Tips for Getting Started with the Azure Data Platform

Practical Tips and Lessons Learned

April 17, 2018

Melissa Coates

Solution Architect, BlueGranite
Microsoft Data Platform MVP

Blog: www.sqlchick.com
Twitter: @sqlchick
Goals for this session:

1. Key concepts for using Azure as a data platform
2. Share suggestions and lessons learned through a series of tips and demos

Lots of important things are out of scope though!
- Database migration techniques
- Security
- Azure Stack & private cloud deployments
- Networking & connectivity
- Troubleshooting, performance tuning & growth
- Installation and configuration details
- Monitoring & alerting

Azure services & features change very frequently, so be sure to verify details.
Tip #1

Know the What & Why for Moving to the Cloud
Azure
A public cloud computing platform and infrastructure for building, deploying, and managing MSFT-specific and third party software and services through a global network of Microsoft-managed datacenters

### Compute
- Virtual Machines
- Azure Container Service
- Functions
- Service Fabric
- Virtual Machine Scale Sets
- Azure Container Registry
- Batch
- Cloud Services

### Networking
- Virtual Network
- Load Balancer
- Application Gateway
- VPN Gateway
- Azure DNS
- ExpressRoute
- Traffic Manager
- Network Watcher

### Storage
- Storage: Blobs, Tables, Queues, Files, Disks
- Data Lake Store
- StorSimple
- Azure Backup
- Site Recovery

### Web & Mobile
- Web Apps
- Mobile Apps
- Logic Apps
- API Apps
- Content Delivery Network
- Media Services
- Search

### Databases
- SQL Database
- SQL Server Stretch Database
- Redis Cache
- DocumentDB
- Data Factory

### Intelligence & Analytics
- HDInsight
- Machine Learning
- Cognitive Services
- Azure Bot Service*
- Power BI Embedded
- Azure Analysis Services

### Monitoring & Management
- Azure Portal
- Azure Resource Manager
- Azure Advisor
- Azure Monitor
- Log Analytics
- Automation
- Scheduler

### Internet of Things & Enterprise Integration
- Azure IoT Hub
- Event Hubs
- Stream Analytics
- Notification Hubs
- BizTalk Services
- Service Bus
- Data Catalog

### Security + Identity
- Security Center
- Key Vault
- Azure Active Directory
- B2C
- Domain Services
- Multi-Factor Authentication

### Developer Services
- Visual Studio Team Services
- Azure Dev/Test Labs
- VS Application Insights
- API Management
- HockeyApp
- Developer Tools
- Service Profiler*
Why the Cloud

<table>
<thead>
<tr>
<th>Common Pros</th>
<th>Common Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Eliminate/reduce data center management</td>
<td>✓ Uptime guarantees</td>
</tr>
<tr>
<td>✓ Elasticity (scaling up/down)</td>
<td>✓ Performance</td>
</tr>
<tr>
<td>✓ Self-service provisioning of services</td>
<td>✓ Security</td>
</tr>
<tr>
<td>✓ Ease of experimentation (agility)</td>
<td>✓ Compliance, regulations, legal</td>
</tr>
<tr>
<td>✓ Faster time-to-market</td>
<td>✓ Sharing of resources (multi-tenancy; noisy neighbors)</td>
</tr>
<tr>
<td>✓ Easier high availability and disaster recovery</td>
<td>✓ Data and intellectual property privacy</td>
</tr>
<tr>
<td>✓ Subscription-based operating expenses (rather than capital expenses with</td>
<td>✓ Vendor lock-in/dependency</td>
</tr>
<tr>
<td>large up-front investment)</td>
<td>✓ Connecting legacy systems (hybrid/on-prem)</td>
</tr>
<tr>
<td></td>
<td>✓ Sprawl of self-provisioned services</td>
</tr>
<tr>
<td></td>
<td>✓ Lack of cloud expertise</td>
</tr>
<tr>
<td></td>
<td>✓ Complexity</td>
</tr>
<tr>
<td></td>
<td>✓ The cloud is not <em>always</em> easier</td>
</tr>
<tr>
<td></td>
<td>✓ The cloud is not <em>always</em> cheaper</td>
</tr>
<tr>
<td></td>
<td>✓ Difficulty to estimate cost up-front</td>
</tr>
</tbody>
</table>

