



Microsoft Azure 自習書シリーズ

Microsoft Azure HDInsight トレーニング

Step-2: HDInsight 環境準備

この自習書では、Microsoft が提供するパブリッククラウドサービスである Microsoft Azure を利用し、HDInsight クラスターの作成から操作までの一連の流れをハンズオン形式で学習体験します。

更新履歴

版数	発行日	更新履歴
第 1 版	2017 年 7 月 XX 日	初版発行
第 2 版	2017 年 9 月 21 日	第 2 版発行
第 3 版	2017 年 11 月 29 日	第 3 版発行
第 4 版	2017 年 12 月 26 日	第 4 版発行
第 5 版	2018 年 1 月 24 日	第 5 版発行

目次

1. はじめに	4
2. (参考) Microsoft Azure ポータルの概要	5
3. 実習環境の準備	19
4. (参考) Microsoft Azure 無料評価版のサインアップ	21
5. リソースグループの作成	24
6. HDInsight クラスターのデプロイ	27
7. Azure Storage Blob へのデータ投入(1)	32
8. Azure Storage Blob へのデータ投入(2)	36
9. (参考) Microsoft Azure に関する情報の入手元	40
20. Microsoft Azure のお問合せ	42

1. はじめに

本自習書をご利用いただきありがとうございます。この自習書では、Microsoft が提供するパブリッククラウドサービスである Microsoft Azure を利用し、HDInsight クラスターの作成から操作までの一連の流れをハンズオン形式で学習体験します。

2. (参考) Microsoft Azure ポータルの概要

Microsoft Azure を管理する主な方法として、次の 3 つがあります。

● Azure 管理ポータル

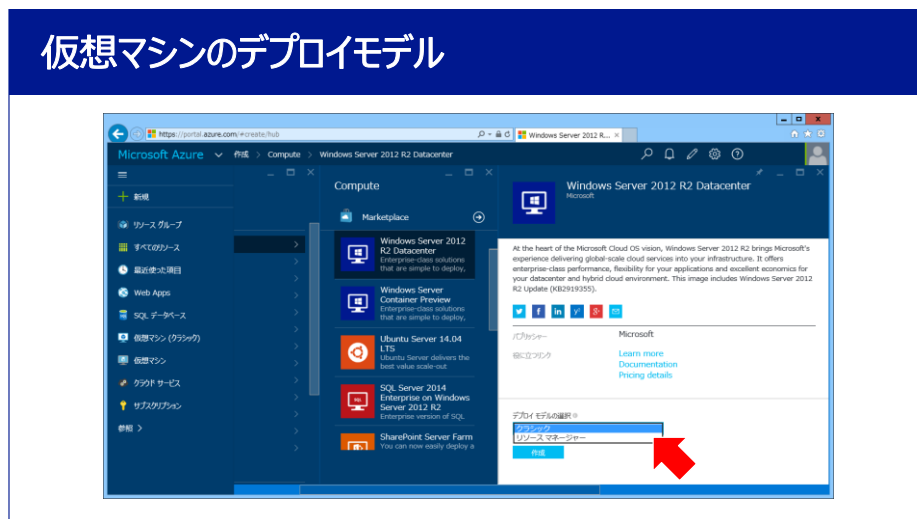
Web ベースの管理ポータルです。Azure リソースマネージャーによるリソース管理の概念が取り入れられています。また、シンプルでカスタマイズ性に優れています。<https://portal.azure.com> からアクセス可能です。管理ポータルは、HTML 5.0 で記述されているため、Internet Explorer をはじめ、Firefox、Chrome、Safari などの主要な Web ブラウザおよびデバイスで動作します。

● Windows PowerShell

スクリプトによる自動処理を実現する Windows PowerShell コマンドレット群です。一部の機能は、Windows PowerShell からのみ操作可能です。Windows PowerShell は、<https://azure.microsoft.com/ja-jp/downloads> より入手可能です。

● コマンドラインインタフェース (CLI)

Microsoft Azure が操作可能なコマンドラインプログラムです。マルチプラットフォームに展開されており、Windows だけでなく、Mac や Linux 用も提供されています。コマンドラインインタフェースは、<https://azure.microsoft.com/ja-jp/downloads> より入手可能です。



Azure 管理ポータルでは、仮想マシンなどの作成時、[デプロイモデル] として [クラシック] または [リソースマネージャー] が選択できます。デプロイモデルにより、仮想マシンの特徴が大きく異なります。

● クラシック

従来の仮想マシンと互換性のある仮想マシンです。「仮想マシン V1」または「IaaS V1」とも呼ばれます。管理は Azure 管理ポータルからおこなえます。

●リソースマネージャー

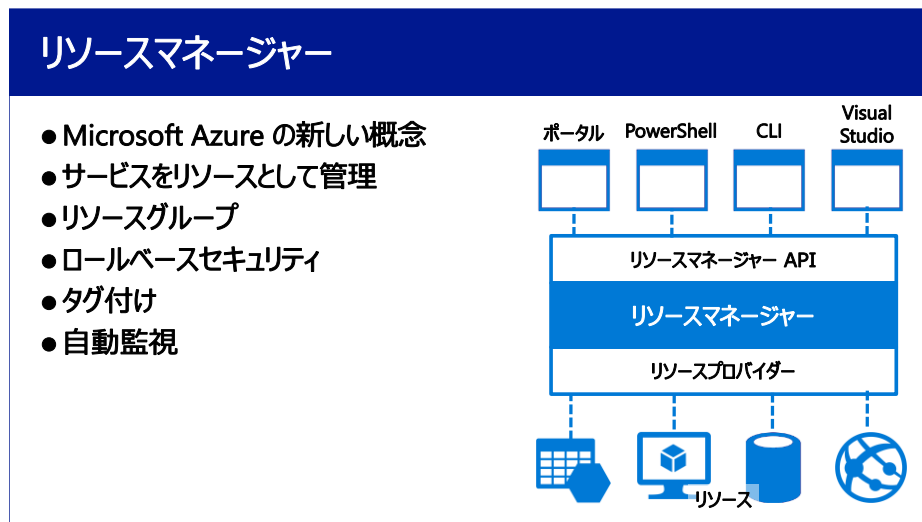
リソースマネージャーを使用した新しい仮想マシンです。「仮想マシン V2」または「IaaS V2」とも呼ばれます。従来の仮想マシンとは互換性がありません。

ワンポイント

Microsoft Azure では、マイクロソフトサポート経由で、サブスクリプション間での仮想マシンの移行をサポートしています。ただし、このサービスは仮想マシン V1 のみが対象であり、仮想マシン V2 は、移行できないことに注意してください。詳細につきましては、次の Blog をご参照ください。

サブスクリプション間のデータ移行手順について

<http://blogs.msdn.com/b/dsazurejp/archive/2013/01/16/windows-azure-subscription-migration.aspx>



リソースマネージャーは、Microsoft Azure の新しい概念であり、クラウドサービスの構成要素をリソースとして分解し、個々に管理します。例えば、仮想マシンを作成すると、仮想マシンのリソースだけでなく、ネットワークインターフェイスやパブリック IP アドレスなどのリソースも一緒に作成されます。また、Azure リソースマネージャーでは、リソースに対して、次の管理機能も提供します。

●リソースグループ

複数のリソースをまとめてグループ化し、一元的に展開、管理、監視をおこないます。

●ロールベースセキュリティ

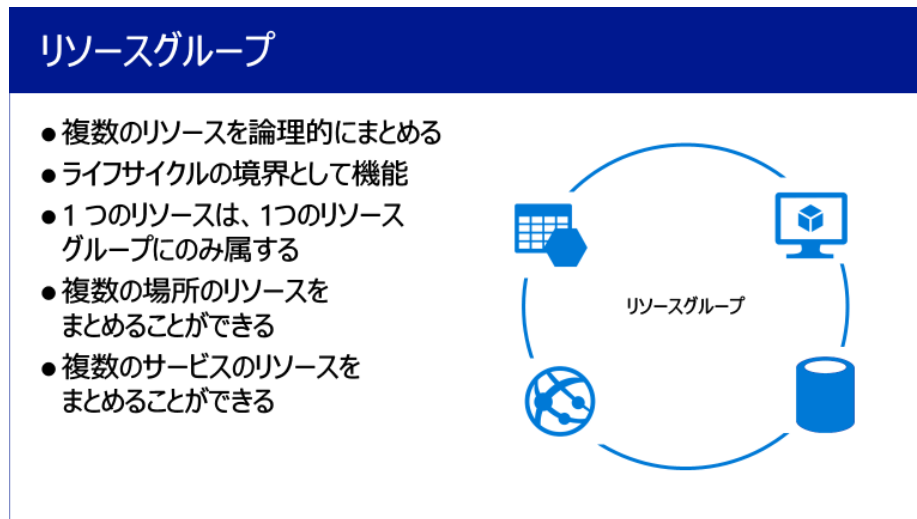
ユーザーやグループに対して、リソースまたはリソースグループごとに役割ベースのアクセス制御を提供します。

●タグ付け

タグは、キーと値のペアです。リソースにタグを付けることで、リソースを整理できます。最大 15 個のタグを付けることができます。特にタグ付けは、課金や管理の目的で使用されます。

●自動監視

パフォーマンスメトリックや各種イベントログを監視し、ストレージアカウントに記録します。また、仮想マシンのブート時の動作の監視（Windows のブート画面の定期的なキャプチャや Linux のブートログの記録）も可能です。



リソースグループは、リソースマネージャーが提供する機能であり、複数のリソースを論理的にまとめるグループです。リソースグループは、展開、更新、削除に至るライフサイクル管理の境界として機能します。リソースグループには次の決まり事があります。

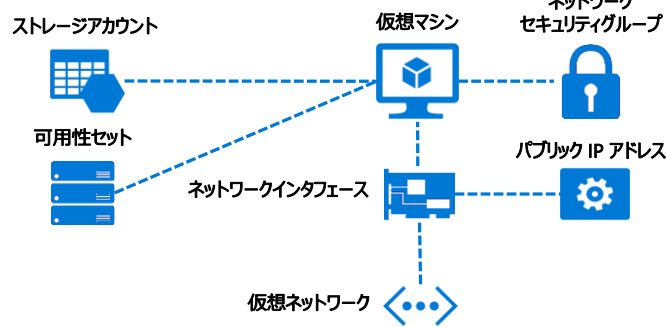
・1つのリソースは、1つのリソースグループにのみ追加できます。1つのリソースを複数のリソースグループに追加することはできません。

・データセンターの異なる複数の場所のリソースをまとめて、1つのリソースグループに追加することができます。

・Web や仮想マシンなど異なるクラウドサービスのリソースをまとめて、1つのリソースグループに追加することができます。

仮想マシンのリソース

- リソースマネージャーを使用した仮想マシンは、複数のリソースから構成される



仮想マシンの作成時、デプロイモデルとしてリソースマネージャーを選択した場合、その仮想マシンは仮想マシン V2 となり、複数のリソースから構成されます。既定では、仮想マシンの作成時、次のリソースが作成されます。各リソースの詳細は、この後、紹介します。

- 仮想マシン

仮想マシン本体のリソースです。サイズやディスク構成などのパラメーターを管理します。

- 可用性セット

可用性セットのリソースです。可用性セットに含まれる仮想マシンを管理します。

- ストレージアカウント

仮想ハードディスク（VHD ファイル）を格納するストレージアカウントのリソースです。アプリケーションからストレージにアクセスする際に必要なアクセスキーを管理します。

- ネットワークセキュリティグループ

仮想マシンのファイアウォールであるネットワークセキュリティグループのリソースです。送信と受信のセキュリティ規則を含みます。

- ネットワークインターフェイス

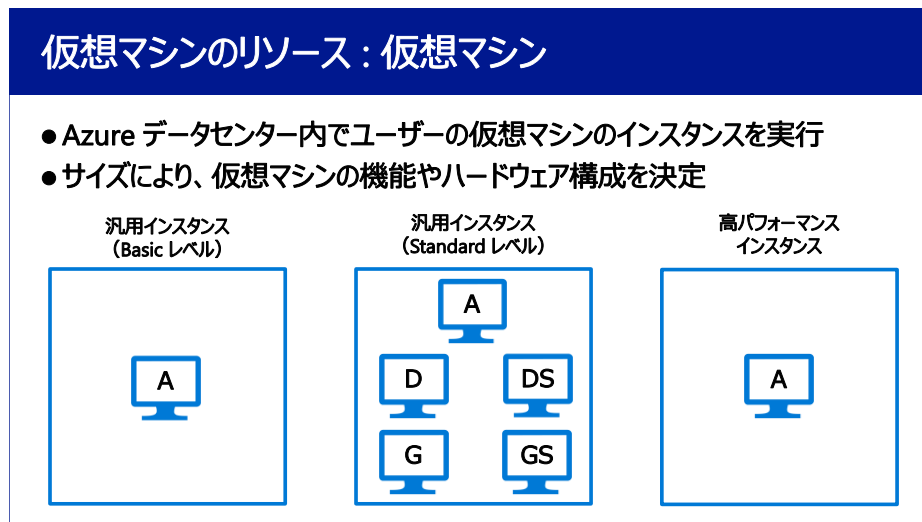
仮想マシンのネットワークアダプタ（NIC）のリソースです。IP アドレスや DNS サーバー、MAC アドレスなどを管理します。

