



Описание семинара «Big Data 10.1 for Developers»

Слушатели семинара смогут лучше понять работу Informatica Big Data Management и узнать, как правильно оптимизировать решения, созданные на его базе для создания эффективных процессов загрузки хранилищ данных. Они смогут использовать механизмы переноса логики обработки на платформу Hadoop, а так же смогут изучить технику работы с NoSQL базами данных и процессы сложных обработок файлов.

В рамках семинара будут рассмотрены необходимые теоретические знания по продукту, а так же разработан набор лабораторных работ, помогающих слушателю лучше понять принципы работы продукта.

Слушатели, успешно прослушавшие данный семинар смогут:

- Работать с “Big Data”;
- Переносить логику вычислений на платформу Hadoop;
- Мигрировать маппинги PowerCenter в Big Data Management и исполнять их на платформе Hadoop;
- Мигрировать и обрабатывать данные на платформе Hadoop используя SQOOP и SQL to Mapping;
- Понимать место Informatica в архитектуре Hadoop;
- Обрабатывать данные на платформе Hadoop используя Informatica polyglot computing;
- Применять Informatica Smart Executor;
- Использовать средства мониторинга для поиска ошибок ;
- Работать со сложными данными, такими как JSON, AVRO и Parquet;
- Понимать как Informatica работает с NoSQL данными.

Семинар, в первую очередь, будет интересен следующим категориям специалистов:

- разработчикам, непосредственно работающим в проектах с Big Data;
- аналитикам, архитекторам и руководителям проектов, участвующим в проектах с Big Data;
- кроме того, к дополнительной целевой аудитории также относятся квалифицированные специалисты, желающие понять суть работы с Big Data и повысить свой профессиональный уровень

Ученики должны предварительно прослушать следующие курсы:

- Informatica Developer Tool 10.1 Big Data Management.

Продолжительность учебного курса составляет 3 дня (24 академических часа)

Каждый посетитель семинара должен иметь при себе ноутбук следующей конфигурации:

- современный процессор (core i3, i5, i7);
- объем оперативной памяти не меньше 4 Гбайт;



Общество с ограниченной ответственностью
«Дата Интегрейшн Софтвр»
127473, г. Москва, ул. Краснопролетарская, 16, стр.3, оф. 3-8
ИНН 7713555858, ОКПО 77352347, ОКАТО 452777568000
Тел. + 7 (495) 645-0201, Факс + 7 (495) 645-0188
info@dis-group.ru

- размер свободного места на жестком диске (на одном разделе) не меньше 10 Гбайт;
- мышь;
- установленная операционная система Windows XP (Professional)/7 (Professional, Ultimate) ;
- языковые настройки операционной системы русские;
- машина должна иметь права локального администратора;
- установленный web-браузер (Internet Explorer 9 и выше, Google Chrome 35.x) ;
- желательно наличие любого SQL средства для выполнение запросов на БД (Oracle).

Программа оказания консультационных услуг по курсу «Big Data 10.1 for Developers»

Module 1: Big Data Integration Course

- *Introduction*
- *Course Agenda*
- *Accessing the lab environment*
- *Related Courses*

Module 2: Big Data Basics

- *What is Big Data?*
- *Hadoop concepts*
- *Hadoop Architecture Components*
- *The Hadoop Distributed File System (HDFS)*
- *Purposes of a Name Node & Secondary Name Node*
- *MapReduce*
- *“Yet Another Resource Manager” (YARN) (MapReduce Version 2)*

Module 3: Data Warehouse Offloading

- *Challenges with traditional Data Warehousing*
- *The requirements of optimal Data Warehouse*
- *The Data Warehouse Offloading Process*

Module 4: Ingestion and Offload

- *PowerCenter Reuse Reports*
- *Importin PowerCenter Mappings to Developer*
- *SQOOP*
- *SQL to Mapping capability*
- *Partitioning and parallelism*



Module 5: Big Data Management

- *Architecture*
- *The Big Data world*
- *Build once, deploy anywhere*
- *The Informatica abstraction layer*
- *Polyglot computing*
- *The Smart Executor*
- *Open source and innovation*
- *Connection architecture*
- *Conections to third Party applications*

Module 6: Informatica Polyglot

- *Computing in Hadoop*
- *Hive MR/Tez*
- *Blaze*
- *Spark*
- *Native*
- *The Smart Executor*

Module 7: Mappings, Monitoring, and

- *Troubleshooting*
- *Configuring and running a mapping in Native and*
- *Hadoop environments*
- *Execution Plans*
- *Monitor mappings*
- *Troubleshoot mappings*
- *Viewing mapping results*

Module 8: Hadoop Data Integration Challenges and Performance Tuning

- *Describe challenges with executing mappings in*
- *Hadoop*
- *Big Data Management Performance Tuning*
- *Hive Environment Optimization*
 - *Mapping Level Tuning*
 - *DIS Level Tuning*
 - *Cluster Level Tuning*
 - *Hadoop environment and cluster tuning*
- *Tips*

Module 9: Data Quality on Hadoop



Общество с ограниченной ответственностью
«Дата Интегрейшн Софтвр»
127473, г. Москва, ул. Краснопролетарская, 16, стр.3, оф. 3-8
ИНН 7713555858, ОКПО 77352347, ОКАТО 452777568000
Тел. + 7 (495) 645-0201, Факс + 7 (495) 645-0188
info@dis-group.ru

- *The Data Quality process*
- *Discover insights into your data*
- *Collaborate and Create Data Improvement Assets*
- *Modify, Manage, and Monitor Data Quality*
- *Self Service Data Quality*
- *Executing Data Quality mappings on Hadoop*

Module 10: Complex File Parsing

- *The Complex file reader*
- *The Data Processor transformation*
- *The Complex file writer*
- *Performance Considerations: Partitioning*
- *Parsing and processing Avro, Parquet, JSON, and XML file*
- *Data Processor Transformation Considerations*

Module 11: Accessing NoSQL Databases

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