



RIGACODINGSCHOOL

DATU ANALĪTIKA



Par mums

- Viena no pirmajām programmēšanas skolām Lietuvā un Latvijā;
- Skolas filiāles Viļņā, Klaipēdā, Kauņā un Rīgā;
- Vairāk nekā 1500 absolventi;
- 80% absolventu, kas vēlas strādāt IT jomā;
- Vairāk nekā 50 sadarbības uzņēmumu Rīgā;
- Skolas lektori – ar pieredzi IT sfērā.



Kursi

- Intensīvs un koncentrēts kurss;
- Visas lekcijas tiek **ierakstītas** un pieejamas privātā YouTube kanālā;
- **Vies-lekcijas** ar sadarbības uzņēmumu speciālistiem;
- Iespēja atkārtot kursu bezmaksas **1 gada laikā**;



Karjeras centrs

- CV un karjeras semināri;
- Palīdzam atrast prakses vietas IT uzņēmumos;
- Studenti tiek vērtēti apmācību laikā;
- Iespēja izmantot skolas telpas ārpus apmācībām;
- Izsniedzam **akreditētu sertifikātu** par 120 akadēmisko stundu apmeklējumu, arī pasniedzēju **rekomendācijas**;



VALDIS



- Izglītība: Maģistra grāds datorzinātnēs
- Pieredze programmēšanā: 20+ gadi
- Specialitāte: grafu teorija sociālo tīklu analizēšanā
- Hobiji: prāta spēles, ritenbraukšana, šahs



Data Lake



Brief History of Data Analysis



- ~ 18,000BC – Uganda, Ishango Bone
- ~ 2400BC – Babylon abacus, libraries
- 300BC – 48AD – Library of Alexandria
- ~ 100-200AD Antikythera Mechanism



Brief History of Data Analysis II



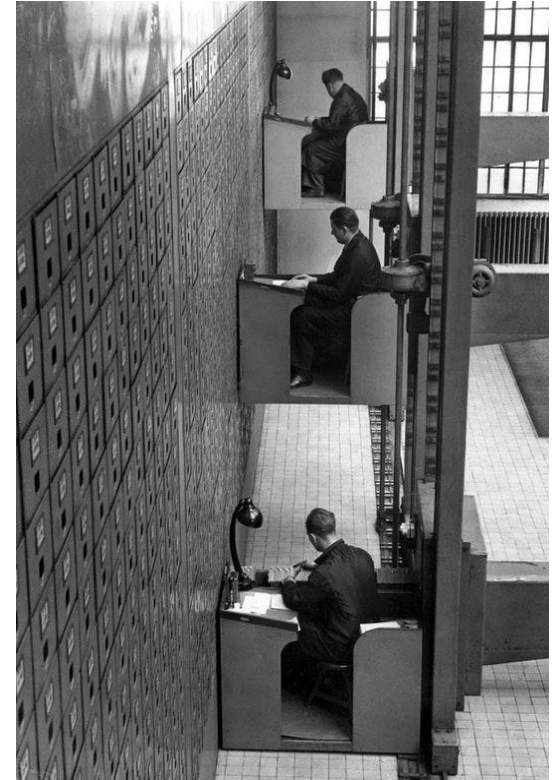
- 1663 – London, J.Graunt mortality analysis
- 1865 – banker H. Furnese business intelligence
- 1880-90 US Census Hollerith Machine -> IBM
- 1928 – F. Pflueger magnetic tape invention



Brief History of Data Analysis III



- 1950s - Flat Files
- 1958 – IBM's Luhn defines Business Intelligence
- 1960s - CODASYL
- 1970s – Codd's relational DBs -> SQL
- 1980s – Data Warehouses / Marts
- 2000s – Big Data / noSQL DBs



BIG DATA LANDSCAPE 2017

INFRASTRUCTURE

HADOOP ON-PREMISE

- cloudera
- MAPR
- IBM InfoSphere
- bluedata
- jethro

HADOOP IN THE CLOUD

- amazon
- Microsoft Azure
- Google Cloud Platform
- IBM InfoSphere
- BigInsights
- oaliscole
- CAZENA
- CenturyLink

STREAMING / IN-MEMORY

- amazon
- databricks
- bonfluent
- stream
- GridGain
- dataArtisans
- ORACLE
- hazelcast
- TECHADOTIA

NOSQL DATABASES

- Google Cloud Platform
- ORACLE
- Microsoft Azure
- MarkLogic
- mongoDB
- DATASTAX
- REDIS
- Couchbase
- influxdata

