Hillview
Budiu, Mihai; Gopalan, Parikshit; Suresh, Lalith; Wieder, Udi; Kruiger, Han; Aguilera, Marcos K.

Published in:
Proceedings of the vldb endowment

DOI:
10.14778/3342263.3342279

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Hillview: A trillion-cell spreadsheet for big data

Mihai Budiu
mbudu@vmware.com
VMware Research

Parikshit Gopalan
pgopalan@vmware.com
VMware Research

Lalith Suresh
lsuresh@vmware.com
VMware Research

Udi Wieder
uwieder@vmware.com
VMware Research

Han Kruiger
Han Kruiger
University of Groningen

Marcos K. Aguilera
maguilera@vmware.com
VMware Research

ABSTRACT

Hillview is a distributed spreadsheet for browsing very large datasets that cannot be handled by a single machine. As a spreadsheet, Hillview provides a high degree of interactivity that permits data analysts to explore information quickly along many dimensions while switching visualizations on a whim. To provide the required responsiveness, Hillview introduces visualization sketches, or vizketches, as a simple idea to produce compact data visualizations. Vizketches combine algorithmic techniques for data summarization with computer graphics principles for efficient rendering. While simple, vizketches are effective at scaling the spreadsheet by parallelizing computation, reducing communication, providing progressive visualizations, and offering precise accuracy guarantees. Using Hillview running on eight servers, we can navigate and visualize datasets of tens of billions of rows and trillions of cells, much beyond the published capabilities of competing systems.

PVLDB Reference Format:
DOI: https://doi.org/10.14778/3342263.3342279

1. INTRODUCTION

Enterprise systems store valuable data about their business. For example, retailers store data about purchased items; credit card companies, about transactions; search engines, about queries; and airlines, about flights and passengers. To understand this data, companies hire data analysts whose job is to extract deep business insights. To do that, analysts like to use spreadsheets such as Excel, Tableau, or PowerBI, which serve to explore the data interactively, by plotting charts, zooming in, switching charts, inspecting raw data, and repeating. Rapid interaction distinguishes spreadsheets from other solutions, such as analytics platforms and batch-based systems. Interaction is desirable, because the analyst does not know initially where to look, so she must explore data quickly along many dimensions and change visualizations on a whim.

Unfortunately, enterprise data is growing dramatically, and current spreadsheets do not work with big data, because they are limited in capacity or interactivity. Centralized spreadsheets such as Excel can handle only millions of rows. More advanced tools such as Tableau can scale to larger data sets by connecting a visualization front-end to a general-purpose analytics engine in the back-end. Because the engine is general-purpose, this approach is either slow for a big data spreadsheet or complex to use as it requires users to carefully choose queries that the system is able to execute quickly. For example, Tableau can use Amazon Redshift as the analytics back-end but users must understand lengthy documentation to navigate around bad query types that are too slow to execute [17].

We propose Hillview, a distributed spreadsheet for big data. Hillview can navigate and visualize hundreds of columns and tens of billions of rows, totaling a trillion cells, far beyond the capability of the best interactive tools today. Hillview uses a distributed system with worker servers that provide storage and computation. It achieves massive data scalability with just a few servers (e.g., with eight commodity servers it supports a trillion spreadsheet cells).

The main challenge facing Hillview is to provide near real-time performance despite having to compute over big data.

To address this challenge, Hillview invokes a common idea in database design: specialize the engine [91]. Rather than using a general-purpose analytics engine, Hillview introduces a new engine specialized to render the tabular views and charts of a spreadsheet. The main technical novelty of the paper is how to accomplish this specialization: we introduce the notion of visualization sketches or simply vizketches, and we propose a new distributed engine to render visualizations quickly using vizketches.

Vizketches combine ideas from the algorithms and computer graphics communities. In the algorithms community, mergeable summaries [2] are approximate computations that compute results over disjoint subsets of the data, that can then be merged to obtain the final result. Mergeable summaries are useful to distribute the computation efficiently with fine control over the accuracy and resolution of the result. A vizketch combines mergeable summaries with a basic principle in computer graphics rendering: compute only what you can display. A vizketch, thus, adjusts its accuracy and resolution to match the display resolution and compute only what can be visually discerned. For example, a vizketch for producing histograms limits the number of bars to ≈100 and computes the height of each bar only to the nearest pixel; these choices reduce the network communication and enable computation over big data. If the user zooms in on the histogram, the vizketch adapts to the new visualization to adjust the histogram buckets and enhance the accuracy of the bars while avoiding the computation of bars that are no longer visible.
Vizketches play a crucial role in Hillview. They parallelize the computation, reduce communication bandwidth, enhance computation efficiency, permit a progressive visualization of results, provide a precise accuracy guarantee, and ensure scalability (§4.4). These benefits are key for a spreadsheet to be able to browse big data at interactive speeds. Furthermore, vizketches can always be computed efficiently. This feature differentiates Hillview from traditional visualization solutions, which let users specify broad declarative queries without exposing their performance to the user, which is problematic for efficiency or usability. That leads to an important question about Hillview: are the queries supported by vizketches rich enough to implement a fully functional spreadsheet? A contribution of this paper is to answer this question positively.

To render visualizations quickly, Hillview introduces a new distributed engine to compute vizketches (Fig.1). Clients access the system via a user interface in a web browser (top of figure), while the dataset is partitioned across a set of worker servers (bottom). The user interface triggers a visualization, such as a histogram on a chosen column. To produce the visualization, the system executes two phases: preparation and rendering. The preparation phase computes broad parameters required to produce a proper visualization—for example, a histogram needs to find the data range and number of items to determine appropriate bucket boundaries and sampling rates. Next, the rendering phase computes the values required for the visualization—for example, the height of each histogram bar. This phase utilizes a vizketch to compute with the minimum accuracy for a good visualization. The rendering phase produces partial results that incrementally update the visualizations, so the client sees an initial visualization quickly while computing the values required for the visualization—for example, the values required for the visualization. The client then sees more precise results. Both preparation and rendering phases use an execution tree to distribute the computation across the workers. The engine provides other important functionality that we describe in the paper: caching computations, distributed garbage-collection, and failure recovery. Furthermore, the engine has a modular design that allows developers to add visualizations easily using new vizketches without dealing with concurrency, communication, and without needing to understand the structure of an existing query optimization engine; in practice supporting for a new storage layer or for a new visualization type can be added in a couple of person-days of work.

