



ainfct

المؤسسة العربية للاستشارات والتدريب وتطوير الأداء
AINFCT | info@ainfct.com | www.ainfct.com

Big Data Technologies and Hadoop Training

فكرة الدورة

تحتاج المؤسسات التي تتعامل مع أحجام كبيرة ومتنوعة وسريعة من البيانات إلى معماريات قادرة على التخزين الموزع والمعالجة المرنة والتحليل واسع النطاق. يقدم هذا البرنامج من AINFCT مساراً فنياً متوسطاً لفهم تقنيات البيانات الضخمة مع تركيز عملي على Hadoop ومكوناته الرئيسية. يتناول البرنامج مفاهيم HDFS وYARN وMapReduce، ثم ينتقل إلى أدوات الاستعلام والمعالجة مثل Hive وPig وSpark، مع ربطها بجوانب الاستيعاب، النمذجة، الجودة، الحوكمة، والأمن. يراعي البرنامج الاحتياجات الفنية للمشاركين الذين يسعون إلى فهم كيفية تصميم مسارات بيانات كبيرة قابلة للتشغيل والمراقبة والتطوير. كما يعرض اتجاهات حديثة مثل بحيرات البيانات والمنصات السحابية دون فصلها عن الأسس المعمارية. يمنح البرنامج قيمة تطبيقية واضحة من خلال الجمع بين الفهم النظري، الأوامر العملية، وتصميم مسار بيانات نهائي قابل للنقاش والتحسين.

أهداف الدورة

فيما يلي الأهداف الرئيسية لهذا البرنامج التدريبي:

- شرح معمارية Hadoop ومكوناته الأساسية

- استخدام HDFS لإدارة البيانات الموزعة
- تنفيذ معالجات بيانات باستخدام MapReduce
- تطبيق استعلامات Hive على بيانات كبيرة
- ربط الجودة والأمن بمسارات البيانات
- تصميم مسار بيانات ضخم متكامل

منهجية الدورة

- شرح معماري مدعوم برسومات ومسارات تشغيلية.
- تطبيقات عملية على أوامر وأدوات Hadoop.
- تحليل حالات فنية لمشكلات الأداء والجودة.
- تمارين تصميم لمسارات استيعاب ومعالجة البيانات.
- مشروع ختامي لبناء تصور مسار بيانات كبير.

أثر الدورة على المنظمة

- يمكن تعزيز قدرات البيانات الضخمة داخل المنظمة من خلال:
- تحسين فهم البنى الموزعة للبيانات
 - رفع كفاءة معالجة البيانات كبيرة الحجم
 - تعزيز حوكمة البيانات داخل المسارات الفنية
 - دعم التكامل بين البيانات والتحليلات المؤسسية

أثر الدورة على المتدربين

يساعد البرنامج المشاركين على:

- فهم Hadoop من منظور معماري وعملي
- استخدام أدوات البيانات الضخمة بثقة
- تشخيص مشكلات الأداء الأولية
- تصميم مسارات بيانات قابلة للتوسع

الشهادات

شهادة معتمدة من AINFCT

الفئة المستهدفة

يناسب هذا البرنامج الفئات الفنية التي تعمل في بيئات البيانات أو التقنية وتحتاج إلى فهم عملي لتقنيات البيانات الضخمة. كما يفيد الفرق التي تشارك في تصميم أو تشغيل مستودعات وبحيرات البيانات.

- مهندسو ومحللو البيانات.
- مطورو نظم البيانات والتكامل.
- مسؤولو قواعد البيانات والأنظمة.
- فرق نكاء الأعمال والتحليلات.
- مختصو التقنية المهتمون بمنظومة Hadoop.

اليوم الأول: مدخل إلى البيانات الضخمة وHadoop

- خصائص البيانات الضخمة الأساسية
- تطور معماريات التخزين والمعالجة
- مكونات نظام Hadoop البيئي
- حالات الاستخدام الفنية الشائعة
- الفرق بين Hadoop والمنصات التقليدية

اليوم الثاني: معمارية Hadoop الأساسية

- دور HDFS في التخزين الموزع
- مفهوم الكتل والنسخ المتماثل
- إدارة الموارد باستخدام YARN
- مبادئ الاعتمادية والتوسع
- تدفق البيانات داخل العنقود

اليوم الثالث: نظام الملفات HDFS

- أوامر التعامل مع HDFS
- رفع الملفات وقراءتها وإدارتها
- الصلاحيات وتنظيم المسارات
- مراقبة السعة والتوزيع
- ممارسات حفظ البيانات الخام

