PLANNING YOUR ATSCALE REPORTING MIGRATION

BY JEN UNDERWOOD

TABLE OF CONTENTS

INTRODUCTION

Mainstream reporting solutions are not scaling to meet digital era demands. Increasing volumes and varieties of data combined with modern business intelligence practices that enable self-service data access is overwhelming both reporting tools and infrastructure. To empower insights-driven, analytical cultures in a new world of big data, organizations are adopting data lake architectures and AtScale semantic layers to enhance capabilities of existing reporting tools.

SPEED. SCALE. SIMPLICITY.

> Business analysts can get rapid insights from massive data sources within their favorite business intelligence tools without waiting or relying on IT.

There are many benefits for both the business and IT to use AtScale with existing business intelligence applications and a modern big data architecture. With AtScale, big data analytics is easy. Businesses do not need to hire deeply technical Hadoop resources to extract value from big data. Business analysts can get rapid insights from massive data sources within their favorite business intelligence tools – Excel, Power BI, Tableau, Qlik, TIBCO Spotfire, SAP Lumira, Microstrategy, and other analytics solutions – without waiting or relying on IT. There are no learning curves, no complex scripting and no need for analysts to change. The business can also delight in better reporting consistency by using AtScale as a centralized source of trusted, shared business logic.

Implementing the right big data analytics infrastructure is vital for future success. It cuts capital expenditures, maintenance and operating costs while also improving reporting speed, scale, simplicity, agility, security and self-service governance. The hidden recurring pains and costs caused by analytical silos can be eradicated.
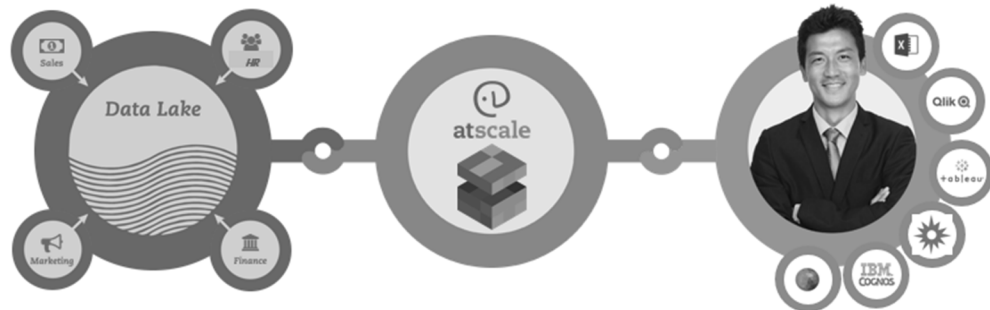

Figure 1 Overview of AtScale

Since AtScale supports standard analytics tools, SQL and MDX languages, inefficient data copying, ETL and silo semantic model processes can be redesigned to use a better unified, single semantic layer approach.

PROBLEMS ATSCALE SOLVES

It is common for large groups to "lift and shift" existing analytics applications to AtScale when 1) current reporting starts exceeding the limits of older environments or 2) when new insights are needed from big data sources that are not possible to do with existing tools. Other popular use cases for making the move include:

- o Delivering optimized big data reporting and analytical scalability

- o Accelerating reporting performance without moving data to a proprietary query engine or data format

- o Eliminating inefficient pre-processing and reporting ETL processes

- o Achieving data source consistency across multiple reporting platforms

- o Simplifying efficient ad-hoc big data queries using Excel or popular BI tools with support for SQL or MDX languages

- o Saving money on cloud query and computing costs

- o Providing self-service BI governance on big data sources

OVERVIEW OF THE REPORT MIGRATION PROCESS

While every organization's journey into big data analytics may be unique, there are common scenarios, environments, tools and project stages. Effective planning, synchronization and communication is key to a successful migration.
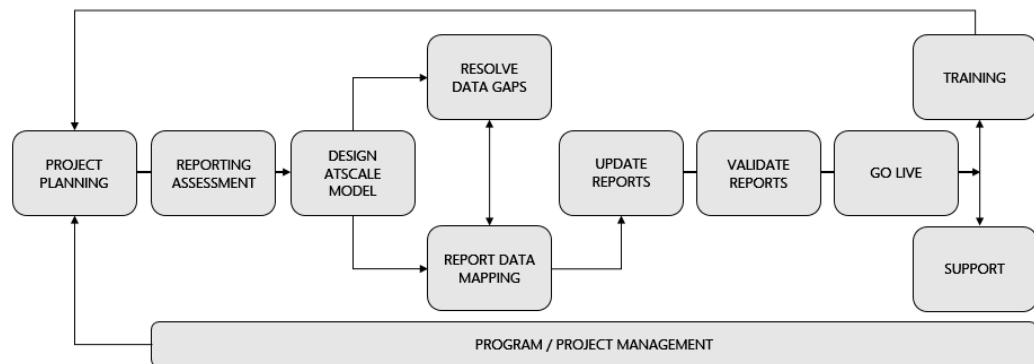

