

Microsoft Professional Program

Big Data 과정 소개서



Transforming Education

“How can you prepare for a job and Career in an Era of the fourth industrial revolution?”

Microsoft Professional Program – Big Data

MPP 란 무엇인가요?

4 차산업혁명시대의 기업과 기관에서 필요로 하는 가장 최신 직무의 수행에 도움이 되는 핵심 기술을 가르치는 코스의 모음입니다. 각 코스는 동영상 강의와 퀴즈, 클라우드 기반의 실습 랩을 포함하고 있어 매우 재미있게 학습할 수 있는 온라인 강좌입니다. 또한 각 코스를 우수한 성적으로 마쳤을 경우 이를 입증할 수 있는 마이크로소프트 공식 인증서를 획득할 수 있어 취업/재취업/창업에 큰 도움이 될 수 있습니다.



Microsoft's mission

Empower every person and every organization on the planet to achieve more.

Microsoft wants to **empower every student today to succeed in tomorrow's job market**, regardless of socio-economic status or ability.

– Microsoft Satya Nadella

Microsoft 는 오늘날의 모든 학생들이 사회 경제적 지위 또는 능력에 관계없이 내일의 직업 시장에서 성공할 수 있도록 힘을 북돋고 싶습니다.

왜 마이크로소프트 Big Data 과정을 학습해야 하나요?

데이터가 점점 더 빠르게 수집되고 분석되면서 대용량 데이터 솔루션을 구축 할 수 있는 전문가에 대한 요구가 높습니다. 기초부터 시작하여 가장 핵심적으로 사용되는 빅 데이터 기술을 습득하여, 빅 데이터의 산업계 및 학계 전문가를 대상으로 하는 Microsoft 의 과정에서 직무를 수행 할 기술을 충분히 배웠음을 입증하십시오.

최고의 대학 및 산업계 전문가들과 협력하여 구축 된 Big Data 의 Microsoft 전문 프로그램 인증서는 Azure Cloud 서비스와 Hadoop 및 Spark 와 같은 오픈 소스 시스템을 사용하여 대규모 데이터 솔루션을 구축하는 데 필요한 기술을 학습하는 데 도움을 줍니다.

과정을 통해 학습할 수 있는 기술

- Analyze and Visualize Data
- Work with NoSQL Data
- Query Relational Data
- Implement a Data Warehouse
- Process Big Data at Rest
- Process Big Data in Motion
- Orchestrate Big Data Workflows
- Build Big Data Analysis Solutions

학습방법안내

빅 데이터의 Microsoft 전문 프로그램 인증서는 세 가지 UNIT으로 구성되어 포괄적인 커리큘럼을 제공하며 대용량 데이터에서 경력에 필요한 기술을 가르치는 과정을 제공합니다. 학습자는 각 학습 단위 내에서 각각의 순서대로 학습을 진행하는 것을 제안하지만, 본인의 관심도와 학습 경험에 따라 자유롭게 과정을 선택하여 학습할 수도 있습니다. 또한 온라인 학습은 자기주도적으로 이루어지지만 교육기관에서 제공하는 Guide에 따라 학습 계획을 세울 수도 있습니다.

MS Certification 인증서 취득

전체 9개 Track 14 과목으로 진행되는 Big Data 과정은 학습자가 목표로 하는 진로와 경력에 맞추어 학습을 완료한 이후 인증서를 취득할 수 있습니다. 우수한 성적으로 학습을 완료하면 academy.microsoft.com 과 연동되어 학습자의 학습 정보를 관리하고, 추후 전세계적으로 채용에 활용되고 있는 LinkedIn 과 연동을 통해 개인의 커리어를 취업과 연동할 수도 있습니다.



Big Data Professional Program - Curriculum Overview



Unit 1 – Data Fundamentals

데이터 과학 기초를 배웁니다. 데이터 쿼리, 데이터 분석, 데이터 시각화 및 통계가 데이터 과학에 실제 활용되는 방법과 같은 주제를 학습 하십시오. 수업을 마치려면 Course 2a 또는 Course 2b 중에서 선택하십시오.

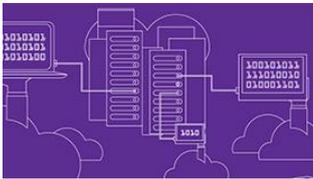
Course 1: Microsoft Professional Program: Introduction to Big Data

Course 2a: Analyzing and Visualizing Data with Power BI

Course 2b: Analyzing and Visualizing Data with Excel

Course 3: Introduction to NoSQL Data Solutions

Course 4: Querying Data with Transact-SQL



Unit 2 – Big Data Processing

이 단원에서는 Azure 서비스 및 오픈 소스 기술을 사용하여 일괄 처리 및 대용량 데이터의 실시간 처리를 구현하는 방법을 학습합니다. Azure HDInsight 에서 Azure 또는 Hadoop 에 초점을 맞춘 경로를 선택하고 코스 6a 및 7a 또는 코스 6b 및 7b 를 선택하여 학습을 진행하십시오.

