



Curso Oficial da Microsoft



*Data Engineering on Microsoft
Azure (DP-203)*



Data Engineering on Microsoft Azure (DP-203)

Curso oficial da Microsoft

Neste curso, o aluno aprenderá sobre a engenharia de dados no que se refere ao trabalho com soluções analíticas em lote e em tempo real usando tecnologias de plataforma de dados Azure. Os alunos começarão por entender as tecnologias básicas de computação e armazenamento usadas para construir uma solução analítica. Os alunos aprenderão como explorar interativamente dados armazenados em arquivos em um data lake. Eles aprenderão as várias técnicas de ingestão que podem ser usadas para carregar dados usando o recurso Apache Spark encontrado no Azure Synapse Analytics ou Azure Databricks, ou como fazer a ingestão usando o Azure Data Factory ou os pipelines do Azure Synapse.

Os alunos também aprenderão as várias maneiras de transformar os dados usando as mesmas tecnologias usadas para ingerir dados. Eles compreenderão a importância de implementar segurança para garantir que os dados sejam protegidos em repouso ou em trânsito. O aluno então mostrará como criar um sistema analítico em tempo real para criar soluções analíticas em tempo real.

Carga horária: 4 dias / 8 noites.

SOLUTION

WWW.KASOLUTION.COM.BR



Módulo 1: Explorar opções de computação e armazenamento para cargas de trabalho de engenharia de dados

Este módulo fornece uma visão geral das opções de tecnologia de computação e armazenamento do Azure que estão disponíveis para engenheiros de dados que criam cargas de trabalho analíticas. Este módulo ensina maneiras de estruturar o data lake e de otimizar os arquivos para exploração, streaming e cargas de trabalho em lote. O aluno aprenderá como organizar o data lake em níveis de refinamento de dados à medida que transformam arquivos por meio de processamento em lote e fluxo. Em seguida, eles aprenderão como criar índices em seus conjuntos de dados, como arquivos CSV, JSON e Parquet, e como usá-los para consulta potencial e aceleração de carga de trabalho.



Lições

- Introdução ao Azure Synapse Analytics
- Descrever Azure Databricks
- Introdução ao armazenamento Azure Data Lake
- Descreva a arquitetura do Delta Lake
- Trabalhe com fluxos de dados usando o Azure Stream Analytics



Laboratório: Explorar opções de computação e armazenamento para cargas de trabalho de engenharia de dados

- Combine streaming e processamento em lote com um único pipeline
- Organize o data lake em níveis de transformação de arquivo
- Índice de armazenamento do data lake para consulta e aceleração da carga de trabalho



Depois de concluir este módulo, os alunos serão capazes de:

- Descrever o Azure Synapse Analytics
- Descrever Azure Databricks
- Descrever o armazenamento do Azure Data Lake
- Descrever a arquitetura do Delta Lake
- Descrever Azure Stream Analytics



Módulo 2: Executar consultas interativas usando pools SQL sem servidor do Azure Synapse Analytics

Neste módulo, os alunos aprenderão como trabalhar com arquivos armazenados no data lake e em fontes externas de arquivos, por meio de instruções T-SQL executadas por um pool de SQL sem servidor no Azure Synapse Analytics. Os alunos consultarão arquivos Parquet armazenados em um data lake, bem como arquivos CSV armazenados em um armazenamento de dados externo. Em seguida, eles criarão grupos de segurança do Azure Active Directory e imporão o acesso aos arquivos no data lake por meio do Controle de Acesso Baseado em Função (RBAC) e Listas de Controle de Acesso (ACLs).



