



How to build event-driven data ingestion pipelines in Azure?

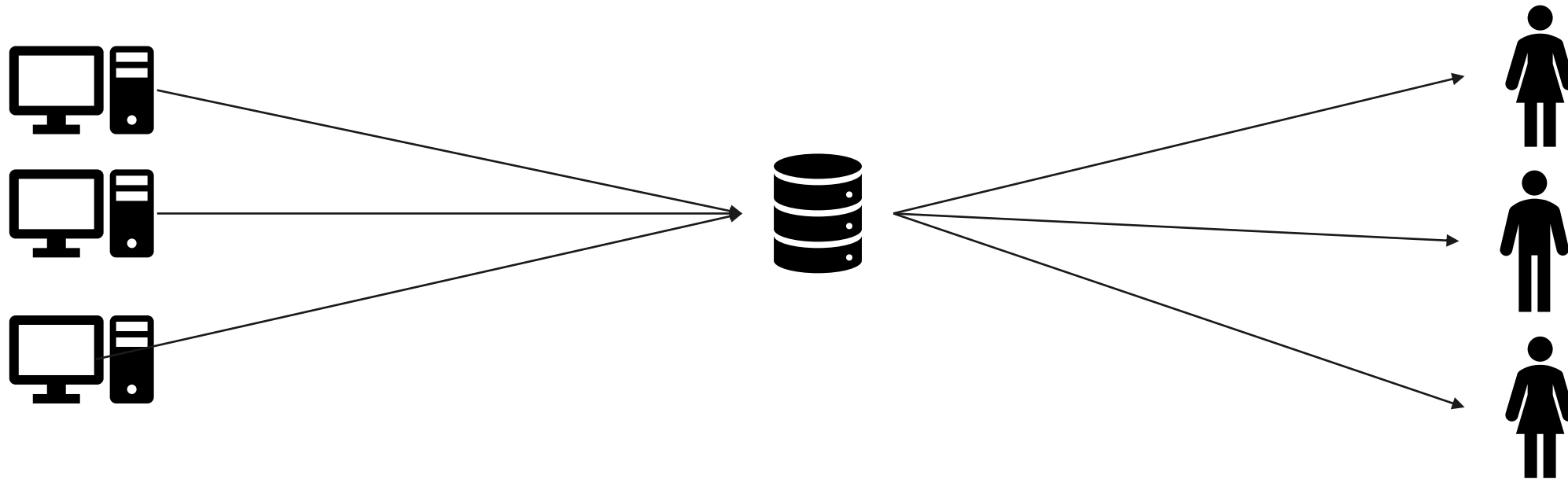
Hassaan Ahmed, Dag König
Microsoft



Customer Problem

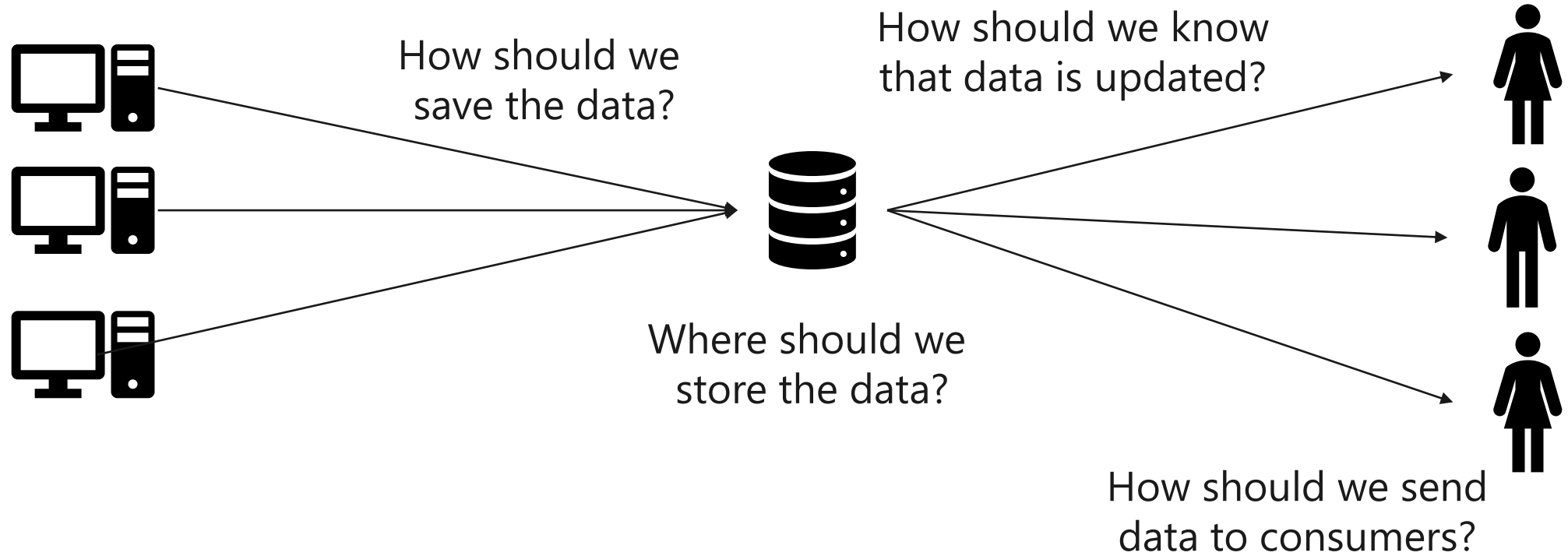
A big retailer company with many "source systems" that want a single truth of their data and be able to send updates to consumers when this data is changed.