The engine of Hillview differs fundamentally from general-purpose query engines in two important ways. First, due to the characteristics of vizketches, Hillview queries are scalable by construction: more specifically, queries are guaranteed to run in time $O(n/c)$, produce results of length $O(\log n)$, using memory of size $O(\log n)$ where $n$ is the number of elements in the dataset and $c$ is the number of worker cores. In addition, many queries run in time $O(1)$. Second, Hillview produces compact results designed to be rendered efficiently on the screen. By contrast, general-purpose engines are not concerned about efficiency renderings; their queries could produce large results that take longer to visualize than to compute (e.g., returning billion points to be plotted)

We evaluate Hillview and its vizketches. We find that Hillview can support tables with 1.4 trillion cells while providing fast response. With this scale and data in memory, operations take 1–15 seconds. Hillview displays an initial partial views even faster, which is incrementally updated until it converges to the final view. With cold data read from an SSD, operations take 2–24 seconds, and an initial view still appears within seconds. For datasets with hundreds of billions of cells, Hillview computes complete answers in under a second for most queries. This is faster than the current approach of connecting a visualization front-end to a general-purpose analytics back-end. We also find that Hillview has broad functionality for answering a wide range of questions. Vizketches are an order of magnitude faster than a popular commercial in-memory database system to compute histograms; and their performance scales linearly or sometimes super-linearly with the number of threads and servers.

To demonstrate the usability of Hillview, we provide a short video and a live demo running on AWS using small EC2 instances (these links are also available in our github repository):

Video: [https://1drv.ms/v/s!AlywK8G1COQjeRQafBqJa3tvk4FQ](https://1drv.ms/v/s!AlywK8G1COQjeRQafBqJa3tvk4FQ)


In summary, in this paper we propose Hillview, a spreadsheet for big data. Hillview makes two novel contributions. First, it introduces vizketches, an idea that combines mergeable summaries with visualization principles; we give vizketches for each chart and tabular view in Hillview, by finding appropriate mergeable summaries and parameters to render information efficiently yet provably accurately. Second, Hillview demonstrates how to efficiently compute vizketches by introducing a new scalable distributed analytics engine that caches computations, performs distributed garbage-collection, and handles failure recovery, while achieving the scalability and speed required for an interactive spreadsheet.

While the above contributions are pragmatic, we believe this work also contains a fundamental contribution. We raise and defend two hypotheses: (1) mergeable summaries are powerful enough to efficiently and accurately visualize massive datasets, and (2) spreadsheets can significantly benefit from a specialized engine. Hillview demonstrates these hypotheses empirically by giving vizketches for many visualizations, by building an engine for vizketches, and by quantifying its benefits. Hillview is an open-source system with an Apache 2 license, available at [https://github.com/vmware/hillview](https://github.com/vmware/hillview).

Due to space limitations, we provide an extended version of this paper [13], with additional details: a formal computational model that captures vizketches, formal definitions of correctness and efficiency, detailed descriptions of vizketches, and correctness proofs.

---

1 Assuming a balanced partition of the data between workers.
2. WHY A NEW ENGINE

In a famous paper, Stonebraker et al. advocate for designing database systems targeted for specific domains, because doing so can dramatically improve performance over one-size-fits-all solutions [91]. This approach has worked well for several domains: data warehousing, stream processing, text, scientific, online transaction processing, etc. More recently, Fisher [38] and Wu et al. [100] point to the need for collaboration between visualization and data management systems. Hillview arises from these insights: we apply the database specialization approach to big data spreadsheets, where existing solutions fall short in scale and performance.

Hillview raises an important question. Data analysts may want to apply rich pipelines to data involving different frameworks, tools, and programming languages. For example, they may use a statistical package in R, then apply a machine learning algorithm in C++, followed by some hand-written scripts in python. How can Hillview integrate in this environment, given Hillview’s specialized engine?

Hillview addresses this concern by adopting a versatile data layer that can connect to other tools in the pipeline. In particular, Hillview can operate directly on data stored in SQL databases, NoSQL systems, JSON files, CSV files, columnar-oriented files such as Parquet or ORC, and other big-data systems (Hadoop+Spark, Impala), without any data transformation overheads. This is because Hillview does not require data ingestion to produce indexes, or repartition data: the efficiency of vizketches permits Hillview to operate on raw data partitioned horizontally in arbitrary ways across servers: there are no requirements that partitions contain contiguous intervals or specific hash values. The only requirements of the data layout is that (1) data be horizontally partitioned ideally with approximately equal-sized partitions available to each worker, and (2) data does not change while Hillview is running. The latter requirement can be met by using a data layer that provides snapshots, immutable data, or by pausing data modifications while Hillview runs. If a processing pipeline meets these requirements, then it is easy to insert Hillview into the pipeline. For example, we can connect the output of a batch-processing system to Hillview for exploration, and then output Hillview visualizations as data files or images that are processed by subsequent tools in the pipeline.

3. GOALS AND REQUIREMENTS

Our main goal is to develop a big data spreadsheet. As a data analytics tool, we are interested in functionality to explore and summarize data, such as navigation, selection, and charts. These are mostly read-only operations—our tool is for analytics exploration rather than transaction processing, data wrangling, cleaning, etc. So, we are less interested in providing interactive editing functionality, but we wish to provide ways to compute new columns from existing ones (e.g., compute a ratio of two columns). We now explain our requirements in more detail.

3.1 Why trillions of cells

Even small and medium companies can generate a trillions cells of data. These companies collect data over time from their servers, where each server might produce logs and metrics hundreds of times per minute, and a data center could have dozens of such servers. For example, 50 servers logging 100 columns at a rate of 100 rows per minute generate in a month 21.6B cells on 216M rows, or 1T cells and 10B rows in 46 months.

3.2 Environment

We target an enterprise computing environment, with tens of commodity server machines in a rack hosted in a private or public cloud. We want to use as few servers as possible, as most companies cannot afford thousands of servers to run a spreadsheet.

3.3 Tabular views functionality

At first thought, it is unclear what a spreadsheet with a billion rows should do. Clearly, paging through all rows is ineffective, but analysts may wish to find patterns and then inspect individual rows.

In our experience browsing big data, we found that a spreadsheet must support at least the following functionality:

- Select data based on rich criteria to produce fewer rows (e.g., rows with timestamps in the past hour).
- Select columns to show (e.g., date and server).
- Sort by a set of columns (e.g., date first, server next).
- Aggregate duplicates and show repetition counts (e.g., selecting just date and server could create millions of repetitions: all entries produced by each server on each day).
- Search free-form text (e.g., server Gandalf) by exact match, substring, regular expressions, case sensitivity, etc.
- Move a page forward or backward.
- Scroll forward and backward using a scroll bar.
- Extract features using tools such as heavy hitters (finds most frequent elements) and Principal Component Analysis [84].

We consider whether this functionality suffices in §7.5 but we expect the list will grow over time, much like conventional spreadsheets have evolved, so we also need a flexible framework that allows us to extend the system.