Lições

- Explorar os recursos de pools SQL sem servidor do Azure Synapse
- Consultar dados no pool usando pools SQL sem servidor do Azure Synapse
- Criar objetos de metadados em pools SQL sem servidor do Azure Synapse
- Proteja dados e gerencie usuários em pools SQL sem servidor do Azure Synapse



Laboratório: Executar consultas interativas usando pools SQL sem servidor

- Consultar dados Parquet com pools SQL sem servidor
- Crie tabelas externas para arquivos Parquet e CSV
- Crie visualizações com pools SQL sem servidor
- Acesso seguro aos dados em um data lake ao usar pools SQL sem servidor
- Configure a segurança do data lake usando Role-Based Access Control (RBAC) e Access Control List



Depois de concluir este módulo, os alunos serão capazes de:

- Compreender os recursos de pools SQL sem servidor do Azure Synapse
- Consultar dados no lake usando pools SQL sem servidor do Azure Synapse
- Criar objetos de metadados em pools SQL sem servidor do Azure Synapse
- Proteger dados e gerenciar usuários em pools SQL sem servidor do Azure Synapse



Módulo 3: Exploração e transformação de dados no Azure Databricks

Este módulo ensina como usar vários métodos Apache Spark DataFrame para explorar e transformar dados em Azure Databricks. O aluno aprenderá como executar métodos DataFrame padrão para explorar e transformar dados. Eles também aprenderão como executar tarefas mais avançadas, como remover dados duplicados, manipular valores de data / hora, renomear colunas e agregar dados.



Lições

- Descrever Azure Databricks
- Ler e gravar dados no Azure Databricks
- Trabalhar com DataFrames em Azure Databricks
- Trabalhe com métodos avançados de DataFrames em Azure Databricks



Laboratório: Exploração e transformação de dados no Azure Databricks

- Use DataFrames no Azure Databricks para explorar e filtrar dados
- Armazene um DataFrame em cache para consultas subsequentes mais rápidas
- Remova os dados duplicados
- Manipular valores de data / hora
- Remover e renomear colunas DataFrame
- Agregar dados armazenados em um DataFrame



Depois de concluir este módulo, os alunos serão capazes de:

- Descrever Azure Databricks
- Ler e gravar dados no Azure Databricks
- Trabalhar com DataFrames em Azure Databricks
- Trabalhar com métodos avançados de DataFrames em Azure Databricks



Módulo 4: Explorar, transformar e carregar dados no Data Warehouse usando Apache Spark

Este módulo ensina como explorar dados armazenados em um data lake, transformar os dados e carregar dados em um armazenamento de dados relacional. O aluno explorará arquivos Parquet e JSON e usará técnicas para consultar e transformar arquivos JSON com estruturas hierárquicas. Em seguida, o aluno usará o Apache Spark para carregar dados no data warehouse e juntar os dados Parquet no data lake com os dados no pool SQL dedicado.



Lições

- Compreenda a engenharia de Big Data com Apache Spark no Azure Synapse Analytics
- Ingerir dados com blocos de anotações Apache Spark no Azure Synapse Analytics
- Transforme dados com DataFrames em Apache Spark Pools no Azure Synapse Analytics
- Integrar pools SQL e Apache Spark no Azure Synapse Analytics



Laboratório: Explorar, transformar e carregar dados no Data Warehouse usando Apache Spark

- Realizar exploração de dados no Synapse Studio
- Ingerir dados com blocos de anotações Spark no Azure Synapse Analytics
- Transforme dados com DataFrames em pools Spark no Azure Synapse Analytics
- Integrar pools SQL e Spark no Azure Synapse Analytics



Depois de concluir este módulo, os alunos serão capazes de:

- Descrever a engenharia de Big Data com Apache Spark no Azure Synapse Analytics
- Ingerir dados com blocos de anotações Apache Spark no Azure Synapse Analytics
- Transformar dados com DataFrames em Apache Spark Pools no Azure Synapse Analytics
- Integrar pools SQL e Apache Spark no Azure Synapse Analytics



Módulo 5: ingerir e carregar dados no data warehouse

Este módulo ensina aos alunos como ingerir dados no data warehouse por meio de scripts T-SQL e pipelines de integração do Synapse Analytics. O aluno aprenderá como carregar dados em pools SQL dedicados do Synapse com PolyBase e COPY usando T-SQL. O aluno também aprenderá como usar o gerenciamento de carga de trabalho junto com uma atividade de cópia em um pipeline do Azure Synapse para ingestão de dados em escala de petabyte.