- パブリック IP アドレス

ネットワークインターフェイスに割り当てられるパブリック IP アドレスのリソースです。DNS 名ラベルも管理します。

- 仮想ネットワーク

仮想ネットワークのリソースです。アドレス空間や DNS サーバーを管理し、サブネットを含みます。



仮想マシンは仮想マシン本体のリソースです。サイズやディスク構成などのパラメーターを管理します。

仮想マシンのサイズは、仮想マシンの機能やハードウェア構成を決定します。仮想マシンのサイズは、仮想マシンの作成時に指定しますが、作成後に変更することもできます。選択可能な仮想マシンのサイズは、大別すると、次の 3 種類です。

● 汎用インスタンス (Basic レベル)

負荷分散、自動スケール、大容量のメモリを必要としない安価な汎用インスタンスです。次の A (Basic) シリーズが選択可能です。A シリーズは、ハードディスクのみ使用できます。このインスタンスは、開発ワークロードやテストサーバーなどで利用します。

A (Basic) シリーズ

サイズ	プロセッサ数	メモリサイズ	ディスク数
A0	1	0.75 GB	1
A1	1	1.75 GB	2
A2	2	3.5 GB	4
A3	4	7 GB	8
A4	8	14 GB	16

● 汎用インスタンス (Standard レベル)

仮想マシンのすべての構成と機能がサポートされる柔軟性の高い汎用インスタンスです。次の A、D、DS、G、GS の各シリーズが選択可能です。このインスタンスは、一般的なワークロードで使用します。

基本となるのが、A (Standard) シリーズです。A (Basic) シリーズと比べると、負荷分散、自動スケール、大容量メモリが利用可能です。A シリーズは、ハードディスクのみ使用できます。

A (Standard) シリーズ

サイズ	プロセッサ数	メモリサイズ	ディスク数
A0	1	0.75 GB	1
A1	1	1.75 GB	2
A2	2	3.5 GB	4
A3	4	7 GB	8
A4	8	14 GB	16
A5	2	14 GB	4
A6	4	28 GB	8
A7	8	56 GB	16

D シリーズは、A シリーズよりもプロセッサが 60%ほど高速化されており、OS ディスクとデータディスクはハードディスクですが、一時ディスクだけは SSD が使用されています。また、バリエーションとして、D_V2 シリーズと DS シリーズがあります。Dv2 シリーズは、オリジナルより高速なプロセッサである E5-2673 v3 が採用されています。DS シリーズは、オリジナルと同じ性能ですが、OS ディスクとデータディスクにハードディスクまたは SSD が選択できます。

D (Standard) シリーズ、Dv2 シリーズ、DS (Standard) シリーズ

サイズ	プロセッサ数	メモリサイズ	ディスク数
D1、D1_V2、DS1	1	3.5 GB	2
D2、D2_V2、DS2	2	7 GB	4
D3、D3_V2、DS3	4	14 GB	8
D4、D4_V2、DS4	8	28 GB	16
D5_V2	16	56 GB	32
D11、D11_V2、DS11	2	14 GB	4
D12、D12_V2、DS12	4	28 GB	8
D13、D13_V2、DS13	8	56 GB	16
D14、D14_V2、DS14	16	112 GB	32

G シリーズは、もっともパフォーマンスの高いインスタンスです。プロセッサとして、D シリーズよりもさらに高速な Intel Xeon E5 v3 ファミリーが採用されています。また、20Gbit/s のネットワークアダプターが搭載されます。D シリーズと同様に、G シリーズは、OS ディスクとデータディスクがハードディスクで、一時ディスクのみが SSD です。GS シリーズは、OS ディスクとデータディスクにハードディスクまたは SSD が選択できます。

G (Standard) シリーズ、GS (Standard) シリーズ

サイズ	プロセッサ数	メモリサイズ	ディスク数
G1、GS1	2	28	4
G2、GS2	4	56	8
G3、GS3	8	112	16
G4、GS4	16	224	32
G5、GS5	32	448	64

●高パフォーマンスインスタンス

より高速なネットワーク、より高速なプロセッサを搭載した高パフォーマンスのインスタンスです。次の A8、A9、A10、A11、H8、H16、H8m、H16m、H16mr、H16r が選択可能です。特にサイズ A8 と A9 には、40Gbit/s の InfiniBand ネットワークアダプターが搭載されています。サイズ A10 と A11 には、最新の Intel Xeon E5 プロセッサが搭載されています。このインスタンスは、高パフォーマンスクラスター、モデリングとシミュレーション、ビデオエンコードなどで利用します。

A (Standard) シリーズ

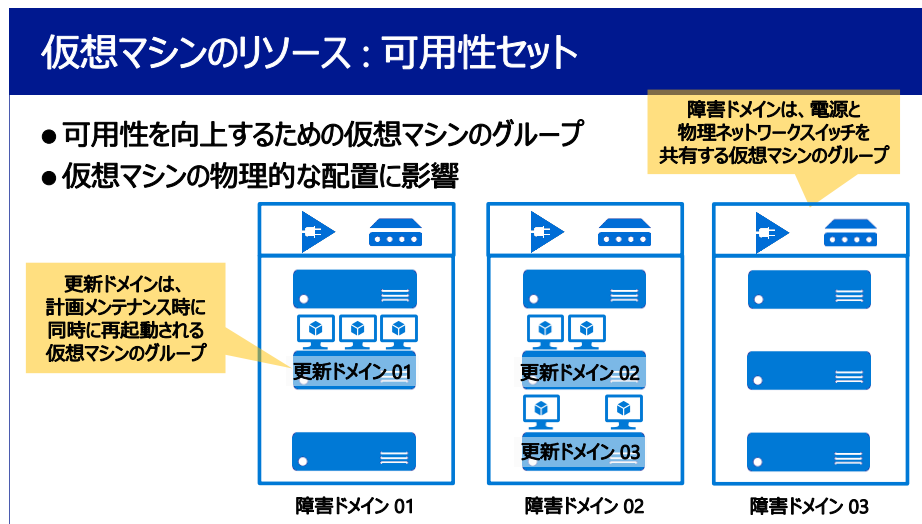
サイズ	プロセッサ数	メモリサイズ	ディスク数
A8	8	56 GB	16
A9	16	112 GB	16
A10	8	56 GB	16
A11	16	112 GB	16

H (Standard) シリーズ

サイズ	プロセッサ数	メモリサイズ	一次ストレージ
H8	8	56 GB	1000GB
H16	16	112 GB	2000GB
H8m	8	112 GB	1000GB
H16m	16	224 GB	2000GB
H16mr	16	224 GB	2000GB
H16r	16	112 GB	2000GB

ワンポイント

使用する場所（Azure データセンター）により、利用可能なサイズは異なります。



可用性セットは、Web ファームなどの同じ役割をもつ複数の仮想マシンを展開する際、可用性を向上させるためのリソースです。可用性セットは、Azure データセンター内での仮想マシンの配置場所を決定します。次の 2 つのパラメーターから構成されます。

●更新ドメイン

Azure データセンターの計画済みメンテナンス時にまとめて再起動される仮想マシンのグループです。複数の更新ドメインを同時に再起動することはありませんが、同じ更新ドメインの仮想マシンは、計画済みメンテナンス時にすべて停止する恐れがあります。仮想マシン V2 の更新ドメインは、1 から 20（既定は 5）まで設定可能です。

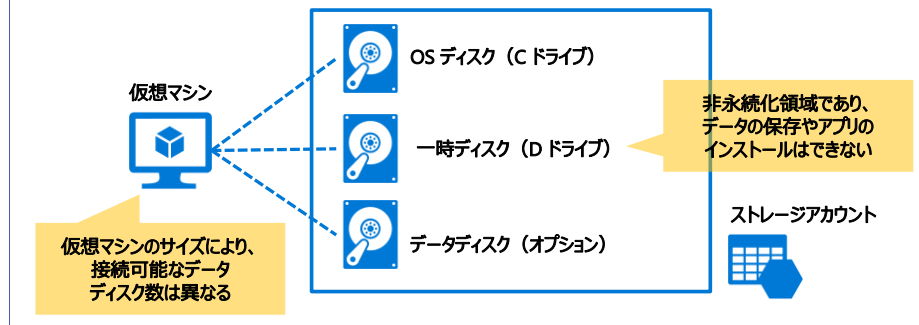
●障害ドメイン

電源とネットワークスイッチを共有する仮想マシンのグループです。同じ障害ドメインの仮想マシンは、ハードウェアの故障時にすべて停止する恐れがあります。仮想マシン V2 の障害ドメインは、2 または 3（既定は 3）が指定可能です。

可用性セットで複数の更新ドメイン、障害ドメインを指定し、複数の仮想マシンをその可用性セットに含めることで、自動的に仮想マシンを分散配置してくれます。これにより、ハードウェアの障害時または計画メンテナンス時でも、サービスを停止することなく運用することが可能となり、Microsoft Azure のサービスレベル契約（SLA） 99.95%が保障されます。

仮想マシンのリソース：ストレージアカウント

- 仮想マシンのディスクを格納
- ディスクはハードディスクまたは SSD が選択可能



ストレージアカウントは、仮想マシンの仮想ハードディスク（VHD ファイル）を格納するストレージのリソースです。アプリケーションからストレージにアクセスする際に必要なアクセスキーも管理します。

仮想マシンには、既定で 2 つの仮想ハードディスクが接続されています。1 つは OS ディスクであり、OS がインストールされています。もう 1 つは、一時ディスクで、OS のページファイルの保存やアプリケーションのキャッシュ領域として利用することを想定しています。

一時ディスクは、特別なディスクです。仮想マシンが実行中であれば、一時ディスク内のデータは、保持されますが、停止や再起動をおこなうと消去される恐れがあるため、アプリケーション本体やアプリケーションデータを格納することはできません。それらは別途、データディスクを追加し、格納する必要があります。

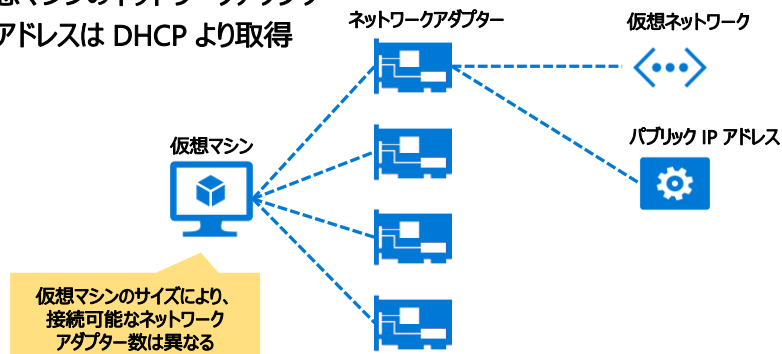
ワンポイント

アプリケーションやアプリケーションデータを OS ディスクに格納することもできますが、OS ディスクは既定で 127GB とサイズが小さく、大量のデータを格納すると空き容量の不足により、OS の動作が不安定になる恐れもあるため、データディスクを追加し、使用することを推奨します。

OS ディスクとデータディスクの種類として、Standard または Premium が選択できます。Standard は、最大 4TB までの任意のサイズのハードディスクです。Premium は、128GB、512GB、1023GB、2048GB、4096GB の各固定サイズの SSD です。なお、Premium は、後述の一部の仮想マシンのサイズでのみ利用可能です。

仮想マシンのリソース：ネットワークアダプター

- 仮想マシンのネットワークアダプター
- IP アドレスは DHCP より取得



ネットワークアダプターは、仮想マシンのネットワークインタフェースカード（NIC）のリソースです。IP アドレスや DNS サーバー、MAC アドレスなどを管理します。

ネットワークアダプターは、仮想ネットワークに接続されます。仮想ネットワークには、標準で DHCP サーバーが用意されています。仮想マシンは DHCP サーバーから動的なプライベート IP アドレスを取得し、同じ仮想ネットワークに接続された他の仮想マシンとの通信を実現します。また、仮想ネットワークを介して、仮想マシンからインターネットへのアクセスも可能です。

