BIG DATA CAMP PROGRAM

1. Big Data essentials

- 1.1 What is Big Data
- 1.2Which problems solves Big Data
- 1.3 Big Data practical aspects
- 1.4 Data Engineer position overview and responsibilities
- 1.5 Intro to Big Data technologies stacks

2. Data Platforms overview

- 2.1 Basic notions and glossary in scope of Data Platforms
- 2.2 Data Pipelines modeling intro
- 2.3 Processing layers and related technologies overview
 - 2.3.1 Data storage
 - 2.3.2 Data processing
 - 2.3.3 Data ingestion
 - 2.3.4 Data querying
 - 2.3.5 Data visualization
- 2.4 Intro to cloud concepts and cloud computing
- 2.5 PaaS / IaaS / SaaS service models

3. Data ingestion & integration

- 3.1 Data exchange types
 - 3.1.1 Ingestion

- 3.1.2 Export
- 3.1.3 Replication
- 3.2 Data sources overview
- 3.3 Data sources integration and prioritization
- 3.4 Batch-driven data integrations
- 3.5 Pull / Push Integration types
- 3.6 Data arrivals control
- 3.7 Realtime-driven data integrations

4. Data Storage fundamentals

- 4.1 Data Storages classification
- 4.2 HDFS /Object Stores key principles
- 4.3 DBMS general concepts
- 4.4 Intro to OLTP, OLAP, DWH
- 4.5 CAP Theorem
- 4.6 Data formats overview
- 4.7 Partitioning / Sharding / Replication concepts
- 4.8 Strong and Eventual data consistency

5. Data modelling fundamentals

- 5.1 Data Lakes and Data Warehouses design concepts
- 5.2 Data modelling in scope of different storage types
- 5.3 Data schemas introduction
- 5.4 Schemas on read / on write overview

- 5.5 Batching/Partitioning strategies
- 5.6 Data granularity basics
- 5.7 Schema-less storages
- 5.8 Late arrivals handling

6. Data Processing intro

- 6.1 Historical context
- 6.2 ETL pipelines introduction
- 6.3 Map-Reduce model basics
- 6.4 Distributed processing model and data locality
- 6.5 Data Processing technologies stack overview

7. Data Processing fundamentals

- 7.1 Resource management basics
- 7.2 Mainstream Data Processing frameworks overview
- 7.3 Batch data processing concepts
 - 7.3.1 Workflow development
 - 7.3.2 Scheduling
 - 7.3.3 Orchestration
- 7.4 Delivery / Processing guarantee semantics
- 7.5 Streaming data processing concepts
 - 7.5.1 Types
 - 7.5.2 Pipelines development
 - 7.5.3 Synchronization

8. Cloud computing

- 8.1 Core AWS cloud services
- 8.2 Resources for building Data Platforms
- 8.3 AWS SDK in scope of data pipelines implementation
- 8.4 AWS CLI management tools
- 8.5 Data pipelines overview in AWS cloud
- 8.6 Logging in AWS cloud solutions
- 8.7 Monitoring in AWS cloud solutions
- 8.8 Continuous Integration / Continuous Deployment intro in AWS

9. Data querying & visualization

- 9.1 Data representation forms
- 9.2 Data visualization intro
- 9.3 Data marts, reports, views notions
- 9.4 DWH concepts in scope of reporting
- 9.5 Business Intelligence tools
- 9.6 Automating reports using Data Dashboards
- 9.7 Data exploration with Notebooks

10. Data governance

- 10.1 Data Governance goals
- 10.2 Data Management procedures
- 10.3 Data Tracing, Lineage, Versioning notions

- 10.4 Data Retention policies
- 10.5 Data Security/Privacy/PII concepts
- 10.6 Data Governance tools and solutions