

Real-Time Pipeline Accelerator

Mulai dengan Spark Streaming, Kafka, dan Cassandra untuk analisis data waktu-nyata. BlueData menyediakan solusi turnkey dengan perangkat lunak dan layanan untuk mempercepat penyebaran Anda dan membangun jaringan data real-time end-to-end dalam hitungan menit. Kecepatan adalah pendorong utama untuk semakin banyak kasus penggunaan analitik, mulai dari deteksi penipuan untuk transaksi keuangan hingga pemantauan Internet of Things (IoT) dengan data yang dihasilkan sensor. Analisis langsung dari aliran data baru ini secara real-time dapat membawa nilai luar biasa baik dalam memberikan keunggulan bisnis yang kompetitif, menghindari potensi krisis, atau menciptakan peluang pendapatan baru. Jika organisasi Anda ingin mengaktifkan pembuatan prototipe dan pengembangan cepat analitik waktu nyata, sekarang ada solusi di tempat untuk membantu Anda memulai dengan cepat dan mudah.

Spark Streaming + Kafka + Cassandra

BlueData memudahkan untuk menyebarkan infrastruktur dan aplikasi Spark di lokasi. Platform perangkat lunak BlueData EPIC dibuat khusus untuk menyederhanakan dan mempercepat penyebaran Spark, Hadoop, dan alat-alat lain untuk analisis Big Data yang memanfaatkan wadah Docker dan infrastruktur virtual. Solusi Accelerator Pipeline Real-Time kami yang baru menyediakan perangkat lunak dan layanan profesional yang Anda butuhkan untuk membangun jaringan pipa data di lingkungan multi-tenant untuk Spark Streaming, Kafka, dan Cassandra. Dengan solusi ini, ilmuwan data Anda akan dapat membuat jaringan pipa terintegrasi untuk menangkap, menganalisis (model / skor), dan menyimpan aliran data waktu-nyata dalam hitungan menit. Sekarang para ilmuwan dan pengembang data Anda dapat fokus pada kasus penggunaan dan jalur pipa, tanpa khawatir tentang kerumitan infrastruktur teknologi seperti Spark, Kafka, dan Cassandra. Dan saat kasing Anda menjadi matang dan meluas dari waktu ke waktu, Anda dapat menggunakan platform EPIC BlueData untuk memperluas jaringan pipa Anda dengan aplikasi dan kerangka kerja pelengkap. Saat ini, ada aliran data yang terus-menerus dalam volume besar yang dihasilkan dari pasar keuangan, sensor, log mesin, media sosial, aplikasi seluler, dan banyak sumber lainnya. Daripada menyusun data ini dan menganalisisnya setelah disimpan, menjadi semakin penting untuk menganalisis data secara waktu nyata sebagai peristiwa sedang terjadi untuk membuat keputusan instan dan mengambil tindakan segera. Data ini seringkali mudah rusak dan dapat kehilangan nilai operasionalnya dalam jangka waktu yang sangat singkat. Kecepatan adalah esensi. Dalam pengertian umum, jalur pipa data terdiri dari semua langkah yang diperlukan untuk menangkap, menyiapkan, dan memproses data Anda untuk analisis. Membangun jalur pipa data untuk analisis waktu nyata memerlukan teknologi dan metodologi khusus:

- Waktu nyata berarti Anda tidak bisa menunggu sehari-hari, berjam-jam, atau bahkan beberapa menit sebelum menjalankan analisis Anda dan menghasilkan wawasan alat-alat tersebut perlu mendukung persyaratan mendasar ini.
- Teknologi dan kerangka kerja untuk analitik waktu-nyata berbeda dari sistem perusahaan yang ada dan kerangka kerja pemrosesan data berorientasi batch.
- Ada beberapa komponen (baik perangkat lunak dan infrastruktur) yang diperlukan untuk menangkap dan menyimpan data streaming, memproses dan menganalisis beberapa stream dalam batch kecil, dan menyimpan hasil yang berharga ini untuk analisis berkelanjutan.
- Merupakan hal yang kompleks dan memakan waktu untuk mengumpulkan semua infrastruktur dan perangkat lunak yang diperlukan, dan sebagian besar organisasi tidak memiliki keterampilan untuk menggunakan dan menyatukan semua komponen ini.