Figure 2 High-Level Reporting Migration Process Overview

Atscale is a virtual OLAP cube not a relational database. It is designed to support a subset of SQL generated by leading BI tools. Although the move to AtScale

typically does not require many changes to reports, there can be query syntax nuances. Also, data sources and ETL processes may change.

WHAT YOU NEED TO SUCCEED

Most groups need to justify, plan, roll out and educate the business for successful adoption of any modern technology. After an organization identifies the need for AtScale and wants to move forward, there are numerous ways to get started. Most organizations begin with a small proof-of-concept reporting pilot project and then develop a prioritized list of other projects to migrate to the new architecture.

To use reporting tools with AtScale, the source data needs to be available within your data lake or other big data source. In this white paper, we will share tips on how to identify reporting source data but we do not delve into the specifics of how to move or load data into big data sources. Please refer to the **AtScale Planning Toolkit** for recommended reading, utilities and other resources to help your technical team expedite that prerequisite data loading step.

In the following sections, we will walk through how to assess your environment to migrate your reports to use AtScale. We'll start with planning and wrap up with suggested next steps. With the AtScale Intelligence Platform, you can finally add the power of big data analytics to the business intelligence tools that you already know and love.
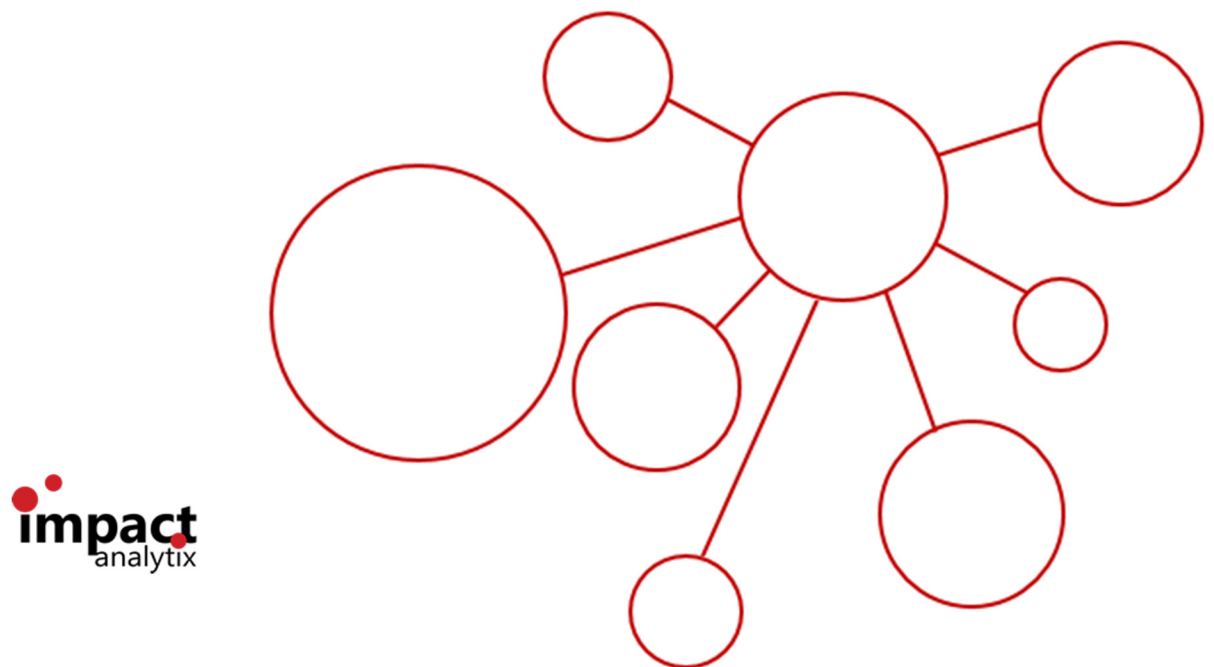
BUILD YOUR CASE

Business intelligence on big data is neither a product nor a system. It is a continuously evolving architecture and a collection of reporting applications. With AtScale, you can start small, secure quick wins and then incrementally migrate more over time. You don't need to move everything all at once.

