CONCEPT

دورة Apache Kafka لتدفق البيانات اللحظي: Connect Kafka والتقاط تغييرات البيانات CDC

10 – 14 مايو 2027

لندن - فندق أربع نجوم في وسط لندن



دورة Apache Kafka لتدفق البيانات اللحظي: Kafka Connect والتقاط تغييرات البيانات CDC

المرجع: 23971_1453 **التاريخ:** 10 – 14 مايو 2027 **المكان:** لندن - فندق أربع نجوم في وسط لندن **الرسوم:** SAR 25300

مقدمة:

دورة Apache Kafka لتدفق البيانات اللحظي، بمحاورها في Kafka Connect والتقاط تغييرات البيانات CDC، دورة مدتها أسبوع لمهندسي البيانات ومطوري الأنظمة الخلفية ومهندسي المنصات، تنتهي بخط تدفق جاهز ينقل تغييرات قواعد البيانات إلى منصة التحليلات مع دليل تشغيله. ما زالت منشآت كثيرة تنقل بياناتها التشغيلية بدفعات ليلية أو عبر تكاملات هشة بين كل نظامين، فتصل مؤشرات الاحتيال وأرصدة المخزون وأحداث العملاء متأخرة ساعات، ويحتاج كل مستهلك جديد إلى تغذية مخصصة. يكتب المرشحون الشيفرة أو يديرون منصات البيانات في عملهم، ويقضون كل يوم في مختبرات عملية على عنقود Kafka قيد التشغيل. وتقدم كور كونسبت هذه الدورة في تدفق البيانات اللحظي.

أهداف الدورة:

- تصميم توزيع الموضوعات والأقسام والنسخ المتماثلة على عنقود Kafka يعمل بنمط KRaft لإدارة البيانات الوصفية بما يحقق أهداف الترتيب والمتانة والإنتاجية
- ضبط المنتجين ومجموعات المستهلكين لتسليم الرسائل مرة واحدة على الأقل أو مرة واحدة بالضبط عبر خاصية عدم التكرار والمعاملات والتحكم في تثبيت الإزاحات
- التقاط تغييرات صفوف قواعد البيانات بموصلات المصدر في Debezium العاملة على Kafka Connect وتوجيهها إلى الموضوعات وأنظمة الوجهة
- حوكمة صيغ الأحداث بسجل المخططات ومحولات Avro وProtobuf وقواعد التوافق التي تتيح للمنتجين تطوير رسائلهم دون كسر المستهلكين
- بناء مهام معالجة تدفقات ذات حالة في Kafka Streams وApache Flink باستخدام نوافذ وقت الحدث وعلامات التقدم وربط التدفقات بالجداول
- تأمين منصة التدفق ومراقبتها وضبط أدائها بتشفير TLS ومصادقة SASL وقوائم صلاحيات ACL وتنبيهات تأخر المستهلكين وإعادة توزيع الأقسام

الفئة المستهدفة:

- مهندسو البيانات المسؤولون عن خطوط الاستيعاب التي يجب أن تنقل التغييرات التشغيلية خلال ثوان بدلا من دفعات ليلية
- مطورو الأنظمة الخلفية المسؤولون عن خدمات تنشر أحداث الأعمال أو تستهلكها بين الفرق
- مهندسو المنصات وموثوقية المواقع المسؤولون عن تشغيل عناقيد Kafka وتأمينها وتوسيعها
- مهندسو التكامل المسؤولون عن استبدال الربط المباشر بين الأنظمة بموضوعات أحداث مشتركة
- مهندسو التحليلات المسؤولون عن تغذية لوحات المتابعة وفهارس البحث ومخازن الخصائص من البيانات الحية

محاور الدورة:

اليوم 1: البنية الموجهة بالأحداث وأسس سجل Kafka

- مقارنة البنية الموجهة بالأحداث بالتكامل القائم على الطلب والاستجابة
- نموذج سجل الإيداع في Kafka بالموضوعات والأقسام والإزاحات
- أدوار الوسطاء ونصاب وحدات التحكم في KRaft لإدارة البيانات الوصفية
- معامل النسخ المتماثل والنسخ المتوافقة وسلوك انتخاب القائد
- حصر حالات استخدام التدفق وربطها بمتطلبات زمن الاستجابة

اليوم 2: المنتجون ومجموعات المستهلكين ودلالات التسليم

- تجميع رسائل المنتج ومهلة الانتظار وخيارات خوارزميات الضغط
- تصميم مفاتيح التقسيم وضمانات الترتيب ومخاطر الأقسام المكتظة
- إعادة توازن مجموعات المستهلكين والإسناد التعاوني والعضوية الثابتة
- استراتيجيات تثبيت الإزاحات للتسليم مرة على الأكثر ومرة على الأقل
- المنتجون عديمو التكرار ومعاملات Kafka للتسليم مرة واحدة بالضبط

اليوم 3: موصلات Kafka Connect والتقاط التغييرات بأداة Debezium وإدارة المخططات

- عمال Kafka Connect الموزعون والمهام وإعدادات المحولات
- موصلات المصدر في Debezium وقراءة سجلات معاملات قواعد البيانات
- تحويلات الرسالة المفردة وتوجيه أحداث التغيير إلى الموضوعات
- موضوعات سجل المخططات مع محولات Protobufg Avro
- قواعد التوافق الخلفي والأمامي لتطور مخططات الأحداث