Course 5: Delivering a Data Warehouse in the Cloud

Course 6a: Processing Big Data with Azure Data Lake Analytics

Course 6b: Processing Big Data with Hadoop in Azure HDInsight

Course 7a: Processing Real-Time Data Streams in Azure

Course 7b: Implementing Real-Time Analytics with Hadoop in Azure HDInsight



Unit 3 – Big Data Analysis Solutions

이 단원에서는 주기적인 데이터 처리 및 이동 작업을 자동화하고 빅 데이터에 예측 분석을 적용하는 workflow 솔루션을 생성하는 방법을 학습합니다. Unit 2 에서 선택한 프로그래밍 언어로 계속 학습을 진행하십시오. 단원을 수료하려면 과정 9a, 과정 9b 또는 과정 9c 를 선택하십시오.

Course 8: Orchestrating Big Data with Azure Data Factory

Course 9a: Developing Big Data Solutions with Azure Machine Learning

Course 9b: Analyzing Big Data with Microsoft R

Course 9c: Implementing Predictive Analytics with Spark in Azure HDInsight

Course 1 : Microsoft Professional Program: Introduction to Big Data



- 예상 학습 기간: 3 주
- 1 주일에 3~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 입문
- 동영상 강의: 영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Big Data](#) 과정에 포함되어 있습니다.

Big Data 분석 솔루션을 구축하는 데 필요한 사항에 대해 알아보십시오.

이 과정은 Microsoft의 Big Data 커리큘럼의 첫 번째 시작 점입니다. 커리큘럼을 시작하고, 학습 일정을 계획하고, 교육 과정을 따라 빅 데이터 시나리오에 대한 데이터 및 몇 가지 기본 개념과 기술 작업에 대해 소개합니다.

무엇을 배우나요?

- How the Microsoft Big Data curriculum works
- An introduction to data formats, technologies, and techniques
- Fundamentals of Databases
- Basic principles for working with Big Data

강의 계획서

- Module 1: Introduction
- Module 2: Data Basics
- Module 3: Fundamentals of Databases
- Module 4: Introduction to Big Data

강사 소개



Graeme Malcolm

Senior Content Developer Microsoft Learning Experiences

Course 2a : Analyzing and Visualizing Data with Power BI



- 예상 학습 기간: 6 주
- 1 주일에 2~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 입문
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Big Data](#) 과정에 포함되어 있습니다.

클라우드 기반으로 제공되는 Power BI 는 기업 내에서 필요로 하는 데이터를 직관적으로 분석하고 분석 결과를 시각화 하여, 이를 목적에 맞게 공유하는 데 큰 인기를 얻고 있습니다.

이 과정에서는 Microsoft 의 Power BI 제품 팀에서 직접 개발한 동영상 강의와 데모, 퀴즈 및 실습을 통해 학습을 진행하게 됩니다. Power BI 를 통해 데이터를 연결하고 가져 오는 방법부터 Power BI Desktop 을 사용하여 보고서를 작성한 다음 Power BI 클라우드 서비스에 게시 할 수 있습니다. 또한 웹 및 모바일 장치에서 사용자를 위한 대시보드를 만들고 공유하는 방법을 배울 수 있습니다.

선행 학습

엑셀, 데이터베이스, 혹은 텍스트 기반의 데이터 작업 경험

강의 계획서

1 주

- 비즈니스 인텔리전스, 데이터 분석 및 데이터 시각화의 주요 개념 이해
- Marketo, Salesforce 및 Google Analytics 와 같은 서비스에서 데이터 가져 오기 및 대시보드 자동 생성
- 데이터 연결 및 가져오기를 수행 한 다음 해당 데이터를 구성하고 변환
- 비즈니스 계산을 통한 데이터의 가치 증대

2 주

- 데이터 시각화 및 보고서 작성
- 업데이트 일정에 따라 자동 보고서 새로 고침
- 보고서 및 자연언어 쿼리를 기반으로 대시보드 만들기
- 조직 전체에 대시보드 공유
- 모바일 앱에서 대시보드 활용

3 주

- Power BI 내에서 엑셀 보고서 활용
- 대시보드 및 보고서에서 사용할 수 있는 사용자 지정 시각화 만들기
- 보고서 및 대시보드 작성을 위해 그룹 내에서 협업
- 조직의 필요에 따라 대시보드를 효과적으로 공유

4 주

- Power BI 를 사용하여 데이터에 대한 실시간 연결 탐색
- SQL Azure, HD Spark 및 SQL Server Analysis Services 에 직접 연결
- Power BI Development API 소개
- Power BI 의 사용자 지정 비주얼 활용 방법

무엇을 배우나요?