3.4 Visualization functionality

We are also interested in obtaining various visualizations of columns we choose. But we face a problem with big data: graphs with billions of points can produce useless black blobs and other clutter. We want to support visualizations that can avoid this problem [86, 33], such as histograms, stacked histograms, and heat maps (Figure 2). These visualizations generalize charts, such as x-y plots and bar charts (subsumed by heat maps); and pie charts (subsumed by heavy hitters [133]). We also want to extend the system with future new visualizations.

For each visualization, we want to inspect the value of individual points, change parameters (e.g., # buckets in histogram) and, if applicable, understand trends, correlations, and swap axes. Furthermore, we want to zoom in parts of the data, by regenerating the visualization for a subset of its data as determined by a mouse selection.

3.5 Other features

Data types. We want to support integers, floating-point numbers, dates, free-form text, and strings describing categorical data.

Map functions. We want to produce a new column by combining existing ones using user-defined map functions (e.g., a ratio of two columns).

Data sources. We want to read data from a variety of common sources (comma-separated files, SQL databases, row columnar files such as ORC, future formats, etc).

4. VIZKETCHES

Key to providing the required performance of Hillview, vizketches are a simple concept that combine the idea of mergeable summaries (or sketches) from the algorithms community with

---

This requirement is common in data warehousing and analytics systems.
<table>
<thead>
<tr>
<th>Name</th>
<th>What it shows</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDF</td>
<td>Distribution of variable</td>
<td># events before noon</td>
</tr>
<tr>
<td>Histogram</td>
<td>Frequency of variable for each bucket</td>
<td># events per hour of day</td>
</tr>
<tr>
<td>Stacked histogram</td>
<td>Frequency of first variable and frequency of second variable grouped by first</td>
<td># events of each type per hour of day</td>
</tr>
<tr>
<td>Normalized stacked hist.</td>
<td>Ditto but bars normalized</td>
<td>% of events of each type per hour of day</td>
</tr>
<tr>
<td>Heat map</td>
<td>Frequency of two variables</td>
<td># events for each server and hour</td>
</tr>
<tr>
<td>Trellis plots</td>
<td>Arrays of the other plots grouped by one or two variables</td>
<td># events for each server and hour, for each datacenter</td>
</tr>
</tbody>
</table>

Figure 2: Some clutter-free visualizations for large datasets. Visualizations cover a single variable (column) or multiple variables, up to four.

the principle of visualization-driven computation from the graphics community.

### 4.1 Background

**Mergeable summaries.** Intuitively, a summarization method computes a compact representation ("summary") of a large dataset, which can then answer approximate queries on the dataset. A summarization method is mergeable if the summary can be obtained by merging many summaries computed independently over parts of the dataset. More precisely, a mergeable summarization method consists of two functions `summarize(D)` and `merge(S, S')`. The first takes a dataset `D` and returns a summary; the second merges two summaries and returns another summary. A summary is small compared to `D`—typically by many orders of magnitude—and it can approximate queries on `D` (the allowable queries depend on the choice of summarization method). Summaries of two separate datasets can be merged via the `merge` function:

\[
\text{summarize}(D_1 \cup D_2) = \text{merge}(\text{summarize}(D_1), \text{summarize}(D_2))
\]

where `D_1` and `D_2` are multisets and \( \cup \) is multiset union. There are summarization methods for many types of queries, such as histograms, heavy hitters, heat maps, and PCA. Many summarization methods are sketches from the streaming algorithms literature, and so the community sometimes mixes these two concepts. However, a summarization method can also use sampling, which can be more efficient because it does not scan all data. The summarization method has two accuracy parameters: an error \( \varepsilon \) and an error probability \( \delta \), with the guarantee that an approximation computed from a summary has error at most \( \varepsilon \) with probability \( 1 - \delta \). For a more formal description of our computational model, we refer the reader to Appendix A of the extended version of this paper.

**Visualization-driven computation.** In computer graphics, rendering is an expensive operation that must be optimized. To do that, a basic principle is to drive the computation based on what will be visualized and its resolution, taking into consideration the limits of human perception and the lossy channels of displays. This principle is behind many graphics techniques, such as ray tracing, culling, and imposters.

### 4.2 Basic idea

A vizketch is a mergeable summary designed to produce a good visualization. More precisely, a vizketch method targets a specific visualization (e.g., a histogram) with a given display dimension (width and height in pixels). The vizketch method consists of the
two functions of a mergeable summary, summarize and merge, with parameters carefully chosen to achieve two goals: the summary is small, and it permits a good rendering of the visualization.

Small summary means that its size depends only on the length of the description of the visualization, not on the input size. More precisely, visualizations are inherently limited by the finiteness of their renderings, so they have a short description (e.g., a histogram is described by its bucket boundaries and heights). The length of this description is a lower bound on the size of the summary. We seek summaries whose size is polynomial in this length, rather than the data set size. The key hypothesis behind Hillview is that visualizations always admit vizketches with such small summaries. This hypothesis is not obvious; it can be formalized with proper definitions of the computational model, visualizations, etc., but this is outside the scope of this paper. Instead, Hillview supports this hypothesis empirically: we give vizketches for many visualizations, by adapting techniques from the sketching/streaming literature.

Good rendering means two things. First, the rendering has a bounded error with high probability (e.g., histogram bars are off by at most 1 pixel). Second, the rendering is not cluttered (e.g., there are at most 50 buckets for a histogram when the screen width is 200 pixels). The precise requirements are carefully chosen for each type of visualization. These choices are made so that a human can consume the information effectively without perceiving any errors in the approximation.

To use vizketches, Hillview defines a computation tree whose nodes are assigned to the servers (Figure 1). Hillview assumes that the data is stored on a distributed storage layer, and is shredded into small chunks, which are distributed to the tree leaves. The sharding can be arbitrary: chunks need not be sorted or partitioned by a specific key.

To perform a visualization, each leaf independently runs the vizketch’s summarize function on the shards that it has; this function might choose to sample or scan the data in the chunk. The summaries are then merged along the computation tree, using the vizketch’s merge function. The root receives the final summary, which reflects a view of the entire dataset and produces the rendering of the visualization for the client.

Vizketches parallelize the computation across threads and servers, while reducing computation and network bandwidth to only what is necessary for a good rendering. They can also provide partial results for progressive visualizations, in addition to other benefits (§4.4). We now describe specific vizketches.

4.3 Algorithms

Hillview uses a large number of vizketches. Some produce graphs (histograms, stacked histograms, heat maps, trellis plot); others produce information for the spreadsheet tabular view (next items, quantile for scroll bar, find text, heavy hitters). We describe a few here; others are omitted due to space limitations but they follow a similar approach and can be found in Appendix B of [13]. Vizketches have rigorous guarantees of correctness, which we present in Appendix C of [13].

