

IBM SPSS Analytic Server
Versão 3.1.1

Visão geral

IBM

Nota

Antes de utilizar estas informações e o produto suportado por elas, leia as informações em “Avisos” na página 5.

Informações sobre o Produto

Esta edição se aplica à versão do 3, liberação 1, modificação 1 do IBM SPSS Analytic Server e a todas as liberações e modificações subsequentes até que seja indicado de outra forma em novas edições.

Índice

Visão Geral (Overview)	1	Avisos	5
Arquitetura	2	Marcas Registradas	7
Spark e Analytic Server	2		
Novidades na versão 3.1.1	3		

Visão Geral (Overview)

O IBM® SPSS Analytic Server é uma solução para analítica de Big Data que combina a tecnologia IBM SPSS com sistemas Big Data e permite trabalhar com interfaces com o usuário do IBM SPSS familiares para resolver problemas em uma escala previamente inalcançável.

Porque a Analítica de Big Data É Importante

Os volumes de dados coletados pelas organizações estão crescendo exponencialmente; por exemplo, negócios financeiros e varejistas têm todas as transações dos clientes por um ano (ou dois anos, ou dez anos), provedores telco têm Call Data Records (CDR) e leituras de sensor de dispositivo e as empresas de Internet têm os resultados de crawls da web.

A analítica de big data é necessária onde existe:

- Um grande volume de dados (terabytes, petabytes, exabytes), principalmente quando há uma mistura de dados estruturados & dados não estruturados
- Dados que são alterados/acumulados rapidamente

A analítica de big data também ajuda quando:

- Um grande número (milhares) de modelos está sendo construído
- Modelos são construídos/atualizados frequentemente

Desafios

As mesmas organizações que coletam grandes volumes de dados costumam ter dificuldades ao fazerem uso deles por vários motivos:

- A arquitetura dos produtos de analítica tradicionais não é adequada para cálculo distribuído, e
- Algoritmos estatísticos existentes não são projetados para trabalharem com big data (esses algoritmos esperam que os dados cheguem até a eles, mas os big data são muito caros para se mover); dessa forma
- Executar uma analítica moderna em big data requer novas qualificações e profundo conhecimento de sistemas de big data. Poucos analistas têm essas qualificações.
- As soluções em memória trabalham para solucionar problemas de médio porte, mas não se adaptam muito bem a dados realmente grandes.

Solução

Analytic Server fornece:

- Uma arquitetura centrada em dados que alavanca sistemas de big data, como Hadoop Map/Reduce, como dados em HDFS.
- Uma interface definida para incorporar novos algoritmos estatísticos projetados para acessar dados.
- Interfaces com o usuário do IBM SPSS familiares que ocultam os detalhes dos ambientes de big data para que os analistas possam se concentrar na análise dos dados.
- Uma solução escalável para problemas de qualquer porte.