- Connect, import, shape, and transform data for business intelligence (BI)
- Visualize data, author reports, and schedule automated refresh of your reports
- Create and share dashboards based on reports in Power BI desktop and Excel
- Use natural language queries
- Create real-time dashboards

강사 소개



Will Thompson

Program Manager, Power BI
Microsoft



Kim Manis

Program Manager Microsoft



Jonathan Sanito

Senior Content Developer
Microsoft



Miguel Llopis

Senior Program Manager
Microsoft

Course 2b : Analyzing and Visualizing Data with Excel



- 예상 학습 기간: 6 주
- 1 주일에 2~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Business & Management
- 레벨: 중급
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Big Data](#) 과정에 포함되어 있습니다.

엑셀은 데이터를 분석하고 시각화 하는데 가장 널리 사용되는 솔루션 중 하나입니다. 향상된 시각화 기능과 보다 정교한 비즈니스 로직을 반영하여 더 많은 데이터를 분석 할 수 있는 방법을 학습하게 됩니다. 이 데이터 사이언스 과정에서는 Microsoft의 Excel Product Team 전문가가 Excel 2016에서의 최신 기능을 소개합니다.

서로 다른 소스에서 데이터를 가져 오는 방법, 데이터 원본 간에 매시업을 만드는 방법, 분석을 위해 데이터를 준비하는 방법에 대해 학습합니다. 데이터를 준비한 후 DAX 계산 엔진을 사용하여 비즈니스 결과를 표현하는 방법을 알아보십시오. Power BI 클라우드 서비스를 활용하여 데이터를 시각화 하고 공유 할 수 있는 방법을 확인한 후, 대시보드에서 사용하거나 또는 일반 영어로 된 명령문을 사용하여 쿼리할 수 있으며, 모바일 장치에서도 사용할 수 있습니다.

선행 학습

테이블, 피벗 테이블 및 피벗 차트와 같은 엑셀 분석 도구에 대한 이해 또한 DB의 데이터 및 텍스트 파일 작업에 대한 선행 지식이 도움이 됩니다. (DAT205x: Introduction to Data Analysis using Excel)

강의 계획서

1 주

Office 응용 프로그램을 설치하여 실습 환경을 설정하십시오. 이미 워크 시트 / 그리드 데이터에 준비되어 있는 데이터에 대해 피벗 테이블, 피벗 차트 및 슬라이서와 같은 도구를 사용하여 엑셀에서 데이터 분석을 수행하는 방법을 학습하십시오. Power Pivot 추가 기능을 사용하여 엑셀 데이터 모델, 콘텐츠 및 구조를 탐색하십시오. 또한 컬럼의 데이터를 측정하기 위한 DAX(Data Analysis Expressions)에 대해서 배웁니다.

2 주

쿼리(Excel 2013 및 Excel 2010의 Power Query 추가 기능)에 대해 알아보고 단일 플랫폼 테이블에서

엑셀 데이터 모델을 작성하십시오. SQL 데이터베이스에서 여러 테이블을 가져 오는 방법과 가져온 데이터에서 엑셀 데이터 모델을 만드는 방법에 대해 알아보십시오. 텍스트 파일의 데이터와 SQL 데이터베이스의 데이터 사이에 매시업을 만드는 방법을 학습합니다.

3 주

각 셀에 대해 계산할 측정 값을 만드는 방법, 계산을 위해 컨텍스트를 필터링하는 방법 및 여러 가지 고급 DAX 함수를 탐색하는 방법에 대한 세부 내용을 학습할 수 있습니다. 고급 텍스트 쿼리를 사용하여 서식이 지정된 엑셀 보고서에서 데이터를 가져 오는 방법을 배우십시오. 또한 표준 사용자 인터페이스 이상으로 난이도가 높은 쿼리를 배웁니다.

4 주

Excel 에서 멋진 시각화를 만드는 방법을 살펴보십시오. 큐브 함수를 사용하여 전년 대비 비교를 수행하게 됩니다. 타임라인, 계층구조 및 슬라이서를 만들어 시각화를 보다 향상시킵니다. 엑셀이 Power BI 와 함께 어떻게 작동하는지 알아볼 수 있습니다. 엑셀 통합 문서를 Power BI 서비스에 업로드하는 방법을 배웁니다. 마지막으로 모바일 플랫폼에서 엑셀 사용법을 살펴보게 됩니다.

무엇을 배우나요?

- Gather and transform data from multiple sources
- Discover and combine data in mashups
- Learn about data model creation
- Explore, analyze, and visualize data

강사 소개



Dany Hoter

Solutions Architect DataRails
Microsoft



Jonathan Sanito

Senior Content Developer
Microsoft

Course 3 : Introduction to NoSQL Data Solutions



- 예상 학습 기간: 3 주
- 1 주일에 2~3 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Big Data](#) 과정에 포함되어 있습니다.