NEWSQL DATABASES

- SAF
- Clustrix
- Pivotal
- cockroach Labs
- memsql
- splice
- Voltron
- chudata
- Trifolium
- deepdb
- paradigm4

GRAPH DBS

- IBM
- ORACLE
- NEPHEUS
- action
- kognitio
- XSQL
- dremio

MPP DBS

- TERADATA
- VERTICA
- IBM
- Microsoft Azure
- Pivotal
- InfoWorks
- InfoSphere

CLOUD EDW

- amazon
- Microsoft Azure
- Google Cloud Platform
- IBM
- ORACLE
- Microsoft Azure
- Pivotal
- InfoWorks
- InfoSphere

DATA TRANSFORMATION

- talend
- pentaho
- alteryx
- TRIFACTA
- IBM
- StreamSets
- UNIFI

DATA INTEGRATION

- Informatica
- logiXML
- TEALUM
- enigma
- ALZANI
- glity
- Stitch

DATA GOVERNANCE

- informatica
- IBM
- shyghy
- collibra
- Alation
- Waterline

MGMT / MONITORING

- amazon
- New Relic
- APDYNAMICS
- oatlio
- WAVEFRONT
- WAVEFRONT
- WAVEFRONT
- WAVEFRONT
- WAVEFRONT
- WAVEFRONT

STORAGE

- amazon
- Google Cloud Platform
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

CLUSTER SERVICES

- amazon
- Google Cloud Platform
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

APP DEV

- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

CROWDSOURCING

- amazon
- Upwork
- WorkFusion
- Upwork
- WorkFusion
- Upwork
- WorkFusion
- Upwork
- WorkFusion
- Upwork
- WorkFusion

HARDWARE

- Google TPU
- ARM
- Graphcore
- MYTHIC
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

ANALYTICS

DATA ANALYST PLATFORMS

- Microsoft
- pentaho
- alteryx
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

DATA SCIENCE PLATFORMS

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

BI PLATFORMS

- Microsoft
- pentaho
- alteryx
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

VISUALIZATION

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

VERTICAL ANALYTICS

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

STATISTICAL COMPUTING

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

DATA SERVICES

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

MACHINE LEARNING

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

HORIZONTAL AI

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

SPEECH & NLP

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

SEARCH

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

LOG ANALYTICS

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

SOCIAL ANALYTICS

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

WEB / MOBILE / COMMERCE ANALYTICS

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

APPLICATIONS - ENTERPRISE

SALES

- Microsoft
- pentaho
- alteryx
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

MARKETING - B2B

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

MARKETING - B2C

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

CUSTOMER SERVICE

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

HUMAN CAPITAL

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

LEGAL

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

FINANCE

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

ENTERPRISE PRODUCTIVITY

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

BACK OFFICE AUTOMATION

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

SECURITY

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

APPLICATIONS - INDUSTRY

ADVERTISING

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

EDUCATION

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

GOVERNMENT

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

FINANCE - LENDING

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

FINANCE - INVESTING

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

REAL ESTATE

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

INSURANCE

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

HEALTHCARE

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

LIFE SCIENCES

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

TRANSPORTATION

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

AGRICULTURE

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

COMMERCE

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

OTHER

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

CROSS-INFRASTRUCTURE/ANALYTICS

amazon, Google Cloud Platform, Microsoft, IBM, SAP, Hewlett Packard Enterprise, SAS, vmware, TIBCO, TERADATA, ORACLE, NetApp

OPEN SOURCE

FRAMEWORK

- Spark
- SQL
- presto
- SLANDATA
- AVRIL
- DRILL
- Google Cloud Dataflow

QUERY / DATA FLOW

- Spark
- SQL
- presto
- SLANDATA
- AVRIL
- DRILL
- Google Cloud Dataflow

DATA ACCESS

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

COORDINATION

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

STREAMING

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

STAT TOOLS

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

AI / MACHINE LEARNING / DEEP LEARNING

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

SEARCH

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

LOG ANALYSIS

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

VISUALIZATION

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

COLLABORATION

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

SECURITY

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

DATA SOURCES & APIS

HEALTH

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

IOT

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

FINANCIAL & ECONOMIC DATA

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

AIR / SPACE / SEA

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

PEOPLE / ENTITIES

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

LOCATION INTELLIGENCE

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

OTHER

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

DATA RESOURCES

INCUBATORS & SCHOOLS

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

RESEARCH

- IBM
- KNIME
- dataiku
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE
- Microsoft Azure
- IBM
- ORACLE