A vizketch is parameterized by the target display resolution, and produces calculations that are just precise enough to render at that resolution.

Histograms. We are given a numerical column (or a value that can be readily converted to a real number, such as a date) with range \([x_0, x_1]\), a number \(B\) of histogram bars, and their maximum pixel height \(V\). The histogram vizketch (Figure 3(b)) divides the range \([x_0, x_1]\) into \(B\) equi-sized intervals, one per bin. To maximize use of screen, we should scale the bars so that the largest one has \(V\) pixels. We furthermore allow an error of .5 pixels in the estimation of the height of each bar. We prove in Appendix C of [13] that to obtain this error with probability < 1 − δ, it is sufficient to sample \(n = O(V^2 B^2 \log(1/δ))\) items from the dataset. Notice that this function is independent on the dataset size, and depends only on the screen size. The summarize function outputs a vector of \(B\) bin counts, and the merge function adds two vectors.

Heat map. We are given two columns \(X\) and \(Y\) with ranges \([x_0, x_1]\) and \([y_0, y_1]\), and the pixel dimensions \(H \times V\). A heat map (Figure 3(d)) defines bins in two dimensions, where each bin consumes \(b \times b\) pixels (\(b\) is 2 or 3, depending on the screen resolution). Thus, we have \(B_x = H/b\) and \(B_y = V/b\) bins for \(X\) and \(Y\). The density of the data in a bin is represented by a color scale. If we use \(c\) distinct colors, the required accuracy for each bin density is 1/2c. This requires a target sample size \(n = O(c^2 B_x^2 B_y^2 \log(1/δ))\). The summarize function samples data with the target rate, counting the number of values that fall in each bin. It outputs a matrix of \(B_x \times B_y\) bin counts. The merge function adds two such matrices.

Next items. This vizketch is used to render a tabular view of the spreadsheet given the current row shown at the top \(R\) (or \(R = \infty\) to choose the beginning of the dataset). We are also given a column sort order, and the number \(K\) of rows to show. This vizketch returns the contents of the \(K\) distinct rows that follow \(R\) in the sort order. The summarize function scans the dataset and keeps a priority heap with the \(K\) next values following row \(R\) in the sort order. The merge function combines the two priority heaps by selecting the smallest \(K\) elements and dropping the rest.

Heavy hitters. A vizketch to find heavy hitters works by sampling. Let \(K\) be the maximum number of heavy hitters desired. The basic idea is to sample with a target size \(n\) (determined below), and select an item as a heavy hitter if it occurs with frequency at least \(3n/4K\). A statistical calculation shows that by picking \(n = K^2 \log(K/δ)\), with probability \(1 − δ\) we can obtain all elements that occur more than \(1/K\) of the time and no elements that occur fewer than \(1/4K\) of the time. This method is particularly efficient if \(K\) is small. We employ several other algorithm for finding heavy hitters, described in Appendix C of [13].

\footnote{Sampling can be used only if the map from count to color is linear.}
4.4 Benefits

Vizketches bring many benefits to Hillview. In the list below, the parentheses indicate from where the benefit is inherited: S means sketches/mergeable summaries, V means visualization-driven computation, and S+V means the combination of both.

- **Parallel computation** (S). Servers and cores within servers independently compute on different parts of the data, and the result is merged.
- **Bandwidth efficiency** (S+V). When a server finishes its computation, it communicates only a compact summary to be merged.
- **Computation efficiency** (S+V). Some computations can done over a small sample of data based on the required accuracy.
- **Progressive visualization** (S). As servers complete their computation, the system computes a partial summary that gradually progresses to the final result. This ensures that slow servers and tail latencies do not hinder interactivity. Users can cancel a visualization after seeing partial results.
- **Accurate visualization** (S+V). The resulting visualization has a precise accuracy guarantee.
- **Scalability** (S+V). As we add more data, vizketches can sample more aggressively to enhance efficiency while achieving the required accuracy.
- **Easy to obtain** (S). There is a rich literature on mergeable summarization methods and sketches of various types (histograms, heat maps, heavy hitters, etc); these sketches can often be converted into vizketches through a relatively simple analysis that translates the accuracy of the sketch into the required accuracy of the visualization, as illustrated above.
- **Modularity** (S). New visualizations can be added to Hillview by defining new vizketches as two simple functions (§4.1) without the developer worrying about distributed systems aspects.

5. DESIGN AND ARCHITECTURE

We now explain in detail the design and architecture of Hillview, starting with its high-level design choices (§5.1), followed by a detailed description in the subsequent sections.

5.1 Design choices

We now explain the key design choices of Hillview, which derive from the power and characteristics of vizketches.

- **Distribute computation while minimizing server coordination.** To answer a query, Hillview launches a computation tree to efficiently distribute the query to worker servers and aggregate the results according to the vizketch computations.
- **Storage-independence.** Hillview can access data in a wide variety of formats (SQL, NoSQL, text, JSON, etc), with few restrictions on how data is partitioned (§2), and without the need to pre-compute indexes or perform extract-transform-load. As a result, Hillview does not require any pre-processing to ingest data. This is beneficial to integrate Hillview into a diverse analytics pipeline (as explained in §2), and this is possible because the efficiency and parallelization of vizketches permits forgoing data conversions, repartitioning, and pre-computations.
- **Sample data in a controlled manner.** Sampling improves efficiency but introduces error. Vizketches allow Hillview to sample while bounding the error to what we can perceive.
- **Modular algorithms.** Programmers who write vizketch algorithms do not have to worry about concurrency, communication, or fault-tolerance; they just implement single-threaded code, and the architecture handles all such issues in a uniform and transparent manner.

5.2 Architecture

Figure 1 shows the architecture of Hillview. Hillview is designed as a cloud service accessible to clients through a web interface. A web browser handles user interaction with the spreadsheet and renders the charts incrementally as computation results arrive. To produce a visualization, a web server launches the required computation as one or more execution trees. Communication happens only along the edges of the tree, and is restricted to small messages: queries in one direction and summaries in the other. Each tree is rooted at the web server, followed by one or more layers of aggregation nodes, and several leaf nodes. The leaf nodes perform the actual computation over disjoint partitions of the dataset.

The execution of each tree is based on the `summarize` and `merge` functions (§4.2) of the mergeable summary. A tree executes in two phases.

The first phase initiates the computation from the root down the tree to each leaf, and causes the leaf nodes to apply the `summarize` method on their data partition. To parallelize execution within a server, each server runs multiple leaf nodes: there is a thread pool that serves leafs with work to do. To facilitate this process, the data partition within a server is divided into micro-partitions of 10-20M rows, each micro-partition assigned to a leaf.