실시간 분석, 개인화된 포털사이트, IoT 및 모바일 애플리케이션과 관련된 일부 시나리오는 관계형 데이터베이스보다 NoSQL 스토리지 및 컴퓨팅 솔루션으로 더 잘 처리될 수 있습니다. Microsoft Azure에는 여러 가지 NoSQL (또는 "Not Only SQL") 비 관계형 데이터 저장소 옵션이 있습니다. NoSQL 데이터베이스는 일반적으로 여러 서버에 배포되고 분할되도록 구축됩니다. 또한 높은 가용성을 위해 수평 확장하고 반 구조화된 데이터와 구조화 되지 않은 데이터를 처리 할 수 있을 만큼 유연하게 설계되었습니다. 끊임없이 진화하는 데이터 모델을 가지고 있고 빠르게 움직이기를 원한다면, 이것은 바로 이 데이터베이스들이 가지고 있는 특징입니다.

실습 과정과 평가 및 최종 시험을 통해 학습 과정에서 NoSQL이 시간이 지남에 따라 어떻게 발전해 왔는지 전문가를 통해 배울 수 있습니다. Azure에서 비 관계형 데이터 저장소 옵션을 탐색하고 응용 프로그램에서 이를 사용하는 방법을 살펴보십시오. 다양한 스토리지 옵션으로 데이터를 생성, 저장, 관리 및 액세스하는 방법에 대해 알아보십시오. 또한 Azure Table Storage, DocumentDB, MongoDB 등을 자세히 살펴보십시오. 다양성 (신속하게 진화하는 스키마 또는 시나리오), 볼륨 (데이터 저장 측면에서의 확장) 및 속도 (많은 사용자를 지원해야 하는 처리량)에 대해 자세히 알아보십시오. Azure에서 NoSQL 옵션으로 직접 학습하며 실습할 수 있습니다.

선행 학습

- 관계형 데이터베이스 기본 사항
- T-SQL 쿼리
- HTTP API 및 Request에 대한 기본적인 이해

무엇을 배우나요?

- NoSQL 기본 사항
- Microsoft Azure의 NoSQL 옵션
- DocumentDB, Azure 테이블 저장소 및 MongoDB 사용을 위한 핵심 기술
- NoSQL 스토리지의 성능에 액세스하고 성능을 향상시키는 기타 기술

강사 소개



Pete Harris

Senior Content Developer

Microsoft's Learning
Experiences



Sidney Andrews

Microsoft Certified Trainer,
Cloud Applications Consultant

SeeSharpRun.NET



Andrew Liu

Program Manager

Microsoft



Micheleen Harris

Data Scientist

Microsoft

Course 4 : Querying Data with Transact-SQL



- 예상 학습 기간 : 6 주
- 1 주일에 4~5 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 한글/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Big Data](#) 과정에 포함되어 있습니다.

트랜잭트 SQL 은 데이터베이스로 작업을 하는 데이터 전문가 및 개발자에게 필수적인 기술입니다. 전문가의 동영상 강의, DEMO, 실습을 통해 가장 처음 SELECT 구문에서부터 트랜잭션 방식의 다양한 논리 구현을 학습합니다. 여러 개의 코스 모듈을 학습하며 Microsoft SQL Server 또는 Azure SQL Database 의 데이터 쿼리 및 업데이트 등 트랜잭트 SQL 의 주요 영역을 경험할 수 있습니다. 특히 이 과정은 실습환경을 제공하고 있으며 Azure Cloud 기반의 데이터베이스를 쉽게 배포할 수 있는 샘플 데이터베이스를 사용하므로 SQL 을 설치 및 구성하지 않아도 트랜잭트 SQL 을 학습할 수 있습니다.

선행 학습

데이터베이스 및 IT 시스템에 대한 기본적인 이해

무엇을 배우나요?

- Create Transact-SQL SELECT queries
- Work with data types and NULL
- Query multiple tables with JOIN
- Explore set operators
- Use functions and aggregate data
- Work with subqueries and APPLY
- Use table expressions
- Group sets and pivot data
- Modify data
- Program with Transact-SQL
- Implement error handling and transactions

강사 소개



Geoff Allix
Microsoft Certified IT
Professional for SQL Server
Content Master



Graeme Malcolm
Senior Content Developer
Microsoft Learning
Experiences

Course 5 : Delivering a Data Warehouse in the Cloud



- 예상 학습 기간: 6 주
- 1 주일에 2~3 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Big Data](#) 과정에 포함되어 있습니다.

