


Private Databricks에 ADF 연결 (작성중)

| | |
|----------|---|
| ➤ 벤더 |  <u>Databricks</u> |
| Σ 작성시간 | 2024-01-31 14:59 |
| ⌚ 작성일 | @2024년 1월 31일 오후 2:59 |
| ⌚ 최종 수정일 | @2024년 1월 31일 오후 4:01 |
| Σ 최종수정시간 | 2024-01-31 16:01 |
| 👤 저자 | ② 김석원 |

Private Databricks 구성 & ADF로 Databricks 접근

Azure to Databricks with Private Link

Private Link를 사용하여 Private 네트워크 구성하기

목차

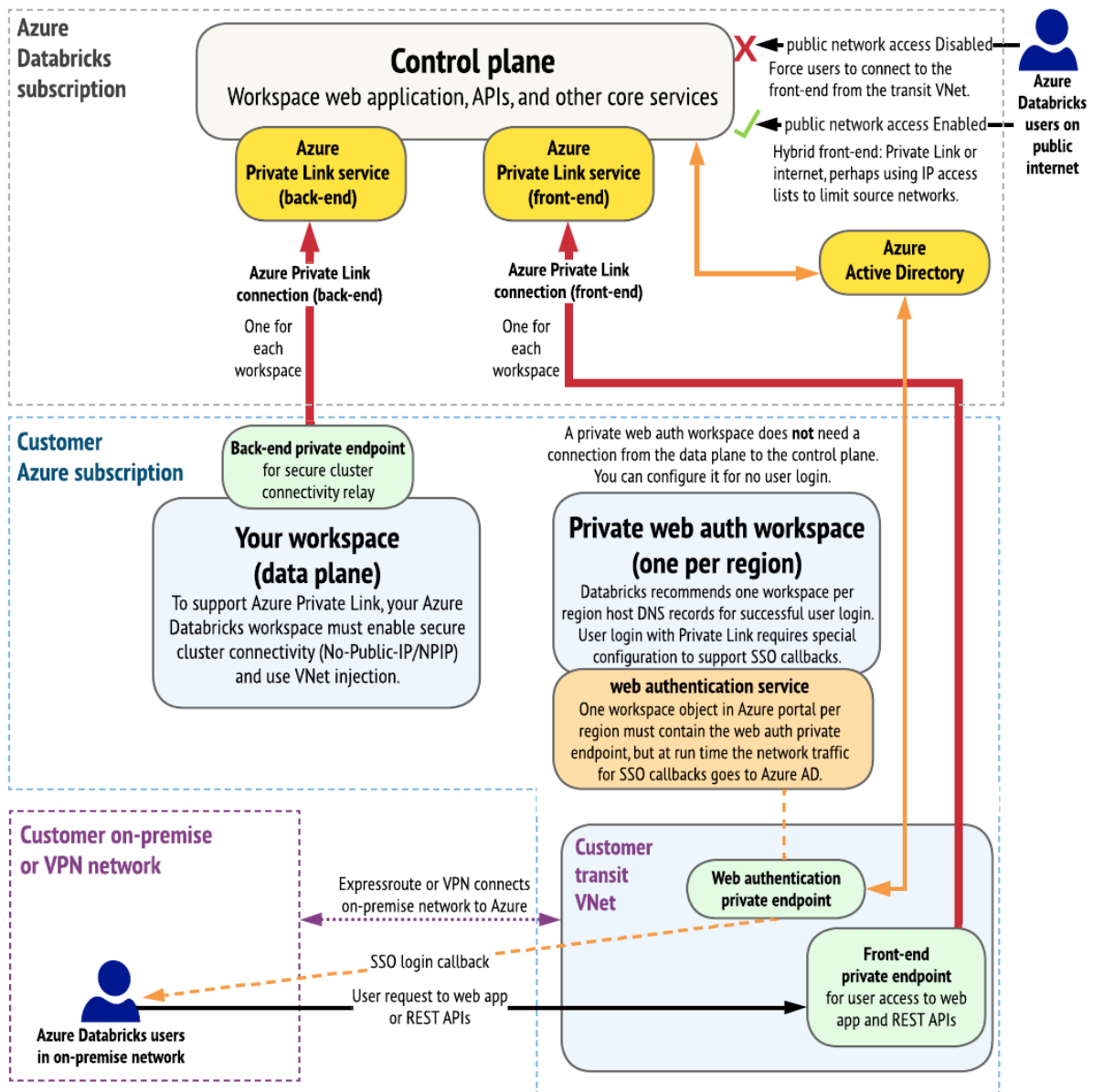
1. Vnet 만들기
2. Databricks만들기
3. 외부 접속을 위한 Auth용 Browser Privat link만들기
4. VM만들어 외부에서 접속 하기
5. ADF 배포
6. Integration Runtime 만들기
7. 데이터브릭스 접속 연결

