



ainfct

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

Advanced Data Analytics with Spark

فكرة الدورة

تتزايد حاجة المؤسسات إلى معالجة كميات كبيرة من البيانات بسرعة وموثوقية، مع القدرة على تحويل النتائج التحليلية إلى قرارات تشغيلية قابلة للتنفيذ. يركز هذا البرنامج التدريبي من AINFCT على بناء فهم متقدم لبيئة Apache Spark واستخداماتها في التحليل الموزع، إعداد البيانات، تحسين الأداء، وتنفيذ نماذج تحليلية قابلة للتوسع. يتناول البرنامج العلاقة بين بنية Spark، إدارة الذاكرة، Spark SQL، DataFrames، المعالجة الدفعية، والتحليل المتقدم ضمن بيئات بيانات حديثة. كما يوازن بين المفاهيم الفنية والتطبيقات العملية، بما يساعد المشاركين على تطوير حلول تحليل بيانات أكثر استقراراً وقابلية للتوسع. يقدم البرنامج قيمة مهنية واضحة من خلال ربط أدوات التحليل الموزع بمخرجات أعمال دقيقة ومنظمة.

أهداف الدورة

فيما يلي الأهداف الرئيسية لهذا البرنامج التدريبي:

- تحليل البيانات واسعة النطاق باستخدام Spark بفعالية
- تصميم مسارات تحليل موزعة قابلة للتوسع

- تحسين أداء الاستعلامات والمعالجات التحليلية
- تطبيق Spark SQL في سيناريوهات متقدمة
- إدارة جودة البيانات داخل بيئات موزعة
- تقييم مخاطر الأداء والاعتمادية

منهجية الدورة

- عروض تفاعلية تربط المفاهيم بالحالات العملية.
- تمارين تطبيقية فردية وجماعية على أدوات ونماذج العمل.
- مناقشات موجهة لتحليل التحديات المؤسسية ذات الصلة.
- دراسات حالة قصيرة لتعزيز نقل المعرفة إلى الواقع العملي.
- مراجعة ختامية للمخرجات وخطط التطبيق الشخصية.

أثر الدورة على المنظمة

يمكن تعزيز القيمة المؤسسية للبيانات من خلال:

- تحسين سرعة معالجة البيانات الكبيرة
- رفع كفاءة منصات التحليل المؤسسية
- دعم قرارات مبنية على بيانات موسعة
- تقليل اختناقات الأداء في التحليل

أثر الدورة على المتدربين

يساعد البرنامج المشاركين على:

- إتقان مفاهيم Spark المتقدمة
- تطوير حلول تحليل موزعة
- تشخيص مشكلات الأداء بكفاءة
- توظيف Spark في حالات عملية

الشهادات

شهادة معتمدة من AINFCT

الفئة المستهدفة

يناسب هذا البرنامج المشاركين الذين يعملون في مجالات البيانات أو التحليل أو التحسين المؤسسي، ويحتاجون إلى تطوير معرفة عملية قابلة للتطبيق في بيئات العمل.

- مختصو إدارة وتحليل البيانات.
- محللو الأعمال والعمليات.
- مشرفو التحول الرقمي وأنظمة المعلومات.
- فرق الجودة والامتثال والحوكمة.
- المهنيون المرشحون لأدوار بيانات متقدمة.

موضوعات الدورة

اليوم 1: مدخل متقدم إلى تحليلات البيانات الموزعة

- دور Spark في بيئات البيانات الحديثة
- الفروقات بين المعالجة المركزية والموزعة
- مكونات منظومة Spark الأساسية
- حالات الاستخدام المؤسسية للتحليلات الكبيرة
- متطلبات الحوكمة والجودة في التحليل

اليوم 2: بنية Apache Spark ومفاهيم التنفيذ

- Spark Driver وExecutors وآلية التنفيذ
- DAG والمهام والمراحل داخل Spark
- إدارة الذاكرة والتخزين المؤقت
- التعامل مع الفشل وإعادة التنفيذ
- تأثير البنية على الأداء

اليوم 3: إعداد بيئة العمل ومصادر البيانات

- تهيئة بيئات Spark المحلية والسحابية
- الاتصال بملفات CSV وParquet وJSON
- قراءة البيانات من قواعد البيانات
- إدارة مخططات البيانات Schema
- ضبط إعدادات الجلسات التحليلية