Menerapkan Trinity Real-Time

Untuk para ilmuwan dan pengembang data yang bekerja dengan jalur pipa data waktu nyata, tumpukan Spark-Kafka-Cassandra dengan cepat muncul sebagai tempat terbaik untuk memulai. Trinitas baru ini menghasilkan persyaratan utama secara real-time analisis data:

- Spark: mesin pemrosesan data dalam memori yang cepat, dan teknologi open source Apache yang paling cepat berkembang. Spark Streaming adalah perpanjangan dari API Spark inti; memungkinkan integrasi data waktu nyata dari aliran peristiwa yang berbeda.

- Kafka: sistem pengiriman pesan untuk menangkap dan mempublikasikan aliran data. Dengan Spark Anda dapat mengambil data dari

Kafka, filter yang mengalir ke kumpulan data yang lebih kecil, menambah data, dan kemudian mendorong kumpulan data yang disempurnakan

ke penyimpanan data persisten.

- Cassandra: data ini perlu ditulis ke database operasional yang dapat diskalakan dan tangguh seperti Cassandra untuk kegigihan, pengembangan aplikasi yang mudah, dan analitik waktu-nyata.

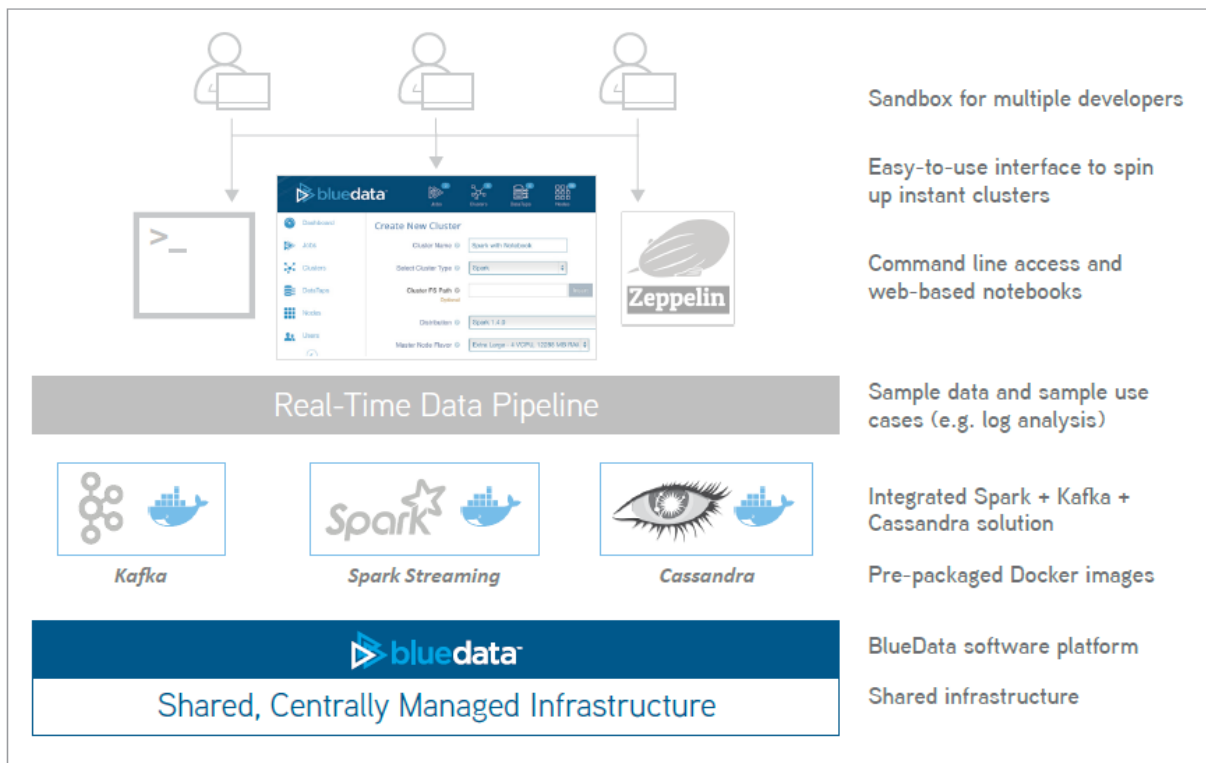
Kerangka kerja open source ini sangat kuat dan cocok untuk persyaratan membangun pipa data real-time. Namun, mungkin perlu berminggu-minggu bahkan berbulan-bulan bagi tim Anda untuk menggenjot produksinya dan memulai. Misalnya, Anda mungkin perlu melatih setidaknya satu atau dua anggota tim tentang Spark, Kafka, dan Cassandra (biasanya satu minggu pelatihan atau lebih, untuk setiap kerangka kerja). Anda perlu membangun integrasi pipa antara kerangka kerja yang berbeda ini dan mengujinya secara internal pada infrastruktur Anda (membutuhkan setidaknya beberapa minggu lagi atau lebih).