Databricks 내부적으로 돌아가는 원리

- control plane(Databricks web접속)
 - Public network access(공용 네트워크) public internet에서 접속
 - Private network access를 이용한 web auth 접속
- Data plane - private link 이용
 - public subnet - private link 를 이용한 databricks workspace 접속(frontend)
 - private subnet - private link 를 이용한 databricks cluster 접속(backend)

Azure Private Link network flow

You can add private connectivity for front-end (user access to control plane), for back-end (data plane access to control plane), or for both

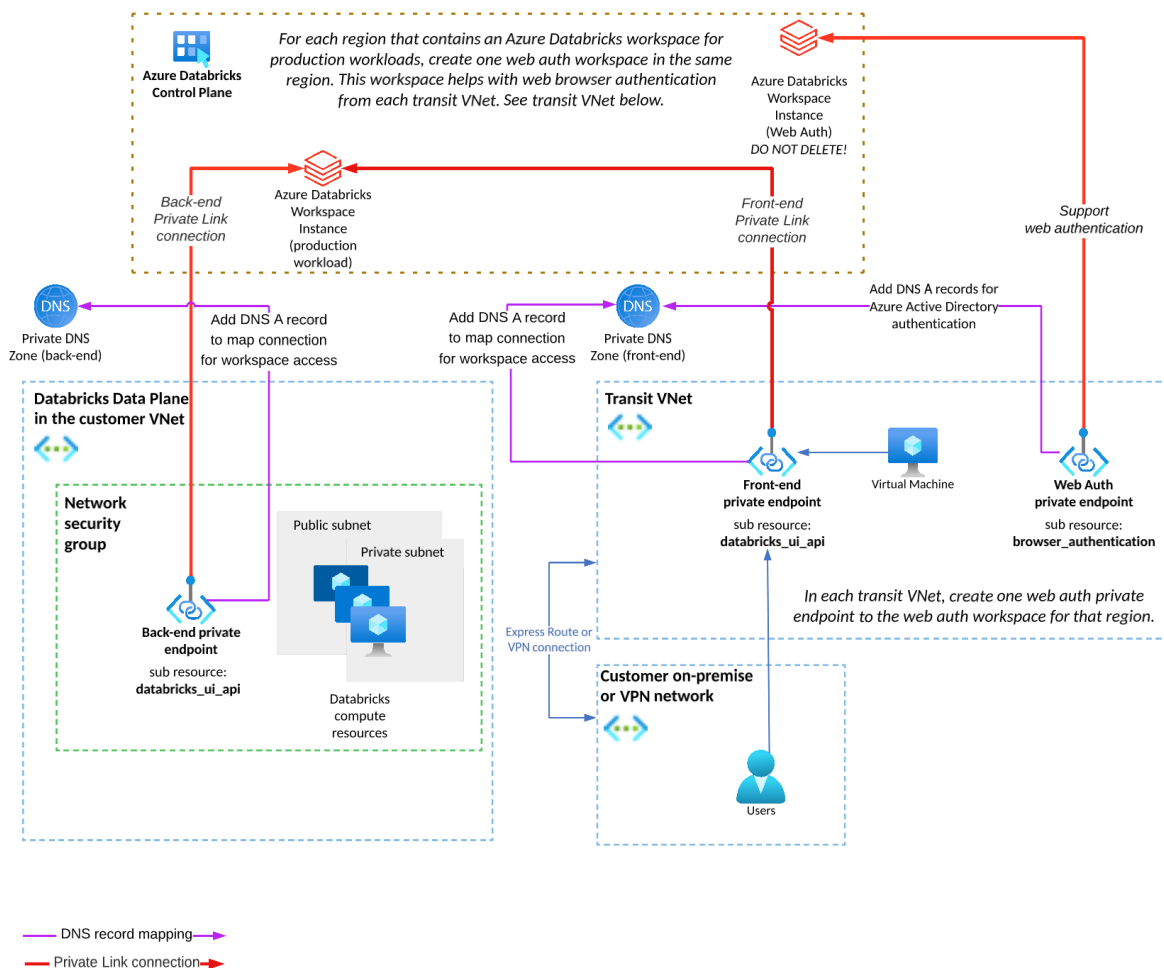


Private Link 용 subnet 구성요소 정의

1개의 Vnet 와 3개의 subnet이 필요함 이들은 private endpoint로 통신

- private link용 subnet(default) - databricks-ui용, web-auth용
- public subnet - databricks frontend용
- private subnet - databricks backend용

Azure Private Link network objects



Vnet 만들기

이름 정의 → Ip주소 (Subnet) 정의 → 만들기

홈 > MSP-심종성-220503~230530-5dc91456-4b49-48af-85ff-99660683e5cc-05 | 리소스 그룹 > RG-hanh

Virtual network

Microsoft

★ 4.1 (24 Marketplace 등급) | ★ 4.1 (16 외부 등급)

플랜
Virtual network

Resource Manager과(와) 함께 배포 (클래식(으)로 변경)

개요 플랜 사용 정보 + 지원 리뷰

Create a logically isolated section in Microsoft Azure with this networking service. You can securely connect it to an IPsec connection. Virtual Networks make it easy for you to take advantage of the scalable, on-demand infrast applications on-premises, including systems running on Windows Server, mainframes, and UNIX.

