# Project

Big data and distributed processing

# Themes

- 1. Application that uses distributed database for managing data.
- 2. Application that uses Apache Hadoop for data analysis.

# Distributed Database Application

Create Application that will use distributed database, for example:

- Distributed system for cinema reservation.
- Distributed system for house renting.
- Distributed system for hospital.
- Distributed system for library.
- Distributed system for flight management

- Possible own theme after discussion.

# Distributed Database Application - Requirements

- One or two persons
- Console Application (1 person) or GUI (pair).
- At least 2-3 node working in project.
- Application can work on every node in cluster
- In project you have to be able to Make reservation, update reservation, see reservation and see who made it.
- (Only for Pairs ) Cancellation of reservation and ability to cancel more then one at a time.

# Distributed Database Application - Requirements

- Error handling and no big delays while processing data.
- The ability to automatically generate a large load of data (stress test), the created system must handle it.
- Stress Test 1: The client makes the same request very quickly.
- Stress Test 2: Two or more clients make the possible requests randomly.
- Stress Test 3: Immediate occupancy of all seats/reservations on 2 clients.
  (Main point is to have 2 parties try to make as many as possible reservation at a time and result should be that both are able to make have reservations, not 1 party take all reservations.)
- Stress Test 4: (only for pairs) constant cancellations and seat occupancy.
- Stress Test 5: (only for pairs) Make large group cancellation of many reservations.

# Apache Hadoop for data analysis

- Find (or create) a large Dataset.
- Make an analysis, see what's in data and what you can take from it.
- Look for at least 5 interesting points in data that you want user to browse from.
- Create an application that user can browse this data. (The accessibility of shown data is part of grade)
- Example: From Car Price dataset you could get:
  -Brands of car and prices
  -Fuel Types and prices
  -Mileage and Brands
  -Number of Doors and Brands
  -Number of previous owners and price

  -Now application should be able to make user select one of such options and make "queries" to browse this data.
- *Recommendation (not necessary for project but will give higher grade) Have application make 1 recommendation, for example which car is best to buy from this brand.

# Requirements

- Description (readme file in git describing in a few sentences what the project does and how to run it)
- Short report (1-2 pages, description, database schema, problems encountered)
- Attach the database schema (for distributed database application)
- Submission by the end of the semester. Late submission will result in a lower grade.
- Sharing the project with the teacher, preferably in the form of a git repository
- GitLab:@AdamGodzinski
- GitHub:@AdamGodzinski