

Список на теми за изработка на проектни задачи по предметот  
**НЕСТРУКТУРИРАНИ БАЗИ НА ПОДАТОЦИ И XML**  
учебна 2022/2023 година

РБ	Наслов	Бр. групи	Бр. студ.
<b>NoSQL</b>			
1	<b>Key-value бази на податоци</b> Да се импортираат дадени податоци и истите да се користат во NoSQL базата на податоци. Препорачани технологии: Riak, Voldemort, Oracle NoSQL	5	3-5
2	<b>Document бази на податоци</b> Да се импортираат дадени податоци и истите да се користат во NoSQL базата на податоци. Препорачани технологии: MongoDB, CouchDB	5	3-5
3	<b>Column-family бази на податоци</b> Да се импортираат дадени податоци и истите да се користат во NoSQL базата на податоци. Препорачани технологии: Hadoop, Cassandra	5	3-5
4	<b>Graph бази на податоци</b> Да се импортираат дадени податоци и истите да се користат во NoSQL базата на податоци. Препорачани технологии: Neo4J, Infinite Graph	5	3-5
<b>Останато - сопствен предлог на тема</b>			
5	<b>Останато</b> Постои можност да се пријават и самостојни теми	5	2-4

## Дополнително објаснување:

Изработката на проектните задачи ќе биде практична. Резултатот ќе биде прикажан во форма на пишан документ и презентација која ќе се користи при јавната одбрана на проектната задача. За време на одбраната на проектната задача треба да имате подготвено презентација од 10-тина минути во која ќе ја образложите работата што сте ја завршиле. За секоја од групите даден е препорачаниот број на студенти што ќе бидат вклучени во работата.

### 1. NoSQL:

Тимот ќе треба да се подели на два дела. Едната група ќе работи со една БП, а другата група со друга БП.

- 1) ИНСТАЛАЦИЈА: Се инсталира конкретната база на податоци. Сите подесувања и/или дополнителни инсталации треба да се документираат во рамките на текстот од проектната задача.
- 2) ИМПОРТИРАЊЕ НА ПОДАТОЦИТЕ: Се анализираат дадените податоци. Пожелно е да се направи некој модел за подобро да се разберат податоците и меѓусебните врски. Се дизајнира соодветниот “модел” кој ќе се користи за импортирање на податоците во БП. Може да се пробаат две различни нивоа на агрегација на податоците. Целата постапка треба да се документира во рамките на текстот од проектната задача.
- 3) КОРИСТЕЊЕ НА ПОДАТОЦИТЕ: Се избираат неколку сценарија за пристап до податоците. За секое од сценаријата се документираат соодветните наредби за пристап до БП и се воочуваат ограничувањата и/или проблемите што настануваат. Сценаријата за начинот на користење на податоците треба да се документираат во рамките на текстот од проектната задача.
- 4) СПОРЕДБА НА ПЕРФОРМАНСИТЕ: Пожелно е на крај да се споредат перформансите што ги нудат двете конкретни БП (од двата под-тима од тимот), со што би се добиле препораки во начинот на користење на едната и другата БП. Евентуално може да се проба (доколку е можно) да се споредат перформансите на извршување на еден процесор и повеќе процесори.

За секој од проблемите ќе ви бидат дадени соодветни податочни множества што ќе треба да се искористат при практичната реализација на проектната задача. На вас е оставено од веќе понудените податочни множества, да го изберете она што ќе ви биде најинтересно за анализа и работа.

## Структура на пишаниот елаборат:

### 1. Насловна страница

лого на Универзитетот и на факултетот (може да се превземе од <http://finki.ukim.mk/downloads>), наслов на проектната задача, членови на тимот (име, презиме и број на индекс), датум

### 2. Содржина

### 3. Вовед

краток опис на проблематиката, евентуален преглед на литература што бил користен за решавање на проблемот

### 4. Методологија

овде ќе се објасни постапката за добивање на решението за проблемот кој се решава. Ако станува збор за NoSQL базите на податоци, ќе се објасни моделот на податоци кој се користи, потоа се презентира начинот на агрегација (пожато е да објасните неколку различни модели за агрегација), за на крај да се стигне до имплементацијата на предложените прашања што ќе се поставуваат до податоците. За NoSQL базите на податоци потребно е да се смисли множество од прашалници кои ќе се поставуваат во базата на податоци. Ова множество треба да се движи во интервалот 6-10 прашалници, при што треба да имате неколку категории прашалници: едноставни (филтрирања), сложени (спојување на податоци од повеќе инстанци) и доста сложени (агрегирани извештаи)

### 5. Добиени резултати

Се објаснува податочното множество што се анализира. Се презентираат добиените резултати од спроведените експерименти. Доколу станува збор за NoSQL бази на податоци, тогаш пожелно е да се прикажат графици со перформансна анализа на спроведените експерименти (мерење на времињата на извршените прашалници). Дополнително, ако се прави споредбена анализа на различни бази на податоци, тогаш истите прашалници се извршуваат врз истото податочно множество импортирано во различни бази на податоци и се дава дискусија за добиените резултати

### 6. Заклучок

што било сработено во проектната задача, може да се даде одредена дискусија, како и можности за подобрување и идна работа

### 7. Користена литература

список на референци што биле користени во текот на изработката на проектната задача