



Scaling Up Adhoc Analytics Productivity With Scribble



Executive Summary

This whitepaper discusses the state-of-the-art adhoc analytic process in data-driven businesses, and showcases how the Scribble tool can scale up the productivity of this process, with a use case abstracted from an existing retail client of Scribble. With the growing need for adhoc analytics, Scribble can save more than 20% of analytics cost for businesses.

What Scribble is About

Scribble is a data automation platform. We think deeply about organizations' ability to access and learn from their own data, and survive in a fast moving marketplace. We believe that organizations cannot go data-driven until they are able to ask lots of questions against their data. We are bringing search, modeling, and systems techniques to drive one key metric - cost/question - to 1/100th of what it is today.

With Scribble, the questions that can be asked of the data now, and the exploration that can be done, increase dramatically. This means faster insights that lead to better business decisions, and so, to measurable improvements to the bottomline.

A large proportion of business questions look like adhoc queries against databases. Most of these databases support SQL. Our flagship tool, **Scribble**, provides a **near natural language interface that lets users rapidly build complex ANSI-compatible SQL queries in a fraction of the time** it would otherwise take. The platform enables our users, usually business analysts or data engineers, to instantly reuse any previous work done on the platform by themselves or others, create audit trails for their code, organize work into investigations, and integrate with third-party applications - all this with near zero deployment, training, and integration costs.

The Adhoc Analytics Process Today

This section details our conception of the data analytics space, along with the opportunities to improve upon existing processes and technologies.

Analytics comes in three forms - **dashboarding** that captures the state at a given moment (e.g. sales of a particular product), **adhoc analytics** that help explain business performance or identify opportunities to increase business value (e.g. why there is a change in sales since last month), and **advanced statistics** that identifies fundamental drivers of the state (e.g. price elasticity of the product).

Scribble is focused on the middle tier, of adhoc questions, be it examinations of what happened in the past or designing experiments to increase business value in the future. The process consumes



Scaling Up Adhoc Analytics Productivity With Scribble

enormous resources and effort while also being long drawn out. That said, it is this capability of asking questions and having them answered reliably and quickly, that gives businesses an edge over their competition when it comes to identifying opportunities for increasing growth and cutting losses.

A typical adhoc analytics process starts with a business question and ideally ends with an actionable insight, but the process requires analysts to have knowledge of the tools (languages, softwares), methods (mathematics) and business context, to produce useful output. **There are a number of opportunities to improve upon the state-of-the-art, cutting across all stages of the analytical process:**

STAGE	DESCRIPTION	CHALLENGE/OPPORTUNITY
FRAMING	Business relevant question	Recall of organizational knowledge
DISCOVERY	Identify relevant information	*Simplicity, discoverability of artifacts (data, definitions, notes, queries)
ACCESS	Query data, search datasets	*Simplicity, efficiency, recall, accuracy, auditability, governance of access
MODELING	Apply statistics and other pattern discovery mechanisms	Method selection, automation
COMMUNICATION	Sensemaking and storytelling	Soundness, newer definitions, continuous process learning
AUDIT & SHARING	Test methodological robustness, share with peers	*Efficiency, repeatability

* Solved by Scribble

Solution approaches to these problems such as Business Intelligence (BI), data lakes, data management, and workflow systems usually have long deployment cycles and significant complexity. There is an opportunity for creative low complexity solutions as well.

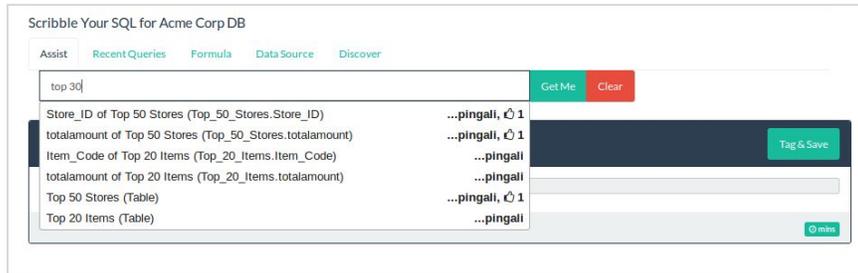
How Scribble Works

Scribble saves time and effort in the analytic process in three ways:

