

БЛОК2 Программа Алгоритмы параллельной обработки данных Map-Reduce
Преподаватель **Гаврилов Е.С., Перов В.А.**

Таблица 1

Наименование программы	Количество часов	Форма обучения	Стоимость обучения 1 слушателя, руб.
Алгоритмы параллельной обработки данных Map-Reduce	32+4	Очная	25 000

Описание программы

Алгоритмы параллельной обработки данных Map-Reduce

АННОТАЦИЯ к программе повышения квалификации

Цель программы: качественное изменение профессиональных компетенций, направленных на повышение профессионального уровня в рамках имеющейся у слушателей квалификации в области параллельной обработки данных. Особое внимание уделяется алгоритмам распараллеливания, пониманию механизмов распараллеливания на низком прикладном уровне и практической реализации на движке Apache Hadoop MapReduce.

Результатом обучения является приобретение **знаний и умений** по:

- Паттернам параллельной обработки данных
- Вероятностным структурам данных
- Анализу производительности и оптимизации параллельной обработки данных
- Практике работы с Apache Hadoop MapReduce

Категория слушателей: работники профильных отраслей промышленности, высшего образования, научно-исследовательских организаций, IT-сектора (аналитики, тестировщики, разработчики, менеджеры), знакомые с основами программирования на Java и основами языка SQL.

Объем программы: 36 часов.

Форма обучения: очная

Таблица 2

УЧЕБНЫЙ ПЛАН ПРОГРАММЫ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

№ п/п	Наименование разделов	Трудоёмкость, час	Аудиторные занятия, дистанционные занятия					Аттестация		СР С, час
			Всего, час	Лекции, час	Лабораторные работы, час	Практические занятия, семинары, час	час	Форма контроля		
1	Введение в параллельную обработку данных. Обзор экосистемы Apache Hadoop и дистрибутива Cloudera Hadoop.	4	4	2	2	0	0		0	
2	Файловая система HDFS, планировщик и управление ресурсами Apache YARN. Запуск и мониторинг задач MapReduce	4	4	2	2	0	0		0	
3	Движок MapReduce: API, этапы выполнения задачи. Агрегирование данных.	4	4	2	2	0	0		0	
4	Движок MapReduce:	4	4	2	2	0	0		0	

	алгоритмы соединения таблиц (Join)								
5	Движок MapReduce: сквозная сортировка данных	4	4	2	2	0	0		0
6	Фильтрация, Тор-N. Вероятностные структуры данных. Bloom-фильтр	4	4	2	2	0	0		0
7	Оркестровка задач MapReduce с помощью кода, Apache Oozie.	4	4	2	2	0	0		0
8	Итоговая аттестация	4	4	0	0	0	4	Зачет	0
	Итого	32	32	14	14	0	4		0

Всего 36 часов