

# Powering Up HDInsight with Power BI

**Optimize your data lifecycle.**





## ***Products***

Making Business  
Intelligent



## ***Services***

Consultation,  
Mentoring,  
Solutions



## ***Training***

Online Training,  
Webinars, and  
More

### **Steve Hughes**

*Principal Consultant Lead*

[shughes@pragmaticworks.com](mailto:shughes@pragmaticworks.com)

[www.dataonwheels.com](http://www.dataonwheels.com)

@dataonwheels



# Agenda

Setting Up an HDInsight Cluster

Uploading Files to the Cluster

Creating an External Hive Table

Power BI with Hadoop Files

Power BI with Hive

Wrap Up and Questions

# Setting Up an HDInsight Cluster

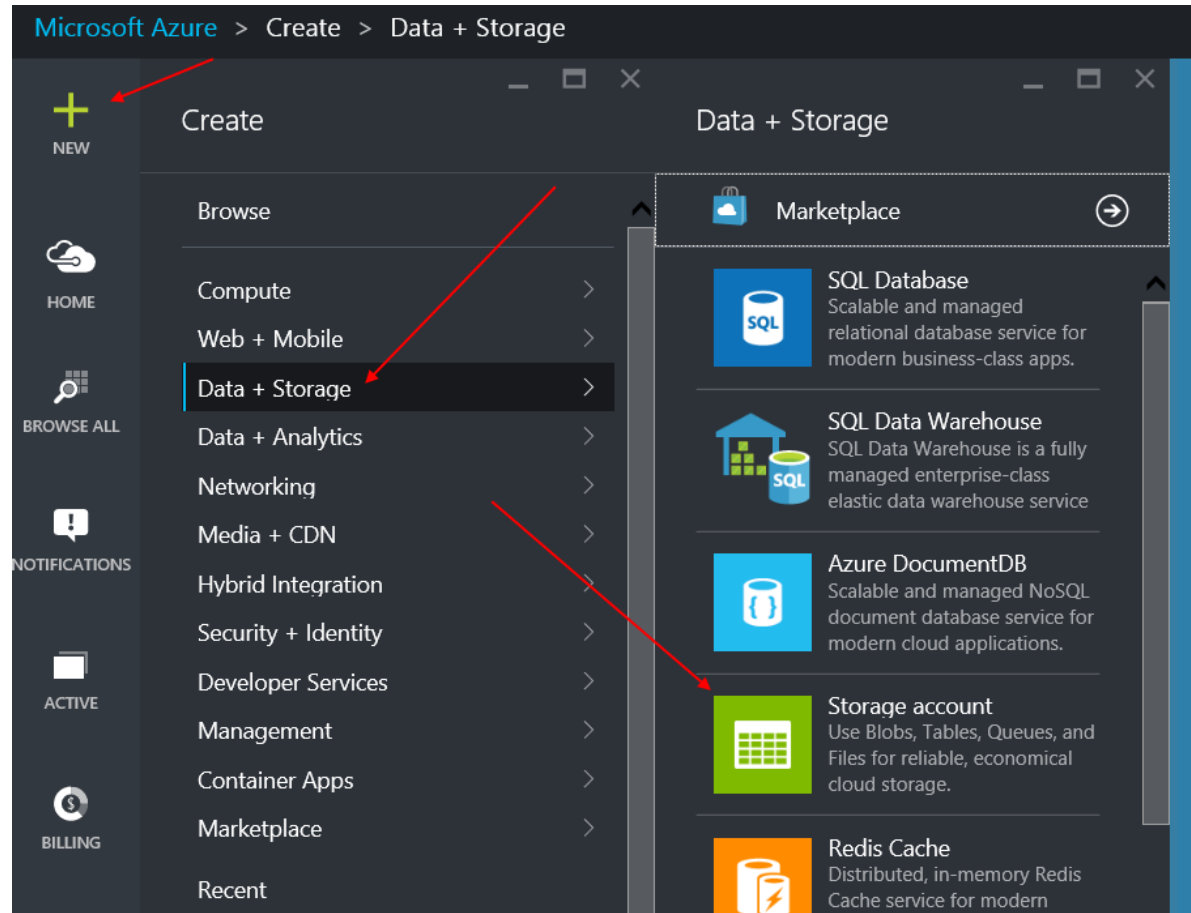
No Scripts Required

Create Azure Storage Account

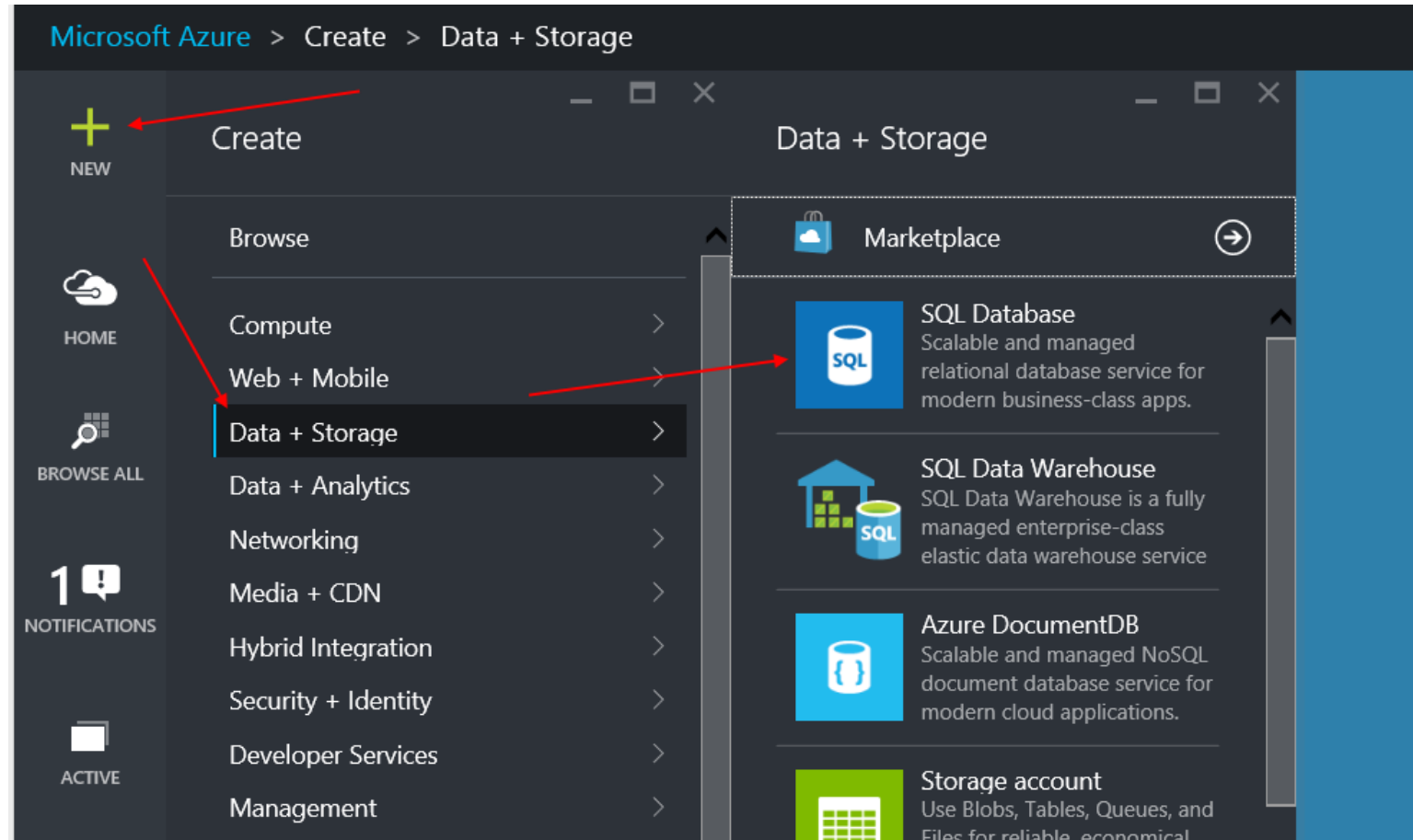
Set Up Azure SQL Database for Metabase

Create HDInsight Cluster

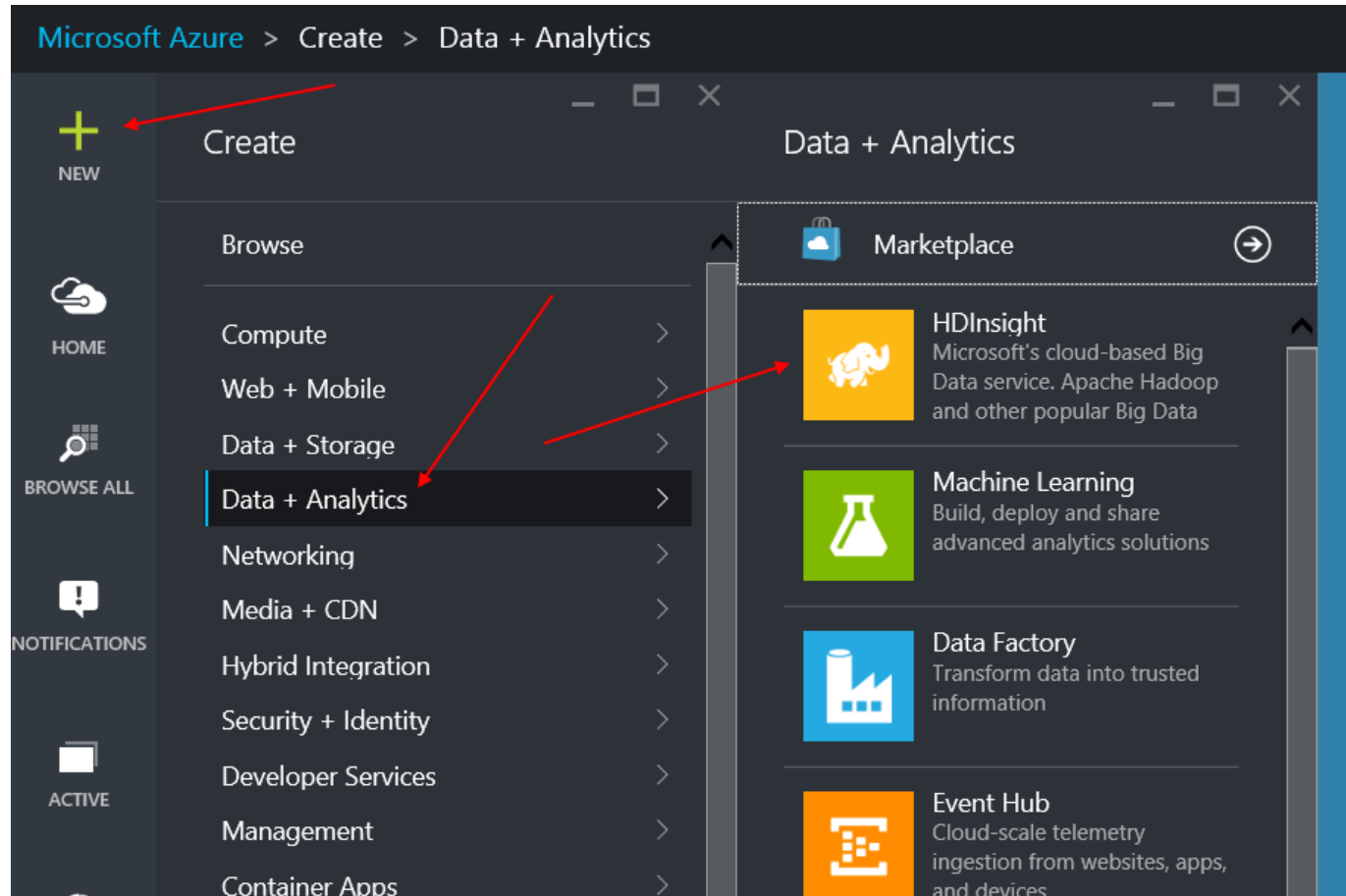
# Create Azure Storage Account



# Add Azure SQL Database Metabase



# Create HDInsight Cluster



# Setting Up the HDInsight Cluster

The screenshot displays the Microsoft Azure