

Kod szkolenia: **BIGDATA/AWS**

Tytuł szkolenia: **Technologie Big Data w chmurze AWS**

Dni: 3

Opis:

Adresaci szkolenia

Szkolenie adresowane jest do architektów, analityków i inżynierów danych pragnących poznać serwisy AWS do przetwarzania danych, zarówno w trybie batch jak i w czasie rzeczywistym.

Cel szkolenia

Cele szkolenia obejmują:

- zapoznanie uczestników z koncepcją Data Lake w chmurze AWS,
- zapoznanie z serwisami do migracji oraz integracji danych do chmury AWS w trybie batch oraz real time,
- zapoznanie z serwisami do przetwarzania danych w chmurze AWS w trybie batch oraz real time,
- zapoznanie z serwisami do przechowywania danych w chmurze AWS (data lake storage, sql stores, no-sql stores),
- zapoznanie z aspektami bezpieczeństwa usług analitycznych

Wymagania

Od uczestników wymagana jest znajomość min. jednej z technologii Big Data w wersji on-premise (np. Hadoop, Kafka, MongoDB) oraz podstawowa znajomość chmury AWS. Szkolenie rozpocznie się od omówienia podstawowych usług AWS (S3, EC2, VPC, IAM, ...), po czym omówione zostaną wszystkie istotne serwisy analityczne.

Parametry szkolenia

3*8 godzin (3*7 godzin netto) wykładów i warsztatów. W trakcie warsztatów użytkownicy wykonają serię zintegrowanych zadań, począwszy od migracji danych do chmury w trybie batch oraz real time, przetwarzanie danych oraz ich wizualizację oraz analizę.

Wielkość grupy: maks. 12

Program szkolenia:



Program szkolenia

1. Przegląd podstawowych usług Amazon Web Services
 - Compute (EC2, Lambda, on demand vs spot instances)
 - Storage (S3, EBS, EFS, Glacier)
 - Networking (VPC, IGW, NGW, sieci publiczne i prywatne, security groups, acls, service endpoints)
 - Monitoring (Cloudwatch, Cloudtrail)
 - Infrastruktura AWS (m.in. regiony, strefy dostępności, direct connect,)
 - Praca z AWS (CLI, SDK, REST API, IDE)
 - Security (IAM)
 - Infrastructure Provisioning (CloudFormation, Terraform)
2. Koncepcja Data Lake
 - Koncepcja Data Lake vs Data Warehouse vs Data Platform
 - Rola poszczególnych serwisów w koncepcji Data Lake w chmurze AWS
3. Migracja danych do chmury
 - Migracja danych relacyjnych za pomocą serwisu AWS Database Migration Service (DMS)
 - Migracja danych nierelacyjnych za pomocą serwisu AWS Batch
 - Streaming danych za pomocą serwisu AWS Kinesis Data Streams
4. Integracja danych w chmurze AWS
 - Integracja serwisów AWS przy wykorzystaniu AWS Lambda
 - Serverless Computing w rozwiązaniach analitycznych
 - Orkiestracja (orchestration) oraz planowanie (scheduling) procesów przetwarzania danych za pomocą AWS Step Functions
 - Przetwarzanie Big Data za pomocą serwisu AWS Elastic Map Reduce (EMR)
 - Przetwarzanie Serverless ETL przy użyciu serwisu AWS Glue ETL
 - Przechowywanie metadanych Data Lake z wykorzystaniem serwisu AWS Glue Data Catalog
 - Kolejki i notyfikacje - serwisy AWS SQS oraz AWS SNS
 - Analityka w czasie rzeczywistym przy użyciu AWS Kinesis Data Analytics
 - Integracja danych real time z innymi serwisami przy użyciu serwisu AWS Kinesis Data Firehose
5. Przechowywanie i serwowanie danych w chmurze AWS
 - Data Lake storage - S3
 - AWS Athena - Serwis serverless do zapytań SQL
 - AWS DynamoDB - baza no-sql
 - AWS Elasticsearch - wyszukiwanie, wizualizacje oraz analizy danych w czasie rzeczywistym
 - Hurtownia danych - AWS Redshift - przetwarzanie danych oraz integracja z AWS Data Lake (Redshift Spectrum)