Buzzword bingo



- Big Data
 - Data Mining (datizrace)
 - Machine Learning – subset of AI
 - Data Science – statistics
-
- Big Data or Pokemon
 - <https://pixelastic.github.io/pokemonorbigdata/>

Random Forests	Neural Network	Reinforcement Learning	Supervised Learning	Cognitive Computing
Caffe	Support Vector Machine	Artificial Intelligence	Python	Cloud
Unstructured Data	Bot	DATA SCIENCE BUZZWORD BINGO (free square)	K-means	GPU
Spark	Data Wrangling	Deep Learning	Ensemble	Machine Learning
Keras	Tensorflow	Big Data	Algorithm	Feature Engineering

Big Data



- Volume
- Velocity
- Variety
- Veracity

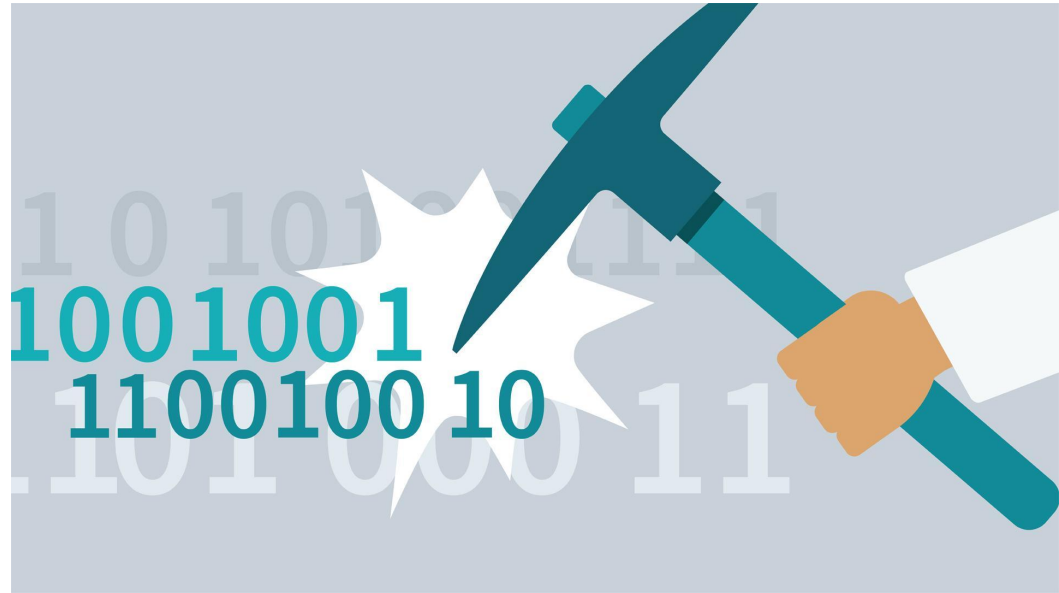


BIG DATA

Data Mining



- Anomalies
- Classification
- Clusters
- Dimension Reduction
- Regression
- Relationship finding
- Summarization / Visualization