portal interface for creating a new HDInsight cluster. The breadcrumb navigation at the top reads: Microsoft Azure > Create > Data + Analytics > New HDInsight Cluster > Node Pricing Tiers.

The left-hand navigation pane shows the 'Data + Analytics' category selected. The main area is divided into three panes:

- Marketplace:** Lists available services including HDInsight (Microsoft's cloud-based Big Data service), Machine Learning, Data Factory, Event Hub, Azure Stream Analytics, and Cloudera Enterprise.
- New HDInsight Cluster:** Contains configuration fields:
  - Cluster Name: `hugheshdinsight` (with a checkmark)
  - Cluster Type: `Hadoop` (dropdown)
  - Cluster Operating System: `Windows Server 2012 R2 Datacenter` (dropdown)
  - Subscription: `Pay-As-You-Go`
  - Resource Group: `shughes-datafactory`
  - Credentials: `Configured`
  - Data Source: `hugheshdinsight (North Central...`
  - Pin to Startboard
- Node Pricing Tiers:** Shows pricing details:
  - Number of Worker nodes: `4` (with a checkmark)
  - Worker Nodes Pricing Tier: `D12 (4 nodes)`
  - Head Node Pricing Tier: `D12 (2 nodes)`
  - Summary table:

WORKER NODES	0.81 x 4 = 3.22
HEAD NODES	0.81 x 2 = 1.61
<b>TOTAL COST</b>	<b>4.84</b>

USD/HOUR (ESTIMATED)  
*Using 24 of 50 total cores in North Central US*
  - Text: "This estimate does not include subscription discounts or storage costs." and "Questions? [Contact billing support.](#)"

Buttons for 'Create' and 'Select' are visible at the bottom of their respective panes.



# Moving to Demos

Exploring your cluster

Uploading files

Creating Hive table

Using Power BI with Files

Using Power BI with Hive

# Wrap Up and References

<http://azure.microsoft.com/en-us/>

<http://www.cloudberrylab.com/free-microsoft-azure-explorer.aspx>

<https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL#LanguageManualDDL-Create/Drop/TruncateTable>

Hive ODBC: <http://www.microsoft.com/en-us/download/details.aspx?id=40886>

Follow Up Blog Post:

<http://dataonwheels.wordpress.com/2015/09/15/powering-up-hdinsight-with-power-bi>

[shughes@pragmaticworks.com](mailto:shughes@pragmaticworks.com)

@dataonwheels