Your goals for the cloud affect the tradeoffs & decisions you will make on control, complexity, cost, performance & security
The Azure Lingo

Subscription
A billing relationship

Resource Group
A logical container for resources

Resource
The individual services

Resource Group

Active Directory
Sync

Azure Active Directory
Tenant which provides directory & identity management

Subscription

Azure Resource Provider
Registration per resource

Azure Resource Manager
Framework to deploy & manage resources

Trust

Office 365
Power BI
Tip #2
Plan for Resource Groups
Very Carefully
Resource Groups - Considerations

Considerations:

- Logical organization by purpose
- Permissions/security boundaries
- Policies & management of resources as a group
- Deployment lifecycle -- RGs are the default scope for declarative ARM deployment scripts
- Location of services
  - Geographic location of data
  - Proximity to business users
  - Co-location of related resources
  - Minimizing latency
  - Minimizing data egress charges
  - Not all resources/services available in every region
  - Backup/recovery considerations
Resource Groups

Suggestions:
✓ Plan very carefully for resource groups
✓ Keep resource groups more narrow than broad
✓ Select the region (location) carefully
Tip #3
Plan Deliberately for Dev, Test, Prod Environments
Separation of Dev, Test, Prod Environments

Most commonly environments are segregated by:
1. Directory, or
2. Subscription, or
3. Resource Groups,
4. A combination of 1 and 3, or 2 and 3
Option: Separate By Directory

Pros:
✓ Clear boundary
✓ Offers the most scalability

Cons:
✓ More infrastructure to manage
✓ A lot of objects intermixed in a subscription - need clear resource group names and/or tags to tell what belongs to which team
✓ More complex AAD configuration
✓ Separate virtual networks & VPN set up across subscriptions
Option: Separate By Subscription

Pros:
✓ Clear boundary

Cons:
✓ Cannot always provision new resources if you’re just an owner at the resource group level
✓ A lot of objects intermixed in a subscription - need clear resource group names and/or tags to tell what belongs to which team
✓ Separate virtual networks & VPN set up across subscriptions
Option: Separate by Resource Group

Pros:
- Isolation of subscriptions by the team who owns/manages
- Co-admin privileges easier to delegate at the subscription level for each team
- Billing segregation
- More subscriptions: might be less likely to hit Azure subscription limits

Cons:
- More risk because Prod is mixed in with Dev & Test (mitigate w/ clear names & diff IDs)
- Separate virtual networks & VPN set up across subscriptions

Directory

Analytics Team Subscription

- Resource Group: Sales Rptg Dev
- Resource Group: Sales Rptg Test
- Resource Group: Sales Rptg Prod
- Resource Group: Performance Rptg Dev
- Resource Group: Performance Rptg Test
- Resource Group: Performance Rptg Prod

Other Team Subscriptions
Demo

Resource Groups, Resources, and Resource Providers
Tip #4
Use Clear Naming Conventions
Naming Conventions

Purpose → Type of Service → Environment

✓ **Type of service** in the name helps with logging/metrics in monitoring scenarios

✓ **Environment as the suffix** makes any concatenations easier within scripts

✓ **Prod is enumerated** because we separate Dev/Test/Prod within one subscription

✓ **No dashes** since all services don’t allow them

✓ **Camel case** if the service allows it; otherwise lower case

Tip #5
Use Tags to Categorize Billing Charges & Other Metadata
Standard Tags

Tags are very helpful for:
✓ Billing or cost center categories
✓ Environment names
✓ Project or system
✓ Purpose or application
✓ Team, department, business unit
✓ Who owns or supports a resource
✓ Release or version #s (ex: testing infrastructure)
✓ Archival date (ex: if needed only temporarily)
✓ Patching or maintenance window or SLA
✓ Which customer it applies to (ex: if an ISV)
✓ etc...