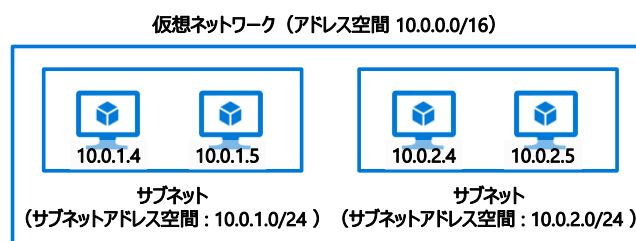
ネットワークアダプターには、プライベート IP アドレスとは別に動的または静的なパブリック IP アドレスを割り当てることもできます。パブリック IP アドレスを割り当てることで、インターネットからの仮想マシンへのアクセスを実現できます。

ワンポイント

仮想マシンには既定で、1 枚のネットワークアダプターが接続されていますが、オプションで複数のネットワークアダプターを接続することもできます。

仮想マシンのリソース：仮想ネットワーク

- 仮想マシンを接続するための論理的なネットワーク
- 仮想ネットワーク内の仮想マシンは自由に通信可能
- 指定した範囲内のプライベート IP アドレスと DNS サーバーアドレスを配布



仮想ネットワークは、仮想マシンを接続するための論理的なネットワークです。1つの仮想ネットワークに接続された仮想マシンは、お互い自由に通信することができます。また、仮想ネットワークには、Azure DHCP サーバーが標準で用意され、仮想マシンに指定した範囲内のプライベート IP アドレスと DNS サーバーのアドレスを配布します。

ワンポイント

DNS サーバーのアドレスには、Azure が提供する DNS サーバーを指定することも、ユーザーが独自に構築した DNS サーバーを指定することもできます。

仮想ネットワーク内には、既定でサブネットが 1 つありますが、オプションでサブネットを追加することもできます。サブネットを追加することで、サブネットごとにアドレス空間を指定することが可能です。

仮想マシンのリソース：ネットワークセキュリティグループ

●仮想マシンで送受信されるトラフィックを制御するファイアウォール

Windows 仮想マシンの既定の受信セキュリティ規則

優先度	名前	ソース	ターゲット	サービス	アクション
1000	default-allow-rdp	任意	任意	TCP/3389	Allow

Linux 仮想マシンの既定の受信セキュリティ規則

優先度	名前	ソース	ターゲット	サービス	アクション
1000	default-allow-ssh	任意	任意	TCP/22	Allow

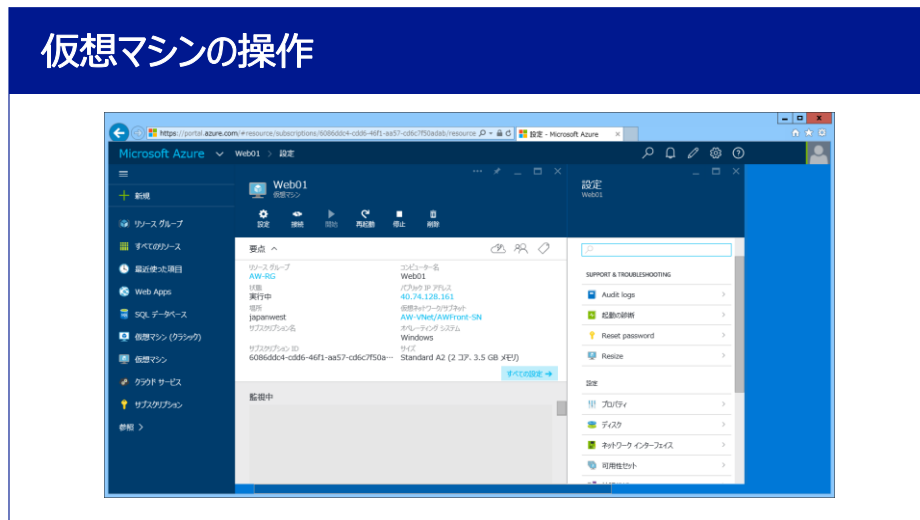
ネットワークセキュリティグループは、仮想マシンのネットワークセキュリティを管理するリソースです。仮想マシンが送受信できるトラフィックを制御するファイアウォール機能を提供します。

ネットワークセキュリティグループは、仮想マシンまたは仮想ネットワークのサブネットに関連付けできます。仮想マシン V2 を作成すると、既定でネットワークセキュリティグループが作成され、仮想マシンに関連付けされています。

ネットワークセキュリティグループは、受信セキュリティ規則と送信セキュリティ規則から構成されます。Windows 仮想マシンの場合、受信セキュリティ規則として、リモートデスクトップ接続を許可する規則が作成済みです。Linux 仮想マシンの場合は、受信セキュリティ規則として、SSH 接続を許可する規則が作成済みです。

例えば、仮想マシンに IIS をインストールし、インターネットから Web アクセスを許可したい場合は、TCP/80 を許可するカスタムの受信セキュリティ規則が必要です。なお、ネットワークセキュリティグループは、ステートフルパケットインスペクション方式（受信パケットに対応した送信パケット

は動的に許可する方式) のため、別途、送信セキュリティ規則を作成する必要はありません。



アプリケーションのインストールや更新プログラムの適用など、仮想マシンを管理する際、仮想マシンの停止や再起動などの操作が必要となります。これらの基本操作は、Azure 管理ポータルからおこなえます。Azure 管理ポータルでは、仮想マシンの次の操作が可能です。

●設定

ディスクの追加やサイズの変更などの仮想マシンの設定を変更できます。一部の設定の変更には、仮想マシンの再起動が必要となります。

●接続

仮想マシンの OS が、Windows の場合のみ、リモートデスクトッププロトコル (RDP) ファイルをダウンロードできます。このファイルを実行することで、仮想マシンのパブリック IP アドレスを調べなくても、インターネット経由でリモートデスクトップ接続が可能です。

●開始

仮想マシンが停止している場合、仮想マシンを開始できます。

●停止

仮想マシンをシャットダウンし、停止します。この方法で、仮想マシンを停止した場合、リソースは解放されるため、仮想マシンは課金されません。なお、リモートデスクトップ接続や SSH 接続で、仮想マシンにアクセスし、そこでシャットダウンをした場合は、リソースは解放されないため、仮想マシンの課金は継続されますので注意してください。

●再起動

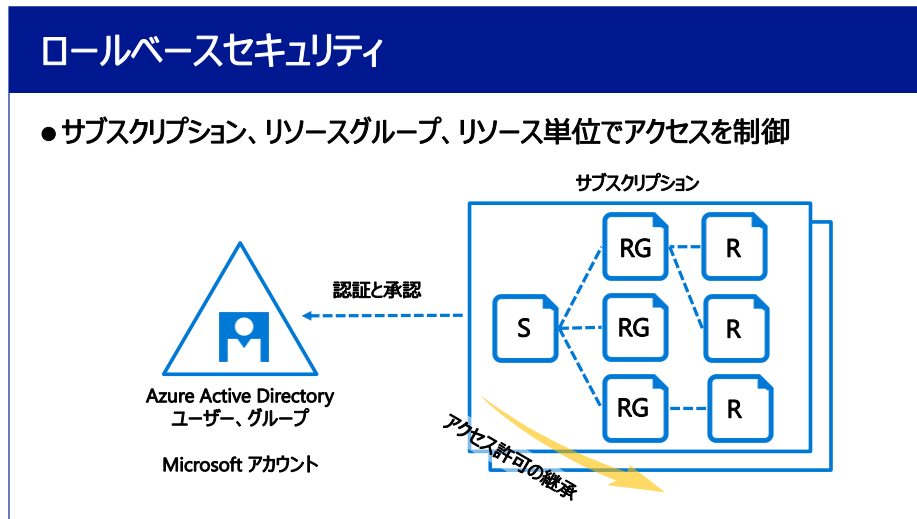
仮想マシンを再起動します。

●削除

仮想マシンを削除します。実行中の仮想マシンを削除することもできます。

ワンポイント

仮想マシンの削除で、削除されるのは、仮想マシンのリソースだけであり、ネットワークインターフェイスやパブリック IP アドレスなどの他のリソースは削除されません。また、ストレージアカウント内の仮想マシンの VHD ファイルも削除されません。必要に応じて、別途、これらのリソースを削除する必要があります。



リソースマネージャーでは、Azure Active Directory のユーザーやグループ、Microsoft アカウントに対して、役割ごとのアクセス許可を割り当てることができます。これを「ロールベースセキュリティ (Role-based access control、RBAC)」と言います。

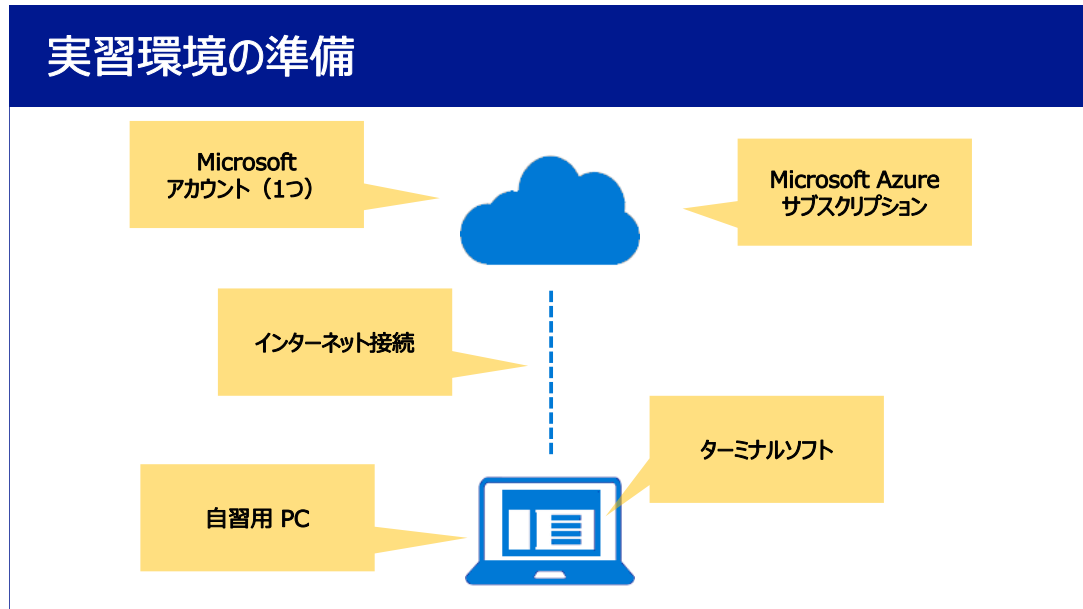
ロールベースセキュリティで割り当て可能な役割は、リソースの種類により異なりますが、例えば、仮想マシンの場合、次の役割があります。

役割	説明
所有者	リソースへのアクセスを含め、すべてを管理できます。
共同作成者	すべてを管理できますが、リソースへはアクセスできません。
閲覧者	すべてを表示できますが、変更はできません。
DevTest ラボユーザー	すべてを表示できます。仮想マシンの接続、起動、再起動、シャットダウンができます。
ネットワーク作成協力者	ネットワークを管理できますが、それらへのアクセスを管理することはできません。
ユーザーアクセス管理者	Azure のリソースへのユーザーのアクセスを管理できます。
仮想マシン作成協力者	仮想マシンを管理できますが、それらへのアクセスを管理することはできません。また、接続先の仮想ネットワークやストレージアカウントも管理できません。

役割ごとのアクセス許可は、リソースだけでなく、リソースグループやサブスクリプションにも割り当て可能です。リソースグループやサブスクリプションにアクセス許可を割り当てた場合、サブスクリプション、リソースグループ、リソースの順にアクセス許可が継承されます。

3. 実習環境の準備

この自習書の手順は、次の実習環境を準備することで、実際に試すことができ、理解を深めることができます。



●自習用 PC

Windows 10 などの自由に操作可能な Windows コンピューター（物理コンピューター）が 1 台必要です。また、Web ブラウザとして Internet Explorer/Google Chrome/Firefox も使用します。

●ターミナルソフト

Windows コンピューターには、Tera Term や Putty などの SSH 接続可能なターミナルソフトをインストールしておきます。なお、本自習書では、オープンソースで開発されている Tera Term を使用しています。Tera Term は、<https://osdn.jp/projects/ttssh2/> よりダウンロード可能です。

●インターネット接続

Windows コンピューターは、インターネットに接続されている必要があります。企業内で実習をおこなう場合は特に注意が必要です。多く企業のインターネット接続では、ファイアウォールが介在します。その場合は実習で必要なリモートデスクトップ接続や SSH 接続のプロトコルがブロックされていないことを確認しておく必要があります。