Use Virtual Network to:

홈 > Virtual network >

가상 네트워크 만들기

기본 사항 IP 주소 보안 태그 검토 + 만들기

CIDR 표기법으로 된 하나 이상의 주소 집두사로 지정된 가상 네트워크의 주소 공간입니다(예: 192.168.1.0/24).

IPv4 주소 공간

10.0.0.0/16 10.0.0.0 - 10.0.255.255(65536개 주소)

☐ IPv6 주소 공간 추가 ⓘ

서브넷의 주소 범위가 CIDR 표기법(예: 192.168.1.0/24)으로 되어 있습니다. 이 주소 범위는 가상 네트워크의 주소 공간에 포함되어야 합니다.

+ 서브넷 추가 - 서브넷 제거

| 서브넷 이름 | 서브넷 주소 범위 | NAT 게이트웨이 |
|---|-------------|-----------|
| <input type="checkbox"/> SN-PrivateLink | 10.0.0.0/24 | - |

서브넷의 어드레스도 인터넷 액세스에는 NAT 게이트웨이를 사용하는 것이 좋습니다. NAT 게이트웨이를 배포하여 가상 네트워크를 만든 후에 서브넷에 할당할 수 있습니다. [간세한 정보](#)를

- IP 범위: 먼저 기본 IP 범위를 제거한 다음 IP 범위를 추가합니다 10.28.0.0/23.
- public-subnet 범위가 있는 서브넷을 만듭니다 10.28.0.0/25.
- private-subnet 범위가 있는 서브넷을 만듭니다 10.28.0.128/25.
- private-link 범위가 있는 서브넷을 만듭니다 10.28.1.0/27.

- 기본 ip주소를 10.28.0.0/23으로 생성
- private-link라는 subnet을 10.28.1.0/27로 추가하여 생성
- public-subnet, private-subnet도 만들어야함

Databricks 만들기

순서도 : 이름 정의 → Databricks Subnet 정의 → Private Link연결 → 만들기

이름 정의

Azure Databricks 작업 영역 만들기 ...

기본 사항 네트워킹 고급 태그 검토 + 만들기

프로젝트 세부 정보

배포된 리소스와 비용을 관리할 구독을 선택합니다. 폴더 같은 리소스 그룹을 사용하여 모든 리소스를 정리 및 관리합니다.

| | |
|------------|--|
| 구독 * ① | MSP-심종성-220503~230530-5dc91456-4b49-48af-85ff-99660683e... ▼ |
| 리소스 그룹 * ① | RG-hanhwa-test ▼ |

[새로 만들기](#)

인스턴스 정보

| | |
|--------------|------------------------|
| 작업 영역 이름 * | DB-hanhwa-test ✓ |
| 지역 * | Korea Central ▼ |
| 가격 책정 계층 * ① | 프리미엄(+ 역할 기반 액세스 제어) ▼ |

i 작업 영역에 권장되는 가격 책정 계층을 선택했습니다. 필요에 따라 계층을 변경할 수 있습니다. ✕

Databricks subnet 정의

속성 정의

- 보안클러스터 : 공용 IP가 아닌 Private으로 클러스터 설정
- Vnet : 자신이 정의한 Vnet에 내용 배포
 - Public Subnet : 맨처음 사진에서 Network구성에서 Private Link안에 Databricks Front와 연결하는 subnet주소
 - Private Subnet : 맨처음 사진 Network 구성에서 Private Link안에 Databricks Back과 연결하는 Subnet 주소

Azure Databricks 작업 영역 만들기 ...

자체 VNet(가상 네트워크)에 Azure Databricks 작업 영역 배포 ☒ 예 ☐ 아니요

가상 네트워크 * ① VN-hh-test

가상 네트워크에 새로운 서브넷 두 개가 만들어집니다.

사용자를 대신하여 Azure Databricks에 대해 두 서브넷의 암시적 위임이 수행됩니다.

퍼블릭 서브넷 이름 * public-subnet

퍼블릭 서브넷 CIDR 범위 * ① 10.0.0.0/25

프라이빗 서브넷 이름 * private-subnet

프라이빗 서브넷 CIDR 범위 * ① 10.0.0.128/25

공용 네트워크 액세스 허용 ① ☐ 사용 ☒ 사용 안 함

공용 네트워크를 통해 작업 영역에 액세스할 수 없습니다.

필수 NSG 규칙 ① Azure Databricks 규칙 없음

프라이빗 엔드포인트는 사용자-작업 영역과 데이터 평면(클러스터)-컨트롤 플레인에 모두 필요합니다.

프라이빗 엔드포인트

Private Link 연결

- Private Endpoint 연결 : 외부와 Private Link(public, private subnet)과 연결하기 위해서 설정

Azure Databricks 작업 영역 만들기 ...