اليوم 4: DataFrames والتحويلات الأساسية

- إنشاء DataFrames ومعالجة الأعمدة
- استخدام التحويلات والإجراءات Actions
- تنظيف القيم المفقودة والشاذة
- دمج وتصفية مجموعات البيانات
- تطبيق قواعد جودة أولية

اليوم 5: Spark SQL للتحليل المتقدم

- إنشاء الجداول والعروض المؤقتة
- تنفيذ الاستعلامات التحليلية المعقدة
- استخدام التجميعات والنوافذ التحليلية
- تحسين خطط تنفيذ SQL
- مقارنة SQL مع DataFrame API

اليوم 6: إدارة البيانات شبه المهيكلة

- تحليل JSON والهياكل المتداخلة
- تسطيح الحقول متعددة المستويات
- معالجة Arrays و Structs
- تحويل البيانات إلى صيغ تحليلية
- ضبط المخططات الديناميكية

اليوم 7: تحسين الأداء في Spark

- فهم Catalyst Optimizer
- استخدام Partitioning بفعالية
- إدارة Shuffle وتقليل تكلفته
- تطبيق Caching و Persisting بوعي
- قراءة خطط التنفيذ وتفسيرها

اليوم 8: معالجة البيانات الدفعية Batch Processing

- تصميم وظائف تحليل دُفعية
- تنظيم خطوات الاستخراج والتحويل
- تطبيق قواعد التحقق قبل التحليل
- إدارة مخرجات المعالجة
- توثيق نتائج التشغيل

اليوم 9: التحليل الإحصائي داخل Spark

- حساب المؤشرات الإحصائية الأساسية
- تحليل العلاقات والارتباطات
- استخدام العينات Sampling
- بناء ملخصات تحليلية واسعة
- تفسير النتائج ضمن سياق الأعمال

اليوم 10 : التكامل مع Python وPySpark

- استخدام PySpark للتحليل العملي
- تنظيم الشيفرة التحليلية
- إنشاء وظائف قابلة لإعادة الاستخدام
- التعامل مع UDFs بحذر
- إدارة الأخطاء أثناء التنفيذ

اليوم 11 : التمثيل والتحضير للنمذجة

- إعداد الخصائص Features للتحليل
- ترميز المتغيرات وتصنيفها
- توحيد المقاييس والقيم
- تقسيم البيانات للتدريب والاختبار
- التحقق من جاهزية البيانات للنمذجة

اليوم 12 : مقدمة إلى Spark MLlib

- مفاهيم خطوط النمذجة Pipelines
- تطبيق نماذج تصنيف أساسية
- تطبيق نماذج انحدار عملية
- تقييم النماذج بمقاييس مناسبة
- ضبط المعلمات بشكل منهجي

اليوم 13: المراقبة والتشخيص التشغيلي

- استخدام Spark UI في التحليل
- تتبع المهام البطيئة والفاشلة
- تحليل استهلاك الموارد
- تحديد أسباب بطء التنفيذ
- توثيق نتائج التشخيص

اليوم 14: الحوكمة والأمن في بيئات Spark

- إدارة الوصول إلى البيانات
- حماية البيانات أثناء المعالجة
- تتبع Lineage للبيانات
- التوافق مع سياسات الاحتفاظ
- تطبيق ضوابط الجودة والامتثال

اليوم 15: مشروع تطبيقي متكامل

- تصميم حالة تحليل بيانات كبيرة
- بناء مسار معالجة باستخدام Spark
- تحسين الأداء بناءً على القياسات
- إعداد مخرجات تحليلية قابلة للعرض
- مراجعة الدروس والتحسينات

course _daily_ schedule

يمتد البرنامج لمدة 15 أيام تدريبية، بواقع 4 ساعات يومياً. يتضمن كل يوم مدخلاً مفاهيمياً، تطبيقات عملية، مناقشات موجهة، وتمارين مراجعة لضمان ترسيخ المفاهيم وربطها بسياق العمل.

course_assessment

يعتمد التقييم على المشاركة، التمارين التطبيقية، المناقشات، ومخرجات الورش العملية. يحصل المشاركون الذين يستكملون متطلبات الحضور والمشاركة على شهادة إتمام/حضور صادرة عن AINFCT.

course_key_competencies

- التحليل الموزع للبيانات
- تحسين أداء Spark
- هندسة مسارات البيانات
- تحليل البيانات الكبيرة
- جودة البيانات التشغيلية
- إدارة موارد المعالجة

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