●Microsoft Azure サブスクリプション

Microsoft Azure サブスクリプションは、Microsoft Azure を使用するための権利です。すでに Microsoft Azure サブスクリプションをお持ちの場合は、そのサブスクリプションを利用することができます。まだ、Microsoft Azure サブスクリプションをお持ちでない場合は、実習用に <https://azure.microsoft.com/ja-jp/pricing/free-trial/> より、1 か月間の無料評価版をサインアップし、使

用することができます。

ワンポイント

Microsoft Azure の 1 か月間の無料評価版のサインアップには、本人確認のため、電話番号（固定電話または携帯電話）およびクレジットカードの情報が必要です。なお、無料評価版の利用は 1 回までとなっており、過去すでに利用された方は無料評価版にサインアップいただけません。有償のサブスクリプションに切り替えていただきますようお願いいたします。

●Microsoft アカウント

Microsoft アカウントは、マイクロソフトが提供するクラウドサービスを利用するための ID です。この自習書では、Microsoft Azure サブスクリプションの利用権を Microsoft アカウントに割り当てます。Microsoft アカウントは <http://www.microsoft.com/ja-jp/msaccount/default.aspx> より、無償で登録できます。この自習書をすべての手順を試すには、1 つの Microsoft アカウントが必要です。

ワンポイント

新しく、Microsoft Azure の 1 か月間の無料評価版のサブスクリプションをサインアップする場合、同時に Microsoft アカウントを登録することもできます。

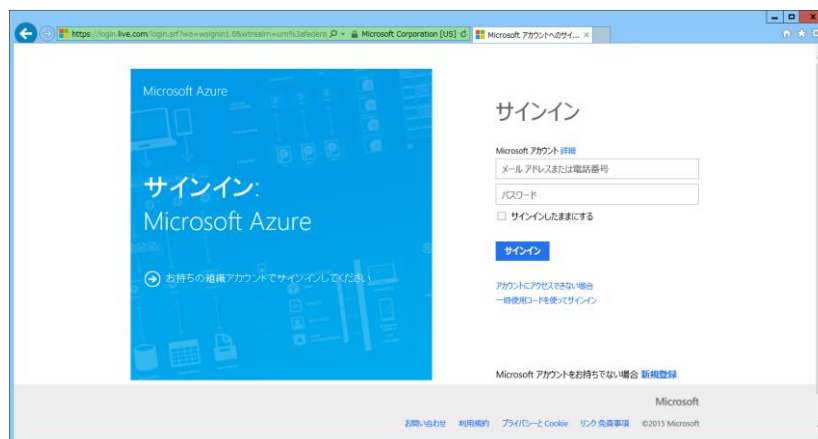
4. (参考) Microsoft Azure 無料評価版のサインアップ

この手順はオプションです。Microsoft Azure サブスクリプションをお持ちでない場合、次の手順を実行し、Microsoft Azure の 1 ヶ月間の無料評価版のサブスクリプションを取得してください。

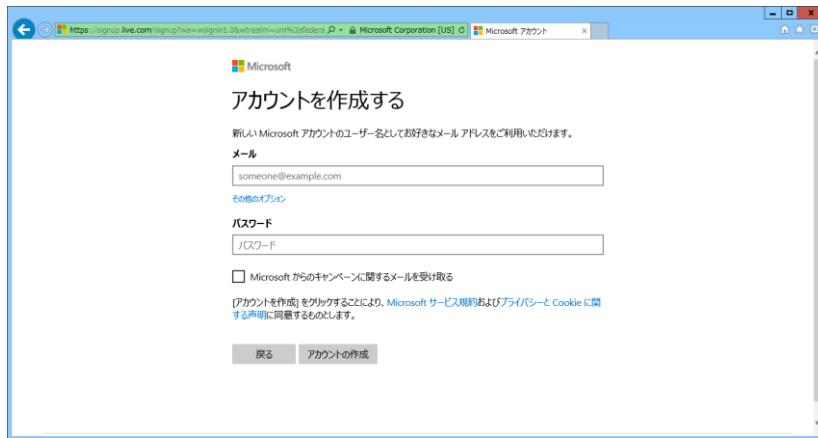
1. Web ブラウザを起動し、<https://azure.microsoft.com/ja-jp/pricing/free-trial/> にアクセスします。
2. [1 か月間の無料評価版] が表示されます。[今すぐ試す] をクリックします。



3. [サインイン] が表示されます。Microsoft アカウントのメールアドレスとパスワードを入力し、[サインイン] をクリックします。まだ、Microsoft アカウントをお持ちでない場合は、[新規登録] をクリックします。



4. [新規登録] をクリックした場合、[アカウントを作成する] が表示されます。自分のメールアドレス（自分が日常的に使用しているメールアドレス）と任意のパスワードを指定し、[アカウントの作成] をクリックします。なお、利用できるメールアドレスを持ちでない場合、[その他のオプション] をクリックし、新しい Outlook メール（～@outlook.jp）を作成することもできます。



ワンポイント

ここで登録した Microsoft アカウントが、Microsoft Azure サブスクリプションの管理者（サブスクリプション所有者）になります。

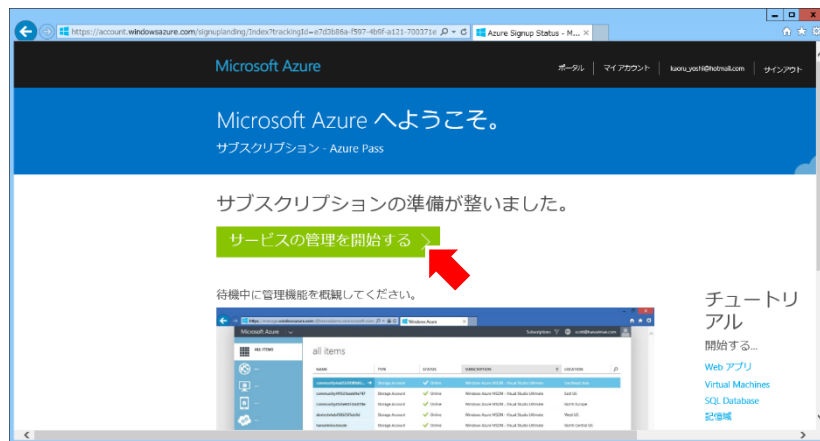
5. 「サインアップ」が表示されます。サインアップに必要な情報を指定します。なお、サインアップには、ショートメッセージ（SMS）またはコールバックを利用した電話による確認とクレジットカードの登録による確認が必要です。確認が完了したら、「サインアップ」をクリックします。



ワンポイント

登録したクレジットカードが勝手に使用されることはありません。

6. 「Microsoft Azure へようこそ」の「準備が完了するまでしばらくお待ちください」が表示されます。しばらく待つと、「サブスクリプションの準備が整いました」が表示されます。「サービスの管理を開始する」をクリックします。



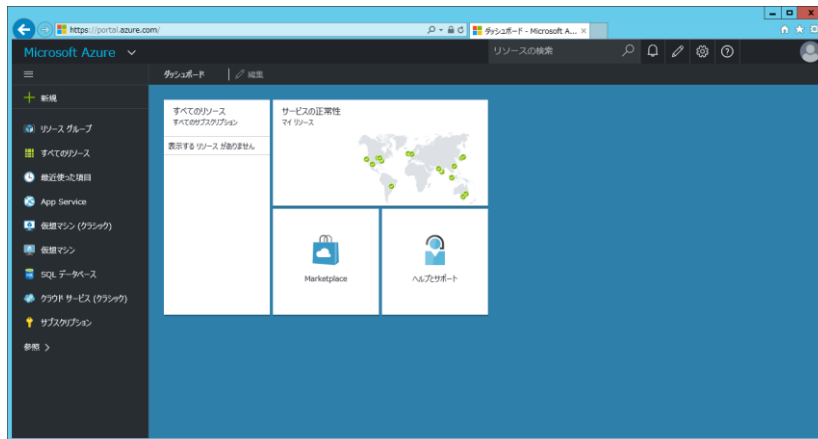
ワンポイント

Microsoft Azure の 1 か月間の無料評価版では、20,500 円相当のクレジットが利用可能です。なお、無料評価版の使用制限に達した場合には、利用している仮想マシンとクラウドサービスの開放がおこなわれます。もし制限にかかった場合には、翌請求月になるまでは利用が制限されます。ストレージサービスについては、読み取り専用としてアクセスが可能です。

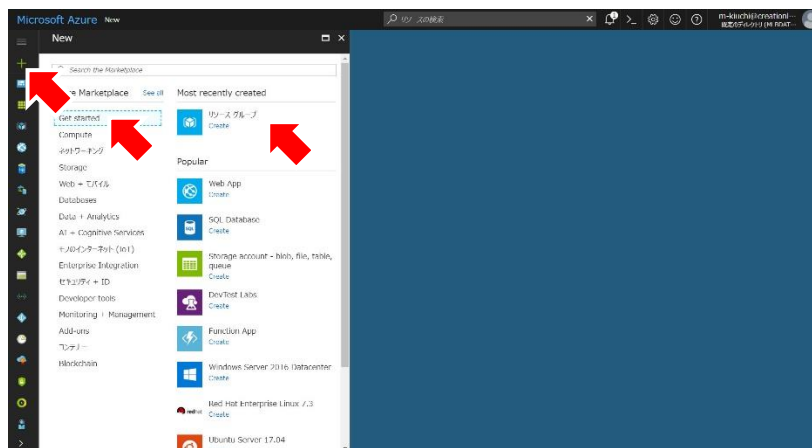
5. リソースグループの作成

はじめに仮想マシンなどのリソースを格納するためのリソースグループを作成します。次の手順では、リソースグループ「HDInsight-Training」を作成します。

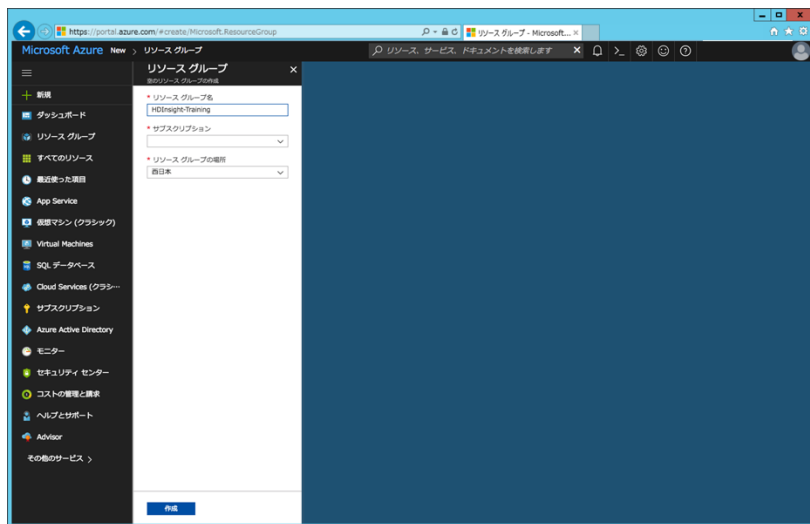
1. 実習用 PC で Internet Explorer / Google Chrome / Firefox などの Web ブラウザを起動して <https://portal.azure.com/> にアクセスし、Microsoft アカウントを指定して、サインインします。
2. Azure 管理ポータルの「スタート画面」が表示されます。



3. 画面左のジャンプバーより「新規」→「Get Started」→「リソースグループ」をクリックします。



4. 「リソースグループ」が表示されます。次のパラメーターを指定し、「作成」をクリックします。



パラメーター	説明	今回の設定
リソースグループ名	リソースグループに付ける任意の名前です。	HDInsight-Training
サブスクリプション	使用するサブスクリプションです。	<既定値>
リソースグループの場所	リソースグループを作成する場所（データセンターの場所）を選択します。	西日本