Try to assign tags right away so you don’t have partial billing without tags assigned

Can’t assign a tag if the resource is paused
Demo

Find Resources Based on Tags
Tip #6

Determine if IaaS vs PaaS vs SaaS Best Meets Your Needs
IaaS vs. PaaS vs. SaaS

**Shared Infrastructure (Lower Cost)**

- Infrastructure as a Service (IaaS)
  - SQL Server in a Virtual Machine

**Dedicated Infrastructure (Higher Cost of Ownership)**

- Platform as a Service (PaaS)
  - Azure SQLDB, Azure SQLDW, CosmosDB
- Software as a Service (SaaS)
  - Power BI, Office 365

**On-Premises**

- Physical Server
- Virtual Server

**Azure Stack (Private Cloud)**

**Easier to Scale**

- Serverless Apps
  - Azure Functions

**More Difficult to Scale**

- Power BI, Office 365
- SQL Server in a Virtual Machine

**More Control (Higher Administration Effort)**

- Physical Server
- Virtual Server

**Less Control (Lower Administration Effort)**

- Cloud Services
- Serverless Apps
- Azure Functions
Comparing the SQL Offerings in Azure (1/2)

**SQL Server in a Virtual Machine (IaaS)**
- Run full workload within an Azure virtual machine, including SQL Server, SSIS, SSAS, SSRS, etc

**Azure SQL Database (PaaS)**
- A relational database-as-a-service (DBaaS)
  - **Non-Managed**
    - A traditional Azure SQLDB deployment (isolated DB)
  - **Managed Instance**
    - Newer - closer feature parity to SQL Server (instance level features)
  - **Elastic Pools**
    - Unpredictable multi-tenant OLTP workloads (NOT for data warehousing workloads)

**Azure SQL Data Warehouse (PaaS)**
- An data warehouse-as-a-service (DWaaS) optimized for performance of large-scale analytical workloads
  - MPP architecture (massively parallel processing)
Comparing the SQL Offerings in Azure (2/2)

<table>
<thead>
<tr>
<th>SQL Server in a Virtual Machine (IaaS)</th>
<th>Azure SQL Database (PaaS)</th>
<th>Azure SQL Data Warehouse (PaaS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best for:</strong></td>
<td><strong>Best for:</strong></td>
<td><strong>Best for:</strong></td>
</tr>
<tr>
<td>✓ Migrating existing solutions</td>
<td>✓ New database solutions</td>
<td>✓ DW with larger data volumes</td>
</tr>
<tr>
<td>✓ Running all SQL Server features</td>
<td>✓ Migrating existing database solutions as-is (managed instances which are in preview)</td>
<td>(ideally 1-4TB+ bare min. of 250GB+ )</td>
</tr>
<tr>
<td>(ex: SSIS, SSRS, SSAS, MDS)</td>
<td>✓ OLTP with scaling &amp; pooling needs (unpredictable workloads)</td>
<td>✓ Dimensional models which can take advantage of CCI</td>
</tr>
<tr>
<td>✓ Running 3rd party software</td>
<td>✓ DW workloads which are small to medium</td>
<td>✓ Ability to scale compute up/down, or pause (elasticity)</td>
</tr>
<tr>
<td>(ex: Oracle)</td>
<td>✓ Reduced administration</td>
<td>✓ Data Lake Store integration (relational + nonrelational data)</td>
</tr>
<tr>
<td>✓ Bring your own license</td>
<td></td>
<td>✓ Reduced administration</td>
</tr>
<tr>
<td>(Software Assurance)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ Isolated dev/test environments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓ Administering all aspects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(highest level of control)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

https://www.blue-granite.com/blog/is-azure-sql-data-warehouse-a-good-fit
Other PaaS Database Offerings in Azure

Cosmos DB

Best for:
✓ Schema-agnostic data model that is write-optimized
  o NoSQL JSON documents
  o Key-value pairs
  o Column-family data
  o Graph data
✓ Scalable, globally distributed apps
✓ Reduced administration

Azure Database
for MySQL and PostgreSQL

Best for:
✓ Open source DB (community editions)
✓ Focus is mostly on a front-end application
✓ Familiar with MySQL or PostgreSQL
✓ Reduced administration

Azure Analysis Services

Best for:
✓ Semantic layer for reporting
✓ Alternative to SQL Server Analysis Tabular Model (otherwise run in an IaaS VM)
✓ Upsizing a data model from Power BI
✓ Sharing some workload with Azure SQLDW or Azure SQLDB
✓ Reduced administration
Many features go first to Azure SQLDB ("cloud first"). However, some features are not available in Azure SQL Database such as:

- PolyBase (alternative: elastic queries)
- SQL Server Machine Learning & R Services
- Change data capture
- CLR
- DB snapshots
- Some T-SQL syntax
- Profiler
- Non-primary filegroups

Full list:
Some features are *not available* in Azure SQL Data Warehouse:
- Primary key, foreign key, unique key
- Computed columns, default values
- Merge statements
- Some data types (ex: xml, json, geography, geometry, hierarchyid)
- Certain update and delete techniques and other T-SQL differences
- Statistics creation and management
- SSDT database project
- Cross-database queries

The recommended data load technique is via PolyBase.