They want to support an unpredictable load, with max of 1500 rec/sec.



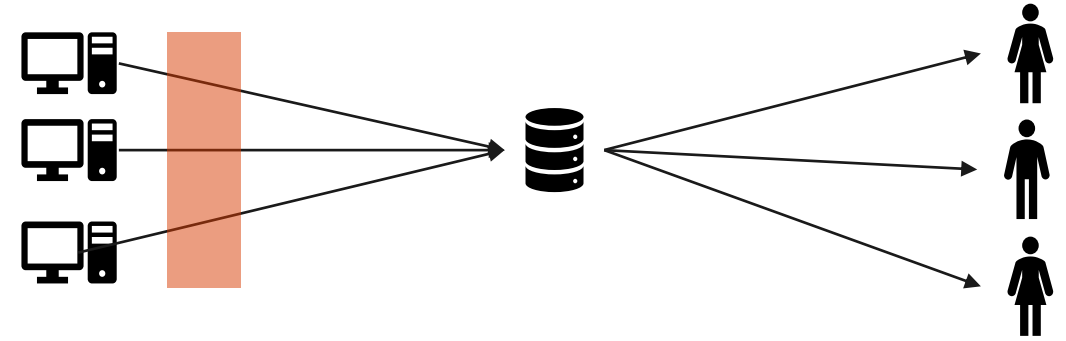
Questions

Ingestion of the data?