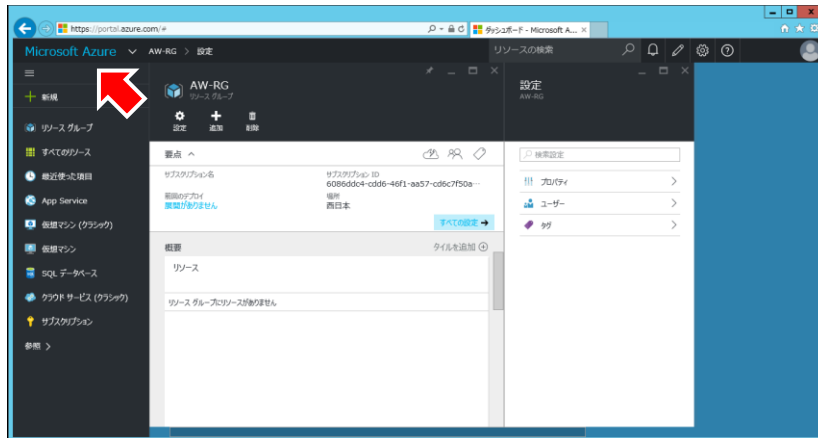


※通知に上のようなメッセージが表れるので「ダッシュボードにピン留めする」をクリックする

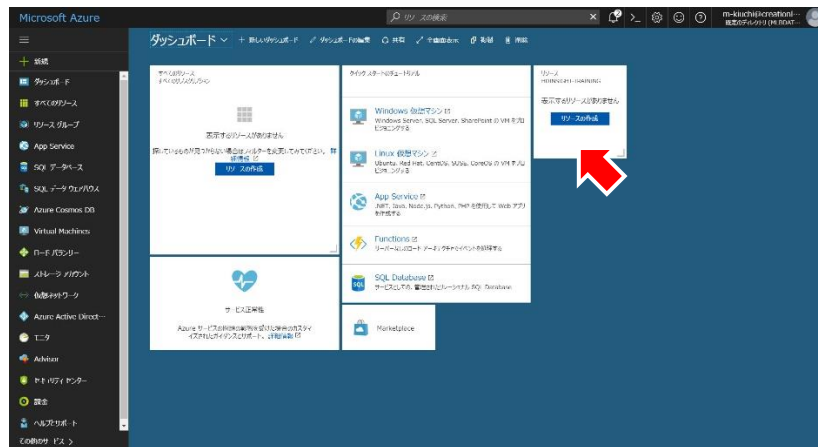
ワンポイント

Microsoft Azure の 1 ヶ月 評価版のサブスクリプションの場合、一部制限があり、選択できる場所（データセンターの場所）や Azure Marketplace の利用、Oracle などのサードパーティ製のアプリケーションの利用などが限定されます。例えば、データセンターとして東日本は選択できません。そのため、この自習書では、すべて西日本を使用します。

- リソースグループが作成され、[HDInsight-Training] と [設定] が表示されます。画面左のジャンプバーより、[Microsoft Azure] をクリックします。



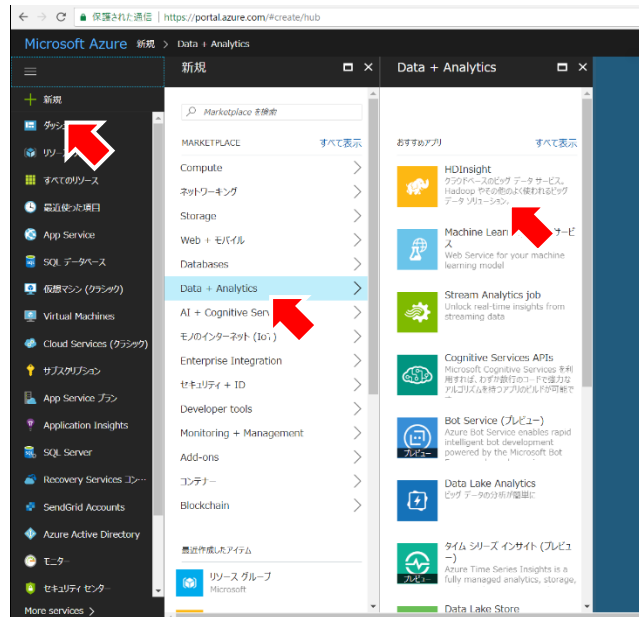
6. 「スタート画面」が表示されます。「HDInsight-Training」のタイルがピン留めされていることを確認します。



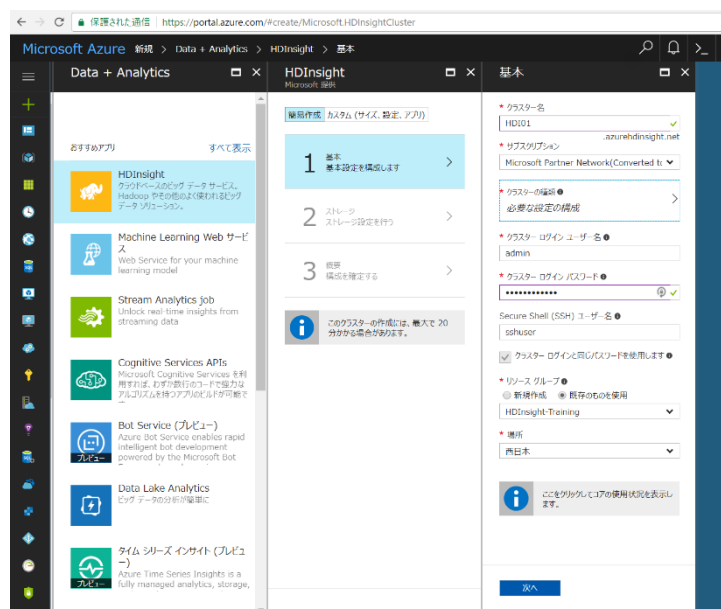
6. HDInsight クラスターのデプロイ

次の手順では、先に作成した「HDInsight-Training」リソースグループ内に、HDInsight クラスター「HDI01」を作成します。

1. Azure 管理ポータル画面左のジャンプバーより、[新規] をクリックし、[Data + Analytics] → [HDInsight] をクリックします。



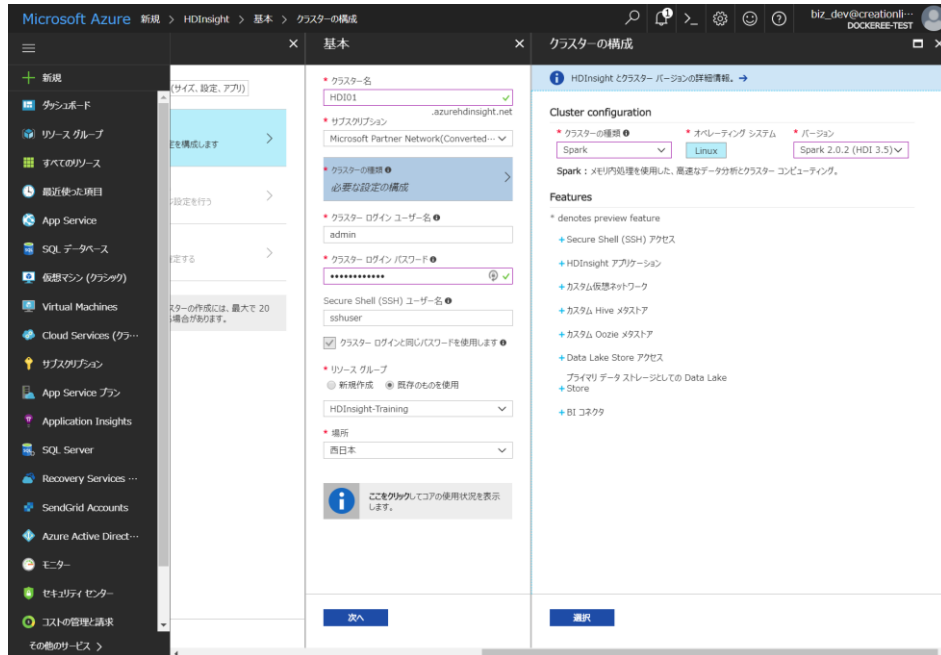
2. HDInsight デプロイのためのブレードが新たに表示されます。次のパラメーターを指定し、[次へ] をクリックします。



パラメーター	説明	今回の設定
クラスター名	HDInsight クラスターに付ける任意の名前です。(注意: HDInsight クラスター名は Azure シ	HDI01

	システムで一意の名前にする必要があります。 エラーが出た場合は設定値を変えてください)	
サブスクリプション	使用するサブスクリプションです。	<既定値>
クラスターの種類	(後述)	(後述)
クラスターログインユーザー名	HDInsight クラスターにログインするためのユーザー名です	<既定値(admin)>
クラスターログインパスワード	HDInsight クラスターにログインするためのパスワードです	Pa\$\$w0rd1234
Secure Shell(SSH)ユーザー名	HDInsight クラスターの個々のノードにログインするためのユーザー名です	<既定値(sshuser)>
<input checked="" type="checkbox"/> クラスターログインと同じパスワードを使用します	クラスターログインと SSH ログインを同じパスワードにするかどうかを指定します	<既定値(チェックを入れる)>
リソースグループ	HDInsight クラスターを管理するためのリソースグループを指定します	“既存のものを使用” 先ほど作成したリソースグループ “HDInsight-Training” を選択
場所	HDInsight クラスターがデプロイされるリージョンを指定します	西日本

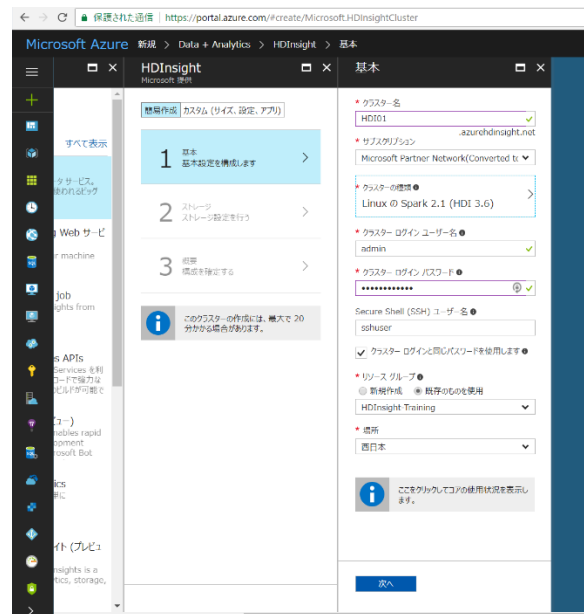
[クラスターの種類]をクリックし、[クラスターの種類]ブレードを表示します。次のパラメータを指定し、[選択]ボタンを押します。



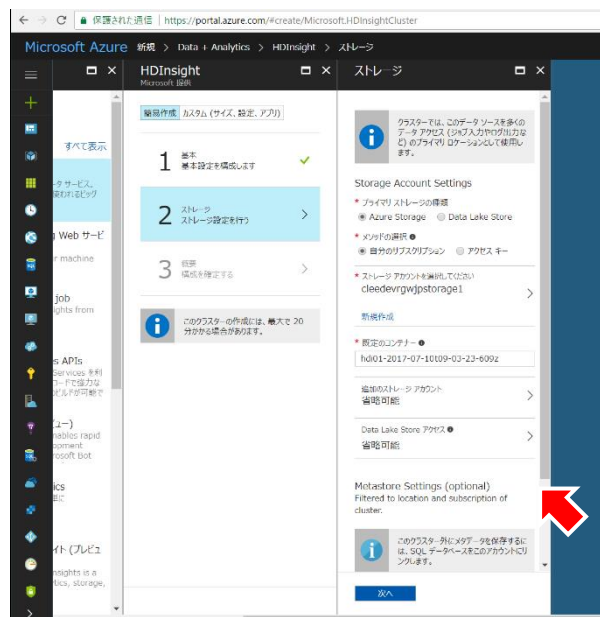
パラメーター	説明	今回の設定
クラスターの種類	デプロイする HDInsight クラスターの種類を	Spark

	選択します。選択する種類によってデプロイされるパッケージが変化します。	
オペレーティングシステム	使用するオペレーティングシステムです。	<既定値(Linux)>
バージョン	デプロイする HDInsight クラスターのバージョンを選択します	Spark 2.0.2(HDI 3.5)

設定後のブレードの内容は以下のようになっています。[次へ]をクリックします。



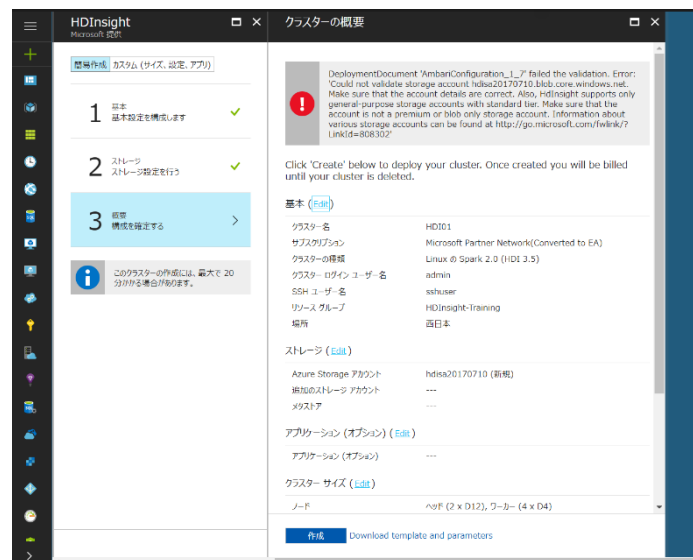
3. [ストレージ] ブレードが表示されます。次のパラメータを指定し、[次へ]ボタンを押します。