Since a migration to AtScale requires cooperation of resources across departments and potential reporting process changes, it is important to approach your project properly. Do not proceed in isolation. Be aware of upstream and downstream dependencies from reporting users, applications, and systems.

SECURE PROJECT SPONSORSHIP

Executives make investment decisions with the bottom line in mind. Your business case should be both business need and cost-justified supported by an estimated return on investment (ROI), payback period and risk assessment.

> Your business case should be both business need and cost-justified supported by an estimated return on investment (ROI).

Validation for business intelligence applications is primarily driven by business opportunity or fear rather than business need alone. Clearly define the benefits of expediting time to insights, solving business problems, uncovering new opportunities or exploiting new ways to maximize revenues, reduce costs and diminish risks. As you socialize AtScale to garner support, share relevant case studies and present how this initiative aligns to your overall organizational strategy. What metrics are used to evaluate stakeholder performance? How will AtScale help them succeed?

CALCULATE BUSINESS VALUE

If there is no business problem to solve, it is going to be difficult for you to win approval. A proven method for justifying the AtScale investment is to tie it directly to a business problem of measurable proportion. All decision-support initiatives should fulfill at least one of business benefit.

- o Increase revenue or market share by more rapidly identifying new markets or opportunities, improving cross-sell, efficiently using limited resources, and responding faster to times changes in the market.

- o Improve customer experience and satisfaction.

- o Save money by reducing analytics infrastructure costs, simplifying reporting, reducing requests for customized reports, and minimizing expensive, inefficient development or maintenance costs.

PLAN YOUR PROJECT

After securing sponsorship, it is time to start planning your project. You'll establish a cross-functional project team, start thinking through the requirements, options, exploring potential quick wins, and begin scoping the level of effort.

ASSEMBLE THE TEAM

Your AtScale project team should include key constituents from both business and IT whose input and support you'll need. The ideal team would include a mixture of resources with different project perspectives, skills and abilities. Typically, AtScale projects involve the following team members.

- o **Executive Sponsor** from the business

- o **Project Manager** to communicate updates, coordinates activities and track ongoing, incremental progress

- o **Business Reporting Subject Matter Experts,** data analysts, data modelers or business intelligence professionals that understand current reporting systems, data sources or future big data analytics needs

- o **Other Technical Professionals** from the departments that own ETL processes, application data sources, big data analytics infrastructure and your data lake or other big data source

Larger organizations might also include representation from a Business Intelligence Center of Excellence, Data Governance Stewards, Enterprise Architects, Security Officers, and Support Staff.

BUSINESS ROLE	ATSCALE ROLE	ATSCALE ROLE DESCRIPTION	TRAINING	SKILLSET REQUIRED
Infrastructure / Data Management	Install AtScale on Big Data Source	Provide Secure Data Foundation	Administrator Training	• Big Data Cluster Management • Data Ingestion • Data Security
Data Modeler / Technical Analyst	AtScale Cube Designer	Design, Develop & Publish AtScale Cubes	Cube Designer Training	• Relational Data Modeling Experience • OLAP Knowledge • Business Data Understanding
Report Developers	AtScale Runtime Query Users	Access AtScale Cubes with Reporting Tools	Not applicable	• Understand Data • Report Development
Business Users	AtScale Runtime Query Users	View AtScale Cubes with Reporting Tools	Not applicable	• Ability to View or Run Reports

Figure 3 Common AtScale Project Team Roles

UNDERSTAND THE 10 STEPS TO SUCCESS

To estimate the scope of your project, your team will review the current reporting environment, select a report migration approach, design an AtScale semantic model and update your reports. Throughout that process, a migration checklist should be created that includes training and ongoing support activities.

It is a best practice to start with small quick wins to gain momentum and then incrementally migrate more reports over time.

Project time can vary widely depending on the assessment results and scope of the current BI or data visualization tool implementation. Are there only a few front-end reports or many reports and integration points? Is there an ecosystem of ETL, data models and embedded business logic within those reports to review? Regardless of size, it is a best practice to start with small quick wins to gain momentum and then incrementally migrate more reports over time.

Although every project and reporting environment may not require all the migration steps, here is list of the most common AtScale report migration activities that your team will perform.