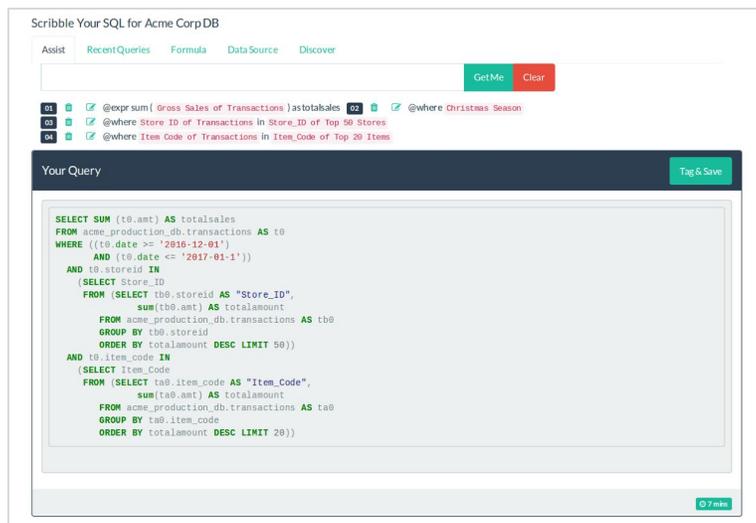
First, **rapid query creation** - Scribble provides a simple search bar-like interface where users enter what data they're looking for. **Scribble instantly turns this near natural language input into a ready-to-run SQL query.**



Scaling Up Adhoc Analytics Productivity With Scribble



The user's search is guided by an intelligent dropdown



Scribble takes the user from data search to ready query instantly

Executing this request is cumbersome today. The analyst must first rebuild these definitions and corresponding SQL snippets. It is not unusual for the analyst to look through past queries stored either in the code repository or the execution history, and manually extract the relevant bits. These snippets are then stuck in an editor for cleaning and inclusion with the rest of the code. Then the analyst looks through table definitions to see which columns to access and possibly stackoverflow for SQL syntax. After a couple of tries, the analyst may have a working query. **Without Scribble, this will take in the order of 20 minutes.**

In Scribble, finding known definitions is simple. You start typing, and scribble suggests the most relevant options after looking at every single query and definition. You simply select one of them. **With Scribble, it takes just a couple of minutes to complete the SQL.** It will be a working query at every point that the user searches for data or for a sub-query. When you store and tag the new query, you will have the complete lineage to show who built the sub-queries and how the final query was arrived at.



Scaling Up Adhoc Analytics Productivity With Scribble

<pre>(SELECT Store_ID FROM (SELECT tb0.storeid AS "Store_ID", sum(tb0.amt) AS totalamount FROM acme_production_db.transactions AS tb0 GROUP BY tb0.storeid ORDER BY totalamount DESC LIMIT 50)) AND t0.item_code IN (SELECT Item_Code FROM (SELECT ta0.item_code AS "Item_Code", sum(ta0.amt) AS totalamount FROM acme_production_db.transactions AS ta0 GROUP BY ta0.item_code</pre>		
Dependency	Description & Version	Last touched
Query	transactions for christmas season(v1)	an hour ago by pingali
Query	Christmas Season(v1)	an hour ago by pingali

Scribble provides an auditable trail for every query to enable the easy reproduction of numbers

Impact on Bottomline

The advanced analytics usecases are growing rapidly as indicated in a recent survey by the Eckerson Group.

Power users dominate the use of cloud BI and analytics solutions, driving more complex use cases that include ad-hoc analysis (57%) and advanced report and dashboard creation (55%). Casual users are 20% of all cloud BI and analytics, with their most common use being for reporting and dashboards (76%) (Source: BARC and Eckerson Group Study 'BI and Data Management in the Cloud: Issues and Trends', January 2017 via Forbes)¹

As the need for adhoc analytics grows, Scribble provides a clear path to RoI. This return comes from a number of factors, including lowering the hiring cost in terms of absolute numbers of data analysts - through an increase in productivity of existing teams. Scribble also ups productivity from a skill perspective, turning junior analysts into intermediate or advanced analysts, reducing the cost per hire. Thirdly, through its ease of use and collaborative features, it reduces training costs of existing and new hires. Fourthly, Scribble's creative use of metadata significantly reduces the complexity and delay to seeing initial impact. Finally, Scribble dramatically shortens the time to get from adhoc business question to insight for every query, increasing the potential positive impact on the bottomline daily.

A survey of analytics practitioners, including data analysts and heads of analytics, carried out by Scribble, provided data on the cost of stages of an adhoc analytics, including a one-time deployment of a relevant tool. Below, this data is juxtaposed against the reduction that Scribble found under these various stages in different pilots carried out by Scribble Data. On average, **Scribble improved productivity by 22%**. Every organization and team has a different distribution of cost and impact. So this should be considered indicative of the potential.