パラメーター	説明	今回の設定
プライマリストレージの種類	HDInsight クラスターから接続可能なストレージの種類を選択します。	<既定値(Azure Storage)>

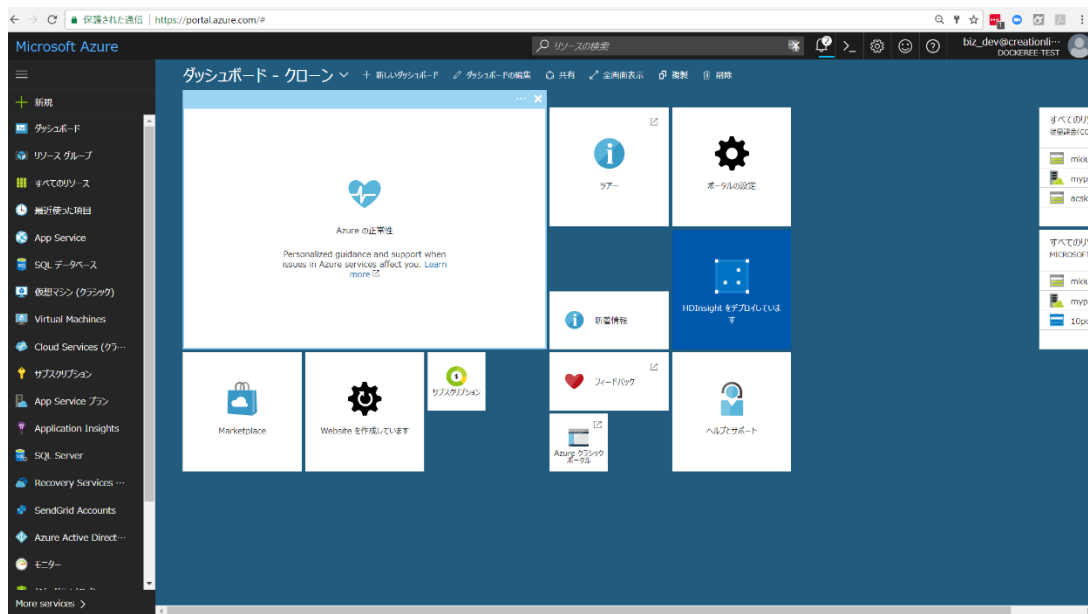
メソッドの選択	接続可能なストレージアカウントを一覧表示するための方法を選択します	<既定値(自分のサブスクリプション)>
ストレージアカウントを選択してください	HDInsight クラスターから接続するストレージアカウントを選択します(注意: ストレージアカウント名は Azure システムで一意の名前にする必要があります。エラーが出た場合は設定値を変えてください)	"新規作成"をクリックし、ストレージアカウント名"hdisa20170710"を入力
既定のコンテナ	HDInsight がデータを保存するストレージコンテナ名を指定します	hdi01
追加のストレージアカウント	他に接続するストレージアカウントがある場合は選択します	<既定値(省略可能)>
Data Lake Store アクセス	デプロイする HDInsight クラスターが Data Lake Store に接続する場合は選択します	<既定値(省略可能)>
Hive 用の SQL データベースを選択する	Hive metastore 用のデータベースに、HDInsight 外のデータベースを指定する場合にはここに Database URL を指定します	<既定値(空欄)>
Oozie 用の SQL データベースを選択する	Oozie 用のデータベースに HDInsight 外のデータベースを指定する場合にはここに Database URL を指定します	<既定値(空欄)>

4. [クラスターの概要] ブレードが表示されます。設定を確認し、[作成]ボタンをクリックします。

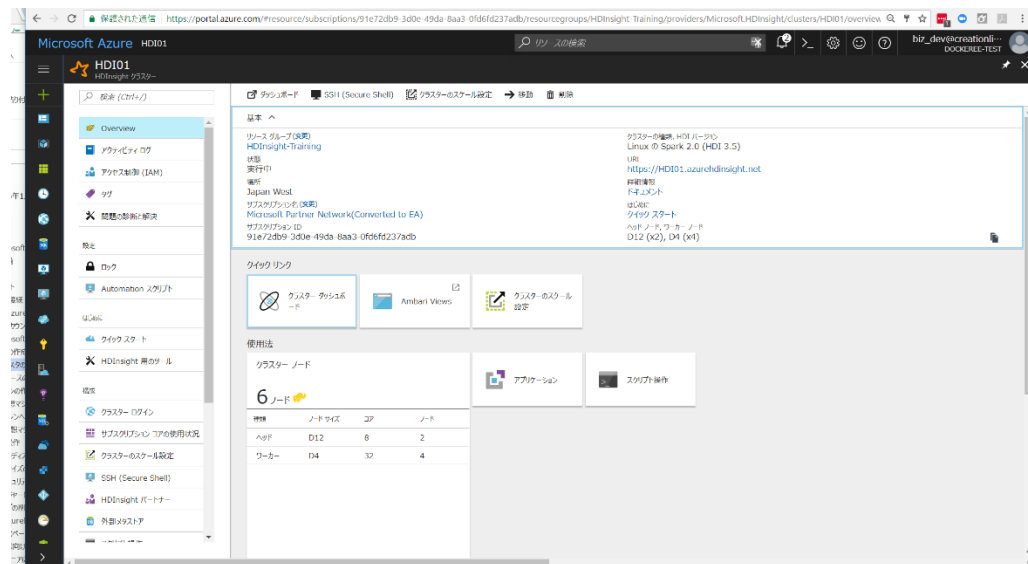


5. ダッシュボード画面に戻り、デプロイが開始されます。完了までには 20~30 分ほどかかります

Microsoft Azure 自習書シリーズ 仮想マシンの作成と操作



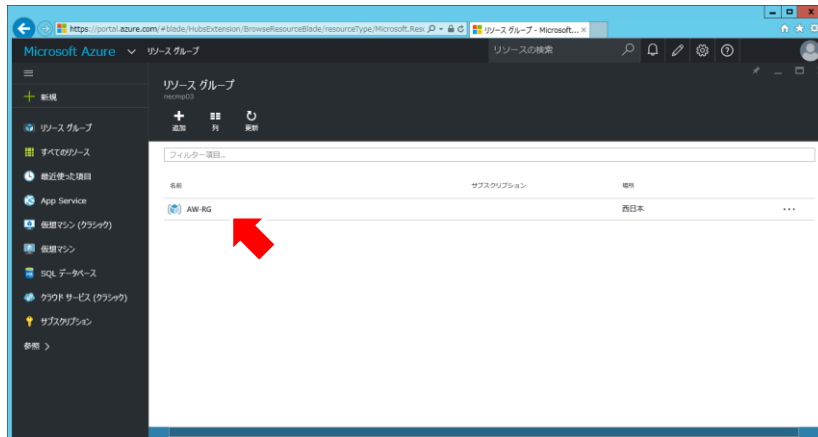
6. デプロイが完了すると、自動的に [HDI01] ブレードが開き、使用を始めることができます



7. Azure Storage Blob へのデータ投入(1)

次の手順では、リソースグループ「HDInsight-Training」に作成されたストレージアカウント「hdisa20170710」にデータを投入する方法を解説します。

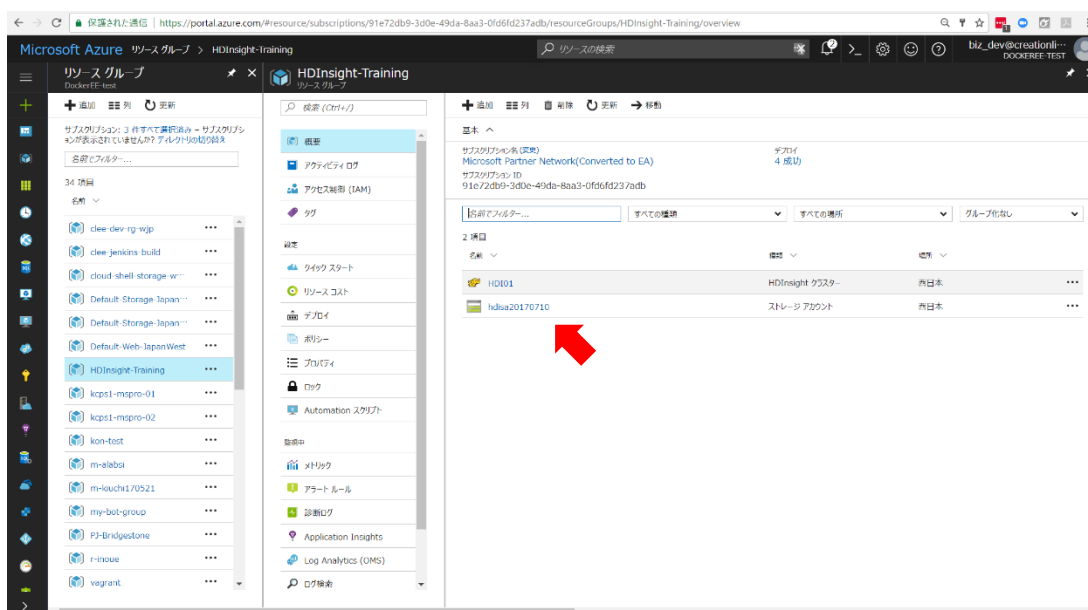
1. Azure 管理ポータル画面左のジャンプバーより、[リソースグループ] をクリックします。
2. [リソースグループ] が表示されます。一覧から「HDInsight-Training」をクリックします。



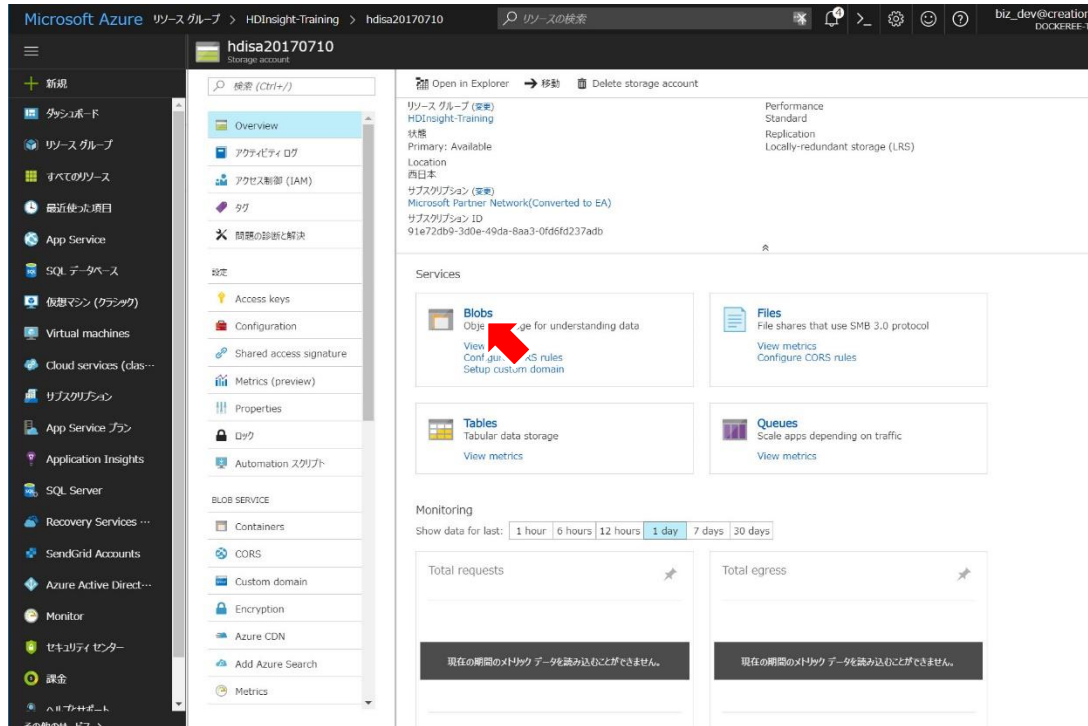
ワンポイント

Microsoft Azure は、いくつかのリソースグループを自動的に作成します。そのため、作成した記憶のないリソースグループも存在します。

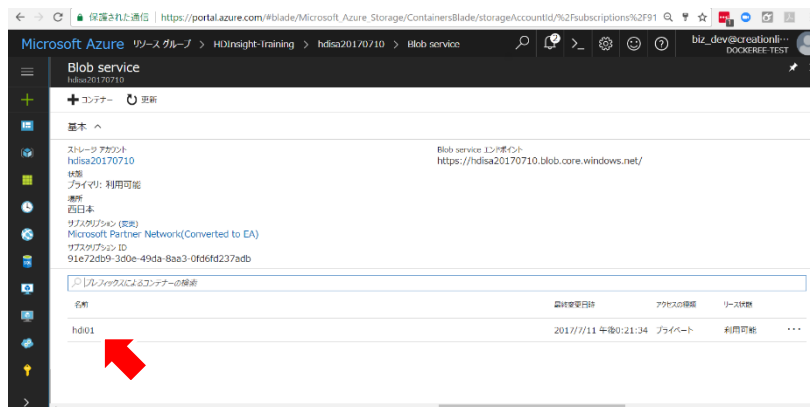
3. 「HDInsight-Training」ブレードが開き、管理されているサービス(HDInsight, ストレージアカウント)が表示されます。ストレージアカウント「hdisa20170710」をクリックします。