1. ASSESS CURRENT REPORTING ENVIRONMENT

2. EVALUATE DATA SOURCES

3. IDENTIFY LOCATION OF BUSINESS LOGIC

4. DESIGN AN ATSCALE SEMANTIC MODEL

5. MAP SOURCE TO DESTINATION DATA

6. RESOLVE DATA SOURCE GAPS

7. UPDATE REPORT DATA SOURCE CONNECTIONS

8. UPDATE ANY CALCULATION SYNTAX DIFFERENCES

9. TEST AND VALIDATE YOUR REPORTS

10. COMMUNICATE, EDUCATE AND SUPPORT REPORT USERS

In the following section, we will provide an overview of each one of these steps to jump start your migration.

MAKE THE MOVE

As you plan to make the move, you'll want to review the current and target architecture to gain a high-level understanding of the reporting landscape before and after goals. Rather than "lift and shift", consider using this opportunity to truly optimize existing reporting processes to fully enjoy the value AtScale offers across entire analytics lifecycle.

> Rather than "lift and shift", consider using this opportunity to truly optimize existing reporting processes.

REPORT MIGRATION TASKS

1. ASSESS CURRENT REPORTING ENVIRONMENT

Where do you start? How will you determine what reports to run on the AtScale Intelligence Platform? What does it take to migrate? How do you prioritize? To answer these questions, your team will perform an assessment. Often a specific reporting area to improve has already been identified. In other cases, you'll need to find a suitable proof of concept project and then develop a prioritized list of other viable candidate projects.

To identify suitable projects, assessment team members will review report performance metrics and understand usage. They will collect and itemize the following types of information. In other steps, you will delve deeper into data sources and reports.

- o Current and future state reporting solution architecture

- o ETL process flow overviews

- o Data model entity relationship diagrams

- o Analytics and business intelligence tool and user landscape

- o Report area owners, reporting tools used, report names, semantic data models, data sources, used tables and columns. To optimize future reporting, also note if there are any unused elements. Within reports, note the location of business logic while reviewing queries. Check for query redundancy and complexity. Cross reference currently used query language compatibility with AtScale supported query functions to identify where report or query syntax changes might be needed.

- o Note upstream and downstream reporting users, process or application dependencies that would need to be made aware of any changes
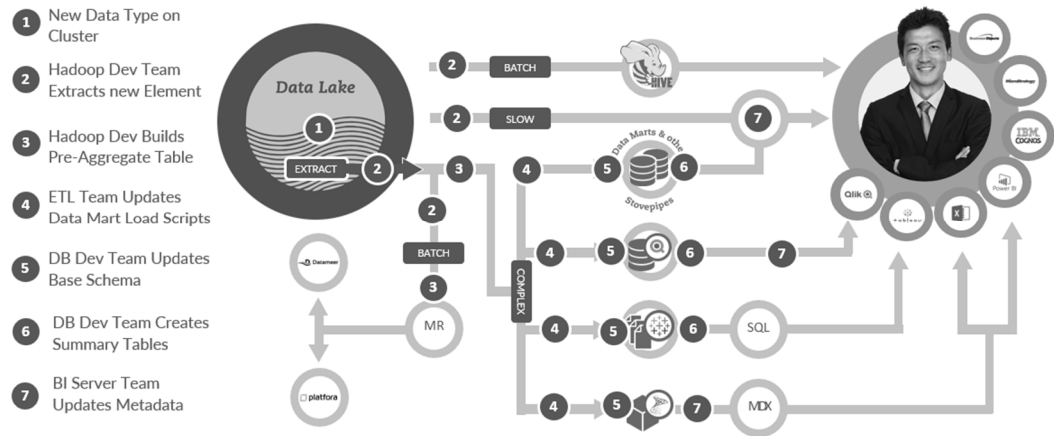
Figure 4 Example Current Reporting Solution Architecture

