

## ПЛЮС7 ФОРМИТ: BIG DATA ДЛЯ РАЗРАБОТЧИКОВ

Пройдя данный тренинг, участники познакомятся с обновленным интерфейсом разработчика – Плюс7 ФормИТ и узнают о основных особенностях разработки для платформы Hadoop.

Материал курса построен на основании уникальных материалов и методик разработанных DIS Group, лидером в области интеграции данных на территории России и стран СНГ

**Слушатели познакомятся со** средствами разработки, контроля и управления, а так же основными методами работы с платформой Hadoop.

В результате, слушатели смогут:

- Извлекать данные из реляционных источников и плоских файлов;
- Применять часто используемые трансформации;
- Использовать динамические маппинги;
- Применять Smart Executor;
- Использовать средства мониторинга для поиска ошибок;
- Работать со сложными данными, такими как JSON, AVRO и Parquet;
- Понимать как Плюс7ФормИТ работает с NoSQL данными.

Тренинг, в первую очередь, будет интересен:

- Разработчикам, непосредственно работающим с продуктом Плюс7 ФормИТ Big Data;
- Аналитикам и руководителям проектов, участвующим в проектах с Big Data;
- Квалифицированные специалисты, желающим понять суть работы с Big Data и повысить свой профессиональный уровень.

Продолжительность тренинга составляет 4 дня (32 академических часа).

## Каждый посетитель тренинга должен иметь при себе ноутбук следующей комплектации

### Компьютер PC или Mac:

- не ниже CPU: 1.3 GHz
- не менее 2 GB RAM
- Размер экрана не менее 1024 x 768 или выше

### Операционная система:

- Windows XP и выше
- Mac OS X
- Linux

### Поддерживаемые браузеры (не ниже):

Internet Explorer 10.0, Firefox 15, Safari 6.0, Opera 12.0, Chrome 16

Для наиболее полного усвоения программы участникам тренинга предварительно рекомендуется иметь опыт работы с СУБД и общее понимание концепции реляционной моделей данных. Так же, предполагается, что участники тренинга имеют представление о алгоритмическом программировании и владеют английским языком на уровне чтения технической документации.

## Программа тренинга по курсу **ПЛЮС7 ФОРМИТ: BIG DATA ДЛЯ РАЗРАБОТЧИКОВ**

### **Модуль 1: Вступление**

- *Начало работы с инструментом разработки*
- *Краткий обзор Big Data*
- *Архитектура*
- *Обзор интерфейса*

### **Модуль 2: Работа с физическими данными**

- *Объекты*
- *Представление о физических объектах данных*
- *Использование подключений к реляционным БД*
- *Подключение к плоским файлам*
- *Синхронизация объекта flat file data object*

### **Модуль 3: Просмотр данных**

- *Возможности средства просмотра данных*
- *Поиск ошибок*
- *Просмотр логов*
- *Мониторинг активностей*

### **Модуль 4: Разработка маппингов и трансформаций**

- *Концепция маппингов и трансформаций*
- *Основные трансформации*
- *Разработка и валидация маппинга*
- *Использование трансформаций в маппингах*

### **Модуль 5: Работа с динамическими схемами и маппингами**

- *Концепция динамических маппингов*
- *Разработка и запуск динамических маппингов*
- *Шаблоны*

### **Модуль 6: Параметры**

- *Концепция работы с параметрами*
- *Использование параметр - файла*
- *Использование набора параметров (parameter set)*

### **Модуль 7: Процессы (Workflow)**

- *Концепция Workflow*
- *Создание workflow*
- *Конфигурация Workflow*

### **Модуль 8: Работа с приложениями (Application)**

- *Концепция работы с Application*
- *Создание и развертывание Application*
- *Остановка и повторное развертывание Application*

### **Модуль 9: Обзор курса Big Data Integration**

- *Вступление*
- *Программа тренинга*
- *Доступ к учебным средам*
- *Дополнительные тренинги*

### **Модуль 10: Основы Big Data**

- *Что такое Big Data?*
- *Концепция Hadoop*
- *Основные компоненты архитектуры Hadoop*
- *The Hadoop Distributed File System (HDFS)*
- *Назначение Name Node & Secondary Name Node*
- *MapReduce*
- *“Yet Another Resource Manager” (YARN) (MapReduce Version 2)*

### **Модуль 11: Загрузка данных в хранилище**

- *Основные проблемы в традиционных хранилищах данных*
- *Требования к оптимальному хранилищу данных*
- *Процессы загрузки данных*

### **Модуль 12: Загрузка и выгрузка данных**

- *PowerCenter Reuse Reports*
- *Импорт маппингов PowerCenter в Developer*
- *SQOOP*
- *SQL to Mapping capability*
- *Партиции и параллельная загрузка*

### **Модуль 13: работа с Big Data**

- *Архитектура*
- *Мир Big Data*
- *Создай один раз запускай везде*
- *The Плюс7ФормИТ abstraction layer*
- *Polyglot computing*
- *The Smart Executor*
- *Open source и инновации*
- *Архитектура подключений*
- *Использование сторонних приложений*

#### **Модуль 14: Плюс7ФормИТ Polyglot**

- *Управление в Hadoop*
- *Hive MR/Tez*
- *Blaze*
- *Spark*
- *Native*
- *The Smart Executor*

#### **Модуль 15: Маппинги, мониторинг, поиск ошибок**

- *Конфигурирование и запуск маппингов в Native и Hadoop средах*
- *Исполнение планов*
- *Мониторинг маппингов*
- *Поиск ошибок в маппингах*
- *Просмотр результатов*

#### **Модуль 16: Проблемы Hadoop Data Integration и оптимизация производительности**

- *Описание проблем с запуском процессов в Hadoop*
- *Тюнинг процессов Big Data Management*
- *Оптимизация среды Hive*
  - *Оптимизация на уровне маппинга*
  - *Оптимизация на уровне сервиса DIS*
  - *Оптимизация на уровне кластера*
  - *Оптимизация на уровне кластера Hadoop*
- *Рекомендации*

#### **Модуль 17: Качество данных в Hadoop**

- *Процесс Data Quality*
- *Исследование потенциала данных*
- *Совместная работа разных ролей проекта*
- *Модификация, управление и мониторинг Data Quality*
- *Self Service Data Quality*
- *Запуск маппингов Data Quality на Hadoop*

#### **Модуль 18: Обработка сложных файлов**

- *Чтение сложных файлов*
- *Трансформация Data Processor*
- *Запись сложных файлов*
- *Производительность: Партиционирование*
- *Обработка и парсинг: Avro, Parquet, JSON, and XML file*
- *Договоренности при работе с Data Processor Transformation*



Общество с ограниченной ответственностью «Дата Интегрейшн Софтвер»  
Юридический адрес: 123022, г.Москва, ул. 2-я Звенигородская, д.13, стр. 37, этаж 6, помещение I, комната 10  
Адрес для корреспонденции: 125284, г. Москва, Ленинградский проспект, д. 31А, стр. 1, Бизнес-центр  
«МонАрх»  
ИНН 7713555858, ОКПО 77352347, ОКАТО 45277556000  
Тел.: + 7 (495) 645-0201, Факс: + 7 (495) 645-0188

## Модуль 19: Работа с NoSQL БД

- *CAP Theorem*
- *HBase*
- *MongoDB*
- *Cassandra*