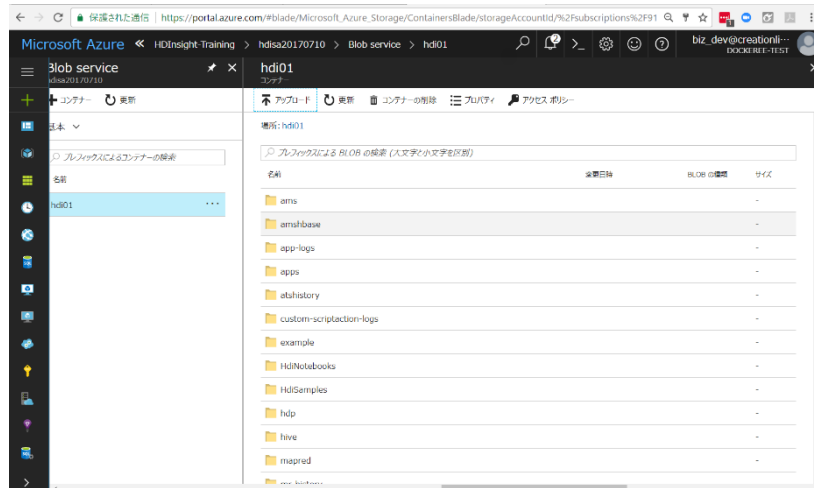
4. [hdisa20170710] ブレードが開きます。[BLOB] をクリックします。



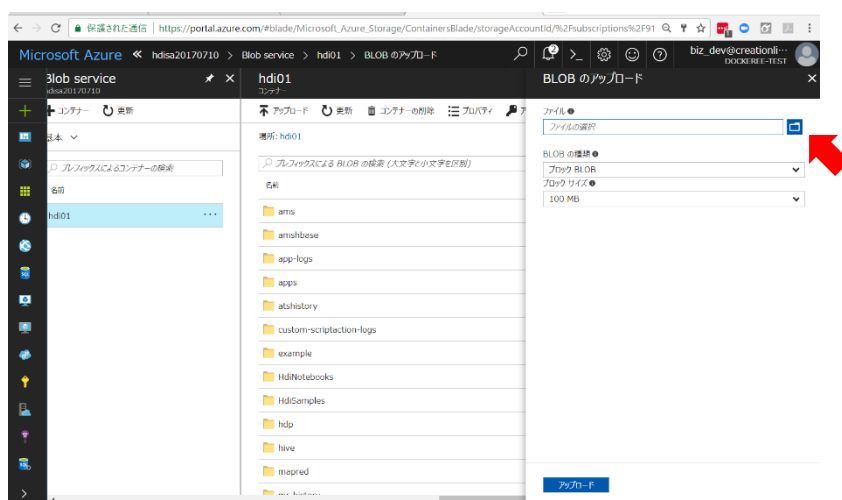
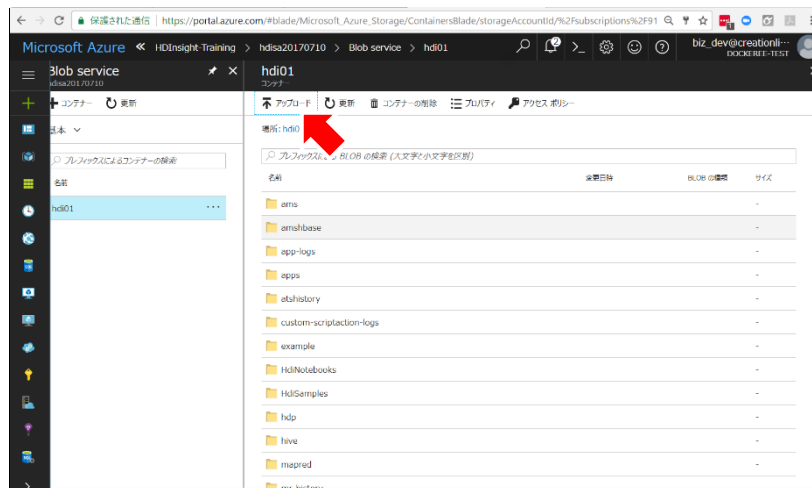
5. [Blob service] ブレードが開きます。[hdi01] コンテナをクリックします。



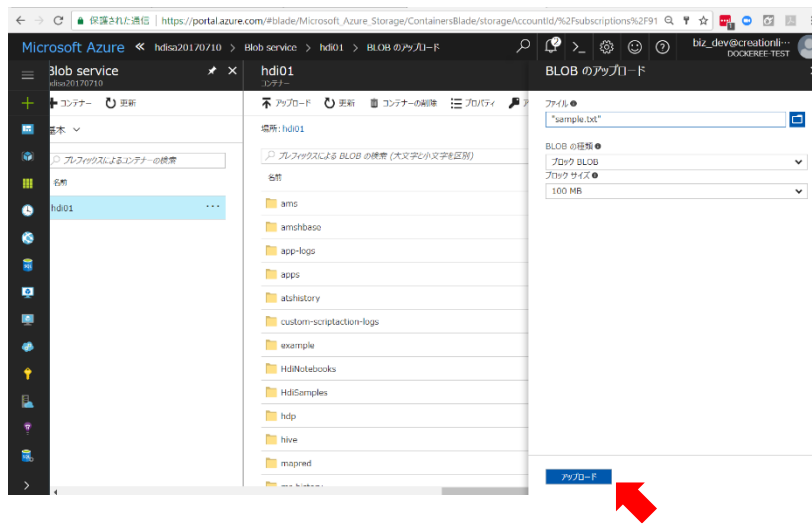
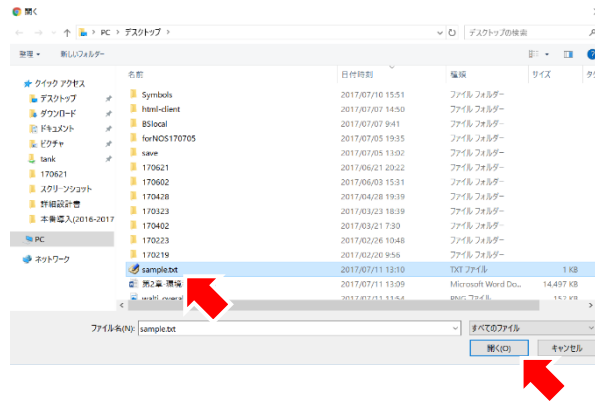
6. [hdi01] コンテナブレードが開きます。ここから内容を確認したりファイルをアップロードすることができます。



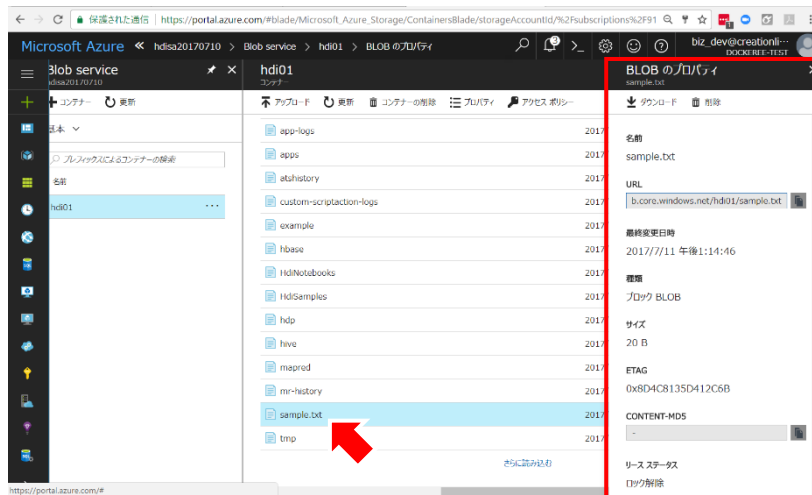
7. それでは試しに1つファイルをアップロードしてみます。[hdi01] コンテナブレードの [アップロード] ボタンをクリックし、PC のファイルを1つアップロードしてみます。



Microsoft Azure 自習書シリーズ 仮想マシンの作成と操作



8. アップロードしたファイルをクリックすると、その情報を確認することができます。

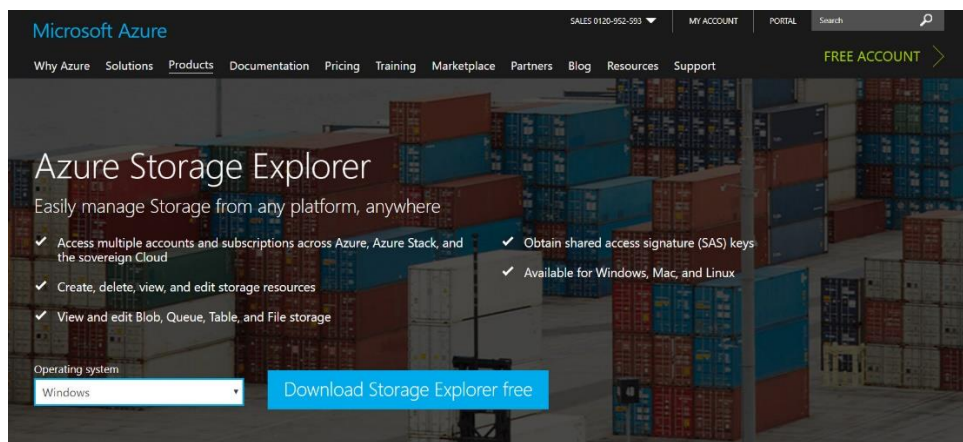


8. Azure Storage Blob へのデータ投入(2)

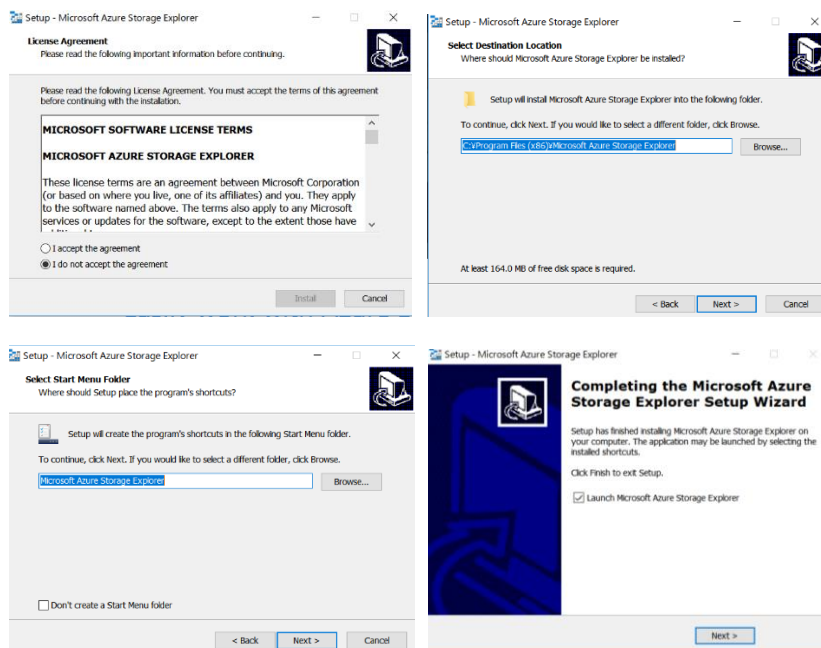
Azure Storage Blob にファイルをアップロードするには、Web ブラウザを使用した方法以外にも方法があります。

ここでは Microsoft Azure Storage Explorer を使用したアップロードを紹介します。

1. 新しいブラウザウィンドウを開き、<http://storageexplorer.com/> を開きます。画面中央の「Operating system」から自分の OS を選択してから "Free Download Storage Explorer free" ボタンをクリックするとインストーラが自動的にダウンロードされます。