Supporting a complex architecture of reporting silos is expensive and challenging. Over time, these types of situations undermine analytics success since the average time to make a change can take six weeks or longer. If you have this type of architecture today, your organization will tremendously benefit from investing in the migration to an agile implementation with AtScale Intelligence Platform.
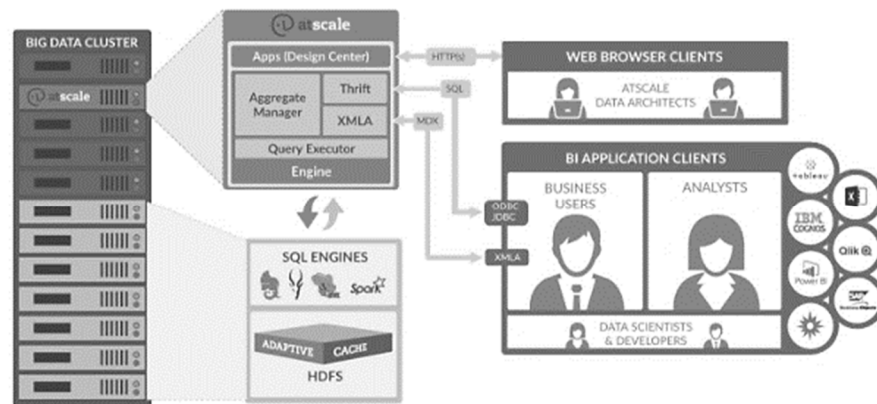

Figure 5 Example Future State Reporting Solution Architecture with AtScale

In the example AtScale design, inefficient processes are eliminated with flexible schema-on-read capabilities. The open data model that supports complex calculations can be used with numerous analytics tools delivering one version of the truth.

After reviewing the current environment, you should have a better understanding of high-level tasks that will need to be performed to develop a prioritized list of possible projects and select a first one to migrate.

2. EVALUATE DATA SOURCES

When you review reporting data sources, it is crucial to assess if you have access to the underlying data. Most mainstream reporting tools today connect directly to structured relational or OLAP data source types. It is less common to have reports from unstructured types but it is entirely possible that they might if you have already been querying big data sources.

Note numerous traditional and self-service BI tools store or transform source data into a proprietary data models. Be sure to identify what data source is being used and if you can get to the raw source data. Here is a brief list of the most commonly used reporting tool proprietary data models.

REPORTING TOOL	DATA FILE EXTENSIONS	NOTES
Excel	.odc	Connection only, data and formulas are stored in Excel cells or the data source.
Power BI	.pbix	Proprietary data source, locate original source data. Power BI files are a zipped reports and data file. You can rename the file to .zip to extract the data file from it. Utilities might be available to extract the source data from this file type.
Tableau	.hyper, .tde, .tdsx, .tsds	Proprietary data source, locate original source data. Refer to the Tableau Data Source (.tds) to locate the data sources if available. Packaged Tableau Workbooks (.tbwx) are a zipped workbooks and data file. You can rename the file to .zip to extract the data file from it. Utilities might not be available to extract the source data from this file type.
Qlik	.qvd	Proprietary data source, locate original source data. Utilities might not be available to extract the source data from this file type.
TIBCO Spotfire	.stdf	Proprietary data source, locate original source data. Utilities might not be available to extract the source data from this file type.
SPSS	.sav	Proprietary data source, locate original source data. Note open source ODBC connectors are available that can extract data from this data source type.
SAS	.sas, .sas7bdat	Proprietary data source, locate original source data. Note open source ODBC connectors are available that can extract data from this data source type.
IBM Cognos	.ppx, .cub, .ppr	Proprietary data source, locate original source data. Utilities might be available to extract the source data from this file type.
SAP Lumira	.lums	Proprietary data source, locate original source data.
SAP Business Objects Universe	.unv	Proprietary data source, locate original source data. Utilities might be available to extract the source data from this file type.

Figure 6 Common Reporting Tool Proprietary Data Models

As part of your report data source evaluation, look at where data source relationships are defined. Also review if any ETL or data cleansing was performed within the reporting tool. Some tools contain embedded ETL processes that happen after data source connection to union, pivot, split, join and further prepare data sources before the report is created. These steps might also happen outside of the tool entirely or partially within the reporting tool. Note any proprietary scripting that might require moving, recoding or removing to improve ongoing maintenance. ETL and data preparation work for reporting can account for approximately 80% of the total work effort. Recreating the visualizations is the easy, fast part.

> Note any proprietary scripting that might require moving, recoding or removing to improve ongoing maintenance.

For Tableau users, the **AtScale Tableau Workbook Analyzer** is an invaluable resource that will expedite your reporting assessment and Tableau reporting migration steps to AtScale. This resource can be provided by your AtScale account team upon request.

3. IDENTIFY LOCATION OF BUSINESS LOGIC

Another essential step is to check for custom reporting tool queries, stored procedures, definitions of hierarchies, sets, groups, calculations KPIs, and other business logic that might need to be updated for AtScale migration. As you inventory reports and look at the amount of embedded business logic, label the level of anticipated migration complexity: no work required, minimal work, simple updates or complex updates.

