power BI gives you the answer