The second phase, in its most basic form, executes from the leafs toward the root, causing each node to aggregate results from its children through the `merge` method. Thus, ultimately the root node combines the output of all nodes, and the result can be rendered.

When processing large datasets in a distributed system, there may be variation in the processing times across servers and partitions. If the root had to wait for all other nodes to finish, its completion would be disrupted by any stragglers, affecting the interactive experience of users. To address this problem, nodes periodically propagate partially merged results of the vizketch without waiting for all children to respond. Thus, the root receives partial results and sends them to the client UI before it gets the final results. The web browser then renders results as they arrive, so that users can see a progression of the computation. Hillview shows a progress bar that reflects the number of leafs that have completed. Users can cancel the computation based on the partial results they see.

There is a trade-off between the freshness of the partial results and the bandwidth savings produced by aggregating partial results. After receiving a result from a child node, aggregation nodes wait for 0.1 seconds and aggregate all results that arrive within this interval. This provides frequent updates to the UI; the increase in communication costs is modest because all vizketch results are small by construction.

Hillview allows users to cancel computations (e.g., because a partial visualization is satisfactory). This is done by interrupting
an execution tree with a high priority cancellation message that bypasses the queuing mechanisms in the communication between tree nodes. This message causes tree nodes to do two things: remove work for that computation that was previously enqueued, and ignore requests for micropartitions not yet started. We currently do not stop ongoing computations on a micropartition.

5.4 Data input, caching, and data output

Unlike most database systems, Hillview reads data repositories without pre-processing, repartitioning, or other optimizations. This is possible because the computational engine of Hillview—based on vizketches—makes few assumptions about the data. The assumptions are that repositories do not change while they are accessed (this can be provided by using storage snapshots or controlling write access) and data is horizontally partitioned, ideally with approximately equal-sized partitions available to each worker, so that data can be read in parallel. When a worker needs a column, it reads it completely from the data repository taking advantage of fast sequential access and columnar access if the repository supports it (SQL, Parquet, ORC). Once data is read, it is kept in an in-memory cache; the cache purges entries not used for a while (currently 2 hours).

Hillview uses two types of caching: data and computation. The first is an in-memory cache of the raw data in the data repositories. The data cache is organized by column to provide data locality, since vizketches tend to operate on relatively few columns.

The computation cache stores results produced by mergeable summaries; these results are small, allowing a large number of results to be cached. This is useful for mergeable summaries that provide auxiliary functionality, such as column statistics, which are used repeatedly and are deterministic. The computation cache is indexed by what mergeable summary was used and what dataset was operated on.

Hillview can save a derived table (§5.6) to a data repository, by having each worker store its partition of the data. This is implemented through a special vizketch with a summarize function that writes a data record to the repository and returns an error indication, while the merge function combines error indications.

5.5 Vizketch modularity and extensibility

The inherent structure of vizketches permits Hillview to cleanly separate them from the rest of its architecture so that developers can implement new vizketches without the hard concerns of distributed systems (communication, coordination, fault tolerance, etc) or data storage. Specifically, to support a new vizketch, a developer needs to implement the following things: (1) a serializable type for the vizketch summary, (2) an implementation of the summarize and merge functions of the vizketch; these all operate on the in-memory columnar representation of the data, and are independent on the storage layer, (3) code to render the vizketch summary as a visualization in the user interface of the spreadsheet in the browser, (4) code to trigger the visualization through a user interface action, and (5) a function to connect the user interface action to the invocation of the vizketch in the root node. None of these functions are concerned with concurrency (they are single-threaded), and most of them can be implemented with only tens of lines of code—the sole exception is (3), which requires more code to provide the graphical functionality. We quantify the effort to for step (2) in §7.4.

5.6 Data transformations

Users may wish to generate new data from existing data as part of the data exploration process. Users can do that externally to Hillview through other analytics tools, and then import the results into Hillview for inspection (§7). Alternatively, Hillview provides some support for deriving new data through two common operations: selection (filtering) and user-defined map operations (§7).

Selection permits a user to create a new table that contains a subset of the rows of another table (e.g., rows where the year column is 2019). A particularly useful selection operation in a spreadsheet is to zoom in part of a graph, which corresponds to choosing the rows within the zoom window. To provide this functionality, Hillview allows mergeable summaries to work on subsets of rows of the dataset. More precisely, a table can be derived from other tables by choosing a subset of the rows. To save space, the tables share common data and store a “membership set” data structure that identifies which rows are contained in the table. This membership set data structure has different implementations, depending on the density of the filtered data. Dense tables that contain most rows store a bitmap, while sparse tables store a hashset of the rows indexes. This information is kept locally for each data partition.

When executing the summarize method, some vizketches work by sampling a subset of rows. We must ensure that sampling is efficient (it does not require reading each row) but it is also correct (it picks rows uniformly at random). For sparse tables, we generate the first sample by choosing a random row number for the first element; we generate the following samples by returning the next elements in sorted order of their hash values. For dense tables we walk randomly the bitmap in increasing index order.

User-defined maps permits a user to create a column from existing ones (e.g., add two columns), where the map is an arbitrary function. Some map functions are built-in (e.g., converting strings to integers); additional functions can be written by users in Javascript. To support this functionality, Hillview creates a new table with the new column populated by running the map function at the leafs of the execution tree. Currently, this data is stored only in the in-memory caches; if the cached data is reclaimed, the column is recomputed on demand. We believe this is a reasonable choice for a spreadsheet, since derived columns tend to be short-lived.

5.7 Memory management

Early versions of Hillview used a distributed garbage-collection protocol to handle memory management. This protocol was complex and fragile (for example, loss of network messages could cause memory leaks). In the current version we have simplified memory management by aggressively using only soft state: all in-memory data structures are disposable, including at leaf-, aggregation- and root nodes. The only requirement to implement this architecture is for the storage layer to provide an API to read a particular snapshot of a dataset; in this way, in-memory data is reconstructed by reloading the original snapshot. We use the Partitioned Data Set architecture from Sketch [14] to represent distributed objects; unlike sketch, all remote references are “soft”— they may not point to valid data structures.

Each machine performs independently garbage-collection; a caching layer maintains a working set of recently accessed objects in memory. In-memory cached objects at leaf nodes can be reconstructed by reading data from disk; tables obtained from filtering (§5.6) or by deriving new columns (§5.6) can be regenerated by re-executing the operation that created them in the first place.

When the root node attempts to access a remote object on a leaf which no longer exists the leaf reports an error. The root node then re-executes the query that produced the missing object. This may
require re-executing other queries, that produced the source objects; the recursion ends when data is read from disk.

To enable query re-execution, the root node maintains a redo log with all executed operations. The redo log is the only persistent data structure maintained by Hillview (recall that the storage layer is not part of Hillview).