Itemize is the report is applying role-based security, using click actions, filters, controls, parameters, advanced analytics, or other custom modules. Often in the front-end of self-service reporting tools, business users will customize business logic, calculations in a proprietary reporting tool script language. If reporting tools script functions are used, verify that they will work with Hive SQL.

4. DESIGN AN ATSCALE SEMANTIC MODEL

The next step in the process is creating a dimensional model for AtScale's virtual, multi-dimensional vROLAP engine. Dimensional modeling is a well-known, proven reporting design pattern based on Kimball methodology that improves report query performance without affecting data integrity. It is a familiar standard in the business intelligence and analytics market.

Dimensional models, also referred to as star or snowflake schemas contain facts surrounded by descriptive data called dimensions. Facts contains numerical values of what you measure such as sales or user counts that are additive, or semi-additive in nature. Fact tables also contain the keys/links to associated dimension tables.

Dimensions contain the descriptive details about the facts. Common dimensions are date, type, and location. Dimension tables are usually denormalized and have many columns with fewer rows of data than fact tables except for "monster dimensions" ie. person/user. The values in a dimension table are typically used as rows and columns in reports.
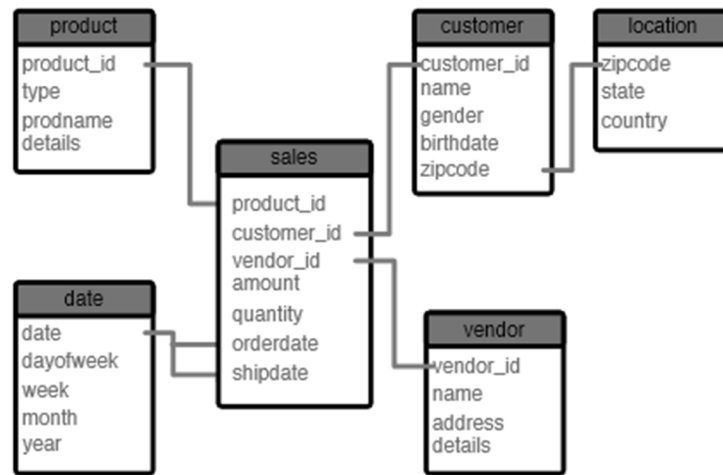

Figure 7 Dimensional Design

> AtScale machine-learning algorithms intelligently optimize BI query workloads and dynamically build aggregates based on user queries.

Your AtScale semantic model should contain dimensions, facts, measures, calculations, and hierarchies. Unlike legacy OLAP, you will not need to design aggregate tables. AtScale machine-learning algorithms intelligently optimize BI query workloads and dynamically build aggregates based on user queries.

The process for dimensional modeling is extremely iterative in nature. It begins with identifying the reporting requirements for a business process. In the real world, it would not be uncommon to rapidly experiment with a basic dimensional model and then continually update it as more information is collected or users begin exploring reports.

To learn how to design a dimensional model, Kimball Group's The Data Warehouse Toolkit book and the free IBM Redbook called Dimensional Modeling: In a Business Intelligence Environment are excellent resources. For Tableau users, the **AtScale Tableau Workbook Analyzer** automates the design of an AtScale project and semantic model from scanning your Tableau files. This resource can be provided by your AtScale account team upon request.

5. MAP SOURCE TO DESTINATION DATA

Using your AtScale dimensional model from step 4, cross reference the reporting source data. An excellent example of data mapping is available in the classic Kimball Data Warehouse Toolkit reporting series.

COLUMN NAME	DISPLAY NAME	DESCRIPTION	DATA TYPE	SOURCE SYSTEM	ATSCALE
VendorUID	VendorID	Unique Key for Vendor Master	Integer	SAP ERP	Available
DUNS	D&B Number	Dunn and Bradstreet Number	String/Text	D&B	Not Yet Available

Figure 8 Example Data Map

6. RESOLVE DATA SOURCE GAPS

After you have mapped reporting data sources and identified missing data that is not yet in AtScale, discuss as a project team how you want to mitigate that issue. Most of the time, the data management team will add that data in an ingestion process. Alternatively, you might opt to work around data gaps by continuing to query missing AtScale data from the current data source if that is feasible and practical to do. If the data it is not being used, you could update reports to no longer include it with a note of that change. Be sure to check with report stakeholders prior to removing data from reports.