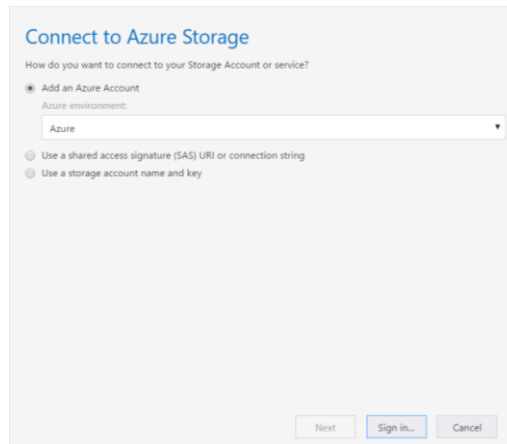


2. ダウンロードが完了したら、ダブルクリックして開き、指示に従ってインストールを完了します。



3. インストールが完了すると、自動的に Azure Storage Explorer が起動し、ログイン画面が表示されます。“Add an Azure Account” を選択し、プルダウンメニューから “Azure(既定値)” を選択しま

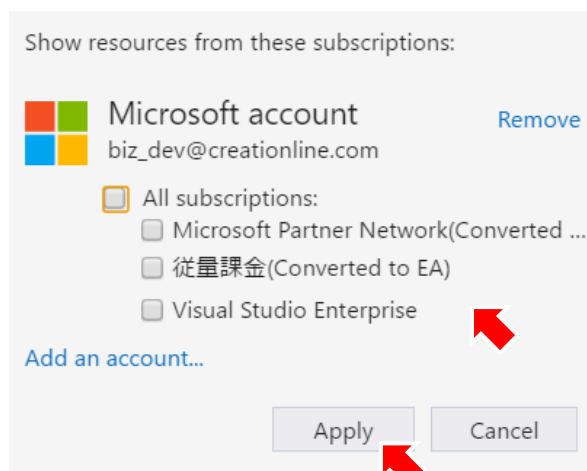
す。[Sign In...] ボタンをクリックします。



4. 別ウィンドウで Azure へのサインイン画面が表示されます。HDInsight, ストレージアカウントを作成したものと同一アカウントでサインインします。

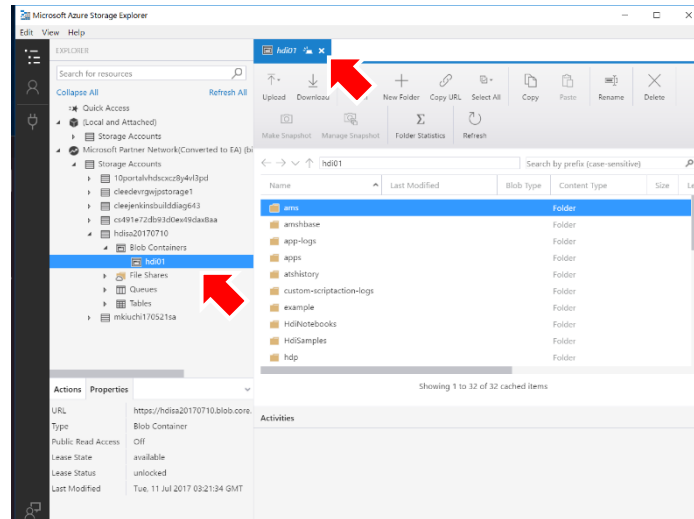


5. サインインが完了すると、複数の Azure サブスクリプションを持っている場合は使用できるサブスクリプションを選択する画面が開きます。使用したいサブスクリプションにチェックを入れて、[Apply] をクリックします。

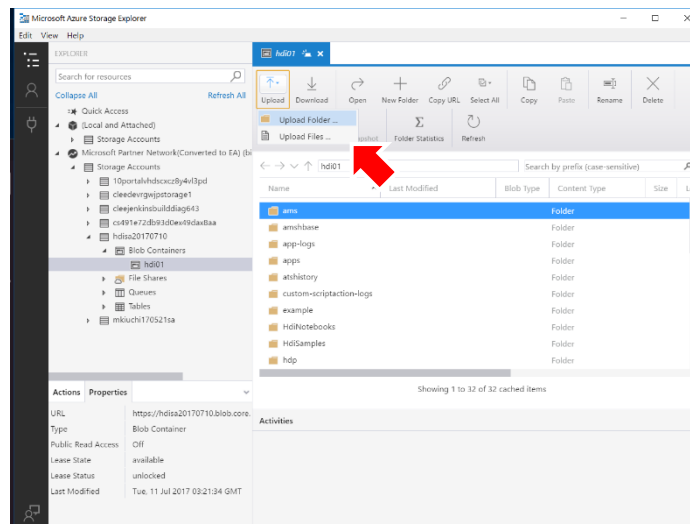


6. エクスプローラー風の画面に遷移します。作成したストレージアカウント（上記例では"hdisa20170710"）を展開すると、Blob コンテナが見つかります。選択するとコンテナの内容が

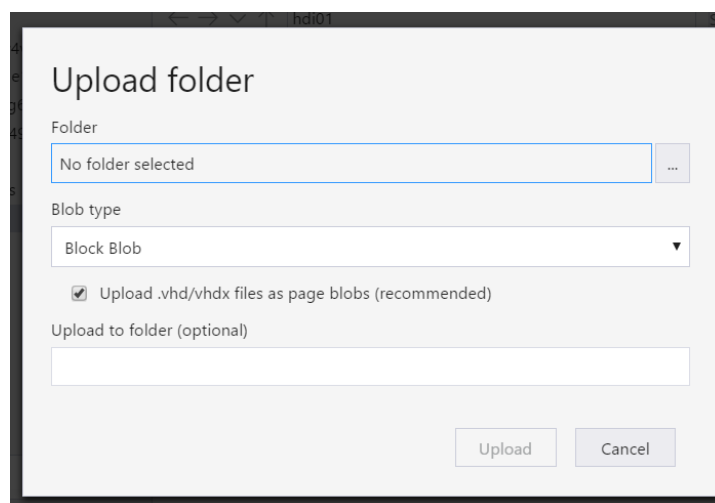
右側のペインに表示されます。



7. それでは実際に Blob にまとめてデータをアップロードします。Storage Explorer 上部のメニューから [Upload] - [Upload folder...] を選択します。

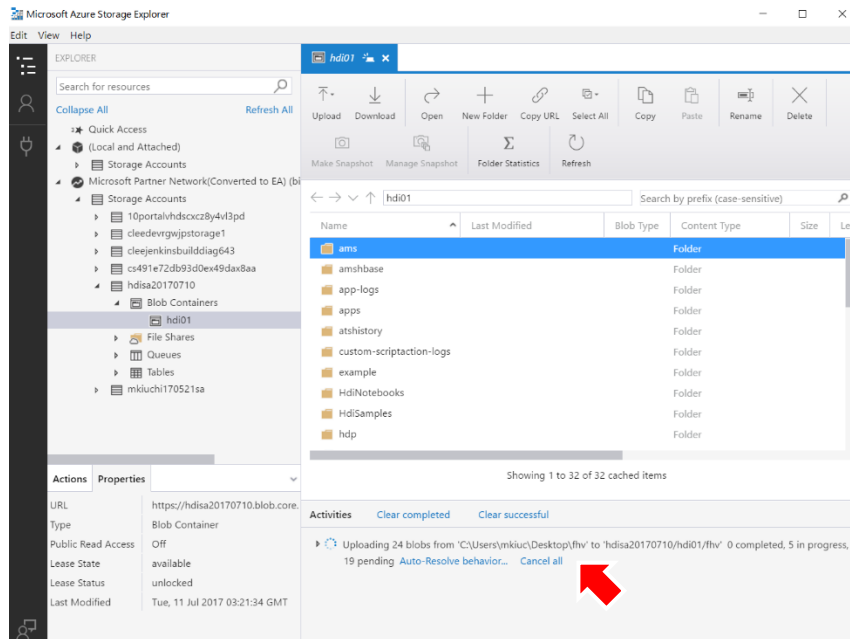


8. アップロード画面が表示されます。画面内の要素はそれぞれ以下の意味を持っています。ここでは例として、ローカル PC の "fhv" フォルダをアップロードします。

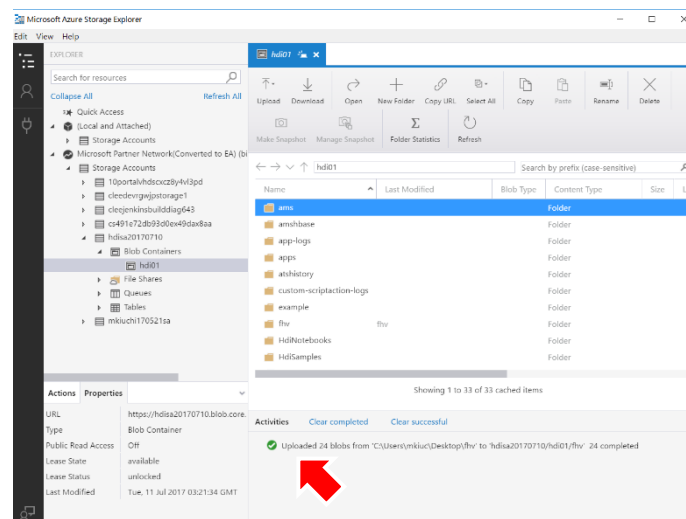


パラメーター	説明	今回の設定
Folder	ローカル PC のフォルダを選択します。	任意のフォルダ
Blob type	アップロードする Blob タイプを指定します	<既定値(Block Blob)>
Upload to folder	アップロード先のフォルダを指定します	<既定値(ローカルフォルダと同じ名前)>

9. 「Upload」 ボタンを押してアップロードを開始すると、Storage Explorer 下側のペインに進行状況が表示されます。



10. アップロードが完了すると、Storage Explorer 下側ペインの進行状況は以下のように変化します。



9. （参考）Microsoft Azure に関する情報の入手元

Microsoft Azure に関する最新の情報は、次の Web サイトから入手できます。



● Azure の公式ページ（各国共通）

製品情報、価格、技術情報など、Azure に関するすべての情報への入口です。Azure をお使いのお客様は、右上の「ポータル」をクリックすると Azure のポータルにログインすることができます。ポータルでは、Azure 上で構築したアプリケーションの管理、課金状況の確認などができます。

<http://azure.microsoft.com/ja-jp/>

● サポートエンジニアによるブログ サイト

よくお問い合わせをいただく技術・課金・サポートに関する内容をまとめたものです。サポートに問い合わせる前に、まずはここをご参照ください。

<http://blogs.msdn.com/b/dsazurejp/>

<http://blogs.msdn.com/b/jpsql/>（SQL データベース / SQL Server）

● MSDN フォーラムの Azure フォーラム

Azure に関する技術的な質問に対して、これまでの投稿から情報を入手するのみならず、自分の質問を投稿し、他のユーザーや MVP (Most Valuable Professional) からの回答を得られることが期待できます。開発者（DEVELOPER）サポート / 標準（STANDARD）サポート契約ではカバーしていない“How To”や“仕様”に関する質問も、このフォーラムをご活用ください。

<https://social.msdn.microsoft.com/Forums/ja-JP/home?category=azure>

20. Microsoft Azure のお問合せ

Microsoft Azure では、お問い合わせ内容に応じて窓口を用意しております。お問い合わせの内容に応じて各窓口をご利用ください。なお、Premier のお客様は、Premier 窓口をご利用いただけます。

お問合せ内容	料金	窓口名	連絡方法	お問い合わせ方法
製品や機能の概要、価格、ライセンスなどの情報の収集や購入前相談	無償	Cloud Direct	電話、メール	「Cloud Direct」で検索してください。
課金、サブスクリプションに関するお問い合わせ、請求書払いへの変更、クォータ増加の依頼	無償	Microsoft Azure 新ポータル	Azure 新ポータル※1	※3
技術的なお問い合わせ	有償	Microsoft Azure 新ポータル※2		

※1：電話窓口はありません。

※2：有償のサポートプランが必要です。

※3：お問い合わせ方法につきましては、次の Blog をご参照ください。

サポートにお問い合わせする方法について

<http://blogs.msdn.com/b/dsazurejp/archive/2013/10/31/10462044.aspx>

なお、ポータルサイトにアクセスできない場合は、次のどちらかの方法でお問い合わせください。

窓口名 / 連絡方法	お問い合わせ方法
カスタマー インフォメーションセンター 電話番号：0120-41-6755	営業時間：平日 9:00 – 17:30 窓口担当者に、「Azure について問い合わせがしたい」とお伝えください。Azure 担当者より折り返しご連絡いたします。
Web フォーム http://azure.microsoft.com/ja-jp/support/options/	「アカウントのサインインまたはサインアップで問題が発生する場合」からお問い合わせを発行してください。