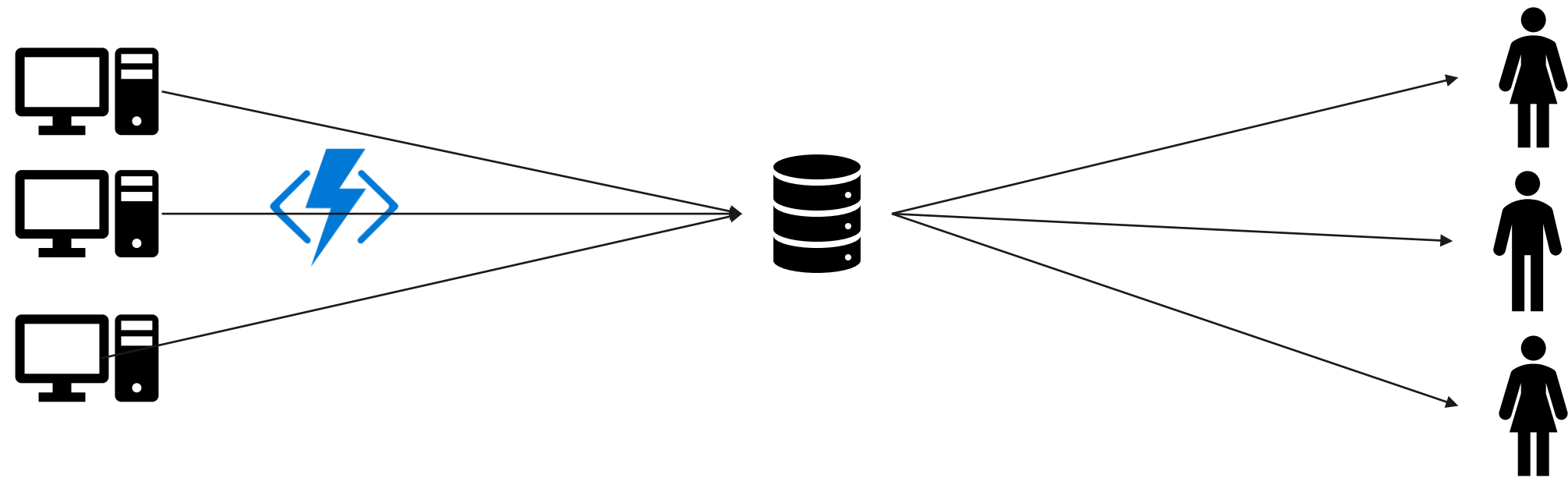
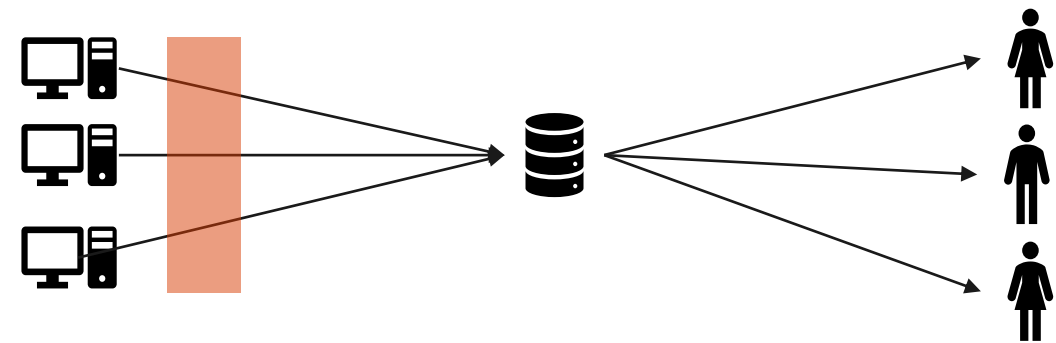


Ingestion of Data

- Hosting of Endpoint
 - IOT Hub / Event Hub / Service Bus
 - Serverless
 - Web App / API App
 - Functions
 - Service Fabric
 - Containers