7. UPDATE REPORT DATA SOURCE CONNECTIONS

After resolving data source gaps, you can begin the process of updating your reports to use your new AtScale semantic model. Before you can change report connections to AtScale, please make sure that you have installed the proper version of the Hive ODBC driver that is compatible with your reporting system. See Installing Hive ODBC Drivers in your AtScale documentation. After installing the needed drivers, configure the data source connection within the reporting tool using the security credentials provided by your AtScale administrator.

> Make sure that you have installed the proper version of the Hive ODBC driver that is compatible with your reporting system.

Note both reporting authors, viewers and consumers of reports might also need to install Hive ODBC depending on how reporting tools are being used. Most self-service BI desktop tools do require local data source drivers to be installed to remotely connect and explore data sources.

8. UPDATE ANY CALCULATION SYNTAX DIFFERENCES

After you connect your report to AtScale, you may notice errors in your reports. Most report data fields such as dimensions and measures should not be the source of those migration errors if the data is available within AtScale. Reporting errors are usually related to subtle query syntax differences, a missing script function, or other unavailable element that is not in your AtScale model.

Keep in mind that Atscale is a virtual OLAP cube with a live connection to your source data; it is not a relational database. See the Atscale guide "Atscale Unsupported Tableau Features and Query Patterns" for more detailed information. AtScale is designed to support a subset of SQL generated by leading BI tools but not all functions are currently available. If you have worked with another OLAP solution in the past such as Microsoft Analysis Services, Microstrategy, IBM Cognos TM1 or Hyperion, then using that knowledge will help you understand how Atscale works.

> Aggregate functions should be updated in reports to fully enjoy the speed and power of AtScale's adaptive cache.

- o In Atscale, connections are "live" and measures are pre-aggregated. This approach takes advantage of Atscale's adaptive cache aggregation technology and consistent semantic layer.

- o Measures are already aggregated by the cube designer in Atscale (and not in Tableau). This approach takes advantage of Atscale's aggregation technology and consistent semantic layer. Tableau's documentation states that Tableau's Calculated Fields do not support aggregation functions or binned data (on multidimensional data sources) because the data are already aggregated by the OLAP cube.

- o Calculated fields are translated into Hive SQL. Several functions such as time may be slightly different. For example, if you are using the NOW() function in Excel, it uses the local machine clock. When migrated to Atscale, that same function, NOW(), will use Hive's Unix _timestamp() function.

- o Some functions may not be available such as CEILING, FLOOR, LOG, ROUND, SIGN, SQRT, ASCII, and SPACE.

- o Aggregate functions such as COUNT, COUNTD, MEDIAN, PERCENTILE, STDEV, VARIANCE should be updated in reports to fully enjoy the speed and power of AtScale's adaptive cache.

- o Other known functions found you might need to update for reporting with AtScale are identified in the online docs at <insert link to more resources.> For Tableau users, ask your account team for the white paper called **Tableau Best Practices and Support Matrix for Tableau WB Authors**

9. TEST AND VALIDATE YOUR REPORTS

Congratulations! You are almost done migrating your first set of reports to the AtScale Intelligence Platform. Already you should have noticed much faster, impressive query and report rendering times in the previous report update steps. You might also want to collect before and after report run time or big data query performance benchmarks.

Now it is time to share your progress with the business stakeholders that use these reports and can validate the numbers on them are accurate. To test and validate reports, select a testing team of detail oriented users that are the first to report bugs and/or create their own reports. These types of power users per se are often fantastic, enthusiastic report testers. Power user testers can help you identify and often also resolve any discrepancies before you roll out your new reports to a larger audience.

Power users are often fantastic, enthusiastic report testers that can help you identify and resolve any reporting discrepancies.

Be sure to test a wide range of reports with data ranges over large and small timeframes. Check historical report values match the values in your new reports. Don't overlook the importance to test date ranges exceeding one year to verify time intelligence and any organizational changes over time have been retained. For additional tips, please refer to the **AtScale Planning Toolkit.**

10. COMMUNICATE, EDUCATE AND SUPPORT REPORT USERS

To ensure success and adoption of your new AtScale migration, be sure to notify your users and educate them on how to get started building or updating their own reports. Most organizations that roll out AtScale do embrace a "Train the Trainer" approach. After reporting area leads learn the AtScale, dedicated internal training sessions, Lunch and Learns and an internal online portal containing support contacts, AtScale documentation, short demonstration videos and other self-service learning materials are highly recommended.