데이터 웨어하우스의 스토리지 및 처리 기능을 수개월이 아닌 수 분 내에 확장해야하는 경우 클라우드 기반 대규모 병렬 처리 솔루션이 필요합니다.

이번 컴퓨터 사이언스 과정에서는 Microsoft 의 Azure SQL Data Warehouse 또는 SQL DW 를 사용하여 데이터를 배포, 디자인, 데이터 로드 하는 방법을 배우게 됩니다. 데이터 분포, 압축 메모리 내 인덱스, Big Data 를 위한 PolyBase 및 elastic scale 에 대해 학습합니다.

참고 : 이 과정의 실습 요소를 완료하려면 Azure 구독이 필요합니다. 무료 Azure 평가판 가입을 신청할 수 있습니다 (확인을 위해서는 유효한 신용 카드가 필요하지만 Azure 서비스에 대해서는 비용이 청구되지 않습니다). 일부 지역에서는 무료 평가판을 사용할 수 없습니다.

선행 학습

- 데이터베이스 개념 및 사용자의 리포팅과 분석에 필요한 SQL 쿼리 구문에 대한 이해도
- 기술적 문제를 해결할 때 적극적이고 인내심을 갖고 학습 할 의지가 필수

강의 계획서

모듈 1 : SQL Data Warehouse 소개

이 모듈에서는 Microsoft Cloud 의 Data Warehouse 인 Azure SQL Datawarehouse 를 소개합니다. 대용량 병렬 처리, SQL DW 를 프로비저닝 하고 구성하는 방법에 대해 배웁니다.

모듈 2 : Data Warehouse 설계 및 쿼리

이 모듈은 테이블 디자인, 분할, 인덱스 및 통계를 다룹니다. 그것은 신속성 있는 쿼리와 쿼리 모니터링을 위한 도구를 소개합니다.

모듈 3 : Integrating and Ingesting Data

이 모듈은 Azure Data Factory, Polybase 및 Azure Stream Analytics 로 데이터를 SQL DW 로 로드하는 방법을 다룹니다. 또한 Azure Machine Learning 과 통합하고 Power BI 로 데이터를 시각화 합니다.

모듈 4 : Data Warehouse 관리

이 모듈에서는 SQL DW 작업 부하 및 성능의 모니터링 및 관리, 보안, 확장 및 백업 관리에 대해 다룹니다.

최종 시험

최종 시험은 성적의 30 %를 차지하며 주간 퀴즈와 함께 종합 점수를 결정합니다. 이 과정을 수료하고 인증서를 취득하려면 총점 70 % 이상을 획득해야 합니다.

무엇을 배우나요?

이 과정에서는 다음과 같은 이론과 기법을 배우게 됩니다.

- 클라우드 기반 Data Warehouse 를 위한 대규모 병렬 처리 아키텍처 선택
- 많은 노드에서 테이블의 데이터를 효율적으로 배포하기 위해 테이블과 인덱스를 설계
- 다양한 Source 에서 데이터를 로드하고, PolyBase 를 사용하여 쿼리하고, 데이터 보안 및 복구를 수행하고, 빅 데이터 환경에 통합

강사 소개



Theresa Iserman
SQL Server Premier Field
Engineer
Microsoft



Scott Klein
CTO
Cloud and Device

Course 6a : Processing Big Data with Azure Data Lake Analytics



- 예상 학습 기간: 4 주
- 1 주일에 3~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 고급
- 동영상 강의: 영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Big Data](#) 과정에 포함되어 있습니다.

규모에 맞게 데이터를 저장하고 처리하고 싶습니까? 이 데이터 분석 과정은 Azure Data Lake 기술을 사용하여 Azure 클라우드의 성능을 빅 데이터에 적용하는 방법을 설명합니다.

Azure Data Lake Store 에서 데이터를 관리하고 Azure Data Lake Analytics 에서 U-SQL 작업을 실행하여 구조화 된 데이터 소스와 구조화되지 않은 데이터 소스로부터 통찰력을 얻는 방법을 익히십시오.

참고 :이 과정을 완료하려면 Microsoft Azure 구독이 필요합니다. <http://azure.microsoft.com> 에서 무료 평가판 구독을 신청하거나 기존 구독을 사용할 수 있습니다. 이 학습과정에서는 실습 활동을 완료하는데 필요한 자원 비용을 무료(최소화)로 이용하도록 설계되었습니다.

강의 계획서

- Module 1: Getting Started with Azure Data Lake Analytics
- Module 2: Using a U-SQL Catalog
- Module 3: using C# Functions in U-SQL
- Module 4: Monitoring and Optimizing U-SQL Jobs

무엇을 배우나요?