Ingestion of Data



Load Testing

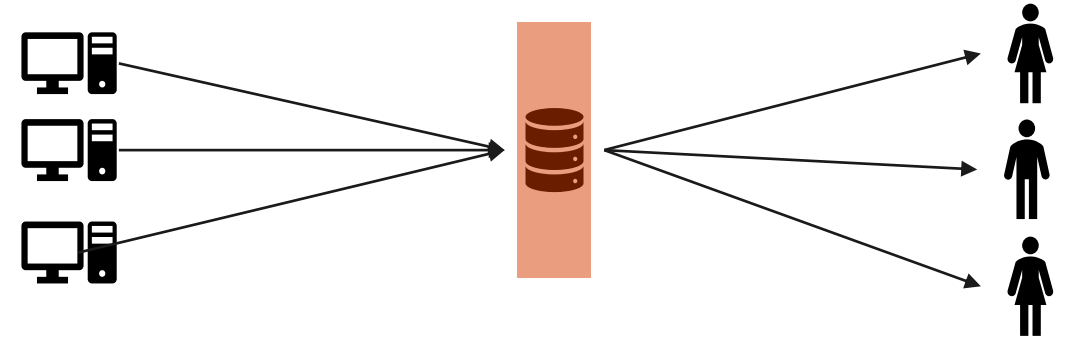
- WRK

<https://github.com/wg/wrk>

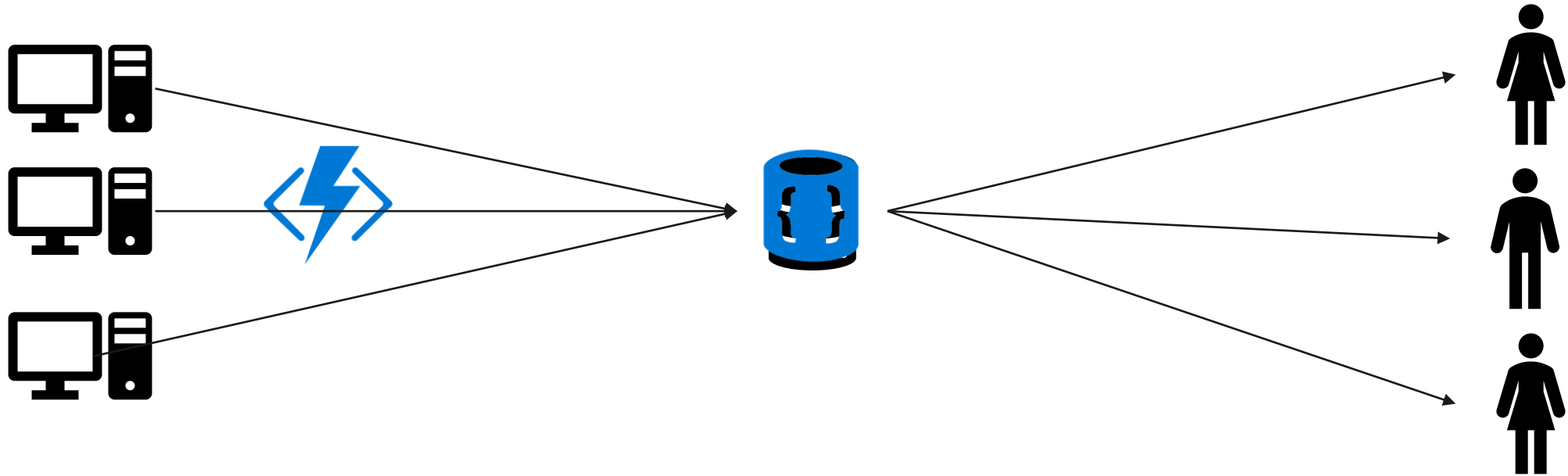
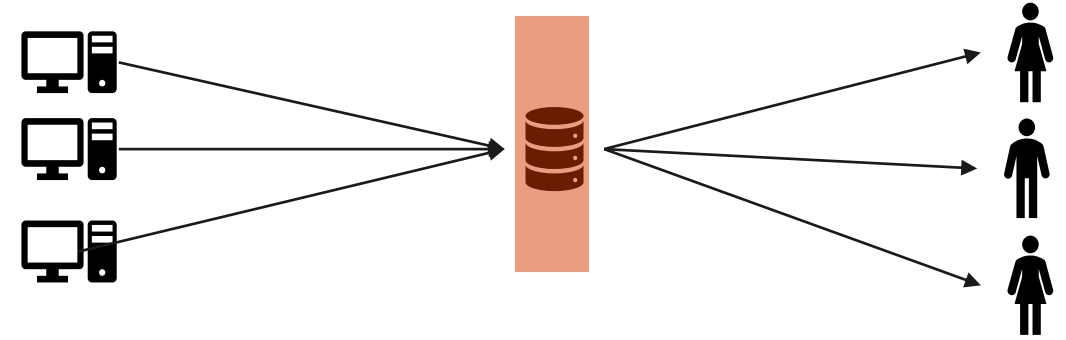
- Azure Container Instances (ACI)

Store the Data

- Managed
 - SQL
 - SQL Database, MySQL, MariaDB, PostgreSQL
 - NoSQL
 - CosmosDB (SQL API, Mongo API, Graph API, Cassandra API, Table API)
 - Azure Storage
 - Big Data
 - Azure Data Lake
- Unmanaged
 - SQL
 - SQL
 - ...
 - NoSql
 - MongoDB
 - ...
 - Big Data
 - ...