¹<https://www.forbes.com/sites/louiscolombus/2017/02/26/business-intelligence-and-analytics-in-the-cloud-2017/>



Scaling Up Adhoc Analytics Productivity With Scribble

ELEMENT	INDICATIVE COST ²	REDUCTION BY SCRIBBLE	FINAL EFFORT
INFRASTRUCTURE	2%	-	2%
INITIAL SETUP	5%	30%	2%
TRAINING	3%	25%	2%
FRAMING	5%	-	5%
DISCOVERY	20%	25%	15%
ACCESS	30%	40%	18%
MODELING	20%	10%	18%
COMMUNICATION	10%	-	10%
AUDIT & SHARING	5%	25%	4%
TOTAL	100%		78%

Hassle-free Implementation

Scribble is a cloud-based tool, designed to be entirely non-intrusive, and **users can get going with it in minutes**. Scribble has no access to the business data. It works by building its own virtual database using only metadata. Scribble also does not provide users the ability to see any data that they do not have access to. The output of Scribble is a query that users run separately using their existing database credentials, and is governed by existing data governance and data security within their organization. Specifically, the steps for implementing the tool are as follow:

1. Users provide (or upload themselves) the schema of the database(s) that they're looking to query. We recommend upto 15 tables at a time.
2. The users themselves, or a service provider, then customizes the columns of each table to add synonyms or meaningful names to enable them to be easily found
3. Users create an account on the Scribble site and get added into groups along with their peers, or can use the tool by themselves without being added into a group
4. With that, users can get started building their queries.

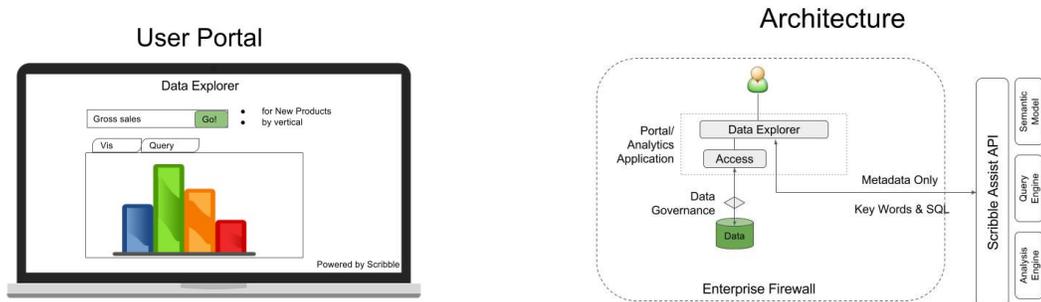
The Scribble service can also be consumed via REST API. In an advanced implementation model, the customer could integrate Scribble into a web application, mobile application, or analytical portal via a

² Based on Scribble survey of practitioners for cost distribution of adhoc analytics, and pilot users for impact assessment



Scaling Up Adhoc Analytics Productivity With Scribble

gateway. The gateway handles application-level integration, data accesses, and any additional needs that the customer has. This means that **users can go from a natural language data search in Scribble to actually seeing the data, all through a single interface.** Data never leaves the network. Thus, Scribble can work with any organization-specific experience design, data governance, resource scheduling, and other enterprise capabilities. Also, this style of implementation enables the organization to avoid expensive upgrade cycles and costs. Scribble has capable system integration partners should a deeper integration with seamless experience be required.



Scribble can be integrated into an enterprise application easily

Feature Set

A quick summary of the features is as follows:

CHALLENGE	DETAIL	IMPACT DIMENSION	RELEVANT FEATURE
HUMAN LIMITATIONS	Looking up schema	Time	✓ Humanized interface
	Looking up SQL syntax	Time	✓ Automatic SQL generation
	Extensive typing	Errors	✓ Autosuggest
HIGH COORDINATION COST	Limited reuse	Time	✓ Replay, Sharing, Voting
	Non-standard definitions	Errors	✓ Built In catalog
	Unreadable code	Errors	✓ Automatic formatting
NON-STANDARD PROCESS	Difficult to audit	Errors	✓ Automatic audit logs
	Manual execution	Time	✓ API
INTEGRATION	Inability to embed adhoc analytics into applications	Functionality	✓ API



Security

Scribble is **highly secure and enterprise ready**. It belongs to the relatively new class of safer metadata-driven services that operate directly not on data but on metadata (data about data):

1. Scribble requires **NO access to data**.
2. All interactions with service only use SSL/HTTPS
3. Authentication and authorization mechanisms exist to limit visibility of metadata within an organization, and definitely across organizations
4. All interactions can be achieved through an API and can be logged if needed
5. Additional mechanisms can be deployed via a gateway if needed,

Scribble's approach dramatically reduces the time to deployment without any additional risk.

Talk to Us

Scribble has these and a number of other exciting capabilities in the works addressing the emerging needs of data users. Reach us at contact@scribbledata.io, and we'd be happy to discuss any aspects of data at your company, or see if the Scribble tool can help add business value to you.