More info:
https://www.blue-granite.com/blog/is-azure-sql-data-warehouse-a-good-fit
Tip #7
Get an Azure SLA if Possible
How to Get a Service Level Agreement (SLA)

SQL Server in a Virtual Machine (IaaS)

SLA on the VM if:
1. Include the VM in an Availability Set - Guarantee: 99.95%
   or
2. Use Premium Storage for all Data and O/S disks - Guarantee: 99.9%

Premium Storage is how you get an SLA on a ‘single instance’ VM. You might consider Standard storage in Dev/Test, and Premium in Prod.

Azure SQL Database (PaaS)

SLA on the database:
✓ Guarantee: 99.99% for Basic, Standard, and Premium tiers that customers will have connectivity

Azure SQL Data Warehouse (PaaS)

SLA on the database:
✓ Guarantee: 99.99% for Basic, Standard, and Premium tiers that client operations will succeed

Tip #8
Take Advantage of Efficiencies Offered by Cloud Services
Starting With a VM Image

Expect to make changes and customizations to the VM image, such as:
- Domain service accounts for each service
- Disable unused services
- Disk structure + permissions
  - G: Data
  - L: Logs
  - T: TempDB
- Enable disk encryption

SQL Server changes to VM image:
- Enable the DAC
- Default DB file locations
- Fill factor
- Optimize for ad hoc workloads
- Max degree of parallelism
- Cost threshold for parallelism
- Startup parameters; trace flags
- Expand # of TempDB files
- Time of SSIS maintenance job etc...

By default, there is one data drive on F: provisioned at 1TB. Stay away from D:\ - it's the temp drive.

As-is before any customizations
Automatic Patching

Windows and SQL Server patching

**PaaS**: Done for you (no opt-out or delay).

**IaaS**: You can optionally allow MSFT to handle. There’s also a new VM “Update Management” feature.

SQL Backups: IaaS

IaaS: You can optionally set up scheduled SQL Server backups.

Suggestions:
- Be sure that this Azure storage account is set to be encrypted since the “automaticbackup” container contains certificates & keys
- Provision the storage account with backups in another region, or with geo-redundant storage
**SQL Backups: PaaS**

**Azure SQL Database (PaaS)**

- Basic service: retained 7 days
- Standard service: retained 35 days
- Premium service: retained 35 days

- Full backups: weekly
- Differential: every few hours
- T-log: every 5-10 minutes

**Azure SQL Data Warehouse (PaaS)**

- Local snapshots: every 4-8 hours & retained 7 days
- Geo-redundant snapshot: retained 24 hours
  (Backups occur when DW is online -- not paused)

**Longer-term backup retention:**
- Not currently an option

---

*The tradeoff here: less administration, but also less control*
VM Backups

VM backups are managed in the Recovery Services Vault.

To ensure the VM backups are also encrypted, Azure disk encryption is required.
Basic Monitoring

Diagnostic Logs

Data Lake Store

Activity Log

Metrics

Turn on diagnostics to collect the following data:

- Audit
- Requests
- Alert Metrics
Tip #9
Be Aware of Cost Savings Techniques
Some services support pausing or stopping compute resources when it is idle. Some services support scaling compute resources up & down to support varying demands.

<table>
<thead>
<tr>
<th>Ability to Pause/Stop to Save Compute $:</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Virtual Machine</td>
</tr>
<tr>
<td>✓ SQL Data Warehouse</td>
</tr>
<tr>
<td>✗ SQL Database</td>
</tr>
</tbody>
</table>
Co-Location to Avoid Data Egress Charges

Data egress charges are incurred when data leaves an Azure region. No charges to ingest data into Azure. Data egress charges are very cheap – but they can easily add up at scale.
## Scale Lower in Dev/Test Environments

In a subscription model, developers and admins can have a huge influence on cost.
Automatic Shutdown of VMs

Dev & Test

Shutdown:
✓ Auto-shutdown feature within the VM properties

Start:
✓ VM is manually restarted when needed by one of the developers

Production

Shutdown & Start:
✓ A machine which needs to be restarted by a certain time (ex: DB engine+SSIS) can be done with:
  o Azure Automation (PowerShell)
  o Microsoft Flow
  o Azure Functions
  o etc...