Lições

- Use as práticas recomendadas de carregamento de dados no Azure Synapse Analytics
- Ingestão em escala de petabyte com Azure Data Factory



Laboratório: ingerir e carregar dados no data warehouse

- Executar ingestão em escala de petabyte com Azure Synapse Pipelines
- Importar dados com PolyBase e COPY usando T-SQL
- Use as práticas recomendadas de carregamento de dados no Azure Synapse Analytics



Depois de concluir este módulo, os alunos serão capazes de:

- Use as práticas recomendadas de carregamento de dados no Azure Synapse Analytics
- Ingestão em escala de petabyte com Azure Data Factory



Módulo 6: Transforme dados com Azure Data Factory ou Azure Synapse Pipelines

Este módulo ensina os alunos a construir pipelines de integração de dados para ingerir de várias fontes de dados, transformar dados usando fluxos de dados de mapeamento e realizar movimentação de dados em um ou mais coletores de dados.



Lições

- Integração de dados com Azure Data Factory ou Azure Synapse Pipelines
- Transformação livre de código em escala com Azure Data Factory ou Azure Synapse Pipelines



Laboratório: Transformar dados com Azure Data Factory ou Azure Synapse Pipelines

- Execute transformações sem código em escala com Azure Synapse Pipelines
- Crie pipeline de dados para importar arquivos CSV mal formatados
- Criar fluxos de dados de mapeamento



Depois de concluir este módulo, os alunos serão capazes de:

- Realizar integração de dados com Azure Data Factory
- Execute a transformação livre de código em escala com o Azure Data Factory



Módulo 7: orquestrar a movimentação e transformação de dados em Azure Synapse Pipelines

Neste módulo, você aprenderá como criar serviços vinculados e orquestrar a movimentação e transformação de dados usando blocos de anotações no Azure Synapse Pipelines.



Lição

- Orquestrar a movimentação e transformação de dados no Azure Data Factory



Laboratório: orquestrar a movimentação e transformação de dados em Azure Synapse Pipelines

- Integrar dados de notebooks com Azure Data Factory ou Azure Synapse Pipelines



Depois de concluir este módulo, os alunos serão capazes de:

- Orquestrar a movimentação e transformação de dados em Azure Synapse Pipelines



Módulo 8: Segurança ponta a ponta com Azure Synapse Analytics

Neste módulo, os alunos aprenderão como proteger um espaço de trabalho do Synapse Analytics e sua infraestrutura de suporte. O aluno observará o SQL Active Directory Admin, gerenciará regras de firewall de IP, gerenciará segredos com o Azure Key Vault e acessará esses segredos por meio de um serviço vinculado ao Key Vault e atividades de pipeline. O aluno compreenderá como implementar segurança em nível de coluna, segurança em nível de linha e mascaramento de dados dinâmico ao usar pools SQL dedicados.



Lições

- Proteja um data warehouse no Azure Synapse Analytics
- Configurar e gerenciar segredos no Azure Key Vault
- Implementar controles de conformidade para dados confidenciais



Laboratório: Segurança ponta a ponta com Azure Synapse Analytics

- Infraestrutura de suporte do Azure Synapse Analytics segura
- Proteja o espaço de trabalho do Azure Synapse Analytics e os serviços gerenciados
- Dados seguros do espaço de trabalho do Azure Synapse Analytics

WWW.KASOLUTION.COM.BR



Depois de concluir este módulo, os alunos serão capazes de:

- Proteger um data warehouse no Azure Synapse Analytics
- Configurar e gerenciar segredos no Azure Key Vault
- Implementar controles de conformidade para dados confidenciais



Módulo 9: Suporte ao Processamento Analítico Transacional Híbrido (HTAP) com o Azure Synapse Link

Neste módulo, os alunos aprenderão como o Azure Synapse Link permite a conectividade perfeita de uma conta do Azure Cosmos DB para um espaço de trabalho do Synapse. O aluno entenderá como habilitar e configurar o link Synapse e, em seguida, como consultar o repositório analítico do Azure Cosmos DB usando Apache Spark e SQL sem servidor.