For official training, AtScale offers several courses to help customers get started and continue to succeed on the big data analytics journey.

CONCLUSION

In this white paper, we discussed how to successfully approach a migration of your existing reporting assets to a more efficient, performant, scalable AtScale Intelligence Platform architecture. We shared a plethora of compelling reasons why organizations are investing time to make this critical move to digital scale, big data analytics. We also provided valuable resources to expedite the 10 steps to success and jump start your big move. With the AtScale Intelligence Platform, you can finally add the power of big data analytics to the analytics and business intelligence tools that you already know and love.

NEXT STEPS

For additional information on planning your project, please contact an AtScale expert. AtScale offers a wide array of professional services to ensure you are successful at bridging the gap between your analytics users and their data. Our Professional Services team helps customers implement use cases, tune Hadoop clusters or improve their working knowledge of the AtScale Intelligence Platform. Whether you are starting your journey or want to expand, AtScale professional services teams are available to guide you along every step of the way.

If you'd like to learn more about AtScale implementations, here are additional recommended resources.

- o AtScale, Inc. http://atscale.com

- o AtScale Public Resources http://atscale.com/resources
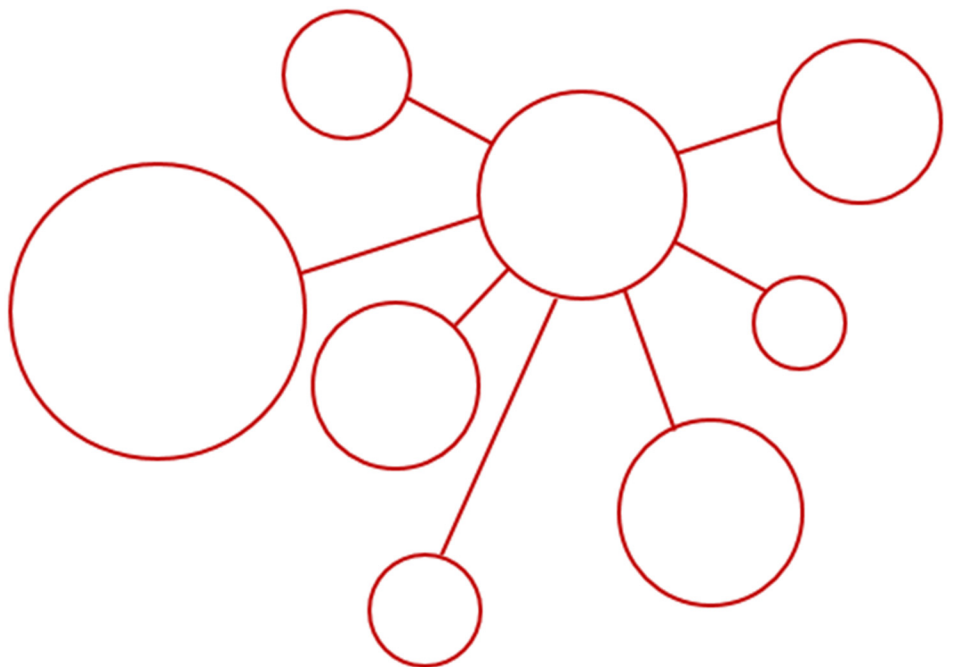
About AtScale, Inc.

AtScale makes BI work on Big Data. With AtScale, business users get interactive and multi-dimensional analysis capabilities, directly on Big Data, at maximum speed, using the tools they already know, own and love – from Microsoft Excel to Tableau Software to QlikView. Built by Big Data veterans from Yahoo!, Google and Oracle, AtScale is already enabling the BI on Big Data revolution at major corporations across healthcare, telecommunications, retail and online industries. To see how AtScale can help your company, go to www.atscale.com/why-atscale.

About the Author

Jen Underwood, founder of Impact Analytix, LLC, is an analytics industry expert with a unique blend of product management, design and over 20 years of "hands-on" advanced analytics development. In addition to keeping a pulse on industry trends, she enjoys digging into oceans of data. Jen is honored to be an IBM Analytics Insider, SAS contributor, and former Tableau Zen Master. She also writes for InformationWeek, O'Reilly Media and other tech industry publications.

Jen has a Bachelor of Business Administration – Marketing, Cum Laude from the University of Wisconsin, Milwaukee and a post-graduate certificate in Computer Science – Data Mining from the University of California, San Diego. For more information, visit www.jenunderwood.com.

impact
analytix