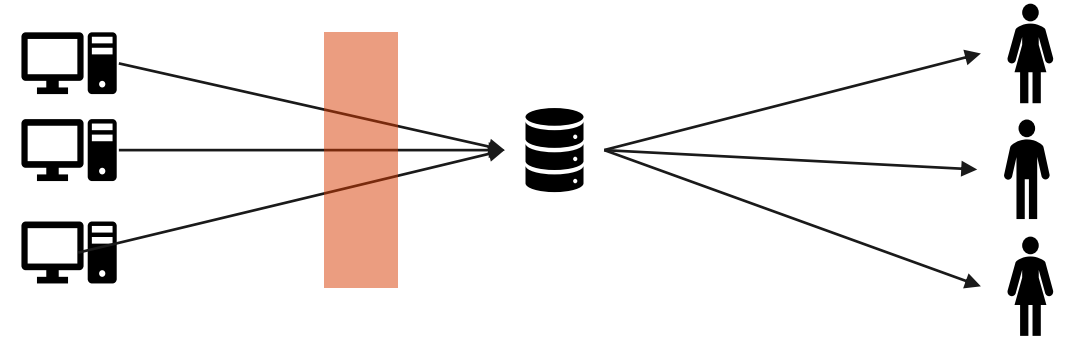
Store the Data



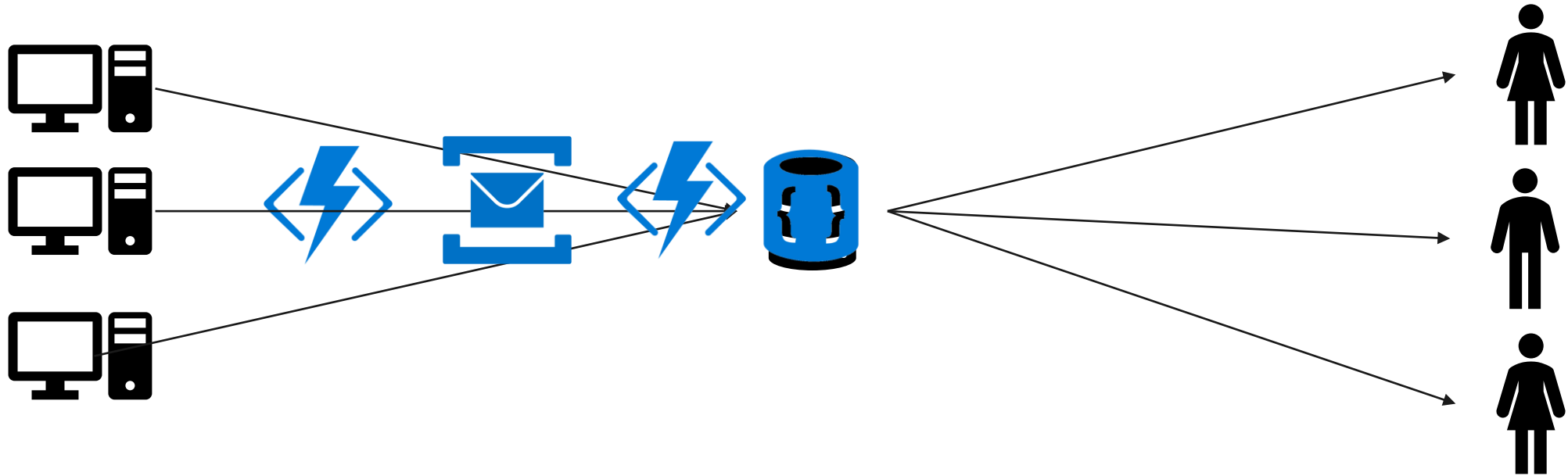
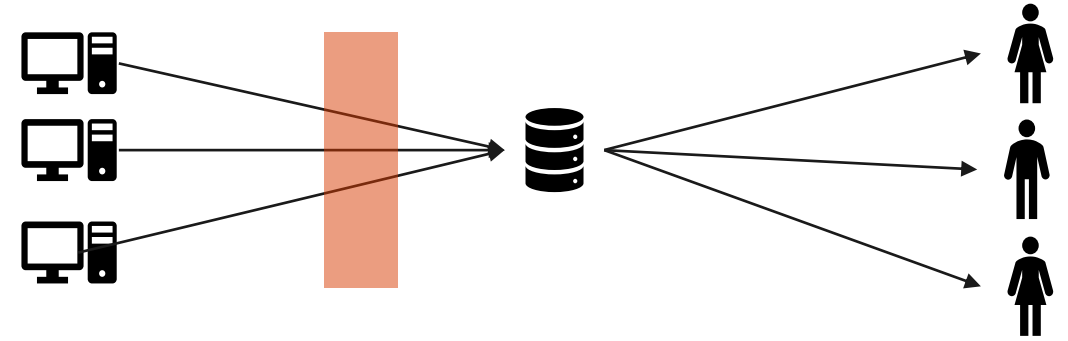
Save the Data