자체 VNet(가상 네트워크)에 Azure Databricks 작업 영역 배포 ☒ 예 ☐ 아니요

가상 네트워크 * ① VN-hh-test

가상 네트워크에 새로운 서브넷 두 개가 만들어집니다.

사용자를 대신하여 Azure Databricks에 대해 두 서브넷의 암시적 위임이 수행됩니다.

퍼블릭 서브넷 이름 * public-subnet

퍼블릭 서브넷 CIDR 범위 * ① 10.0.0.0/25

프라이빗 서브넷 이름 * private-subnet

프라이빗 서브넷 CIDR 범위 * ① 10.0.0.128/25

공용 네트워크 액세스 허용 ① ☐ 사용 ☒ 사용 안 함

공용 네트워크를 통해 작업 영역에 액세스할 수 없습니다.

필수 NSG 규칙 ① Azure Databricks 규칙 없음

프라이빗 엔드포인트는 사용자-작업 영역과 데이터 평면(클러스터)-컨트롤 플라인에 모두 필요합니다.

프라이빗 엔드포인트

프라이빗 엔드포인트를 만들어 이 리소스에 대한 프라이빗 연결을 허용합니다. 자세한 정보

이름 구독 리소스 그룹 지역

프라이빗 엔드포인트를 만들려면 [추가]를 클릭하세요.

+ 추가

검토 > 만들기 < 이전 다음: 고급 >

프라이빗 엔드포인트 만들기

구독 * ① MSP-심종성-220503-230530-5dc91456-4b49-48af-85ff-99660683e5

리소스 그룹 * ① RG-hh-test

위치 * Korea Central

이름 * ① private-endpoint-databricks-ui-api

Azure Databricks 하위 리소스 * ① databricks_ui_api

네트워킹

프라이빗 엔드포인트를 배포하려면 가상 네트워크 서브넷을 선택합니다. 프라이빗 엔드포인트 네트워킹에 대한 자

가상 네트워크 * ① VN-hh-test

서브넷 * ① VN-hh-test/private-link(10.0.1.0/27)

위의 서브넷에 대해 NSG(네트워크 보안 그룹)을 사용하도록 설정한 이 엔드포인트에 대해서만 NSG를 사용하지 않도록 설정됩니다. 서브넷의 NSG가 적용됩니다.

프라이빗 DNS 통합

프라이빗 엔드포인트를 사용하여 비공개로 연결하려면 DNS 레코드가 필요합니다. 프라이빗 엔드포인트를 프라이빗이 아닙니다. 사용자의 DNS 서버를 이용하거나 가상 머신의 호스트 파일을 사용하여 DNS 레코드를 만들 수도 있습니다. 프라이빗 DNS 통합에 대한 자세한 정보 >

프라이빗 DNS 영역과 통합 ① 예 아니요

프라이빗 DNS 영역 * ① (신규) privatelink.azuredatabricks.net

확인 취소

외부 접속을 위한 Auth용 Browser Privat link만들기

순서도 : Private Endpoint만들기 → 이름 정의 → privateendpoint용 subnet정의 → private link 확인

만들기 클릭

홈 > Private Link 센터

Private Link 센터 | 프라이빗 엔드포인트

검색

« + 만들기 ⚙ 보기 관리 ↕ 새로 고침 ↓

개요

보류 중인 연결

프라이빗 엔드포인트

- 프라이빗 링크 서비스
- Azure Arc private link scopes
- Azure Monitor private link scopes

리소스

활성 연결

| <input type="checkbox"/> 이름 ↑↓ | 프라 |
|---|------|
| <input type="checkbox"/> < > bmdb-pb-endp | 10.2 |
| <input type="checkbox"/> < > bmdb-pv-link2 | 10.2 |
| <input type="checkbox"/> < > private-endpoint-databricks-ui-api | 10.0 |

Private Endpoint만들기 → 이름 정의

홈 > Private Link 센터 | 프라이빗 엔드포인트 >
프라이빗 엔드포인트 만들기 ...

기본 사항 리소스 가상 네트워크 DNS 태그 검토 + 만들기

프라이빗 엔드포인트를 사용하여 서비스 또는 리소스에 비공개로 연결합니다. 프라이빗 엔드포인트는 가상 네트워크와 동일한 영역에 있어야 하지만 연결 중인 프라이빗 링크 리소스와 다른 영역에 있을 수 있습니다. [자세한 정보](#)

프로젝트 정보

구독 * ① MSP-실용성-220503~230530-5dc91456-4b49-48af-85ff-99660683e5cc-05

리소스 그룹 * ① RG-hh-test
새로 만들기

엔스틴스 정보

이름 * private-endpoint-web-auth

네트워크 인터페이스 이름 * private-endpoint-web-auth-nic

지역 * Korea Central

홈 > Private Link 센터 | 프라이빗 엔드포인트 >
프라이빗 엔드포인트 만들기 ...