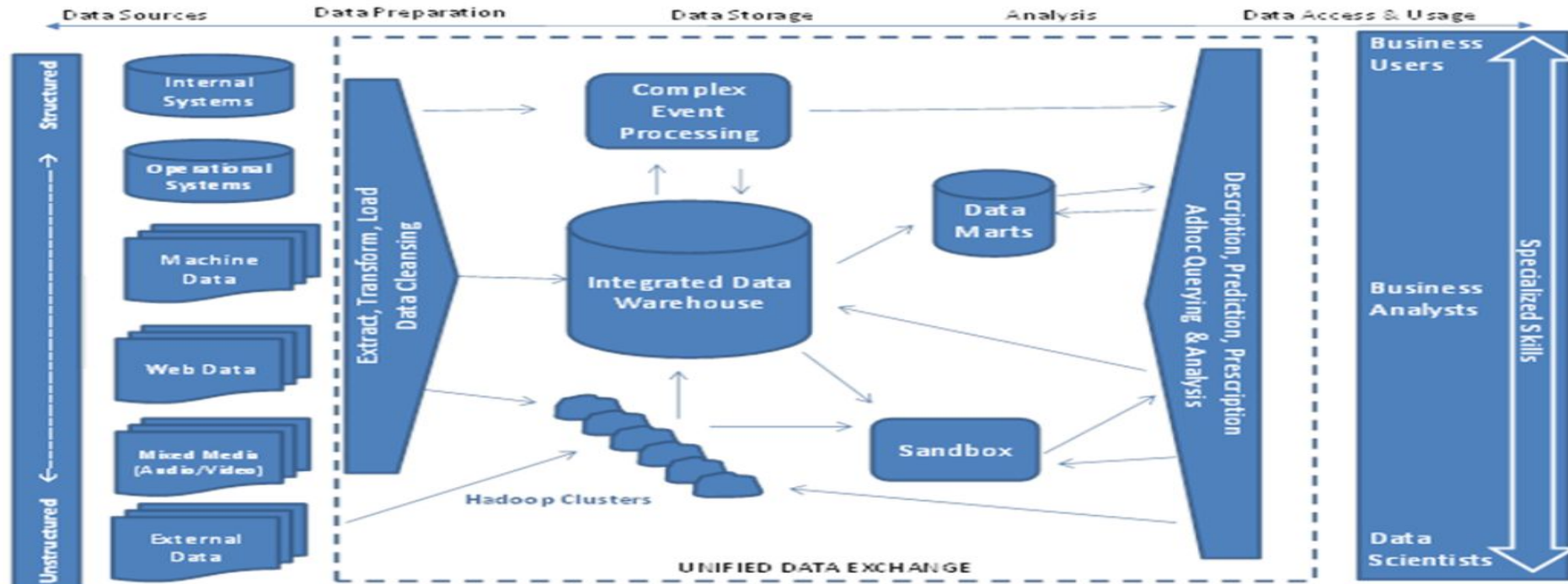
Building a Pipeline



- Data Security
 - Cleanup
 - Organising Database
 - Analysis
 - Visualization – Dashboard
-
- Emphasis on Analysis less on Infrastructure



Full Analysis Framework

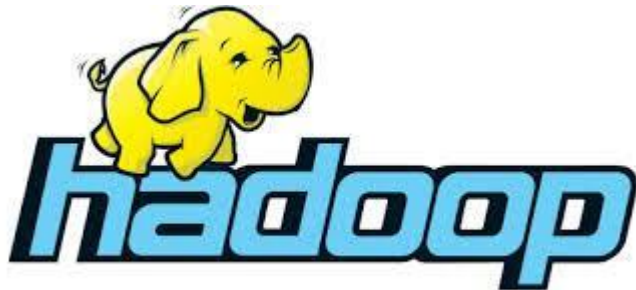


BIG DATA MANAGEMENT and GOVERNANCE: Strategic, Tactical and Operation Levels
(Metadata, Data quality, Access, Use, Ethics, Privacy, and Security Management processes)

Technologies Covered



- Cloudera Distribution
- Apache Hadoop Ecosystem
- Apache MapReduce
- Apache Spark
- Apache Solr