Arquitetura

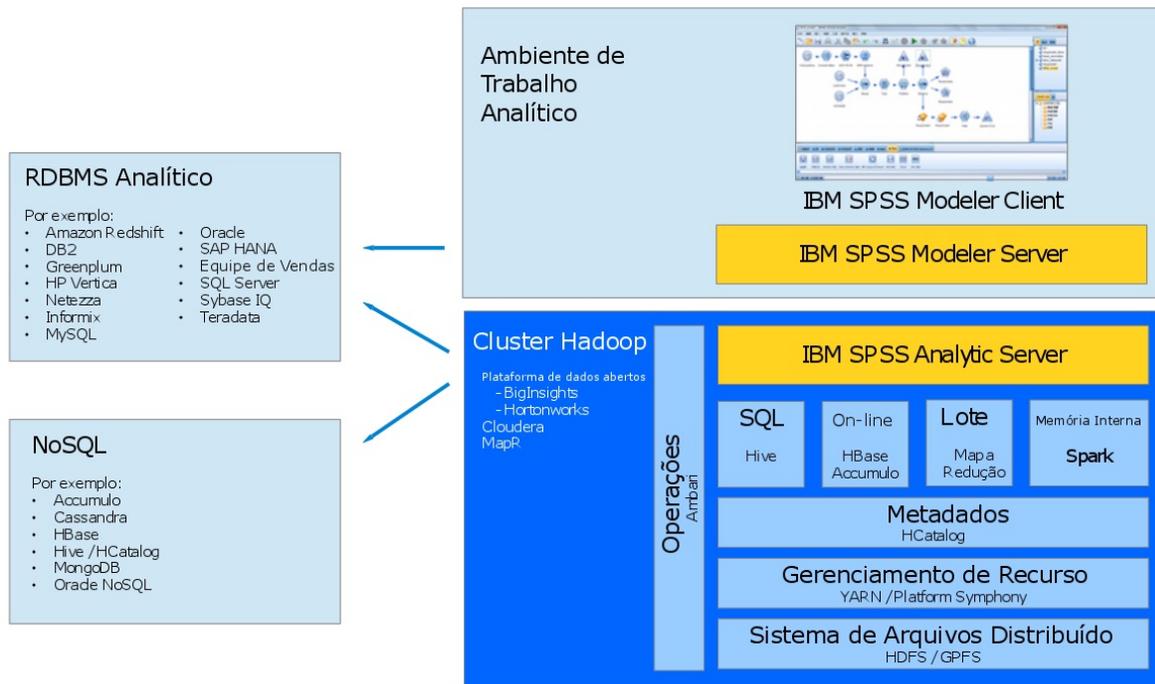


Figura 1. Arquitetura

O Analytic Server está situado entre um aplicativo cliente e uma nuvem Hadoop. Supondo que os dados residem na nuvem, a estrutura geral de tópicos para trabalhar com o Analytic Server está para:

1. Definir as origens de dados do Analytic Server sobre os dados na nuvem.
2. Definir a análise que você deseja executar no aplicativo cliente. Para a liberação atual, o aplicativo cliente é IBM SPSS Modeler.
3. Quando você executa a análise, o aplicativo cliente envia uma solicitação de execução do Analytic Server.
4. O Analytic Server orquestra a tarefa para ser executada na nuvem Hadoop e relata os resultados ao aplicativo cliente.
5. É possível usar os resultados para definir análises posteriores, e as repetições de ciclo.

Spark e Analytic Server

O Analytic Server integra-se com o Apache Spark para aumentar o desempenho.

Quando o Spark é e não é usado

Se o Spark estiver instalado em um serviço do Ambari no cluster do Hadoop, então, o Analytic Server o usa para processar tarefas de big data. As diretrizes a seguir se aplicam para determinar quando o Spark não é usado.

1. Se o conjunto de dados for menor que 128 MB, então, o Analytic Server usa a função MapReduce integrada na JVM do Analytic Server e não utiliza o Spark ou o cluster do Hadoop.
2. Se o Spark não estiver instalado no cluster, então, o Analytic Server usa o MapReduce v2.

3. O Analytic Server usa o MapReduce v2 para construir modelos de PSM. Quando uma tarefa é finalizada com uma construção de modelo de PSM, o Analytic Server usa o Spark para processar a tarefa por meio de todas as etapas que levam à construção de modelo e, em seguida, grava no disco e, então, usa MapReduce para construir o modelo de PSM. Por exemplo, se uma tarefa inclui uma junção seguida por uma construção de modelo de PSM, a junção é executada no Spark e o PSM é executado nos dados associados em MapReduce.

Como o Spark é usado

Quando o serviço do Analytic Server é iniciado e descobre que o Spark está disponível, ele inicializa uma "tarefa de Hadoop do Spark" que permite comunicação entre tarefas distribuídas ao longo do cluster. Essa tarefa é executada contanto que o serviço do Analytic Server seja executado e é usada para todas as execuções do Analytic Server. Essa abordagem melhora o desempenho relativo à orquestração de múltiplas tarefas de Hadoop MapReduce, porque ela elimina a sobrecarga de recarregar todos os componentes do Analytic Server para cada Tarefa de Hadoop.

O Spark é capaz de executar tarefas MapReduce. Isso permite que o Analytic Server use algoritmos "nativos" do Spark como junção, ordem e união onde disponível. Ao mesmo tempo, o Analytic Server pode executar algoritmos existentes de Mapa do SPSS e de Redução no Spark e sem usar diretamente a API do Hadoop.

Novidades na versão 3.1.1

Versão 3.1.1

Plataforma (Platform)

- Suporte para Cloudera 5.11 e 5.12
- Suporte para o Ubuntu Linux 16.04 (com o Hortonworks Data Platform 2.6 e o Cloudera 5.11)
- Cloudera 5.8 e 5.9 não são mais suportados
- Big Insights 4.1, 4.2 e 4.2.5 não são mais suportados
- MapR 5.0 não é mais suportado

Fontes de Dados

- Suporte para Apache Hive 2.1
- MongoDB 2.6 não é mais suportado
- MySQL 5.1 não é mais suportado