기본 사항 리소스 가상 네트워크 DNS 태그 검토 + 만들기

Private Link는 프라이빗 링크 서비스, SQL Server 또는 Azure 스토리지 계정 같은 다양한 Azure 리소스의 프라이빗 엔드포인트를 만드는 옵션을 제공합니다. 이 프라이빗 엔드포인트를 사용하여 연결하려는 리소스를 선택합니다. [자세한 정보](#)

연결 방법 ①

☒ 내 디렉터리에서 Azure 리소스에 연결합니다.
☐ 리소스 ID 또는 별칭을 사용하여 Azure 리소스에 연결합니다.

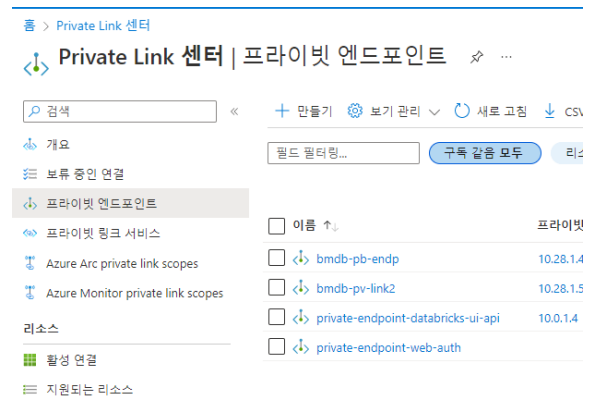
구독 * ① MSP-실용성-220503~230530-5dc91456-4b49-48af-85ff-99660683e5cc-05

리소스 종류 * ① Microsoft.Databricks/workspaces

리소스 * ① DB-hh-test

대상 하위 리소스 * ① browser_authentication

privateendpoint용 subnet정의 → private link 확인



VM만들어 외부에서 접속 하기

순서도 : 이름 정의 → 만들기 → vm접속 → databricks workspace 접속

이름 정의 → 만들기

가상 머신 만들기 ...

기본 사항 디스크 네트워킹 관리 Monitoring 고급 태그 검토 + 만들기

Linux 또는 Windows를 실행하는 가상 머신을 만듭니다. Azure Marketplace에서 이미지를 선택하거나 고유한 사용자 지정 이미지를 사용합니다. [기본] 탭을 완료하고 [검토 + 만들기]하여 기본 매개 변수로 가상 머신을 프로비전하거나, 전체 사용자 지정에 대해 각 탭을 검토합니다. [자세한 정보](#)

프로젝트 정보

배포된 리소스와 비용을 관리할 구독을 선택합니다. 폴더 같은 리소스 그룹을 사용하여 모든 리소스를 정리 및 관리합니다.

| | |
|----------|---|
| 구독 * | MSP-심종성-220503~230530-5dc91456-4b49-48af-85ff-99660683e5cc-05 |
| 리소스 그룹 * | RG-hanhwa-test |

[새로 만들기](#)

인스턴스 정보

| | |
|--------------------|---|
| 가상 머신 이름 * | VM-hanhwa-test |
| 지역 * | (Asia Pacific) Korea Central |
| 가용성 옵션 | 인프라 중복이 필요하지 않습니다. |
| 보안 유형 | 표준 |
| 이미지 * | Windows 10 Pro, version 21H2 - Gen2 |
| | 모든 이미지 보기 VM 생성 구성 |
| VM 아키텍처 | <input type="radio"/> Arm64 <input checked="" type="radio"/> x64 i Arm64는 선택한 이미지에서 지원되지 않습니다. |
| Azure Spot 할인으로 실행 | <input type="checkbox"/> |
| 크기 * | Standard_B2s - 2 vcpu, 4 GiB 메모리 (₩42,692/월) |

검토 + 만들기

< 이전

다음: 디스크 >

가상 머신 만들기 ...

기본 사항 디스크 네트워킹 관리 모니터링 고급 태그 검토 + 만들기

NIC(네트워크 인터페이스 카드) 설정을 구성하여 가상 머신에 대한 네트워크 연결을 정의합니다. 보안 그룹 규칙을 사용하여 포트, 인바운드 및 아웃바운드 연결을 제어하거나 기존 부하 분산 솔루션 뒤에 배치할 수 있습니다. [자세한 정보](#)

네트워크 인터페이스

가상 머신을 만들면 네트워크 인터페이스가 만들어집니다.

| | |
|------------------|--|
| 가상 네트워크 * ① | <div>sw-hyel-vnet</div> <div>새로 만들기</div> |
| 서브넷 * ① | <div>private-link-SN(10.28.1.0/27)</div> <div>서브넷 구성 관리</div> |
| 공용 IP ① | <div>(새로 만드는 중) public-SN-ip</div> <div>새로 만들기</div> |
| NIC 네트워크 보안 그룹 ① | <div><input type="radio"/> 없음</div> <div><input checked="" type="radio"/> 기본</div> <div><input type="radio"/> 고급</div> |
| 공용 인바운드 포트 * ① | <div><input type="radio"/> 없음</div> <div><input checked="" type="radio"/> 선택한 포트 허용</div> |
| 인바운드 포트 선택 * | <div>RDP (3389)</div> |

⚠ 이렇게 하면 모든 IP 주소가 가상 머신에 액세스할 수 있습니다. 이는 테스트 용으로만 권장됩니다. [네트워킹] 탭의 [고급] 컨트롤을 사용하여 인바운드 트래픽을 알려진 IP 주소로 제한하는 규칙을 만듭니다.