5.8 Fault tolerance

Hillview provides fault tolerance by logging operations that initiate each execution tree, and lazily replaying operations to reconstruct node state. When the root node restarts after a failure, it reads the redo log to memory, but does not replay it yet. Replaying occurs only when the user tries to access a dataset that no longer exists, as described in §5.7.

Worker nodes are stateless, so restarting the node after a failure is equivalent to deleting all cached datasets. These datasets are reconstructed by the root node on demand by replaying log operations.

This lazy approach is suitable for a spreadsheet, because most views are short-lived results that a user never accesses again.

For this replay mechanism to work, vizketches must be deterministic, otherwise a restarted node becomes inconsistent with nodes that never crashed. To provide determinism for randomized vizketches (e.g., those that use sampling), the log includes the seed used for randomization.

6. IMPLEMENTATION

Hillview consists of 35000 lines of Java and 16000 lines of TypeScript code. The user interface in the browser is implemented in TypeScript [85], using parts of the D3 JavaScript library [11]. Graphics is done using SVG [28]. The web server runs the Apache Tomcat application server [4]. The browser gets progressive replies from web server using a streaming RPC based on Web Sockets [32]; these RPC messages are serialized as JSON. The cloud service is implemented in Java. We use Java’s type-safe object serialization facilities for sending queries and data between machines. We use the fast collections Java library [34] for efficient data structures, with customizations for faster sampling. For server-side JavaScript we use Oracle Nashorn [73].

We use a variety of open-source libraries to interface with external storage layers (e.g., csv files, various log formats (e.g., RF C5424), JDBC connectors, columnar binary formats such as Parquet or ORC, etc). The communication between back-end machines uses GRPC [44]. The core communication APIs are based on reactive streams, using RxJava [80] [64]. We use RxJava’s Observable datatype for many purposes: (1) It represents a stream of partial results. (2) It offers support for operation cancellation, through its unsubscribe method. (3) It is used for reporting progress to the user for long-running operations, displayed in the form of progress bars, and (4) it is responsible for managing concurrent execution on multi-core machines (using the observeOn(threadpool)); this thread pool is used for all of the workers’ computations. The in-memory tables use as much as possible Java arrays of base types to reduce pressure on the Java GC. String columns use dictionary encoding for compression.

7. EVALUATION

Our evaluation goal is to determine whether Hillview provides interactive performance with large data sets, how Hillview compares to existing systems, how vizketches contribute to that goal, and how effective the spreadsheet is.

Summary. We find the following results:

- Hillview can handle spreadsheets with 130B rows and 1.4T cells using only 8 servers. At the upper range, visualizations can take 20s when loading from disk, but the first partial visualization appears in a few seconds and gets gradually updated. This is much better than existing systems (§7.1). For smaller datasets most response times are on the order of hundreds of milliseconds.
- Vizketches perform well on a single thread and scale well with the number of threads and servers. Vizketches based on sampling scale super-linearly. This performs significantly better than a database system (§7.2).
- Vizketches are key in Hillview: they implement a broad range of functionality of the spreadsheet, to the extent that they are the sole way to access data in the system (§7.3).
- Vizketches are easy to code and do not require an understanding of distributed systems (§7.4).
- Hillview is a spreadsheet with many useful features, able to answer a diverse set of queries effectively (§7.5).

Testbed. Our testbed consists of eight servers running Linux kernel 4.4. Each server has two sockets with 14-core 2.2Ghz Intel Xeon Gold 5120 CPUs, 192 GB of DRAM, two SSDs with 381GB and 1.8TB, connected to a 10 Gbps network. The client web browser runs on a laptop connected to the servers via a 100Mbps network with 1ms ping time to the servers. This setup represents a typical enterprise setting.

Dataset. We use a dataset with US airline flight performance metrics for the past 20 years [71]. Each row has a flight with its origin, destination, flight time, departure and arrival delays, etc. This is a real dataset with numerical, categorical, text, and undefined values. There are 130 million rows and 110 columns, which amount of 58.2 GB of uncompressed data. In some experiments, we scale the dataset by a factor of 5, 10, or 100, by replicating its rows and reading them repeatedly from disk. These datasets are labeled “FlightKx” where K=1, 5, 10, 100 indicates the replication factor (K=1 is the original dataset).

7.1 Hillview end-to-end performance

We measure the end-to-end time that Hillview takes to execute spreadsheet operations for datasets of various sizes.

Baseline. We compare Hillview against the traditional approach for big-data spreadsheets, such as Tableau, which is to connect a visualization front-end to a general-purpose analytics back-end. Our baseline uses a Spark back-end, and we measure only the analytics delay (not the rendering delay), giving an advantage to the baseline. We optimize Spark to our best ability. We write queries in Scala; we pre-load all data to RAM before measuring; and we use the same optimizations for each query as Hillview, including sampling.

Workload. Figure 4 shows the visualizations we are measuring. We picked these operations using two criteria: (1) Each group of operations corresponds to a user action in the spreadsheet (e.g., ask for a histogram, or change the sort order of a tabular view). (2) The operations cover a broad range of vizketches available in Hillview.

Setup. In each experiment, we pick an operation, a dataset size, and a system. The dataset sizes vary from 5x–100x the original data, corresponding to 650M–13B rows of data with 110 columns each, for a total of 71B–1.4T cells. We submit the operation to the system and measure its response time and amount of data received by the root node. For Hillview, we submit the operation at the user interface of the web browser, and we measure two response times
We now consider the base performance of vizketches on one thread, and its scalability over threads and servers. We run each measurement multiple times, and we display the variance of measurements after excluding the fastest and slowest measurements; the variance tends to be small.

\[ \text{Variance} \]

\[ \text{variance after excluding the fastest and slowest measurements} \]

\[ \text{small} \]

\[ \text{generally much slower} \]

\[ \text{Hillview provides acceptable performance for interaction. In our experience using Hillview, we tend to spend significantly more time browsing and analyzing charts than waiting for visualizations (cf §7.5).} \]

\[ \text{Figure 4} \]

\[ \text{Spreadsheet operations. The + indicates serial operations, while & indicates concurrent operation. Numerical data refers to integer or floating point.} \]

\[ \text{at the browser: first partial visualization and final visualization.} \]

\[ \text{For the Spark baseline, we start the measurement when the computation starts, and end the measurement when the query result is obtained.} \]

\[ \text{For Hillview, we consider two cases: data is in memory before the measurement, and data is cold on disk (SSD).} \]

\[ \text{We only consider the case with data in memory.} \]

\[ \text{Figure 5} \]

\[ \text{Results. Figure 5 shows the results for warm data in memory. We could not run Spark with a dataset larger than 5x because it exhausted the memory at the servers: for example, the 10x dataset has 582 GB on-disk but its in-memory representation expands beyond the available aggregate memory in the tested system.} \]