- Azure Data Lake 기술은 U-SQL 작업을 사용하여 데이터를 저장하고 처리함
- U-SQL 카탈로그 개체 생성 및 사용 방법
- 사용자 지정 C # 코드로 데이터 처리 스크립트 확장
- U-SQL 작업 모니터링 및 최적화

강사 소개



Graeme Malcolm

Senior Content Developer Microsoft Learning Experiences

Course 6b : Processing Big Data with Hadoop in Azure HDInsight



- 예상 학습 기간 : 5 주
- 1 주일에 3~5 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Big Data](#) 과정에 포함되어 있습니다.

점점 더 많은 산업 분야에서 빅 데이터를 분석해야하는 과제에 직면해 있습니다. 이 과정에서는 Microsoft Azure HDInsight 의 Hadoop 기술을 사용하여 분석을 위해 데이터를 정제하고 재구성하는 일괄 처리 솔루션을 구축하는 방법을 설명합니다. 이 5 주 과정에서는 HDInsight 에서 Hadoop 과 함께 Hive, Pig, Oozie 및 Sqoop 과 같은 기술을 사용하는 방법을 배우게 됩니다. Windows, Linux 및 Mac OSX 클라이언트 컴퓨터에서 HDInsight 클러스터로 작업하는 방법에 대해 설명합니다.

선행 학습

- 데이터베이스 개념 및 기본 SQL 쿼리 구문 이해도
- 프로그래밍 기초에 대한 지식
- 적극적이고 인내심을 갖고 자발적으로 배울 의향

무엇을 배우나요?

- Provision an HDInsight cluster.
- Connect to an HDInsight cluster, upload data, and run MapReduce jobs.
- Use Hive to store and process data.
- Process data using Pig.
- Use custom Python user-defined functions from Hive and Pig.
- Define and run workflows for data processing using Oozie.
- Transfer data between HDInsight and databases using Sqoop.

강사 소개



Graeme Malcolm

Senior Content Developer Microsoft Learning Experiences

Course 7a : Processing Real-Time Data Streams in Azure



- 예상 학습 기간: 4 주
- 1 주일에 3~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 고급
- 동영상 강의: 한글(7 월)/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Big Data](#) 과정에 포함되어 있습니다.

클라우드에서 실시간 데이터를 캡처하고 처리하고 싶습니까?

이 데이터 분석 과정에서는 Event Hubs, IoT Hubs 및 Stream Analytics 와 같은 Microsoft Azure 기술을 사용하여 실시간으로 IoT (Internet-of-Things) 솔루션을 대규모로 구축하는 방법을 배웁니다.

참고 : 이 과정을 완료하려면 Microsoft Azure 구독이 필요합니다. <http://azure.microsoft.com> 에서 무료 평가판 구독을 신청하거나 기존 구독을 사용할 수 있습니다.

강의 계획서

- Module 1: Getting Started with Azure Event Hubs and IoT Hubs
- Module 2: Using Azure Stream Analytics
- Module 3: Aggregating Data in Temporal Windows
- Module 4: Monitoring a Streaming Solution

무엇을 배우나요?

- Capturing real-time data with Azure Event Hubs and IoT Hubs
- Processing real-time data with Azure Stream Analytics
- Aggregating data in temporal windows
- Monitoring streaming solutions in Azure

강사 소개



Graeme Malcolm

Senior Content Developer Microsoft Learning Experiences

Course 7b : Implementing Real-Time Analytics with Hadoop in Azure HDInsight



- 예상 학습 기간: 4 주
- 1 주일에 2~3 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 중급
- 동영상 강의: 영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Big Data](#) 과정에 포함되어 있습니다.

이 4 주 과정에서는 Microsoft Azure HDInsight 에서 HBase, Storm 및 Spark 와 같은 Hadoop 기술을 사용하여 대기 시간이 짧고 스트리밍하는 Big Data 솔루션을 구현하는 방법을 배우게 됩니다.

참고 : 이 과정의 실습 요소를 완료하려면 Azure 구독 및 Windows, Linux 또는 Mac OS X 클라이언트 컴퓨터가 필요합니다. 무료 Azure 평가판 가입을 신청할 수 있습니다 (확인을 위해서는 유효한 신용 카드가 필요하지만 Azure 서비스에 대해서는 비용이 청구되지 않습니다). 일부 지역에서는 무료 평가판을 사용할 수 없습니다.

이 과정은 HDInsight 를 사용하여 대용량 데이터와 고급 분석 기술을 탐구하는 두 번째 시리즈입니다. DAT202.1x : Processing Big Data with Hadoop in Azure HDInsight 에서 배운 일괄 처리 기술을 기반으로 합니다

선행 학습

- HDInsight 에서 Hadoop 클러스터 및 Hive 에 익숙함
- 데이터베이스 개념 및 기본 SQL 쿼리 구문 이해도
- 기본 프로그래밍 구조 (예 : 변수, 루프, 조건부 논리)에 익숙함, Java 또는 C # 사용 경험은 유용하지만 필수적인 요소는 아님
- 기술 문제를 해결할 때 적극적으로 배우고 자발적으로 배우려는 의지가 필수적

강의 계획서

Module 1: Using HBase for NoSQL Data

Module 2: Using Storm for Streaming Data

Module 3: Using Spark for Interactive Analysis

Module 4: Final Exam

무엇을 배우나요?