اليوم 4: معالجة التدفقات والأمن وتشغيل العنقود

- طوبولوجيا Kafka Streams بعناصر KStream وKTable ومخازن الحالة
- وقت الحدث وعلامات التقدم ومعالجة الأحداث المتأخرة في Apache Flink
- النوافذ المتعاقبة والمتداخلة ونوافذ الجلسات مع ربط التدفق بالجدول
- تشفير TLS ومصادقة SASL وقواعد التفويض بقوائم ACL
- تنبيهات تأخر المستهلكين ومقاييس JMX وضبط إعادة توزيع الأقسام

اليوم 5: مختبر بناء خط تدفق من التقاط التغييرات إلى التحليلات

- المفاضلة بين خدمات Kafka السحابية المدارة والعنقود المدار ذاتيا
- مختبر التقاط قاعدة بيانات الطلبات بأداة Debezium إلى الموضوعات
- مختبر تجميع نافذي في Flink يغذي مخزن التحليلات
- مختبر تمارين الأعطال لفقدان وسيط وإعادة توازن المستهلكين
- استكمال خط التدفق من التقاط التغييرات إلى التحليلات ودليل تشغيله

المهارات التي ستكتسبها:

- هندسة بنية بث الأحداث
- تصميم الموضوعات والأقسام
- ضبط ضمانات التسليم
- هندسة التقاط تغييرات البيانات
- حوكمة تطور المخططات
- معالجة التدفقات ذات الحالة
- تحصين عناقيد Kafka
- مراقبة منصات التدفق

لماذا تحضر هذه الدورة:

- تسليم خط تدفق من التقاط التغييرات إلى التحليلات مع دليل تشغيله إلى قائد منصة البيانات مرجعا لحالة الاستخدام التالية
- تحديد عدد الأقسام وإعدادات النسخ المتماثل وضمانات التسليم والاختيار بين Kafka Streams وApache Flink لكل مهمة معالجة
- تجنب الأحداث المكررة أو المفقودة وتغييرات المخططات الكاسرة وتراكم تأخر المستهلكين والموضوعات غير المؤمنة قبل وصولها إلى بيئة الإنتاج
- توجيه الفرق الأخرى في إعداد الموصلات وفحوص توافق المخططات وأنماط معالجة التدفقات التي جربت في المختبرات

الخاتمة:

بعد العودة إلى العمل يسلم المشارك خط التدفق من التقاط التغييرات إلى التحليلات ودليل تشغيله إلى قائد منصة البيانات وأصحاب قواعد البيانات المصدر. ويعتمد الفريق عليه في الاتفاق على تسمية الموضوعات ومدة الاحتفاظ وتوافق المخططات وإعدادات الأمن، وفي إقرار الجداول التشغيلية التي تبث تاليا والمستهلكين المسموح لهم بقراءتها. وبعد أسابيع من التشغيل تحت حمل فعلي تراجع الوحدة زمن الوصول من الطرف إلى الطرف وذروات تأخر المستهلكين وإعادات تشغيل الموصلات وتغييرات المخططات المرفوضة ونمو أقراص الوسطاء، ثم تعيد ضبط الأقسام والاحتفاظ حيث تبتعد النتائج عن الأهداف.

الأسئلة الشائعة:

ما الذي يحتاج المشارك إلى معرفته قبل دورة Apache Kafka لتدفق البيانات اللحظي؟

يكفي أن يكتب المشارك الشيفرة بلغة Java أو Python أو غيرهما، وأن يستخدم SQL وسطر أوامر Linux، وأن يفهم أساسيات الشبكات. ويحتاج إلى حاسوب محمول يشغل الحاويات، إذ تعمل في المختبرات محليا عناقيد Kafka وقاعدة بيانات مصدر ومهام المعالجة.

بم تختلف دورة Apache Kafka لتدفق البيانات اللحظي عن دورة تحميل البيانات بالدفعات أو تحليل البيانات الضخمة؟

تتعامل هذه الدورة مع الأحداث لحظة وقوعها: الموضوعات وضمانات التسليم والتقاط التغييرات ومعالجة التدفقات ذات الحالة وتشغيل العنقود. أما دورات التحميل بالدفعات والمستودعات فتنقل البيانات وفق جدول، ودورات تحليل البيانات الضخمة تركز على الاستعلام في بيانات مخزنة لا على تشغيل منصة أحداث حية.

لماذا تجمع الفرق بين Apache Kafka والتقاط تغييرات البيانات في تدفق البيانات اللحظي؟

لأن التقاط تغييرات البيانات يقرأ التغييرات المثبتة من سجل معاملات قاعدة البيانات وينشرها أحداثا، فتتلقى الأنظمة الأخرى عمليات الإدراج والتحديث والحذف خلال ثوان دون استعلامات استطلاع متكررة ودون تعديل التطبيق المصدر.

ماذا يعود به المشارك من دورة Apache Kafka لتدفق البيانات اللحظي؟

يعود المشارك بخط تدفق عامل من التقاط التغييرات إلى التحليلات مع إعدادات الموصلات والمخططات ومعالجة التدفقات، إضافة إلى دليل تشغيل يغطي إعدادات الأمن وتنبيهات المراقبة وحدود التأخر وخطوات التعافي من الأعطال.