\[ \text{The top graph shows the response time. We see that for most operations, Hillview performs at least as well as Spark, even when Hillview processes twice the data. We also see that Hillview at 100x can be slow to compute all results: 7.3–15.2s.} \]

\[ \text{However, Hillview produces a partial visualization quickly, which provides a better interactive experience.} \]

\[ \text{The bottom graph shows the amount of data received over the network by the root node (for Hillview) or the master (for Spark); note that the Y axis is log-scale.} \]

\[ \text{Spark consumes an order of magnitude more bandwidth than Hillview, except for O11. This is because Hillview transmits a small amount of data to produce the visualizations.} \]

\[ \text{The exception, O11, is a heatmap, which contains a large number of cells and hence its vizketch carries considerable more data.} \]

\[ \text{We also see that Hillview consumes more bandwidth with a larger dataset. This is because the larger dataset takes longer to complete, and so Hillview transmits partial visualizations; with O11, the total amount of data becomes larger than Spark, but it is still reasonable at 3.5MB.} \]

\[ \text{Figure 6} \]

\[ \text{Results. Figure 6 shows the results for cold data on disk. For 5x and 10x data, visualizations still complete in 3s. For 100x, the delay can be 24s; first visualizations arrive earlier, often within 2.5s (not shown).} \]

\[ \text{In all cases, Hillview provides acceptable performance for inter-} \]

\[ \text{action. In our experience using Hillview, we tend to spend significantly more time browsing and analyzing charts than waiting for visualizations (cf §7.5).} \]

\[ \text{7.2 Vizketch microbenchmark} \]

\[ \text{We now consider the base performance of vizketches on one thread, and its scalability over threads and servers.} \]

\[ \text{We run each measurement multiple times, and we display the variance of measurements after excluding the fastest and slowest measurements; the variance tends to be small.} \]

\[ \text{The first measurement warms up the Java JIT compiler, so it is generally much slower.} \]
Figure 5: End-to-end performance comparison. The top graph shows the response time to produce each visualization, while the bottom graph shows how many bytes the root node received. Here, we ensure the data is in memory before the measurement starts. The bars are labeled with the system name (Spark or Hillview) and the dataset size (5x to 100x corresponding to 650M to 13B rows). Hillview100xF is the time it takes for Hillview to produce the first partial visualization running with 100x data.

Figure 6: End-to-end performance of Hillview when data is not in memory, so it needs to be loaded from SSD. Not shown are first visualizations, which arrive within 2.5s most of the time, and within 4s always. O4 and O6 are omitted because in the spreadsheet these operations never happen with cold data (a prior action loads the data).

Figure 7: Scalability of vizketches as we add more leafs and shards together. Ideal scalability would be constant latency.

the same effect: the sample size remains constant, so the amount of work per server decreases with the number of servers.

7.3 Vizketch applicability

We consider our experience of using vizketch to implement the various spreadsheet functionality, to gain an understanding of the applicability of vizketches to processing data in Hillview.

When we started the project, we did not know if vizketches would suffice or we would need more powerful computation mechanisms. In building the system, however, we found vizketches to be powerful and capable of implementing a broad range of functionality: tabular views, scrolling, simple data transformations, filtering, table summaries, and various visualizations. We eventually realized that we could implement all data visualization functionality of Hillview using vizketches; in fact, Hillview has no other way to visualize data other than vizketches.

7.4 Vizketch coding effort

We now turn our attention to the effort required to write vizketches. We again report on our experience with Hillview. Quantitatively, Figure 9 shows the number of lines of back-end (Java) code required to implement each vizketch in Hillview. We
can see that the code is compact: the largest vizketch takes only 191 lines of code. We found that an expert takes only a few hours to implement and test the code. However, some vizketches involve fairly sophisticated algorithms: selecting or developing those algorithms took considerably longer than implementing them. In general, developing the UI to display the data and provide user interaction requires considerably more effort.

Qualitatively, implementing vizketches never required thinking about distributed systems or concurrency. A developer simply provides the summarize and merge functions, which are purely local, while the rest of Hillview takes care of the distributed systems aspects of running vizketches across many cores and servers. Of course, we had to implement the distributed execution framework for vizketches in Hillview, but this implementation was done once and benefits all vizketches, including future extensions.

7.5 Hillview effectiveness: case study

We next consider the question of how effective Hillview is to browse and answer queries on large datasets. We address this question through a case study.

Workload. A person who is not familiar with Hillview examines the Flights-1x data set and formulates a set of questions (shown in Figure 10) that interests her and that she thinks the dataset answers.

Setup. The experiment is carried out by an operator who is familiar with Hillview well but does not know the questions ahead of time. In each experiment, we show a question to the operator and ask him to answer it using Hillview. Our goal is to understand if the spreadsheet is powerful enough to answer the question and, if so, how easily it can do that. Note that this experiment does not evaluate ease-of-use by beginners, because the operator is an expert. This is intentional: Hillview users are not casual users but data analysts, whose job is dedicated to explore data and so they can obtain the required training.

For each question, we measure the time and number of spreadsheet actions that the operator takes to answer the question. A spreadsheet action consists of choosing an operation on a menu, clicking on the spreadsheet, or dragging the mouse to select a region. For example Q1 can be answered by filtering the main table for column Airline=UA, producing a histogram on DepartureDelay, then going back to the main table and filtering for column Airline=AA, producing a second histogram on DepartureDelay. To answer the question, we hover the mouse over the histograms to find the delay percentiles.

Results. Figure 11 shows the results. Answering a question took at most 6:44 (minutes:seconds), with most questions taking less than 2:30 (all except three). The average and median times are 1:57 and 1:12. Most of the time is the operator thinking about what to do, rather than waiting for the spreadsheet to respond (if the operator knew exactly what to do, all queries could be answered in under 30 seconds). The minimum and maximum number of actions were 1 and 6, with mean and median 3.4 and 3. Queries Q4, Q6 and Q10 did not have completely satisfactory answers because the spreadsheet cannot clearly separate dates (Q4, Q10) or the spreadsheet did not merge and duplicate the destinations (Q6). Question Q20 could not be answered because the dataset does not have the information (e.g., we discovered that it lacks the downed flights on 9/11). We see that Hillview was effective at addressing most queries at small times, showing that (1) Hillview implements enough functionality to be usable and (2) it provides an interactive experience for human timescales.

8. RELATED WORK

Hillview is the first spreadsheet to scale massively with interactive speed. Hillview borrows ideas from the algorithms and computer graphics literature, namely mergeable summaries or sketches) and visualization-driven computation; it uses relies on many techniques from databases (approximate query processing, on-line analytics), big-data analytics, and distributed systems.