Shutting the VM down within Windows does *NOT* stop compute charges. It needs to be de-allocated in Azure to reduce cost.
## Saving Money Running Virtual Machines

<table>
<thead>
<tr>
<th>Low Priority VMs on Azure Batch</th>
<th>Prepaid Reserved VM</th>
<th>“Burstable” B-Series VMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Runs on a pool of VMs; takes advantage of compute surplus in Azure</td>
<td>Upfront commitment for compute capacity at a large discount</td>
<td>Accumulates ‘saved’ credits for unused CPU cycles—uses the credits for bursting</td>
</tr>
<tr>
<td>Useful when:</td>
<td>Useful when:</td>
<td>Useful when:</td>
</tr>
<tr>
<td>✓ Job completion time is flexible</td>
<td>✓ Predictable, long-term workloads</td>
<td>✓ Workload is small but a dedicated VM is still desirable</td>
</tr>
<tr>
<td>✓ Batch job can be parallelized &amp; distributed across VMs</td>
<td></td>
<td>✓ Bursts only occasionally (ex: once a day)</td>
</tr>
</tbody>
</table>

https://docs.microsoft.com/en-us/azure/batch/batch-low-pri-vms

https://blogs.msdn.microsoft.com/uk_faculty_connection/2017/05/14/microsoft-azure-makes-vms-affordable-for-everybody/


# Dev/Test Labs

Offers lower rates for certain Azure services.

Useful for:
- ✓ Learning
- ✓ Experimenting
- ✓ Proof of concept
- ✓ Throwaway work
- ✓ Non-production

Specify policies and thresholds to control cost.

Requires active Visual Studio subscription.

## Pricing Options

<table>
<thead>
<tr>
<th></th>
<th>For Individuals</th>
<th>For Teams</th>
<th>For Teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dev/Test pricing options</td>
<td>Monthly Azure credits for Visual Studio subscribers</td>
<td>Enterprise Dev/Test</td>
<td>Pay-As-You-Go Dev/Test</td>
</tr>
<tr>
<td>Number of instances allowed</td>
<td>One per Visual Studio subscriber</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Monthly credit included*</td>
<td>$50, $100, or $150</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Billing options</td>
<td>No credit card required</td>
<td>Always billed to the Enterprise Agreement</td>
<td>Credit card billing (invoicing available)</td>
</tr>
<tr>
<td></td>
<td>Credit card billing for overages (invoicing available)</td>
<td>Consumes prepaid Azure Monetary Commitment funds, if available</td>
<td></td>
</tr>
<tr>
<td>Billed usage</td>
<td>Overages are billed, after the monthly credit is consumed</td>
<td>All usage is billed</td>
<td>All usage is billed</td>
</tr>
<tr>
<td>Allowed to run Windows 10 Virtual Machines</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Licensing terms</td>
<td>Restricted to dev/test use only</td>
<td>Restricted to dev/test use only</td>
<td>Restricted to dev/test use only</td>
</tr>
<tr>
<td></td>
<td>For use by active Visual Studio subscribers, and by end users providing feedback and performing acceptance tests</td>
<td>For use by active Visual Studio subscribers, and by end users providing feedback and performing acceptance tests</td>
<td>For use by active Visual Studio subscribers, and by end users providing feedback and performing acceptance tests</td>
</tr>
</tbody>
</table>

Efficient Use of Storage Tiers

Cost

Azure Hot Storage
Data which is accessed frequently

Azure Cold Storage
Data which is accessed less frequently

Azure Archive Storage
Data which is not accessed

https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blob-storage-tiers
Virtual Machine Storage