Events (changes)

- Synchronously
 - Direct (the old web architecture)
- Asynchronously
 - Messages
 - Event Hub
 - Service Bus
- Insert / Replace / Upsert



Save the Data



Things to think about when using CosmosDB

- Request Units (RU)
- Indexies
- Partitions
- Consistency Levels

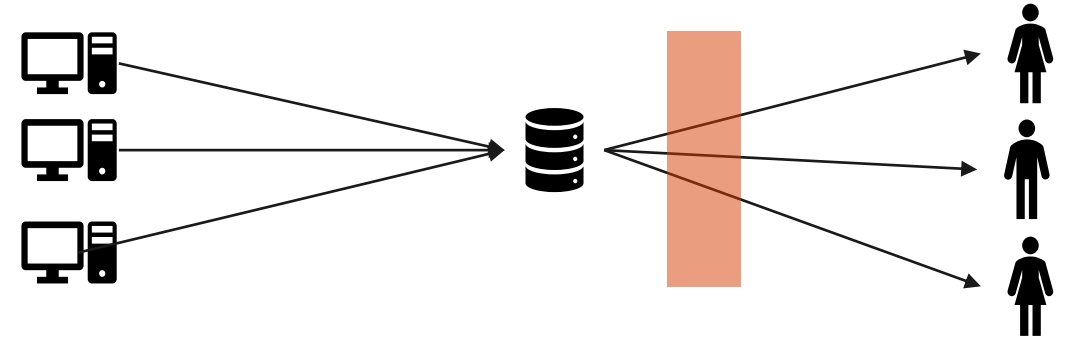
CosmosDB - What does it cost?

RU – Request Units

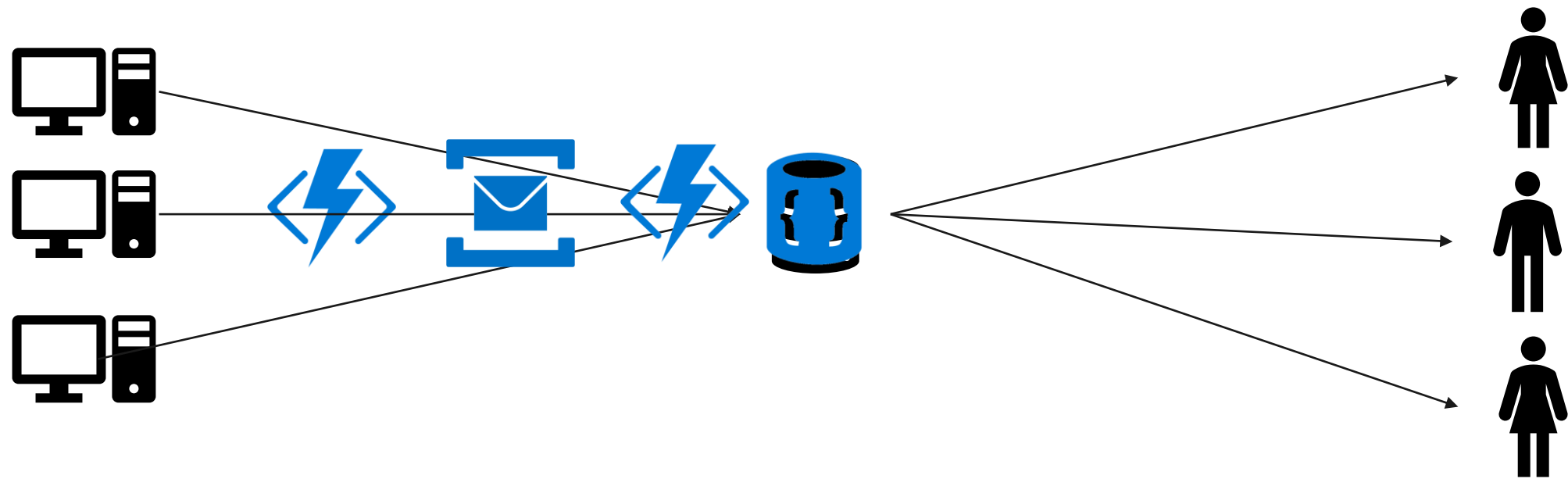
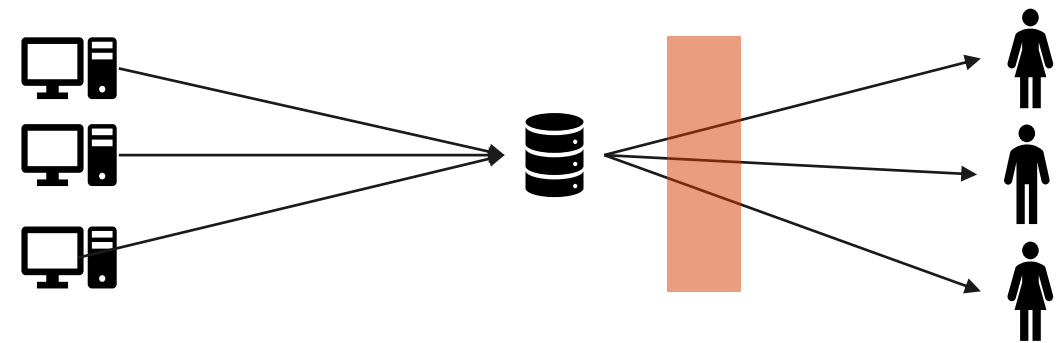
The baseline of **1 request unit** for a **1-KB** item corresponds to a simple **GET** by self link or id of the item.