VM 삭제 시 공용 IP 및 NIC 삭제 ①

☐

vm접속

DB-hanhwa-test - Microsoft Azure | Databricks

Search resources, services, and docs (G+)

DB-hanhwa-test

Azure Databricks Service


Search

Delete

- Overview
- Activity log
- Access control (IAM)
- Tags
- Settings
 - Virtual Network Peerings
 - Encryption
 - Networking
 - Properties
 - Locks
- Monitoring
 - Diagnostic settings
- Automation
 - Tasks (preview)
 - Export template
- Help

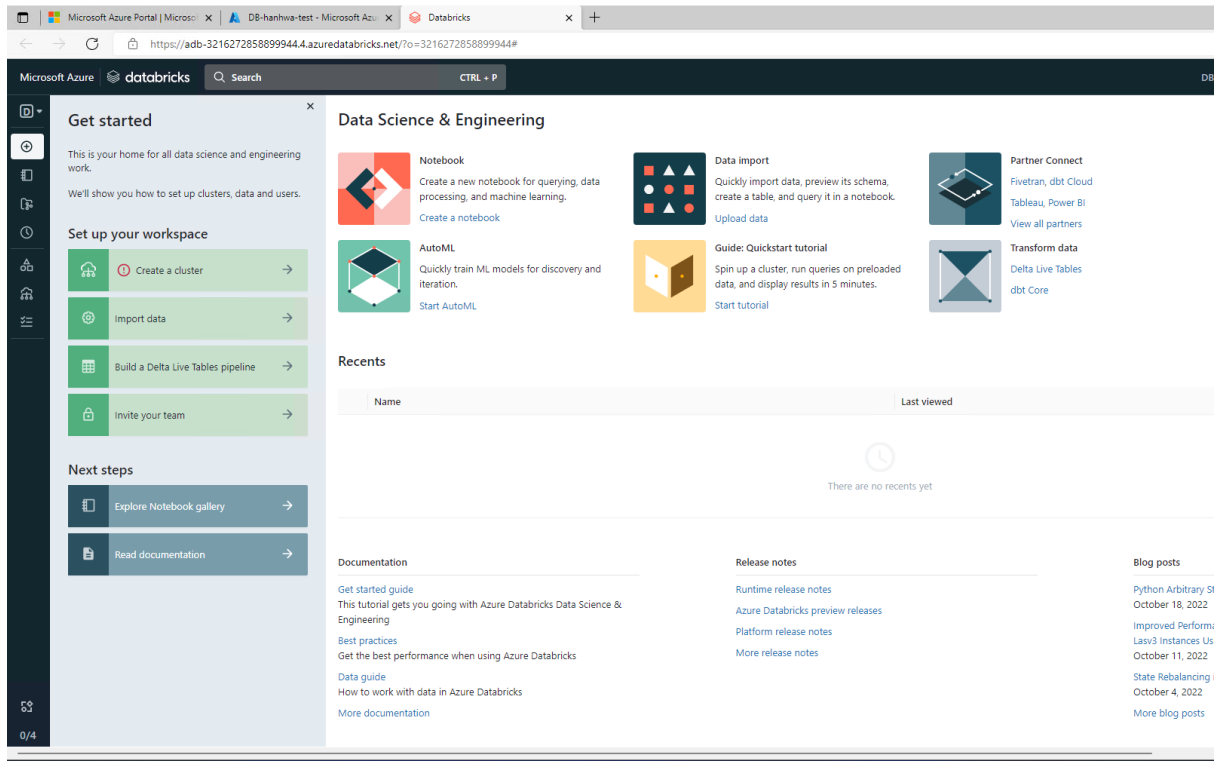
Essentials

| | | | |
|--------------------------|--|------------------------|---|
| Status | Active | Managed Resource Group | databricks-rg-C |
| Resource group | RG-hanhwa-test | URL | https://adb-321 |
| Location | Korea Central | Pricing Tier | premium |
| Subscription | MSP-심종성-220503~230530-5dc91456-4b49-48af-85ff-99660683e... | Enable No Public IP | Yes |
| Subscription ID | 437c091e-6036-4230-a3a1-57bf40149989 | Virtual Network | VN-hanhwa-tes |
| Tags | (edit) Click here to add tags | | |
| See more | | | |



[Launch Workspace](#)

databricks workspace 접속



ADF 배포

데이터팩토리를 간단하게 구성한 후 배포

Data Factory 만들기 ...