Hillview follows Shneiderman’s visualization mantra “overview first, zoom and filter, details on demand”. Fisher identifies principles for interactively visualizing big data (“look at less of it” and “look at it faster”); these principles guided the design of vizketches.
Big data visualization is a broad area; we give an overview of the closest related work below. For more information, we refer the reader to several surveys in the area [22,42,41,8,9]. Compared to published systems, Hillview achieves the best scalability for the amount of resources: we are not aware of any system that can handle a trillion cells with only 8 servers.

**Distributed visualization engines.** Hillview evolved from Sketch [14], which proposes a distributed data exploration library with applications to a performance analyzer and a spreadsheet. VisReduce [50] provides incrementally updated approximations of visualizations computed over progressively larger samples. Vizdom [23] is a simple UI for data manipulation and exploration; it runs on top of the A-WARE smart caching and streaming engine [24] and uses the Tupleware analytics system [22].

**Visualization using big data query engines.** One way to visualize big data is to connect a visualization engine to an analytics engine, such as Hive [93], Impala [60], Presto [78], Dremel [63] (commercialized as BigQuery), Drill [46], PowerDrill [45], Spark [103], Druid [102], or Pinot [49]. This approach has advantages: it reduces design effort by using existing systems, and it leverages the years of effort spent in their optimization. However, this approach does not achieve the speed needed for a spreadsheet: the generality of analytics engine imposes overheads and computes unnecessary results, since there is no integration with the visualization engine. Several systems follow this approach. Microsoft PowerBI [67] using DirectQuery [26] and Polaris/Tableau [90,98,99] provide plug-ins to many analytics engines; as discussed in [17], the users of such systems have to carefully avoid many queries that cannot be answered efficiently. IBM BigSheets [12] computes interactively only over a subset of the data; once the user settles on a query, it is actually run in batch mode using Spark. HadoopVis [32] uses Hadoop to render geo-spatial data. ScalaR [6] uses relational databases; the system in [96] uses MapReduce for mesh rendering and isosurface extraction. SwiftTuna [51,52] uses Spark. OmniSci [72] uses GPUs in one machine for server-side rendering.

Facebook’s Scuba [1] has been used as a back-end to visualization systems. Scuba provides fast response times but with a different trade-off between data scale, responsiveness, and correctness. Scuba computes much more than “what you can see”, since its compute engine is decoupled from its visualization. Thus, queries might return unbounded amounts of data to the visualization engine, hampering real-time responsiveness. To avoid that, Scuba truncates worker responses to 100,000 rows (page 4) and omits workers that do not respond in 10ms (page 4). This can produce arbitrarily incorrect visualizations.

The vizketch computational model is similar to the MPI Reduce [87] primitive used in supercomputing, to the Neptune system [18], to the architecture of log analytics systems such as Splunk [88], and to aggregation networks for sensor networks [79]: these are general-purpose platforms, and not visualization systems.

**Sampling and indexing.** Sampling and indexing are used to accelerate visualization in many systems. [15] considers the problem of sampling a database for minimizing the error for a given set of queries. BlinkDB [6] uses stratified sampling, which is effective, but leaves the burden on users to write appropriate SQL queries and find appropriate error and time bounds. Smart sampling is used by [59,97] uses stratified sampling to accelerate queries in log management systems. [101] uses stratified sampling in Scope to reduce sample sizes while minimizing errors; samples are incrementally maintained. Pangloss [69] uses “optimistic” visualization on sampled data to provide fast results.

The idea of using perceptual limitations to drive sampling appears first in [30,55] uses perceptual limitations and sampling algorithms for specific chart types (e.g., bar charts). M4 [53] uses the screen resolution to rewrite SQL queries to compute reduced results suitable for renderings of line plots; this is extended for other chart types in VDDA [54]. Sample+Seek [29] executes responsive aggregated queries on a single table; it uses measure-biased sampling together with new indexing schemes, specific to the aggregation computed, to minimize errors. G-OLA [104] handles interactive aggregate OLAP queries over massive data sets.

VAS [76] samples data to minimize the visualization errors for scatter-plots. SynopViz [10] and Skydive [43] build hierarchical multi-scale models of the data for browsing linked data sets.

**Progressive analytics.** Hillview visualizations are incrementally updated; this technique is called online aggregation [47] or progressive analytics [89,98,94,35]. There is significant work on this topic. MapReduce Online [20] is based on MapReduce.

EARL [61] uses statistical bootstrapping for providing reliable online early estimates for the output of MapReduce computations. Progressive Insights [89] finds common subsequences in event series of medical records, focusing on its UI design for incremental display and exploration. PIVE [16] adapts computation to limited screen resolution for iterative algorithms (such as clustering or dimensionality reduction). DimXplorer [24] performs progressive computation and rendering of dimensionality reduction operations (such as clustering and PCA); it uses sampling for fast response times. Stat! [5] operates in conjunction with a streaming engine, and presents immediately incremental results. Microsoft’s Tempe [66] runs on top of a streaming engine and provides progressive visualizations.

All above systems lack some of the benefits of vizketches: parallelization, computation efficiency, and bandwidth efficiency (Section 4), which are required for Hillview.

Nanocubes [62], imMens [63], and Hashedcubes [74] improve interactivity by pre-processing the data to build smart indexes. DICE [55], Sesame [56] and ForeCache [7] use sessions and locality to pre-compute views or to reuse computation results between consecutive views. These ideas improve interactivity, but restrict the scope of queries to pre-processed columns (e.g., the user pre-selects a few columns to optimize) or spend significant time for pre-processing. By contrast, Hillview uses no pre-processing or indexes, because we do not know ahead of time which columns the user might choose to explore.

VisTrees are indexes designed to support quick histogram construction for visualizations [31]. Profiler [57] and Foresight [28] propose methods to find abnormality in the data; Hillview could incorporate this functionality, especially Foresight which is based on sketches. NeedleTail [59] uses a small in-memory index to allow fast browsing and displaying any-k records. AQP++ [74] combines approximate query processing with aggregate pre-computation.

9. CONCLUSION

Hillview is a spreadsheet that supports a trillion cells even with a modest number of servers. Hillview introduces a new query execution engine specialized to render tabular views and charts for a spreadsheet. The new engine uses vizketches, a new but simple idea that parallelizes computation and calculates only what is needed for a good visualization. We believe Hillview is a useful tool for humans to explore data; it nicely complements other tools, such as analytics frameworks, which have other uses.
10. REFERENCES


M. Stonebraker, S. Madden, D. J. Abadi, S. avros Harizopoulos, N. Hachem, and P. Helland. The end of an architectural era (it’s time for a complete rewrite). In _International Conference of Very Large Data Bases (VLDB)_, pages 1150–1160. Sept. 2007.