Data is Updated

- Poll
- Push
 - CosmosDB SQL SDK
 - .NET Change Feed Processor SDK
 - Azure Function Change Feed Trigger
 - Soft Deletes

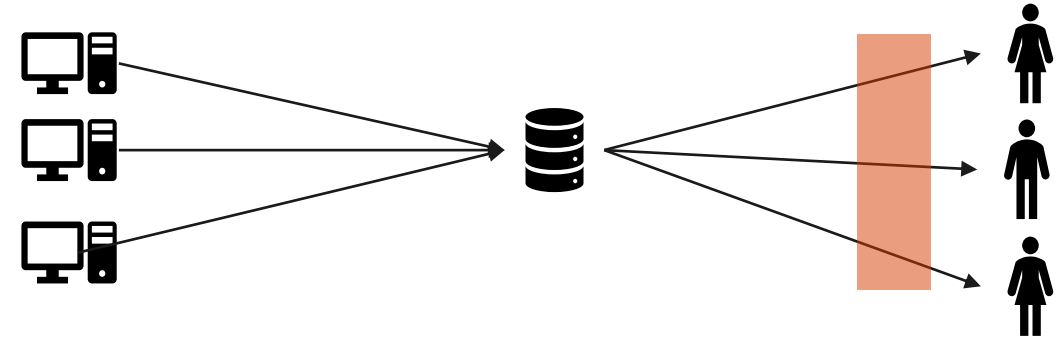


Data is Updated

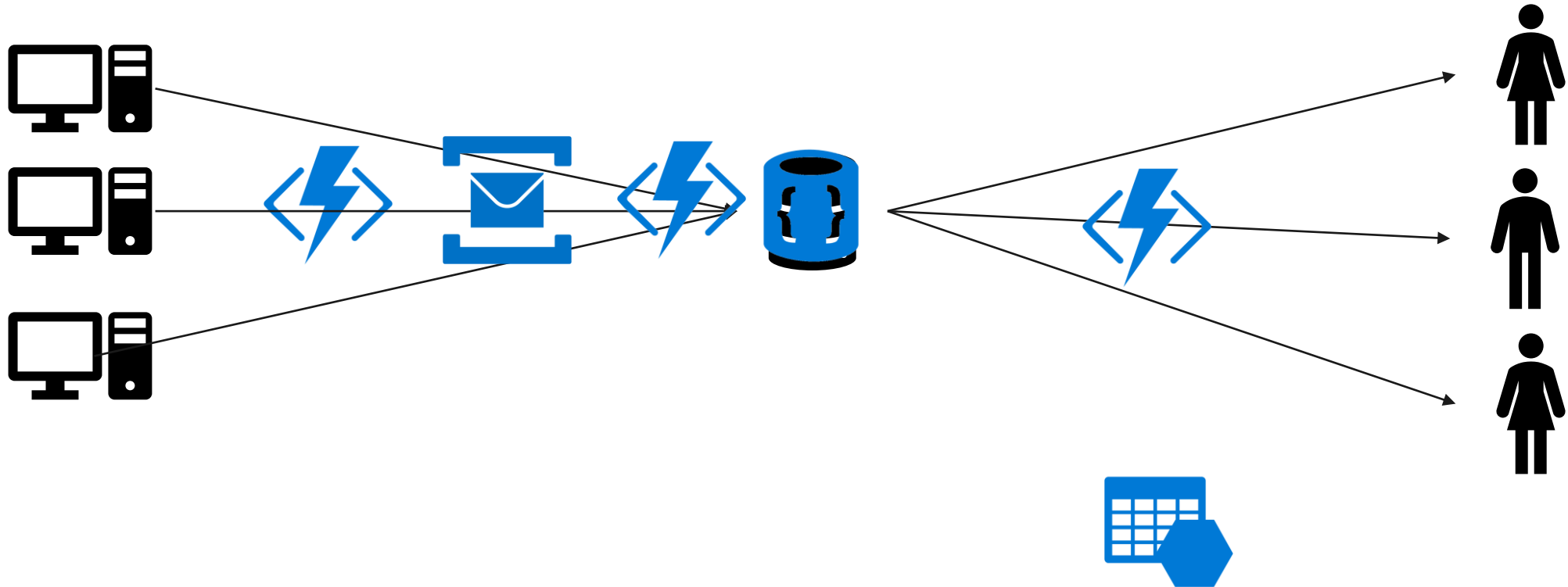
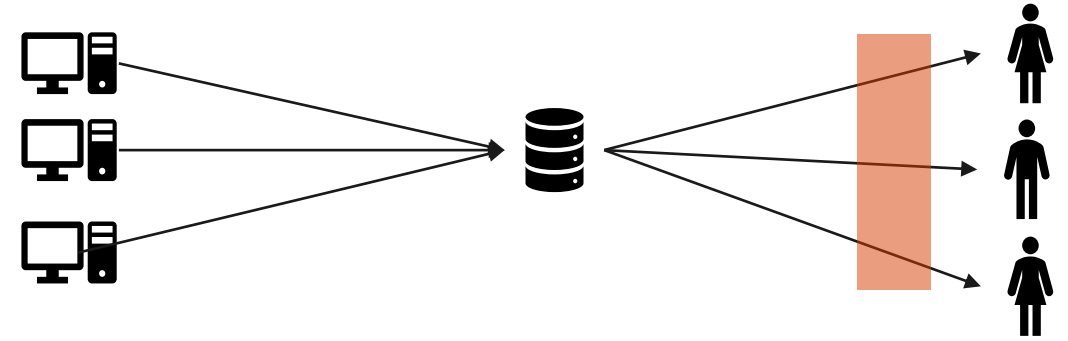


Send to Consumers

- Service Fabric
- Containers
- Serverless
 - Web Job
 - Azure Function
 - Durable Function
 - Logic App



Send to Consumers



Azure Resource Manager (ARM) template

Architecture Patterns

- Lambda Architecture
- CQRS
- Event Sourcing

Resources

- <http://functionlibrary.azurewebsites.net/>
- <https://docs.microsoft.com/en-us/azure/event-grid/compare-messaging-services>
- <https://docs.microsoft.com/en-us/azure/cosmos-db/modeling-data>
- <https://docs.microsoft.com/en-us/azure/azure-functions/functions-compare-logic-apps-ms-flow-webjobs>
- <https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview>
- <https://github.com/Azure-Samples/functions-quickstart>
- <https://docs.microsoft.com/en-us/azure/azure-functions/durable-functions-sequence>
- <https://blogs.msdn.microsoft.com/appserviceteam/2018/02/07/understanding-serverless-cold-start/>

