### **Architecture Overview**

The architecture enables secure analytics access from the **Analytics AWS Account** to production Aurora MySQL databases in the **Game AWS Account** for select employees and ETL systems.

### **Components in the Game AWS Account**

1. **Aurora MySQL Databases**
   * **Primary Databases:** Host production game data.
   * **Read Replicas:** Created to offload analytics queries and minimize impact on primary databases.
   * **Deployment:** Both primary and read replicas are deployed in private subnets within a VPC.
2. **IAM Role for Cross-Account Access**
   * **Role Name:** GameAccountAccessRole
   * **Trust Policy:** Allows IAM role from the Analytics AWS Account to assume this role.
   * **Permissions Policy:** Grants necessary permissions to access the Aurora read replicas and AWS Secrets Manager.
3. **AWS Secrets Manager**
   * Stores database credentials securely.
   * Access is granted to the GameAccountAccessRole.
4. **Network Configuration**
   * **VPC Peering** is established with the Analytics AWS Account.
   * **Security Groups:** Configured to allow inbound MySQL traffic (port 3306) from the Analytics AWS Account.

### **Components in the Analytics AWS Account**

1. **Employees**
   * Managed via **AWS IAM Identity Center**.
   * Added to the **Analytics Team** group.
   * Assigned the **AnalyticsTeamProdReadOnly** permission set.
2. **ETL Systems**
   * Includes AWS Glue jobs, AWS DMS tasks, or custom applications.
   * IAM roles assigned to these systems have permissions to assume the GameAccountAccessRole.
3. **IAM Identity Center Configuration**
   * **Group:** Analytics Team
   * **Permission Set:** AnalyticsTeamProdReadOnly
     + Grants permissions to assume roles in the Game AWS Account.
     + Provisioned to both the Analytics and Game AWS Accounts.
4. **Network Configuration**
   * VPC connected to the Game AWS Account via VPC Peering
   * **Subnets and Security Groups:** Configured to allow outbound traffic to the Aurora read replicas.

### **Data Flow and Access Control**

1. **For Employees**
   * **Authentication:** Via AWS IAM Identity Center with MFA enforced or other IDP used by the organization (Okta, etc.)
   * **Permission set:** The employee selects the desired AWS account and permission set (role). **Account:** Game AWS Account, **Permission Set:** AnalyticsTeamProdReadOnly
   * **Database Access:** Use client tools with SSL/TLS encryption to connect to the Aurora read replicas.
2. **For ETL Systems**
   * **Role Assumption:** Programmatically assume GameAccountAccessRole.
   * **Data Extraction:** Access the Aurora read replicas securely.
   * **Data Loading:** Load transformed data into analytics data stores within the Analytics AWS Account.
3. **Database Replication (Alternative Option)**
   * **AWS DMS Replication:** Continuously replicate data to the Analytics AWS Account.
   * **Benefits:** Provides data ownership and reduces cross-account dependencies.

## **Important Considerations**

### **1. Security and Compliance**

* **Principle of Least Privilege**
  + IAM roles and policies grant only necessary permissions.
  + Avoid using wildcard (\*) permissions; specify exact resources.
* **Multi-Factor Authentication (MFA)**
  + Enforce MFA for all users accessing sensitive resources.
* **Data Encryption**
  + **At Rest:** Use AWS KMS to encrypt data in Aurora MySQL and analytics data stores.
  + **In Transit:** Enforce SSL/TLS for all database connections.
* **Monitoring and Auditing**
  + Enable AWS CloudTrail and Amazon CloudWatch for logging and monitoring.
  + Regularly review logs for unauthorized access attempts or anomalies.

### **2. Network Architecture**

* **Private Connectivity**
  + Use VPC Peering to prevent exposure to the public internet.
  + Configure route tables and DNS settings appropriately.
* **Security Groups and Network ACLs**
  + Restrict inbound and outbound traffic to necessary ports and IP ranges.
  + Regularly audit and update rules as needed.

### **3. Database Replication and Performance**

* **Read Replicas**
  + Offload analytics queries to read replicas to minimize impact on production databases.
* **ETL Process Optimization**
  + Schedule ETL jobs during off-peak hours if possible.

### **4. Access Management**

* **User Groups and Permission Sets**
  + Use AWS IAM Identity Center to manage user access centrally.
  + Assign permission sets with appropriate policies to groups.
* **Role-Based Access Control**
  + Use IAM roles for both employees and ETL systems.
  + Implement strict trust relationships and permissions.
* **Onboarding and Offboarding Procedures**
  + Establish processes to promptly update user access when roles change or employees leave. In the current setup, this would involve managing IAM user accounts and Groups.

### **5. Operational Excellence**

* **Monitoring and Alerting**
  + Set up alerts for replication lag, ETL job failures, and security incidents.
  + Use Amazon CloudWatch dashboards for real-time monitoring.
* **Disaster Recovery**
  + Implement backup and restore strategies for both production and analytics databases.
  + Regularly test recovery procedures.
* **Cost Management**
  + Monitor resource utilization to optimize costs.
  + Be aware of data transfer charges between AWS accounts and regions.

### **6. Data Governance**

* **Data Quality**
  + Implement validation checks in ETL processes to ensure data integrity.
  + Maintain detailed records of data transformations.
* **Audit Trails**
  + Keep logs of data access and modifications.
  + Use AWS Config to track and record configuration changes.