<table>
<thead>
<tr>
<th></th>
<th>Premium Disk</th>
<th>Standard Disk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managed Disk</td>
<td>Billed for disk size reservation</td>
<td>Billed for disk size reservation</td>
</tr>
<tr>
<td>Unmanaged Disk</td>
<td>Billed for disk size reservation</td>
<td>Billed for space used</td>
</tr>
</tbody>
</table>

**Premium (SSD)**
IOPS & throughput are associated with disk size

<table>
<thead>
<tr>
<th>Disk Size</th>
<th>P4</th>
<th>P6</th>
<th>P10</th>
<th>P20</th>
<th>P30</th>
<th>P40</th>
<th>P50</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price per month</td>
<td>$4.81</td>
<td>$9.29</td>
<td>$17.92</td>
<td>$66.56</td>
<td>$122.88</td>
<td>$235.52</td>
<td>$450.56</td>
</tr>
<tr>
<td>IOPS per disk</td>
<td>120</td>
<td>240</td>
<td>500</td>
<td>2,300</td>
<td>5,000</td>
<td>7,500</td>
<td>7,500</td>
</tr>
<tr>
<td>Throughput per disk</td>
<td>25 MB/second</td>
<td>50 MB/second</td>
<td>100 MB/second</td>
<td>150 MB/second</td>
<td>200 MB/second</td>
<td>250 MB/second</td>
<td>250 MB/second</td>
</tr>
</tbody>
</table>

**Standard (HDD)**
IOPS & throughput are based on VM size + tier + # of disks

<table>
<thead>
<tr>
<th>Disk Size</th>
<th>S4</th>
<th>S6</th>
<th>S10</th>
<th>S15</th>
<th>S20</th>
<th>S30</th>
<th>S40</th>
<th>S50</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price per month (promotional)</td>
<td>$0.77</td>
<td>$1.51</td>
<td>$2.95</td>
<td>$5.67</td>
<td>$10.88</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **VM Tier**
  - Max Disk size: 4095 GB
  - Max 8 KB IOPS per disk: Up to 300/Up to 500
  - Max Bandwidth per disk: Up to 60 MB/s/Up to 60 MB/s
Use Scripts to Override Default Storage Sizes

Azure SQL DW
✓ Requires Premium storage
✓ Provisions 1 TB by default if initially created in the portal (or if maxsize is unspecified in a script)

Virtual Machine
✓ Pre-built images provision 1 TB by default

```
CREATE DATABASE BankingSQLSQLDW
COLLATE SQL_Latin1_General_CP1_CI_AS
(
    MAXSIZE = 250 GB
    ,EDITION = 'datawarehouse'
    ,SERVICE_OBJECTIVE = 'DW100'
);
```
Azure Cost Management

A service offering in Azure from Cloudyn to monitor & manage cloud spend. Basic tier is free.

https://docs.microsoft.com/en-us/azure/cost-management/
Azure Advisor

A ‘best practices’ advisor to optimize Azure deployments, including cost optimizations.

https://docs.microsoft.com/en-us/azure/advisor/advisor-overview
Demo

Auto-VM Shutdown + SQLDW Pause + SQLDB Scale
Tip #10
Understand Implications of Storage Decisions for IaaS VMs
**Storage for VMs**

The VHDs (virtual hard disks) behind a VM are in Azure Storage.

- The number of disks allowed depends on the VM size.
- Don’t use the temp disk for anything persistent.
- The VM estimate doesn’t include the O/S disk or data disk(s).
- If you delete a VM in Azure, don’t forget to also go delete the related storage (if appropriate).
Storage Decisions Related to VMs

1. Managed or Unmanaged
2. Premium (SSD) or Standard (HDD)
3. Redundancy (Local, Zone, or Geo-Redundant)
4. Encryption (Storage Service Encryption and/or Disk Encryption)

<table>
<thead>
<tr>
<th></th>
<th>Premium Disk</th>
<th>Standard Disk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managed Disk</td>
<td>Billed for disk size reservation</td>
<td>Billed for disk size reservation</td>
</tr>
<tr>
<td>Unmanaged Disk</td>
<td>Billed for disk size reservation</td>
<td>Billed for space used</td>
</tr>
</tbody>
</table>

Local is recommended per the Best Practices article

Can’t currently reduce a disk size reservation – though it can be increased easily
Organizing Unmanaged Storage for VMs

All data for a single VM is in the same storage account – it is a unit of recovery in case of failure.

