Larry Clark
Strategic Sales Consultant, Tableau Software
Agenda

Big Data in the Tableau Context
Repeating Patterns
Evolving Solutions
Customer Examples
What is Big Data?

Any data set that you must process faster than your current platform has the capability to process data.
What is Big Data?
What is Big Data?
What is Big Data?
What is Big Data?
What is Big Data?
Big Data in the Tableau Context
Big D
The Key Value of Tableau: Visual Analytics
The Effects of Interactive Latency on Exploratory Visual Analysis

Zicheng Liu and Jeffrey Heer

Abstract—To support effective exploration, it is often stated that interactive visualizations cannot tolerate interactive latency. However, the effects of interactive latency on the process and outcomes of exploratory analysis have not been systematically studied. We present an experiment measuring user behavior and knowledge discovery under different latency conditions. We observe that an additional delay of 500ms incurs significant performance penalties. Analyzing verbal data from think-aloud protocols, we find that increasing latency leads to decreased exploration depth, less precise observation, and more calculations. Moreover, we note that higher latencies lead to lower overall performance in a low-latency setting. The effects of interactive latency on exploration strategy and performance have not been systematically studied.

Index Terms—Interaction, latency, exploratory analysis, interactive visualization, search

1 INTRODUCTION

One stated goal of interactive visualization is to enable data analysis at “rates resonant with the pace of human thought” [19, 20]. Goal latencies of two research directions: understanding the rate of cognitive activities in the context of visualization, and supporting these cognitive processes through appropriately designed and performed systems. Latency is a central issue underlying these research problems. Due to the time required for query processing, data transfer, and rendering, data-intensive visualization systems incur delay. It is generally held that low latency leads to improved usability and better user experience. Unsurprisingly, multiple research efforts focus on reducing query and rendering latency for large datasets, which may include billions or more data points. Latencies in state-of-the-art systems can range from 20 milliseconds up to multiple seconds for a task [2, 28, 29].

Despite the shared goal of minimizing latency, the effects of interaction delays on user behavior and knowledge discovery with visualization remain largely unexplored. While previous research on the effects of interactive latency in puzzle solving [4, 17, 35, 36] and search [8] has shown that user behavior changes in response to milliseconds,
Time Scales of User Experience

- <0.1 s
- <1 s
- 1-5 s
- >10 s
Query Latency Matters
Achieving Our Query Latency Goal: Repeating Patterns
Patterns We Have Seen Before: The Evolution of EDW and OLAP

Technology Inflection Points:
- Advanced ETL Software
- Declining Memory Prices
- OLAP software running on commodity SMP systems

Technology Inflection Points:
- Commodity SMP DB systems
- MPP data warehouse based on commodity hardware
- SAN storage
The Pattern Repeats

Technology Inflection Points:
- Distributed computing (Hadoop) on commodity server software (Linux) & SMP hardware
- Declining storage pricing

Technology Inflection Points:
- Cloud based data warehouse servers
- Decoupled storage and compute

Technology Inflection Points:
- Containerization
- Columnar Data Stores
- Decoupled storage and compute
Big Data Architecture – Cold Tier

Data Size

Large data (Unstructured or semi-structured)

- Everything in any Format
- Unknown questions with unknown answer
- Data Mining/Data Science/Data Exploration
- AKA the Data Lake

Prepared data (structured)

Aggregated Data

Performance
Big Data Architecture – Warm Tier

- Large data (Unstructured or semi-structured)
- Prepared data (structured)
  - Structured/Defined meta-data
  - Known questions with unknown answers
  - Data Warehouses/Data marts
- Aggregated Data
Big Data Architecture – Hot Tier

- Large data (Unstructured or semi-structured)
- Prepared data (structured)
- Aggregated Data

- Known questions and known answers
  - Precomputed aggregates
  - Analytical databases/In-Memory/Hyper/Query Accelerators
Technology and marketplace innovations are blurring these layers.
Optimizing the Tableau User Experience with Big Data Solutions
Balance

Data volume
Level of detail

Speed of access
Data prep
Cost
Pattern 1: Directly Query Big Data from Tableau
Pattern 2: ETL from data lake to RDBMS / MPP
Solution; Tableau queries RDBMS
Which data warehouse is the fastest?

Histogram of performance for 99 TPC-DS queries (seconds, log scale) with geometric mean.
Demo: Tableau + MPP
Pattern 3: Big Data Accelerator Technologies
Demo: Tableau + Snowflake
Pattern 4: Tableau Hyper Extracts

Tools required to create extracts:
- Tableau Prep Builder & Conductor
- Third party ETL and Data Prep tools
- Tableau Extract API
Demo: Tableau + Hyper
Customer Examples
Netflix Data Platform

We have REALLY big data
Updated messaging platform with Amazon Kinesis

Added Amazon Athena to our analytics ecosystem in support of discovery and ad-analyses

New configuration helps us maximize benefit while reducing costs without jeopardizing speed or performance

Successfully deployed Tableau Server to ~1,900 internal users in effort to scale self-service analytics
Microsoft Azure

Raw data
- Flat files
- Application data
- Server logs
- Internet APIs

Collect/Store
- Azure Blob Storage

Store/Analyze
- Azure HDInsight
- Spark
- Databricks
- Hive

Data warehouse
- Azure SQL DB or DW

Analysis/Sharing
- Tableau Platform
  - Collaboration
  - Analytics
  - Content Discovery
  - Governance
  - Data Prep
  - Data Access

Note: Tableau Server & Desktop can be deployed on Azure VMs.
Google Cloud Platform

Raw data
- Flat files
- Application data
- Server logs
- Internet APIs

Store/Process
- Cloud Storage
- Cloud DataPrep
- Compute Engine

Databases
- Cloud SQL
- Coming Soon: BigTable/Spanner
- Coming Soon: Dataproc

Data Warehouse
- BigQuery
- Spreadsheets
- Sheets
- Web data

The Tableau Platform
- Analytics
- Collaboration
- Content Discovery
- Governance
- Data prep
- Data access
Snowflake on AWS, Azure, or GCP*

Data Sources

- ELT
- ETL

EDW

Logical Datamarts

Data Lake

Guided Analytics

Deliver rich interactive dashboards to support data-driven decision making for key business processes.

Unconstrained Self-Service Analytics

Empower domain experts to discover new insights from their data.
INTRODUCING THE WORLD’S FIRST ENTERPRISE DATA CLOUD

CLOUdera DATA PLATFORM

Unified control plane

Management Console
Identity | Orchestration | Management | Operations

Analytic experiences

Data Flow & Streaming
Data Engineering
Data Warehouse
Operational Database
Machine Learning

Data Hub & Cloudera Runtime

Data anywhere

Catalog | Schema | Replication | Security | Governance

Any infrastructure

Edge
Private Cloud
Public Multi-Cloud
Hybrid Cloud
What have we not addressed? (Tableau context)

- Cost Details
- Streaming Data
- Data Catalogs / Metadata
- ETL / Data Prep
- AI & Machine Learning Tools
Query Latency Matters
Can we achieve the goal?
Related Sessions

• Hot, Hotter, Hyper: How to Handle Big Data
• Brown Advisory: Journey to the Cloud: Building an Analytics Program with Tableau and Snowflake
• Amazon Redshift and Tableau: Shifting to High Gear
• Next Gen Analytics with Tableau in Google Cloud Platform
• Getting the Most Out of Tableau with Google Cloud Data
• Data Done Right = #Winning (Denodo)
Please complete the session survey in the mobile app

View ‘My Evaluations’ in the menu or find your session under ‘Schedule’
Thank You
TABLEAU CONFERENCE