Proposed architecture (Assume Azure):

1) Streaming (Ingestion)
    - Azure Event Hubs is a real-time data ingestion service that can handle millions of events per second. It is the most suitable service on the Azure platform given the relatively high throughput required for streaming of Dataset A.
2) Stream Processing
    - Azure Databricks is a near real-time processing engine with exactly-once processing guarantees, and supports Spark.
3) Storage
    - Delta Lake on Azure Databricks is compatible with Spark and extends Parquet data with ACID transactions.
4) Serving
    - As specified by the PM, this is to be a dashboard. Power BI is most suitable here due to native integration with Azure

Questions to ask the end user:

1) What is the purpose of the dashboard?
2) Who are the users of the dashboard, if there are any others?
3) What metrics should be displayed on the dashboard?
4) What is the desired refresh frequency?
5) How long should events be stored?
6) In case of duplicate events, which event should be kept?