<table>
<thead>
<tr>
<th>NAME</th>
<th>RESOURCE GROUP</th>
<th>LOCATION</th>
<th>SKU</th>
<th>BILLINGCATEGORY (TAG)</th>
<th>ENVIRONMENTTYPE (TAG)</th>
<th>SUPPORTCONTACT (TAG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERNALREPORTINGRGDEV</td>
<td>InternalReportingRGDev</td>
<td>East US</td>
<td>Standard_LRS</td>
<td>Internal Analytics</td>
<td>Dev</td>
<td>BI &amp; Analytics Team</td>
</tr>
<tr>
<td>bsqlvm1bckstdgdev</td>
<td>InternalReportingRGDev</td>
<td>East US</td>
<td>Standard_LRS</td>
<td>Internal Analytics</td>
<td>Dev</td>
<td>BI &amp; Analytics Team</td>
</tr>
<tr>
<td>bsqlvm1datastdgdev</td>
<td>InternalReportingRGDev</td>
<td>East US</td>
<td>Standard_LRS</td>
<td>Internal Analytics</td>
<td>Dev</td>
<td>BI &amp; Analytics Team</td>
</tr>
<tr>
<td>bsqlvm1diagstdgdev</td>
<td>InternalReportingRGDev</td>
<td>East US</td>
<td>Standard_LRS</td>
<td>Internal Analytics</td>
<td>Dev</td>
<td>BI &amp; Analytics Team</td>
</tr>
</tbody>
</table>

Backups ("bck"), Data, and Diagnostic logging ("diag") are separated.

Read Access Geo-Redundancy (RAGRS) for backup files.

...
VM Disk Performance

Premium (SSD) – IOPS & Throughput are Associated with Disk Size:

<table>
<thead>
<tr>
<th></th>
<th>P4</th>
<th>P6</th>
<th>P10</th>
<th>P20</th>
<th>P30</th>
<th>P40</th>
<th>P50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk Size</td>
<td>32 GB</td>
<td>64 GB</td>
<td>128 GB</td>
<td>512 GB</td>
<td>1 TB</td>
<td>2 TB</td>
<td>4 TB</td>
</tr>
<tr>
<td>Price per month</td>
<td>$4.81</td>
<td>$9.29</td>
<td>$17.92</td>
<td>$66.56</td>
<td>$122.88</td>
<td>$235.52</td>
<td>$450.56</td>
</tr>
<tr>
<td>IOPS per disk</td>
<td>120</td>
<td>240</td>
<td>500</td>
<td>2,300</td>
<td>5,000</td>
<td>7,500</td>
<td>7,500</td>
</tr>
<tr>
<td>Throughput per disk</td>
<td>25 MB/second</td>
<td>50 MB/second</td>
<td>100 MB/second</td>
<td>150 MB/second</td>
<td>200 MB/second</td>
<td>250 MB/second</td>
<td>250 MB/second</td>
</tr>
</tbody>
</table>

Standard (HDD) – IOPS & Throughput are Based on VM Size + Tier + # of Disks:

<table>
<thead>
<tr>
<th></th>
<th>S4</th>
<th>S6</th>
<th>S10</th>
<th>S15</th>
<th>S20</th>
<th>S30</th>
<th>S40</th>
<th>S50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk Size</td>
<td>32 GB</td>
<td>64 GB</td>
<td>128 GB</td>
<td>256 GB</td>
<td>512 GB</td>
<td>1 TB</td>
<td>2 TB</td>
<td>4 TB</td>
</tr>
<tr>
<td>Price per month (promotional)</td>
<td>$0.77</td>
<td>$1.51</td>
<td>$2.95</td>
<td>$5.67</td>
<td>$10.88</td>
<td>$20.48</td>
<td>$40.96</td>
<td>$81.92</td>
</tr>
</tbody>
</table>

https://docs.microsoft.com/en-us/azure/storage/common/storage-premium-storage
https://docs.microsoft.com/en-us/azure/storage/common/storage-standard-storage
Increasing Performance of Standard Disks (1/3)

Storage Spaces (aka Virtual Disks) are multiple Azure data disks in a storage pool to collectively share throughput and IOPS limits.

This aggregation of disks is helpful due to lower Standard limits on IOPS & throughput.

