



Solutions Architect Technical Exercise - Presentation

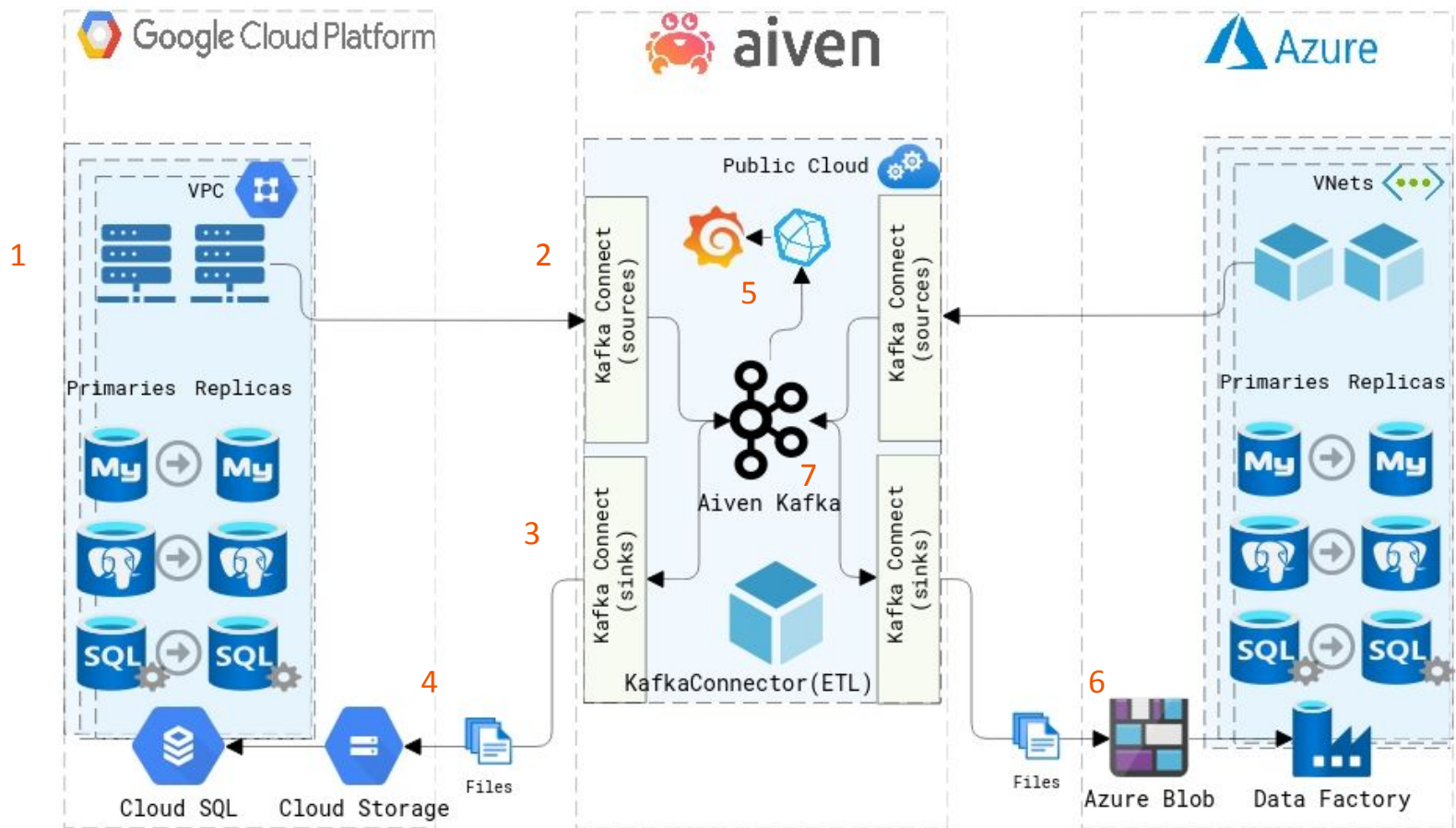
Technical Assignment for Aiven

Vladimir Grigor, vgrigor@gmail.com



Challenges

- Services in GCP and Azure clouds
 - Multiple inputs
 - Serving various internal and external services
- One of destination systems - database for mobile and web app
 - Data is updated once a day
 - Complex and slow (20s) queries
 - Data retention max 7 days
- Peak throughput: write 3MB/s and read 2MB/s
- Solution performance monitoring
- Architecture? Security? Cost and plans?



Recommended architecture:
Syncing services in GCP and Azure via Kafka as a message service



Security / Connectivity

- Connections to all services are protected by SSL/TLS (default), SASL/PLAIN, SASL/SCRAM ¹
- Whitelisting IPs by CIDR
- User/ACL authorization for topic management
- Kafka REST API and Schema Registry (Karapace)
- Kafka Connect
- Aiven services are ISO 27001 certified and SOC 2 compliant ⁴

Monitoring

- Aiven Managed Metrics (Influx + Grafana)²
- Integration to Prometheus, Elasticsearch, Kafka Logs, Datadog and more are built-in³

References:

¹"Kafka authentication types". <https://help.aiven.io/en/articles/4331547-kafka-authentication-types>. Accessed 2021-02-21

²"Getting started with service integrations". <https://help.aiven.io/en/articles/1456441-getting-started-with-service-integrations>. Accessed 2021-02-21

³"Kafka Connect connectors". <https://help.aiven.io/en/articles/1231452-kafka-connect-connectors>. Accessed 2021-02-21

⁴"Aiven security compliance". <https://help.aiven.io/en/articles/3624092-aiven-security-compliance>. Accessed 2021-02-21



Importing data from cloud storage to databases

GCP: GCS -> Cloud SQL:

- <https://cloud.google.com/sql/docs/mysql/import-export/importing#importing-data-from-a-csv-file-in-cloud-storage>
- <https://cloud.google.com/sql/docs/postgres/import-export/importing#importing-data-from-a-csv-file-in-cloud-storage>
- <https://cloud.google.com/sql/docs/sqlserver/import-export/importing#importing-the-sql-file>

Azure: Azure Blob Storage -> Database (via Azure Data Factory):

- <https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-blob-storage>
- <https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-database-for-postgresql>
- <https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-database-for-mysql>
- <https://docs.microsoft.com/en-us/azure/data-factory/connector-sql-server>



Cost and Plans

- Kafka: plan **business-4** (contains Kafka Connect support), ~**500\$/month**
- Kafka Connector: plan **business-4** (3 node, HA), ~**387\$/month**
- InfluxDB: plan **startup-4**, **60\$/month**
- Grafana: plan **startup-1**, **35\$/month**

Total: 982\$/month

References:

"Aiven Kafka pricing". <https://aiven.io/kafka#product-slider>. Accessed 2021-02-21

"How to choose the number of topics/partitions in a Kafka cluster?".
<https://www.confluent.io/blog/how-choose-number-topics-partitions-kafka-cluster/>. Accessed 2021-02-21