Lições

- Projete processamento transacional e analítico híbrido usando o Azure Synapse Analytics
- Configurar o Azure Synapse Link com o Azure Cosmos DB
- Consultar o Azure Cosmos DB com pools do Apache Spark
- Consultar o Azure Cosmos DB com pools SQL sem servidor



Laboratório: Suporte ao Processamento Analítico Transacional Híbrido (HTAP) com o Azure Synapse Link

- Configurar o Azure Synapse Link com o Azure Cosmos DB
- Consultar Azure Cosmos DB com Apache Spark para Synapse Analytics
- Consultar o Azure Cosmos DB com pool SQL sem servidor para o Azure Synapse Analytics



Depois de concluir este módulo, os alunos serão capazes de:

- Projete processamento transacional e analítico híbrido usando o Azure Synapse Analytics
- Configurar o Azure Synapse Link com o Azure Cosmos DB
- Consultar Azure Cosmos DB com Apache Spark para Azure Synapse Analytics
- Consultar Azure Cosmos DB com SQL sem servidor para Azure Synapse Analytics



Módulo 10: Processamento de stream em tempo real com Stream Analytics

Neste módulo, os alunos aprenderão como processar dados de streaming com o Azure Stream Analytics. O aluno irá ingerir dados de telemetria do veículo em Hubs de Eventos e, em seguida, processar esses dados em tempo real, usando várias funções de janelamento no Azure Stream Analytics. Eles enviarão os dados para o Azure Synapse Analytics. Finalmente, o aluno aprenderá como dimensionar o trabalho do Stream Analytics para aumentar o rendimento.



Lições

- Habilite mensagens confiáveis para aplicativos de Big Data usando Hubs de Eventos do Azure
- Trabalhe com fluxos de dados usando o Azure Stream Analytics
- Ingerir fluxos de dados com o Azure Stream Analytics



Laboratório: Processamento de stream em tempo real com Stream Analytics

- Use Stream Analytics para processar dados em tempo real de Event Hubs
- Use funções de janela do Stream Analytics para criar agregados e saída para o Synapse Analytics
- Dimensione o trabalho do Stream Analytics do Azure para aumentar a taxa de transferência por meio do particionamento
- Reparticione a entrada do stream para otimizar a paralelização



Depois de concluir este módulo, os alunos serão capazes de:

- Habilitar mensagens confiáveis para aplicativos de Big Data usando Hubs de Eventos do Azure
- Trabalhar com fluxos de dados usando o Azure Stream Analytics
- Ingerir fluxos de dados com o Azure Stream Analytics



Módulo 11: Criar uma solução de processamento de fluxo com Hubs de eventos e Azure Databricks

Neste módulo, os alunos aprenderão como ingerir e processar dados de streaming em escala com Event Hubs e Spark Structured Streaming no Azure Databricks. O aluno aprenderá os principais recursos e usos do Streaming Estruturado. O aluno implementará janelas deslizantes para agregar blocos de dados e aplicar marcas d'água para remover dados obsoletos. Por fim, o aluno se conectará a Hubs de eventos para ler e gravar fluxos.



Lição

- Processar dados de streaming com streaming estruturado do Azure Databricks



Laboratório: Criar uma solução de processamento de fluxo com Hubs de eventos e Azure Databricks

- Explore os principais recursos e usos de streaming estruturado
- Transmita dados de um arquivo e grave-os em um sistema de arquivos distribuído
- Use janelas deslizantes para agregar blocos de dados no lugar de todos os dados.
- Aplique marca d'água para remover dados obsoletos
- Conecte-se a Hubs de eventos para ler e gravar streams



Depois de concluir este módulo, os alunos serão capazes de:

- Processar dados de streaming com streaming estruturado do Azure Databricks