(Another alternative to increase disk performance is to spread separate database files on different disks.)


\[ IOPS = \# \text{ of requests per second (OLTP workloads)} \]

\[ \text{Throughput (Bandwidth)} = IOPS \times \text{I/O size per specified interval (DW workloads)} \]
Increasing Performance of Standard Disks

Before disk encryption has been configured

Note the single disk performance before it’s collectively shared

Each blob has a lease which associates it with the VM

Disks attached to the VM

Blob storage container
This is a tradeoff: more complexity for greater performance at lower cost.

Stripe size should be 256KB for a data warehousing workload, or 64KB for OLTP workload.
Demo

Azure Storage Account
Tip #11
Know When to Leverage ARM and/or PowerShell
ARM Templates

Goal is for deployments to be modular & repeatable.

Invest time in learning how to do ARM templates so you can make a good decision when to use them.

This is known as “infrastructure as code” which can be checked into source control from a Visual Studio project.

ARM Templates

Use ARM if You Intend To...

✓ Include the configuration of Azure resources in source control ("Infrastructure as Code"), and/or
✓ Repeat the deployment process numerous times, and/or
✓ Automate deployments, and/or
✓ Employ continuous integration techniques, and/or
✓ Utilize DevOps principles and practices, and/or
✓ Repeatedly provision & de-provision testing infrastructure
✓ Do something unsupported by the portal interface

Starting from the automation scripts in the portal is a good way to learn, but you will want to customize & tidy them up.
PowerShell

*Automation is a balance of: Wanting infrastructure in source code vs. little need for deployment repeatability*

PowerShell scripts can be useful for:

- Deploying ARM templates
- Tasks unsupported in the portal
- Deploying specific items
  - Azure Data Factory
  - AAD service principal
  - etc...
- Managing resources
  - Assigning tags
  - Enabling disk encryption
  - Provisioning key vault
  - Defining alerts
  - Start/stop a virtual machine

```powershell
# Input Area
$subscriptionName = 'InsertSubscriptionName'
$resourceGroupName = 'SQLSkillsWaitsLibraryRGDev'
$deploymentName = 'SQLSkillsSQLWaitsDeployment'
$templateFilePath = 'C:\Deployments\SQLWaits\2017-06-07\ARMResources.json'
$parameterFilePath = 'C:\Deployments\SQLWaits\2017-06-07\DevValues.parameters.json'

# ARM template deployment
New-AzureRmResourceGroupDeployment -Name $deploymentName -ResourceGroupName $resourceGroupName -TemplateFile $templateFilePath -TemplateParameterFile $parameterFilePath
```

From:
# Recommendations for ARM and PowerShell

## Use ARM for Provisioning Azure Resources When...

- Recognition of **inter-dependencies** is important
- Speed of deployment is important - ARM is **faster** than PowerShell or CLI (due to **parallel deployment**)

## Use PowerShell For...

- Execution of **ARM template**
- Execution of **misc. or piecemeal deployments**
- **Management** of resources (ex: start/stop a VM) often in conjunction with **Azure Automation**
- Execution of action, resource deployment, or configuration **not supported by ARM**
Demo

Automation Script in Azure Portal
+
Visual Studio ARM Project
Key Takeaways

The **pace of change** in Azure is frequent—plan to keep up with announcements. New features can change a previous decision really fast. New services and features are constant.

Look at **using PaaS** unless you really need an IaaS solution – i.e., the simplest solution that gets the job done. And expect complications, delays, and a learning curve along the way.

Don’t shortcut the initial planning phase – do a **POC** if you can to learn. Don’t assume a feature is definitely available or works as you expect it to. For unfamiliar features or services, consider creating a **sandbox** area to **test & learn** on first.

Plan to invest some time learning **ARM templates & PowerShell**, if you haven’t already.

Pay attention to **opportunities for cost savings**. Developers & administrators have a huge impact on cost in the cloud.
Resources

Conceptual documentation: https://azure.microsoft.com/

Reference documentation: http://msdn.microsoft.com/

Documentation: https://docs.microsoft.com/en-us/azure/

Samples: http://github.com/azure-samples

Templates: https://github.com/Azure/azure-quickstart-templates

Azure Architecture Center: https://docs.microsoft.com/en-us/azure/architecture/