기본 사항 Git 구성 네트워킹 고급 태그 검토 + 만들기

샘플 파이프라인 및 데이터 세트를 사용하여 한 번의 클릭으로 데이터 팩터리를 만드세요. [사용해 보기](#)

프로젝트 정보

배포된 리소스와 비용을 관리할 구독을 선택합니다. 폴더 같은 리소스 그룹을 사용하여 모든 리소스를 정리 및 관리합니다.

구독 * ①

리소스 그룹 * ①

[새로 만들기](#)

인스턴스 정보

이름 * ①

지역 * ①

버전 * ①

Integration Runtime 만들기

통합 런타임

IR(통합 런타임)은 다양한 네트워크 환경에서 다음 데이터 통합 기능을 제공하는 컴퓨팅 인프라입니다. [자세한 정보](#)

+ 새로 만들기 ○ 새로 고침

이름으로 필터링

1~2/2개 항목 표시

| 이름 | 형식 | 하위 형식 | 상태 | 관련 |
|-------------------------------|--------|-------|------|----|
| AutoResolveIntegrationRuntime | Azure | 공용 | 실행 중 | 0 |
| integrationRuntime1 | 자체 호스팅 | --- | 실행 중 | 2 |

Integration Runtime은 활동을 실행하거나 디스패치하기 위해 사용하는 네이티브 컴퓨팅입니다. 통합 런타임이 필요한 기능을 토대로 만들 항목을 선택하세요. [자세한 정보](#)

Azure, Self-Hosted
Perform data flows, data movement and dispatch activities to external compute.

Azure-SSIS
Lift-and-shift existing SSIS packages to execute in Azure.

self hosted로 설정 후 수동 설치의 파일을 받기

통합 런타임 설치

[설정](#) [노드](#) [자동 업데이트](#) [공유](#) [링크](#)

통합 런타임을 Windows 머신에 설치하거나 인증 키를 사용하여 노드를 추가합니다.

이름 ①

integrationRuntime2

옵션 1: 빠른 설치

[이 컴퓨터의 빠른 설치를 시작하려면 여기 클릭](#)

옵션 2: 수동 설치

1단계: [통합 런타임 다운로드 및 설치](#)

2단계: 이 키를 사용하여 통합 런타임 등록

이름 인증 키

Key1

IR@e9eb7f5e-195e-4e38-acdb-dc15d3f01dda@sw-hyel-test-ftory@Serv



Key2

IR@e9eb7f5e-195e-4e38-acdb-dc15d3f01dda@sw-hyel-test-ftory@Serv



ccessible. You can reload Internet Explorer sites with IE mode in Microsoft Edge.

Microsoft Edge



Microsoft Integration Runtime

The Microsoft Integration Runtime is a customer managed data integration infrastructure used by Azure Data Factory and Azure Synapse Analytics to provide data integration capabilities across different network environments.

Important! Selecting a language below will dynamically change the complete page content to that language.

