

Big Data specialisatie traineeship

Week 4

Dag 1:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Dag 2:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Dag 3:

Examen module 1

Zelfstandig doornemen en uitvoeren van oefeningen van Module 02 – Fundamental Big Data

- Security Engines, Cluster Managers and Data Governance Managers
- Visualization Engines and Productivity Portals
- Machine-Level Data Processing Architectural Models
- Shared-Everything and Shared-Nothing Architectures
- Big Data Analytics Logical Architecture
- Big Data Analytics Logical Architecture
- Data Sources and Data Acquisition Layers
- Storage, Processing and Batch Layers
- Realtime Processing, including Event Stream and Complex Event Processing
- Enterprise Data Warehouse and Big Data Integration Approaches (including Series and Parallel)
- Poly Source, including Relational, Streaming and File-based Sources
- Poly Storage, including Automatic Data Replication and Data Size Reduction
- Random Access Storage, including High Volume Binary, Tabular, Linked, Hierarchical and Data Sharding
- Streaming Access Storage, including Streaming Storage and Dataset Decomposition
- Large-Scale Batch Processing, Complex Decomposition and Processing Abstraction

Dag 4:

Klassikale lesdag onder leiding van SQL docent Gerben Hameetman via Microsoft Teams. Jullie gaan je verdiepen in onder andere de volgende onderwerpen:

- Writing Select queries
- Querying Multiple Tables
- Sorting and Filtering Data
- Working with SQL Server Data Types

Dag 5:

Klassikale trainings- en projectdag onder leiding van trainer. Op locatie in Utrecht

Week 5

Dag 1:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Dag 2:

Klassikale lesdag onder leiding van SQL docent Gerben Hameetman via Microsoft Teams. Jullie gaan je verdiepen in onder andere de volgende onderwerpen:

- Using Built-In Functions
- Grouping and Aggregating Data
- Using Windows Ranking, Offset, and Aggregate Functions
- Pivoting and Grouping Sets

Dag 3:

Zelfstandig doornemen en uitvoeren van oefeningen van Module 02 – Fundamental Big Data

- Poly Sink, including Relational Sink, File-based Sink and Automated Dataset Execution
- Big Data Appliance and Data Virtualization
- Architectural Environments, including ETL
- Analytics Engines and Application Enrichment
- Cloud Computing and Big Data Architectural Considerations
- Cloud Delivery and Deployment Models for Hosting Big Data Solutions

Zelfstandig doornemen en uitvoeren van oefeningen van Module 03 – Big Data Analysis & Technology Lab

Dag 4:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Dag 5:

Examen module 2

Zelfstandig doornemen en uitvoeren van oefeningen van Module 03 – Big Data Analysis & Technology Lab

Week 6

Dag 1:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Dag 2:

Zelfstandig doornemen en uitvoeren van oefeningen van Module 04 – Fundamental Big Data Analysis & Science

- Data Science, Data Mining & Data Modeling
- Big Data Dataset Categories
- High-Volume, High-Velocity, High-Variety, High-Veracity, High-Value Datasets
- Exploratory Data Analysis (EDA)
- EDA Numerical Summaries, Rules and Data Reduction
- EDA analysis types, including Univariate, Bivariate and Multivariate
- Essential Statistics, including Variable Categories and Relevant Mathematics
- Statistics Analysis, including Descriptive, Inferential, Covariance, Hypothesis Testing, etc.

- Measures of Variation or Dispersion, Interquartile Range & Outliers, Z-Score, etc.
- Probability, Frequency, Statistical Estimators, Confidence Interval, etc.
- Data Munging and Machine Learning
- Variables and Basic Mathematical Notations

Dag 3:

Examen module 3

Zelfstandig doornemen en uitvoeren van oefeningen van Module 04 – Fundamental Big Data Analysis & Science

- Statistical Measures and Statistical Inference
- Confirmatory Data Analysis (CDA)
- CDA Hypothesis Testing, Null Hypothesis, Alternative Hypothesis, Statistical Significance, etc.
- Distributions and Data Processing Techniques
- Data Discretization, Binning and Clustering
- Visualization Techniques, including Bar Graph, Line Graph, Histogram, Frequency Polygons, etc.
- Prediction Linear Regression, Mean Squared Error and Coefficient of Determination R^2 , etc.
- Clustering k-means, Cluster Distortion, Missing Feature Values, etc.
- Numerical Summaries

Verder werken aan het eindproject

Dag 4:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Dag 5:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Week 7

Dag 1:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Dag 2:

Zelfstandig doornemen en uitvoeren van oefeningen van Module 07 - Fundamental Big Data Engineering

- Big Data Engineering Techniques and Challenges
- Big Data Storage, including Sharding, Replication, CAP Theorem, ACID and BASE
- Master-Slave, Peer-to-Peer Replication, Combining Replication with Sharding
- Big Data Storage Requirements, Scalability, Redundancy and Availability
- Fast Access, Long-term Storage, Schema-less Storage and Inexpensive Storage
- On-Disk Storage, including Distributed File System and Databases
- Introduction to NoSQL and NewSQL
- NoSQL Rationale and Characteristics

Dag 3:

Begeleide lesdag in Discord Open Leercentrum

Examen module 4

Zelfstandig doornemen en uitvoeren van oefeningen van Module 07 - Fundamental Big Data Engineering

- NoSQL Database Types, including Key-Value, Document, Column-Family and Graph Databases
- Big Data Processing Engines
- Distributed/Parallel Data Processing, Schema-less Data Processing
- Multi-Workload Support, Linear Scalability and Fault-Tolerance
- Big Data Processing Requirements, including Batch, Cluster and Realtime Modes
- MapReduce for Big Data Processing, including Map, Combine, Partition, Shuffle and Sort and Reduce
- MapReduce Algorithm Design
- Task Parallism, Data Parallism

Dag 4:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Dag 5:

Werken aan het eindproject en voorbereiding examen 7

Week 8

Dag 1:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Dag 2:

Examen module 7

Werken aan de eindopdracht in de groep

Dag 3:

Klassikale trainings- en projectdag onder leiding van trainer in Discord

Dag 4:

Werken aan de eindopdracht in de groep

Dag 5:

Eindpresentaties op locatie Hoofddorp Young Capital