Aperfeiçoamentos no Desempenho

- Um procedimento de instalação off-line mais automatizado agora está disponível para HDP. Para obter mais informações, consulte Instalação off-line
- Suporte para diversos clusters. O recurso de diversos clusters é um aprimoramento da habilidade de Alta Disponibilidade do IBM SPSS Analytic Server e fornecer um isolamento melhorado em ambientes com diversos locatários. Por padrão, a instalação do serviço do Analytic Server (tanto em Ambari como em ClouderaManager) resulta na definição de um único cluster de servidor analítico.
- Agora é possível configurar filas YARN separadas para cada locatário do Analytic Server. Para obter mais informações, consulte Configurando filas YARN separadas para cada locatário do Analytic Server - HDP ou Configurando filas YARN separadas para cada locatário do Analytic Server - Cloudera.
- O suporte agora é fornecido por meio de amostragem de SQL pushback.
- Os seguintes aprimoramentos do Spark ML estão agora disponíveis:
 - Suporte para bifurcação de k-médias
 - Suporte para XGBoost

- Suporte para regressão isotônica
- O cruzamento da origem de dados do Spark RDD agora pode ser compartilhado com diferentes tarefas de ASL.

- A seção Melhores práticas e recomendações do IBM SPSS Analytic Server foi atualizada para incluir informações específicas do Cloudera.

Para obter informações mais atualizadas sobre os requisitos do sistema, use os relatórios de requisitos do sistema Detalhados no site de Suporte Técnico da IBM: <http://publib.boulder.ibm.com/infocenter/prodguid/v1r0/clarity/softwareReqsForProduct.html>. Nesta página:

1. Digite SPSS Analytic Server como o nome do produto e clique em **Procurar**.
2. Selecione a versão e o escopo do relatório desejados e, em seguida, clique em **Enviar**.

Avisos

Estas informações foram desenvolvidas para os produtos e serviços oferecidos nos EUA. Esse material pode estar disponível a partir da IBM em outros idiomas. No entanto, pode ser solicitado que você tenha uma cópia do produto ou da versão de produto nesse idioma para acessá-lo.

É possível que a IBM não ofereça os produtos, serviços ou recursos discutidos neste documento em outros países. Consulte seu representante IBM local para obter informações sobre os produtos e serviços disponíveis atualmente em sua área. Qualquer referência a um produto, programa ou serviço IBM não significa que apenas produtos, programas ou serviços IBM possam ser usados. Qualquer produto, programa ou serviço funcionalmente equivalente, que não infrinja nenhum direito de propriedade intelectual da IBM poderá ser usado em substituição. Entretanto, a avaliação e verificação da operação de qualquer produto, programa ou serviço não IBM são de responsabilidade do usuário.

A IBM pode ter patentes ou solicitações de patentes pendentes relativos a assuntos tratados nesta publicação. O fornecimento desta publicação não lhe garante direito algum sobre tais patentes. Pedidos de licença devem ser enviados, por escrito, para:

Gerência de Relações Comerciais e Industriais da IBM Brasil
Av. Pasteur, 138-146
Botafogo
Rio de Janeiro, RJ
CEP: 22296-903

Para pedidos de licença relacionados a informações de DBCS (Conjunto de Caracteres de Byte Duplo), entre em contato com o Departamento de Propriedade Intelectual da IBM em seu país ou envie pedidos de licença, por escrito, para:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

A INTERNATIONAL BUSINESS MACHINES CORPORATION FORNECE ESTA PUBLICAÇÃO "NO ESTADO EM QUE SE ENCONTRA", SEM GARANTIA DE NENHUM TIPO, SEJA EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS NÃO SE LIMITANDO A, AS GARANTIAS IMPLÍCITAS DE NÃO-INFRAÇÃO, COMERCIALIZAÇÃO OU ADEQUAÇÃO A UM DETERMINADO PROPÓSITO. Alguns países não permitem a exclusão de garantias expressas ou implícitas em certas transações; portanto, essa disposição pode não se aplicar ao Cliente.

Essas informações podem conter imprecisões técnicas ou erros tipográficos. São feitas alterações periódicas nas informações aqui contidas; tais alterações serão incorporadas em futuras edições desta publicação. A IBM pode, a qualquer momento, aperfeiçoar e/ou alterar os produtos e/ou programas descritos nesta publicação, sem aviso prévio.

Todas as referências nestas informações para websites não IBM são fornecidas apenas para conveniência e não de forma a servir como um endosso a esses websites. Os materiais contidos nesses Web sites não fazem parte dos materiais desse produto IBM e a utilização desses Web sites é de inteira responsabilidade do Cliente.

A IBM pode usar ou distribuir qualquer informação fornecida, da forma que achar adequada, sem incorrer em qualquer obrigação para com o Cliente.

