IBM SPSS Analytic Server
Version 1

Administrator's Guide

IBM

Product Information

This edition applies to version 1, release 0, modification 0 of IBM SPSS Analytic Server and to all subsequent releases and modifications until otherwise indicated in new editions.

Contents

Chapter 1. Overview

IBM® SPSS® Analytic Server is a solution for big data analytics that combines IBM SPSS technology with big data systems and allows you to work with familiar IBM SPSS user interfaces to solve problems on a previously unattainable scale.

Why big data analytics matters

Data volumes collected by organizations are growing exponentially; for example, financial and retail businesses have all customer transactions for a year (or two years, or ten years), telco providers have call data records (CDR) and device sensor readings, and internet companies have the results of web crawls.

Big data analytics is needed where there exists:
* A large volume of data (terabytes, petabytes, exabytes), especially when it is a mixture of structured & unstructured data
* Rapidly changing/accumulating data

Big data analytics also assists when:
* A large number (thousands) of models are being built
* Models are frequently built/refreshed

Challenges

The same organizations that collect large volumes of data often have difficulty actually making use of it, for a variety of reasons:
* The architecture of traditional analytic products are not suited to distributed computation, and
* Existing statistical algorithms are not designed to work with big data (these algorithms expect the data to come to them, but big data is too costly to move), thus
* Performing state of the art analytics on big data requires new skills and intimate knowledge of big data systems. Very few analysts have these skills.
* In-memory solutions work for medium-size problems, but do not scale well to truly big data.

Solution

Analytic Server provides:
* A data-centric architecture that leverages big data systems, such as Hadoop Map/Reduce with data in HDFS.
* A defined interface to incorporate new statistical algorithms designed to go to the data.
* Familiar IBM SPSS user interfaces that hide the details of big data environments so that analysts can focus on analyzing the data.
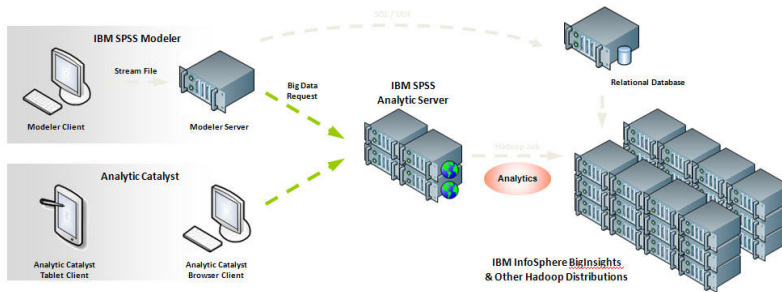* A solution that is scalable to any size problem.

Architecture

Figure 1. Architecture

The Analytic Server sits between a client application and Hadoop cloud. Assuming that the data resides in the cloud, the general outline for working with the Analytic Server is to:

1. Define Analytic Server data sources over the data in the cloud.
2. Define the analysis you want to perform in the client application. For the current release, the client applications are IBM SPSS Modeler and IBM SPSS Analytic Catalyst.
3. When you run the analysis, the client application submits an Analytic Server execution request.
4. The Analytic Server orchestrates the job to run in the Hadoop cloud and reports the results to the client application.
5. You can use the results to define further analyses, and the cycle repeats.

Chapter 2. Tenant management

Tenants provide a high-level division of users, projects, and data sources. Each user accesses the system in the context of a tenant to which they are assigned. You manage tenants, and assign to tenants, in the Analytic Server console.

The view of the Tenants accordion depends upon the role of the user that is logged in to the console:

- The "super user" administrator that is set up during installation is the tenant manager. Only this user can create new tenants and edit the properties of any tenant.
- Users with the Administrator role can edit the properties of the tenant they are logged in to.
- Users with the User role cannot edit tenant properties. The Tenants accordion is hidden from them.

Administrators can access the Projects and Data sources accordions and manage any project or data source for cleanup and administration. See the *IBM SPSS Analytic Server 1.0 User's Guide* for more information.

Left column

The left column displays the existing tenants under the accordion heading. Only the "super user" administrator can use these controls.

- Select a tenant to display its details in the content area and edit its properties. Typing in the search area filters the listing to display only tenants with the search string in their name.
- Click **New tenant** to create a new tenant with the name you specify in the **Add New Tenant** dialog. Names are case-sensitive, ignore leading and trailing white space, and protect against SQL injection.
- Click **Delete tenant** to remove the tenant.

Content area

The content area is divided into *Details*, *Principals* and *Projects* collapsible sections.

Details

    Name An editable text field that displays the name of the tenant. Tenant names must be case-sensitive, ignore leading and trailing white space, and protect against SQL injection.

    Description
        An editable text field that allows you to provide explanatory text about the tenant.

    URL This is the URL to give to users to log in to the tenant through the Analytic Server console, and to use to configure SPSS Modeler server. See the *IBM SPSS Modeler 15.0 update and extensions for IBM SPSS Analytic Server 1.0 Installation Instructions* for details on configuring SPSS Modeler.

Principals

    Principals are users and groups that are drawn from the security provider that is set up during installation. You can add principals to a tenant as Administrators or Users.

- Typing in the text box filters on users and groups with the search string in their name. Select **Administrator** or **User** from the drop-down list to assign their role within the tenant. Click **Add participant** to add them to the list of authors.
- To remove a participant, select a user or group in the member list and click **Remove participant**.

Projects

Projects are versioned based on changes to the file and folder contents. This table lists all of the projects in the tenant, and allows an administrator to specify the maximum number of versions per tenant. Analytic Server automatically deletes the oldest committed project version when the number of versions exceeds the specified number. The values here are the same as those values on the **Automatically clean up when number of versions exceeds** text box on the Versions tab of the Projects accordion, but presented in a summary view.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

All IBM prices shown are IBM's suggested retail prices, are current and are subject to change without notice. Dealer prices may vary.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of The Minister for the Cabinet Office, and is registered in the U.S. Patent and Trademark Office.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

Index

T

IBM ®

Printed in USA