

Big Data Analysis

Practical 10

Objective :

Case study: Use following platforms for solving any big data analytic problem of your choice. (1) Amazon web services,(2) Microsoft Azure, (3)Google App engine

Roll No. & Name : 18bce183 & Prince Prajapati

Submitted to: Prof. Jaiprakash Verma

Real-time Data Access problem -AWS:

In some cases, you can simply map your big data analytics workload into one of the services based on a set of requirements. However, in most real-world, big data analytic workloads, there are many different, and sometimes conflicting, characteristics and requirements on the same data set.

For example, some result sets may have real-time requirements as a user interacts with a system, while other analytics could be batched and run on a daily basis. These different requirements over the same data set should be decoupled and solved by using more than one tool. If you try to solve both of these examples using the same toolset, you end up either over-provisioning or therefore overpaying for unnecessary response time, or you have a solution that does not respond fast enough to your users in real time. Matching the best-suited tool to each analytical problem results in the most cost-effective use of your compute and storage resources.

Big data doesn't need to mean "big costs". So, when designing your applications, it's important to make sure that your design is cost efficient. If it's not, relative to the alternatives, then it's probably not the right design. Another common misconception is that using multiple tool sets to solve a big data problem is more expensive or harder to manage than using one big tool. If you take the same example of two different requirements on the same data set, the real-time request may be low on CPU but high on I/O, while the slower processing request may be very compute intensive.

Decoupling can end up being much less expensive and easier to manage, because you can build each tool to exact specifications and not overprovision.

With the AWS pay-as-you-go model, this equates to a much better value because you could run the batch analytics in just one hour and therefore only pay for the compute resources for that hour. Also, you may find this approach easier to manage rather than leveraging a single system that tries to meet all of the requirements. Solving for different requirements with one tool results in attempting to fit a square peg (real-time requests) into a round hole (a large data warehouse).

Decoupling in Microsoft azure:

Use Azure Queue Storage to build flexible applications and separate functions for better durability across large workloads. When you design applications for scale, application components can be decoupled, so that they can scale independently. Queue storage gives you asynchronous message queueing for communication between application components, whether they are running in the cloud, on the desktop, on-premises or on mobile devices.

Google App engine :

DATA ANALYTICS SOLUTION	HOW GOOGLE CLOUD HELPS YOU TRANSFORM YOUR BUSINESS
Data warehouse modernization	Streamline your migration path and modernize from any on-premises or cloud data warehouse with BigQuery. Solve for today's most challenging demands and seamlessly scale your business with AI built-into BigQuery ML and powerful multicloud capabilities with BigQuery Omni.
Data lake modernization	Empower your teams to securely and cost-effectively ingest, store, and analyze large volumes of diverse, full-fidelity data. Use Dataproc for data lake modernization, ETL, and secure data science. Run SQL queries on your data lake with BigQuery and reduce your storage costs with Cloud Storage.

Streaming analytics	Ingest, process, and analyze event streams in real time to make data more useful and accessible from the instant it's generated. Dataflow unifies streaming and batch data analysis and builds cohesive data pipelines. Pub/Sub ingests hundreds of millions of events per second. BigQuery's streaming API allows real-time streaming for SQL-based analysis.
Business intelligence	Develop a strategy to modernize BI and put data at the center of your business transformation with Looker and BigQuery. Unlock in-memory analysis for sub-second query response with BigQuery BI Engine and analyze billions of rows of BigQuery data in Connect Sheets. Build an interactive dashboard with Data Studio.
Marketing analytics	Gain a more comprehensive picture of the customer journey, predict outcomes, and create personalized experiences. Use BigQuery Data Transfer Service to move your marketing data from Google Marketing Platform to BigQuery, where you can analyze data and build predictive audiences with BigQuery ML. Build an interactive dashboard with Data Studio.
Geospatial analytics & AI	Rethink how geospatial relationships and insights can help you realize a more prosperous and sustainable future for your business. Unlock and accelerate

	innovation with a uniquely powerful, full-platform solution built on category-leading Google products like BigQuery, Earth Engine, Google Maps Platform, and more.
--	--

CONCLUSION:

After implementing this practical now I have clear understanding of AWS, Microsoft Azure and Google App Engine work with big data and what are the tools they are using.