- HBase to implement low-latency NoSQL data stores.
- Storm to implement real-time streaming analytics solutions.
- Spark for high-performance interactive data analysis.

강사 소개



Graeme Malcolm

Senior Content Developer Microsoft Learning Experiences

Course 8 : Orchestrating Big Data with Azure Data Factory



- 예상 학습 기간: 4 주
- 1 주일에 3~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 고급
- 동영상 강의: 영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Big Data](#) 과정에 포함되어 있습니다.

빅 데이터 워크 플로우를 예약하고 관리해야 합니까?

이 데이터 분석 과정에서는 Azure Data Factory 를 사용하여 Hadoop, SQL 및 Azure Data Lake Analytics 와 같은 기술을 사용하여 데이터 이동 및 변환을 조정하는 방법을 설명합니다. 특정 작업을 수행하기 위해 활동을 그룹화 할 수 있는 데이터 파이프 라인을 작성하는 방법을 학습합니다.

참고 : 이 과정을 완료하려면 Microsoft Azure 구독이 필요합니다. <http://azure.microsoft.com> 에서 무료 평가판 구독을 신청하거나 기존 구독을 사용할 수 있습니다. 연구소는 실습 활동을 완료하는 데 필요한 자원 비용을 최소화하도록 설계되었습니다.

강의 계획서

Module 1: Getting Started with Azure Data Factory

Module 2: Scheduling Pipelines

Module 3: Transforming Data in Pipelines

무엇을 배우나요?

- Creating data workflows with Azure Data Factory
- Scheduling data pipelines to orchestrate big data processes
- Applying data transformations in a pipeline with Hive or U-SQL

강사 소개



Graeme Malcolm

Senior Content Developer Microsoft Learning Experiences

Course 9a : Developing Big Data Solutions with Azure Machine Learning



- 예상 학습 기간: 4 주
- 1 주일에 3~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 고급
- 동영상 강의: 영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Big Data](#) 과정에 포함되어 있습니다. 과거는 종종 미래를 예측하는 열쇠가 될 수 있습니다. 역사적 자료의 빅 데이터는 과거 동향을 파악하고 통계 패턴을 적용하여 미래 결과를 예측하는 기계 학습 모델을 구축하는 데 유용한 리소스입니다.

이 과정에서는 Azure Machine Learning 을 소개하고 이를 사용하여 빅 데이터 소스에서 모델을 작성하고 예측 통찰력을 빅 데이터 처리 워크 플로우에 통합하는 데 필요한 기술과 고려 사항을 학습합니다.

선행 학습

- Azure Data Factory 로 데이터 처리 파이프 라인 구축
- Azure Stream Analytics 로 실시간 데이터 처리 솔루션 구축

무엇을 배우나요?

- Azure Machine Learning 을 사용하여 예측 가능한 웹 서비스를 생성하는 방법
- Azure Machine Learning 에서 큰 데이터 소스로 작업하는 방법
- Azure Machine Learning 을 큰 데이터 일괄 처리 파이프 라인에 통합하는 방법
- Azure Machine Learning 을 실시간 대형 데이터 처리 솔루션에 통합하는 방법

강사 소개



Graeme Malcolm

Senior Content Developer
Microsoft Learning
Experiences



Steve Elston

Managing Director Quantia
Analytics, LLC

Course 9b : Analyzing Big Data with Microsoft R



- 예상 학습 기간: 4 주
- 1 주일에 2~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Data Analysis & Statistics
- 레벨: 고급
- 동영상 강의: 한글(8 월)/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Big Data](#) 과정에 포함되어 있습니다.

오픈 소스 프로그래밍 언어 R은 데이터 처리 및 통계 분석을 위해 오랜 기간(특히 학계에서) 인기 있는 언어입니다. R의 장점 중 하나는 간결한 프로그래밍 언어이며 또한 모든 종류의 분석을 수행 할 수 있는 제 3자 라이브러리의 광범위한 저장소가 있다는 것입니다. 이 두 기능을 함께 사용하면 데이터 데이터 사이언티스트가 원시 데이터에서 요약, 차트 및 전체 보고서로 신속하게 만들어 낼 수 있습니다. 그러나 R의 한 가지 부족한 점은 전통적으로 많은 양의 메모리를 사용한다는 것입니다. 왜냐하면 데이터 복사본을 data.frame 개체로 전체적으로 로드해야하기 때문에 데이터를 처리할 때 종종 추가 복사본을 생성하기 때문입니다(copy-on-modify 라고 함). 이것은 R이 학계에 비해 산업계에 더 마지 못해 받아 들여진 이유 중 하나입니다.