Licenciados deste programa que desejam obter informações adicionais sobre este assunto com o objetivo de permitir: (i) a troca de informações entre programas criados independentemente e outros programas (incluindo este) e (ii) a utilização mútua das informações trocadas, devem entrar em contato com:

Gerência de Relações Comerciais e Industriais da IBM Brasil
Av. Pasteur, 138-146
Botafogo
Rio de Janeiro, RJ
CEP: 22296-903

Tais informações podem estar disponíveis, sujeitas a termos e condições apropriados, incluindo em alguns casos o pagamento de uma taxa.

O programa licenciado descrito neste documento e todo o material licenciado disponível são fornecidos pela IBM sob os termos do IBM Customer Agreement, Contrato de Licença do Programa Internacional da IBM ou qualquer contrato equivalente.

Os exemplos de clientes e dados de desempenho mencionados são apresentados apenas com propósitos ilustrativos. Os resultados reais de desempenho podem variar de acordo com as configurações específicas e condições de operação.

As informações relativas a produtos não IBM foram obtidas junto aos fornecedores dos respectivos produtos, de seus anúncios publicados ou de outras fontes disponíveis publicamente. A IBM não testou esses produtos e não pode confirmar a precisão de seu desempenho ou de sua compatibilidade, nem quaisquer reclamações relacionadas aos produtos não IBM. Dúvidas sobre os recursos de produtos não IBM devem ser encaminhadas aos fornecedores desses produtos.

Declarações em relação a direção ou intento futuro da IBM estão sujeitas a mudança ou retirada sem prévio aviso e representam objetivos e metas apenas.

Todos os preços IBM mostrados são preços de varejo sugeridos pela IBM, são atualizados e estão sujeitos a alterações sem aviso prévio. Os preços dos revendedores podem variar.

Estas informações são apenas para fins de planejamento. As informações nesta publicação estão sujeitas a alterações antes que os produtos descritos se tornem disponíveis.

Estas informações contêm exemplos de dados e relatórios utilizados em operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos podem incluir nomes de indivíduos, empresas, marcas e produtos. Todos esses nomes são fictícios e qualquer semelhança com pessoas ou empresas reais é mera coincidência.

LICENÇA DE COPYRIGHT:

Estas informações contêm exemplos de dados e relatórios utilizados em operações diárias de negócios. Para ilustrá-los da forma mais completa possível, os exemplos podem incluir nomes de indivíduos, empresas, marcas e produtos. Todos esses nomes são fictícios e qualquer semelhança com pessoas ou empresas reais é mera coincidência.

Cada cópia ou parte desses programas de amostra ou qualquer trabalho derivado deve incluir um aviso de copyright com os dizeres:

© nome da sua empresa) (ano). Partes deste código são derivadas dos Programas de Amostra da IBM Corp.

© Copyright IBM Corp. _digite o ano ou anos_. Todos os direitos reservados.

Marcas Registradas

IBM, o logotipo IBM e ibm.com são marcas comerciais ou marcas registradas da International Business Machines Corp., registradas em vários países no mundo todo. Outros nomes de empresas, produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atual de marcas registradas da IBM está disponível na web em "Copyright and trademark information" em www.ibm.com/legal/copytrade.shtml.

Adobe, o logotipo Adobe, PostScript e o logotipo PostScript são marcas comerciais ou marcas registradas da Adobe Systems Incorporated nos Estados Unidos e/ou em outros países.

IT Infrastructure Library é uma marca registrada da Central Computer and Telecommunications Agency que agora faz parte do Departamento de Comércio do Governo.

Intel, logotipo Intel, Intel Inside, logotipo Intel Inside, Intel Centrino, logotipo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium são marcas comerciais ou marcas registradas da Intel Corporation ou suas subsidiárias nos Estados Unidos e em outros países.

Linux é uma marca registrada da Linus Torvalds nos Estados Unidos e/ou em outros países.

Microsoft, Windows, Windows NT e o logotipo Windows são marcas registradas da Microsoft Corporation nos Estados Unidos e/ou em outros países.

ITIL é uma marca registrada e uma marca de comunidade registrada da The Minister for the Cabinet Office e está registrada no U.S. Patent and Trademark Office.

UNIX é uma marca registrada da The Open Group nos Estados Unidos e em outros países.

Cell Broadband Engine é uma marca comercial da Sony Computer Entertainment, Inc. nos Estados Unidos e/ou em outros países e é utilizada sob licença.

Linear Tape-Open, LTO, o logotipo LTO, Ultrium e o logotipo Ultrium são marcas comerciais da HP, IBM Corp. e Quantum nos Estados Unidos e em outros países.



Impresso no Brasil