اليوم الرابع: معالجة البيانات باستخدام MapReduce

- فكرة map وreduce
- تدفق تنفيذ الوظائف
- تصميم وظائف معالجة بسيطة

- قراءة السجلات وتجميع النتائج
- قيود MapReduce العملية

اليوم الخامس: Apache Hive للاستعلام

- مفهوم الجداول فوق البيانات
- إنشاء قواعد وجداول Hive
- كتابة استعلامات تحليلية
- التقسيم والتجزئة لتحسين الأداء
- متى يستخدم Hive بفاعلية

اليوم السادس: Apache Pig ومعالجة التدفقات الدفعية

- مفهوم Pig Latin
- بناء خطوات تحويل البيانات
- التجميع والربط والترشيح
- مقارنة Pig مع Hive
- تطبيقات معالجة البيانات شبه المنظمة

اليوم السابع: Apache Spark ضمن منظومة البيانات

- مفهوم المعالجة داخل الذاكرة
- DataFrames و Spark SQL
- مقارنة Spark مع MapReduce
- معالجة بيانات كبيرة بكفاءة
- مؤشرات الأداء الأساسية

اليوم الثامن: استيعاب البيانات إلى Hadoop

- أنماط إدخال البيانات
- استخدام Sqoop للبيانات العلائقية
- استخدام Flume للسجلات

- مبادئ Kafka في التدفق

- إدارة أخطاء الاستيعاب

اليوم التاسع: نمذجة البيانات للبيانات الضخمة

- اختيار صيغ التخزين المناسبة

- Parquet وORC والاستخدامات

- التقسيم حسب الزمن أو المجال

- التوازن بين الأداء والمرونة

- توثيق مخططات البيانات

اليوم العاشر: إدارة الجودة والحوكمة

- تعريف قواعد جودة البيانات

- التتبع lineage وفهم المصدر

- إدارة البيانات الوصفية

- تصنيف البيانات الحساسة

- مواءمة الممارسات مع DAMA-DMBOK

اليوم الحادي عشر: الأمن في Hadoop

- مبادئ المصادقة والتفويض

- دور Kerberos في الحماية

- التحكم بالوصول إلى البيانات

- تشفير البيانات عند الحاجة

- مراجعة السجلات والامتثال

اليوم الثاني عشر: إدارة العنقود والمراقبة

- مؤشرات صحة العقد والخدمات

- إدارة الموارد والوظائف

- متابعة السجلات والتنبيهات

- معالجة الاختناقات الشائعة
- خطط النسخ الاحتياطي والاسترداد

اليوم الثالث عشر: التكامل مع مستودعات البيانات

- ربط Hadoop بالتحليلات المؤسسية
- تصميم مناطق البيانات الخام والمعالجة
- دعم لوحات المعلومات والتقارير
- تغذية نماذج التحليل المتقدمة
- تجنب ازدواجية مصادر الحقيقة

اليوم الرابع عشر: الاتجاهات الحديثة

- بحيرات البيانات والمنصات السحابية
- Data Lakehouse ومفاهيمه
- الحاويات والتشغيل المرن
- المعالجة التدفقية شبه الفورية
- تقييم بدائل Hadoop الحديثة

اليوم الخامس عشر: مشروع تطبيقي ختامي

- تصميم مسار بيانات كبير
- تحميل البيانات وتخزينها
- تنفيذ استعلامات ومعالجات
- إضافة ضوابط جودة وأمن
- عرض التصميم والنتائج النهائية

يمتد البرنامج لمدة 15 يوماً تدريبياً، بواقع 4 ساعات يومياً. يخصص كل يوم لمحور فني محدد، مع مزيج من الشرح، الأوامر العملية، تحليل السيناريوهات، ومراجعة مخرجات التصميم.

course_assessment

يتم تقييم المشاركين من خلال تمارين تشغيلية قصيرة، ومراجعة تصميمات مسارات البيانات، ومشروع ختامي يجمع التخزين والمعالجة والجودة والأمن. يحصل المشاركون في نهاية البرنامج على شهادة حضور أو إتمام من AINFCT.

course_key_competencies

- معمارية البيانات الضخمة
- إدارة HDFS
- معالجة البيانات الموزعة
- استخدام Hive و Spark
- حوكمة البيانات الفنية
- أمن منصات Hadoop

مكتب القاهرة الرئيسي

القاهرة، جمهورية مصر العربية

7 شارع وهران، الطيران، مدينة نصر

201152466358+

info@ainfct.com

ainfct.com

رقم التسجيل الضريبي: 472920235

مكتب مدريد الفرعي

مدريد، إسبانيا

شارع الصحة 3، وسط المدينة، 28013 مدريد

training@ainfct.com

ainfct.com