Membuat laboratorium Spark-Kafka-Cassandra untuk banyak ilmuwan dan pengembang data untuk mendukung pembuatan prototipe cepat, pengembangan, pengujian, dan jaminan kualitas dapat menjadi inisiatif yang menantang dan mahal. Dan ketika Anda mulai tambahkan lebih banyak kasus penggunaan dan pengguna, Anda perlu memperluas infrastruktur, menyediakan lebih banyak perangkat keras, dan mengintegrasikan lebih banyak alat. Solusi Akselerator Pipa Real-Time BlueData dirancang untuk mengatasi tantangan-tantangan ini sehingga menjadi sederhana dan mudah untuk bangkit dan berjalan dengan tumpukan baru ini untuk analitik waktu-nyata.

Real-Time Pipeline Accelerator

Dengan solusi Accelerator Pipeline Real-Time baru dari BlueData, organisasi Anda akan memiliki lingkungan lab multitenant yang siap dijalankan untuk Spark Streaming, Kafka, dan Cassandra.

Gambar di bawah ini menggambarkan contoh lingkungan kotak pasir untuk banyak ilmuwan dan pengembang data (penyewa) dengan platform perangkat lunak BlueData EPIC yang berjalan di atas infrastruktur bersama yang hemat biaya. BlueData menyediakan antarmuka yang mudah digunakan dan dukungan di luar kotak untuk notebook Zeppelin, akses baris perintah, dan alat JDBC lainnya bersama dengan data sampel dan kasus penggunaan untuk jaringan pipa real-time. Pengembang dan ilmuwan data dapat menyediakan sendiri komponen-komponen utama untuk pipa data real-time mereka dalam hitungan menit menggunakan gambar Docker yang telah dikemas sebelumnya untuk Spark, Kafka, dan Cassandra.



Komponen utama dari solusi Accelerator Pipa Real-Time meliputi:

- lisensi berlangganan 1 tahun untuk BlueData EPIC Enterprise hingga 60 core fisik (sekitar 5 node)
- Gambar Docker yang sudah dikonfigurasi sebelumnya untuk Spark, Kafka, dan Cassandra
- Manajer kluster berbasis UI untuk membangun jalur pipa data waktu nyata dengan Spark Streaming, Kafka dan Cassandra
- Contoh mulai cepat dan dokumentasi sampel untuk dua jalur pipa data ujung ke ujung (contoh: Analisis log)
- 5 hari layanan profesional jarak jauh dan konsultasi dari para ahli BlueData