Microsoft R Server (MRS)의 주요 구성 요소는 R 라이브러리에있는 RevoScaleR 패키지입니다. 이 라이브러리는 대규모 데이터 세트를 메모리에 한꺼번에 로드하지 않고도 처리 할 수 있는 일련의 기능을 제공합니다. RevoScaleR은 시간이 지남에 따라 추가되는 다양한 통계 및 머신 러닝 알고리즘을 공유합니다. 마지막으로, RevoScaleR은 우리가 랩톱에서 개발한 코드를 가져 와서 최소한의 노력으로 SQL Server 또는 스파크(인프라가 전혀 다른 곳)에 원격 배포 할 수 있는 메커니즘을 제공합니다. 이 과정에서는 MRS를 사용하여 대규모 데이터 집합에 대한 분석을 실행하고 SQL Server 데이터베이스에 이를 배포하는 방법에 대한 몇 가지 예를 제공합니다. 이 과정을 완료하면 빅데이터 처리를 위해 R을 사용하는 방법을 알게 됩니다.

RevoScaleR은 R 패키지이므로 과정 학습자는 R에 익숙하다고 가정합니다. R 데이터 구조 (벡터, 행렬, 목록, 데이터 프레임, 환경)에 대한 확실한 이해가 필요합니다. dplyr 과 같은 타사 패키지에 익숙한 것도 학습에 도움이 됩니다

선행 학습

- 데이터 사이언스를 위한 R 입문 및 데이터 사이언스를 위한 R 프로그래밍

무엇을 배우나요?

MRS(Microsoft R Server)를 사용하여 대형 datasets 을 읽고, 처리하고, 분석하는 방법을 배우게 됩니다.

- Read data from flat files into R's data frame object, investigate the structure of the dataset and make corrections, and store prepared datasets for later use
- Prepare and transform the data
- Calculate essential summary statistics, do crosstabulation, write your own summary functions, and visualize data with the ggplot2 package
- Build predictive models, evaluate and compare models, and generate predictions on new data

강사 소개



Jonathan Sanito

Senior Content Developer
Microsoft



Seth Mottaghinejad

Data Scientist Microsoft

Course 9c : Implementing Predictive Analytics with Spark in Azure HDInsight



- 예상 학습 기간: 6 주
- 1 주일에 3~4 시간 학습 권고
- 주관기관: Microsoft
- 주제: Computer Science
- 레벨: 고급
- 동영상 강의: 한글(7 월)/영어

과정 소개

이 과정은 [Microsoft Professional Program Certificate in Data Science 및 Big Data](#) 과정에 포함되어 있습니다.

빅 데이터 과학에 대한 학습 준비가 되셨습니까? 이 과정에서는 Microsoft Azure HDInsight 에서 아파치 스파크를 사용하여 대용량 데이터에 대한 예측 분석 솔루션을 구현하는 방법을 학습합니다. 스칼라 또는 파이썬을 사용하여 데이터를 정리 및 변환하고 Spark ML 로 머신 러닝 모델을 작성하는 방법을 배우보십시오.

참고: 이 과정의 실습을 진행하려면 Azure 구독과 Windows 클라이언트 컴퓨터가 필요합니다. 무료 Azure 평가 판 가입을 신청할 수 있습니다 (확인을 위해서는 유효한 신용 카드가 필요하지만 Azure 서비스에 대해서는 비용이 청구되지 않습니다).

선행 학습

- Azure HDInsight 경험
- 데이터베이스 및 SQL 에 대한 지식.
- 프로그래밍 경험.

강의 계획서

Spark를 활용한 데이터 사이언스 개론

Azure HDInsight에서 스파크 클러스터를 시작하고 스파크를 사용하여 파이썬 또는 스칼라를 실행하여 데이터 작업을 수행하십시오.

머신 러닝 시작하기

Spark ML 라이브러리를 사용하여 분류 및 회귀 모델을 작성하는 방법을 배웁니다.

머신 러닝 모델 평가

지도학습 모델을 평가하는 방법과 모델 매개 변수를 최적화하는 방법에 대해 알아보십시오.

추천자 및 비지도 학습 모델

Spark ML을 사용하여 추천 구축 및 클러스터링 모델을 작성하는 방법을 익히십시오

무엇을 배우나요?

- Using Spark to explore data and prepare for modeling
- Build supervised machine learning models
- Evaluate and optimize models
- Build recommenders and unsupervised machine learning models

강사 소개



Graeme Malcolm

Senior Content Developer

Microsoft Learning

Experiences