cloudera



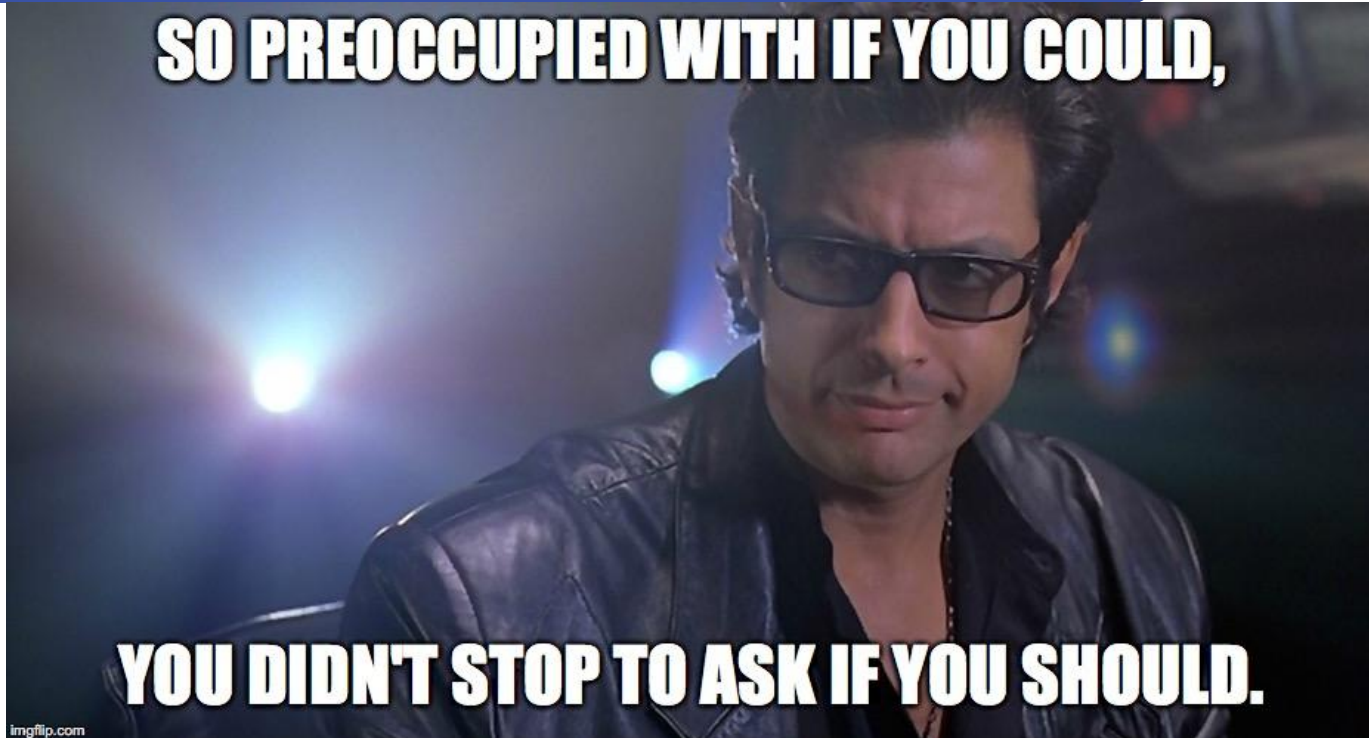
Avoiding Pitfalls



- Overfitting
- Data Dredging / p-hacking
- <https://xkcd.com/882/>



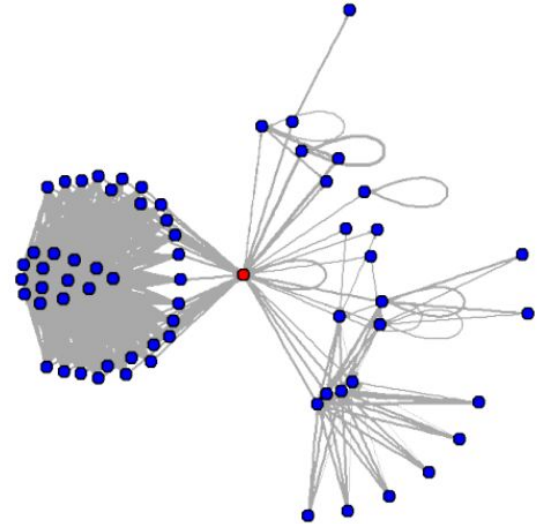
GDPR, ethics



Planned Projects



- Recommendation System / Churn Prediction
- Web Comments Sentiment Analysis
- Network Analysis (possibly Bitcoin Blockchain)
- Visualizations with Tableau



Requirements



- Analytical / Logical Mind
- Helpful but NOT required knowledge:
 - Python (will use Anaconda / Scipy)
 - Comfortable in command line
 - SQL
 - Statistics
- Helpful: a computer with a minimum of 8GB RAM
- <https://www.anaconda.com/download/> (3.6+)



python™

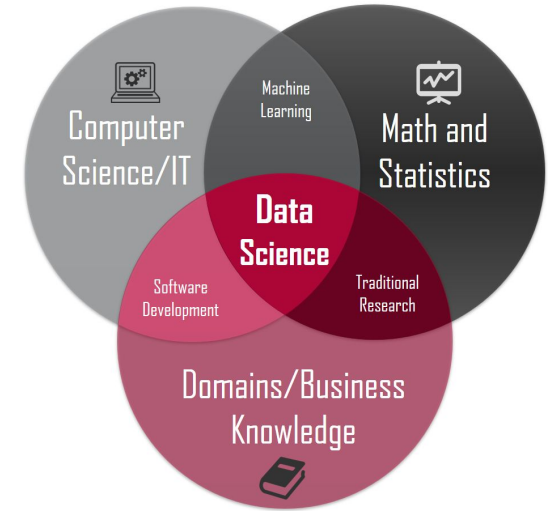


ANACONDA®

Goals



- Access structured/unstructured data
- Clean data
- Apply correct methods for analysis
- Visualize Results



Kādēļ Tu esi šeit?



**Everybody in this country should
learn to program a computer...
because it teaches you how to think**

Steve Jobs, co-founder and CEO of Apple Inc. (1955 - 2011)





Jautājumi?