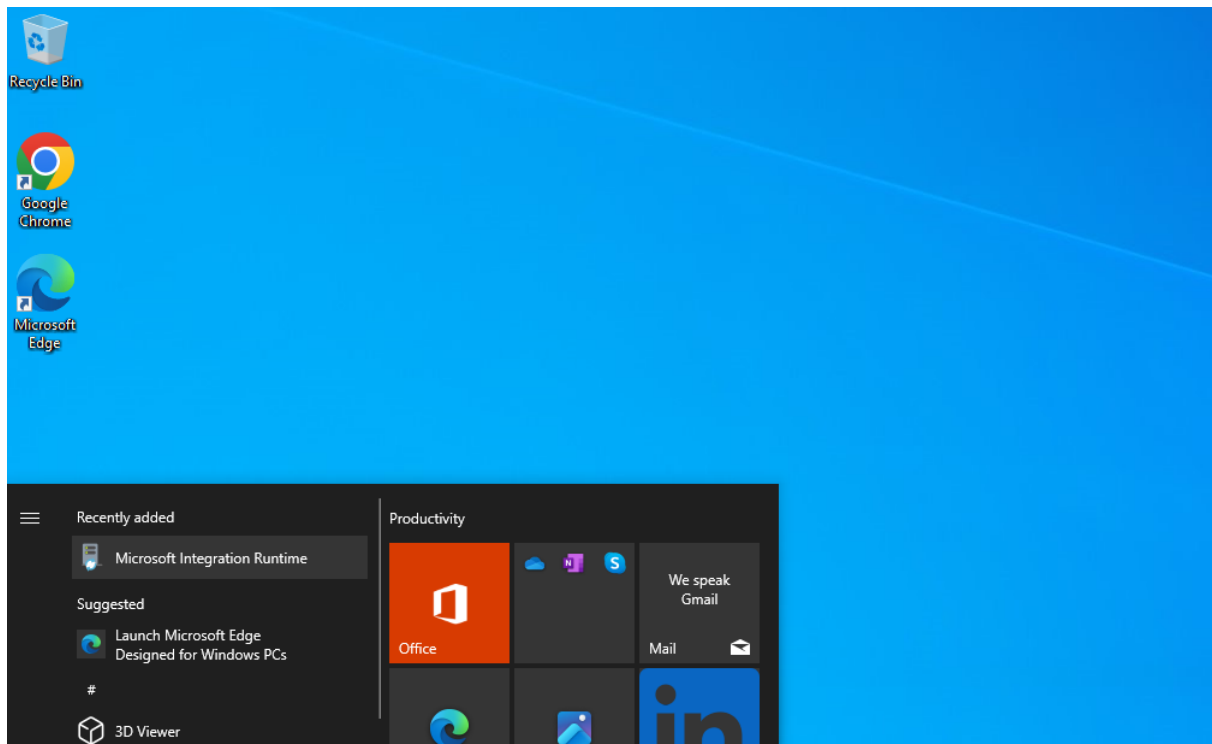
Select language

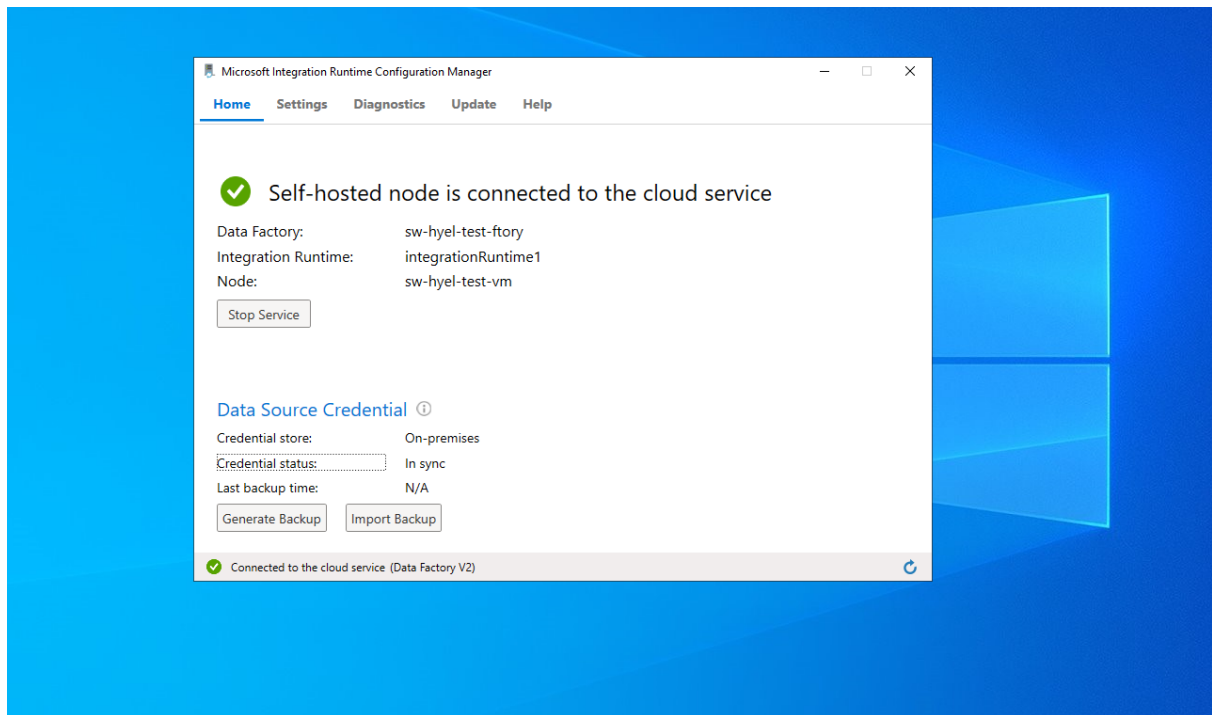
English

Download

상단에 데이터브릭스를 접속하려고 만든 VM에 접속해 다운받은 런타임을 VM에 설치
(하단에 걸어둔 링크의 방법으로 윈도우 가상머신에 파일 옮김)

<https://bosungtea9416.tistory.com/entry/Windows-윈도우-서버에-원격-데스크톱-연결해서-로컬-파일-보내기>





이제 통합런타임 연결에 새로 만든 IR을 선택하면 연결 가능

연결된 서비스 편집


 Azure Databricks [자세한 정보](#) 

이름 *

AzureDatabricks1

설명

통합 런타임을 통해 연결 * ^①

 integrationRuntime1



계정 선택 방법 *

☐ From Azure subscription ☒ Enter manually

Databrick 작업 영역 URL * ^①

https://adb-5243676327643046.6.azuredatabricks.net

인증 형식 *

Access Token



액세스 토큰

Azure Key Vault

액세스 토큰 * ^①

.....

클러스터 선택

☐ 새 작업 클러스터 ☒ 기존 대화형 클러스터 ☐ 기존 인스턴스 풀

기존 클러스터 ID * ^①

0125-063031-lp439167



주석

1 2 3 4 5 6 7 8 9 10