Question #36*Topic 1*

Flowlogistic Case Study -  
  
Company Overview -  
Flowlogistic is a leading logistics and supply chain provider. They help businesses throughout the world manage their resources and transport them to their final destination. The company has grown rapidly, expanding their offerings to include rail, truck, aircraft, and oceanic shipping.  
  
Company Background -  
The company started as a regional trucking company, and then expanded into other logistics market. Because they have not updated their infrastructure, managing and tracking orders and shipments has become a bottleneck. To improve operations, Flowlogistic developed proprietary technology for tracking shipments in real time at the parcel level. However, they are unable to deploy it because their technology stack, based on Apache Kafka, cannot support the processing volume. In addition, Flowlogistic wants to further analyze their orders and shipments to determine how best to deploy their resources.  
  
Solution Concept -  
Flowlogistic wants to implement two concepts using the cloud:  
Use their proprietary technology in a real-time inventory-tracking system that indicates the location of their loads  
  
✑ Perform analytics on all their orders and shipment logs, which contain both structured and unstructured data, to determine how best to deploy resources, which markets to expand info. They also want to use predictive analytics to learn earlier when a shipment will be delayed.  
  
Existing Technical Environment -  
Flowlogistic architecture resides in a single data center:  
✑ Databases  
8 physical servers in 2 clusters  
- SQL Server `" user data, inventory, static data  
3 physical servers  
- Cassandra `" metadata, tracking messages  
10 Kafka servers `" tracking message aggregation and batch insert  
✑ Application servers `" customer front end, middleware for order/customs  
60 virtual machines across 20 physical servers  
- Tomcat `" Java services  
- Nginx `" static content  
- Batch servers  
✑ Storage appliances  
- iSCSI for virtual machine (VM) hosts  
- Fibre Channel storage area network (FC SAN) `" SQL server storage  
- Network-attached storage (NAS) image storage, logs, backups  
✑ 10 Apache Hadoop /Spark servers  
- Core Data Lake  
- Data analysis workloads  
✑ 20 miscellaneous servers  
- Jenkins, monitoring, bastion hosts,  
  
Business Requirements -  
✑ Build a reliable and reproducible environment with scaled panty of production.  
✑ Aggregate data in a centralized Data Lake for analysis  
✑ Use historical data to perform predictive analytics on future shipments  
✑ Accurately track every shipment worldwide using proprietary technology  
✑ Improve business agility and speed of innovation through rapid provisioning of new resources  
✑ Analyze and optimize architecture for performance in the cloud  
✑ Migrate fully to the cloud if all other requirements are met  
  
Technical Requirements -  
Handle both streaming and batch data  
  
✑ Migrate existing Hadoop workloads  
✑ Ensure architecture is scalable and elastic to meet the changing demands of the company.  
✑ Use managed services whenever possible  
✑ Encrypt data flight and at rest  
✑ Connect a VPN between the production data center and cloud environment  
  
SEO Statement -  
We have grown so quickly that our inability to upgrade our infrastructure is really hampering further growth and efficiency. We are efficient at moving shipments around the world, but we are inefficient at moving data around.  
We need to organize our information so we can more easily understand where our customers are and what they are shipping.  
  
CTO Statement -  
IT has never been a priority for us, so as our data has grown, we have not invested enough in our technology. I have a good staff to manage IT, but they are so busy managing our infrastructure that I cannot get them to do the things that really matter, such as organizing our data, building the analytics, and figuring out how to implement the CFO' s tracking technology.  
  
CFO Statement -  
Part of our competitive advantage is that we penalize ourselves for late shipments and deliveries. Knowing where out shipments are at all times has a direct correlation to our bottom line and profitability. Additionally, I don't want to commit capital to building out a server environment.  
Flowlogistic's CEO wants to gain rapid insight into their customer base so his sales team can be better informed in the field. This team is not very technical, so they've purchased a visualization tool to simplify the creation of BigQuery reports. However, they've been overwhelmed by all the data in the table, and are spending a lot of money on queries trying to find the data they need. You want to solve their problem in the most cost-effective way. What should you do?

* A. Export the data into a Google Sheet for virtualization.
* B. Create an additional table with only the necessary columns. **Most Voted**
* C. Create a view on the table to present to the virtualization tool. **Most Voted**
* D. Create identity and access management (IAM) roles on the appropriate columns, so only they appear in a query.

[Hide Solution](https://www.examtopics.com/exams/google/professional-data-engineer/view/8/) [Discussion   **40**](https://www.examtopics.com/exams/google/professional-data-engineer/view/8/)

**Correct Answer:** *C* [🗳️](https://www.examtopics.com/exams/google/professional-data-engineer/view/8/)

*Community vote distribution*

C (60%)

B (40%)